



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Ciências Aplicadas



BRUNA DIAS ARAUJO

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE TENDÊNCIAS DO MERCADO BRASILEIRO DE AÇÕES

LIMEIRA

2022



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Ciências Aplicadas



BRUNA DIAS ARAUJO

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE TENDÊNCIAS DO MERCADO BRASILEIRO DE AÇÕES

Dissertação apresentada à Faculdade de Ciências Aplicadas da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do título de Mestra em Engenharia de Produção e de Manufatura na área de Pesquisa Operacional e Gestão de Processos.

Orientador: Cristiano Torezzan

Co-orientador: Fabiano Boaventura de Miranda

Este exemplar corresponde à versão final da dissertação defendida pela aluna Bruna Dias Araujo, e orientada pelo Prof. Dr. Cristiano Torezzan

LIMEIRA

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Ciências Aplicadas
Ana Luiza Clemente de Abreu Valério - CRB 8/10669

Ar15a Araujo, Bruna Dias, 1996-
Aplicação de técnicas de aprendizado de máquina para predição de
tendências do mercado brasileiro de ações / Bruna Dias Araujo. – Limeira, SP :
[s.n.], 2022.

Orientador: Cristiano Torezzan.
Coorientador: Fabiano Boaventura de Miranda.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade
de Ciências Aplicadas.

1. Aprendizado de máquina. 2. Mercado de ações. 3. Ações (Finanças). 4.
Negociação. I. Torezzan, Cristiano, 1976-. II. Miranda, Fabiano Boaventura de,
1987-. III. Universidade Estadual de Campinas. Faculdade de Ciências
Aplicadas. IV. Título.

Informações Complementares

Título em outro idioma: Application of machine learning techniques to predict Brazilian
stock market trends

Palavras-chave em inglês:

Machine learning

Stock market

Stocks

Negotiation

Área de concentração: Pesquisa Operacional e Gestão de Processos

Titulação: Mestra em Engenharia de Produção e de Manufatura

Banca examinadora:

Cristiano Torezzan [Orientador]

Ana Estela Antunes da Silva

Washington Alves de Oliveira

Data de defesa: 14-10-2022

Programa de Pós-Graduação: Engenharia de Produção e de Manufatura

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0003-4377-7592>

- Currículo Lattes do autor: <https://lattes.cnpq.br/8056847241169050>

Folha de Aprovação

Autor(a): Bruna Dias Araujo

Título: Aplicação de técnicas de aprendizado de máquina para predição de tendências do mercado brasileiro de ações

Natureza: Dissertação

Área de Concentração: Pesquisa Operacional e Gestão de Processos

Instituição: Faculdade de Ciências Aplicadas – FCA/Unicamp

Data da Defesa: Limeira-SP, 14 de outubro de 2022.

BANCA EXAMINADORA:

Prof. Dr. Cristiano Torezzan (orientador)
Faculdade de Ciências Aplicadas - FCA/Unicamp

Prof. Dr. Washington Alves de Oliveira (membro)
Faculdade de Ciências Aplicadas - FCA/Unicamp

Prof. Dr. Ana Estela Antunes da Silva (membro externo)
Faculdade de Tecnologia - FT/Unicamp

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Ciências Aplicadas
Ana Luiza Clemente de Abreu Valério - CRB 8/10669

Ar15a Araujo, Bruna Dias, 1996-
Aplicação de técnicas de aprendizado de máquina para predição de
tendências do mercado brasileiro de ações / Bruna Dias Araujo. – Limeira, SP :
[s.n.], 2022.

Orientador: Cristiano Torezzan.
Coorientador: Fabiano Boaventura de Miranda.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade
de Ciências Aplicadas.

1. Aprendizado de máquina. 2. Mercado de ações. 3. Ações (Finanças). 4.
Negociação. I. Torezzan, Cristiano, 1976-. II. Miranda, Fabiano Boaventura de,
1987-. III. Universidade Estadual de Campinas. Faculdade de Ciências
Aplicadas. IV. Título.

Informações Complementares

Título em outro idioma: Application of machine learning techniques to predict Brazilian
stock market trends

Palavras-chave em inglês:

Machine learning

Stock market

Stocks

Negotiation

Área de concentração: Pesquisa Operacional e Gestão de Processos

Titulação: Mestra em Engenharia de Produção e de Manufatura

Banca examinadora:

Cristiano Torezzan [Orientador]

Ana Estela Antunes da Silva

Washington Alves de Oliveira

Data de defesa: 14-10-2022

Programa de Pós-Graduação: Engenharia de Produção e de Manufatura

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0003-4377-7592>

- Currículo Lattes do autor: <https://lattes.cnpq.br/8056847241169050>

Dedico este trabalho aos meus pais e à minha irmã, que sempre me amaram e me apoiaram incondicionalmente e me deram forças para enfrentar todos os desafios.

Agradecimentos

Agradeço a Deus por minha vida e por todas as oportunidades.

Aos meus pais, Mara e Amauri e à minha irmã, Mayara, que sempre me apoiaram em todos os meus sonhos, me ensinaram e deram todo o suporte necessário para que eu chegasse aqui.

Ao Matheus, por todo carinho e parceria durante as diferentes etapas da minha vida e por sempre acreditar em mim.

Ao Cristiano, pela orientação, confiança e apoio ao longo deste trabalho e, também, durante toda a graduação.

Ao Fabiano, pela coorientação e por todo o suporte e auxílio nos momentos mais difíceis.

À Beatriz, que esteve comigo durante toda a graduação e o mestrado, compartilhando os momentos alegres e desafiadores.

A toda a equipe da FlexSim Brasil pela confiança em meu trabalho e pelo constante incentivo para o meu desenvolvimento pessoal e profissional.

A todas as pessoas que estiveram presentes na minha vida e que, de alguma forma, contribuíram para que eu chegasse até aqui.

Resumo

A previsão do comportamento do mercado financeiro, em específico o mercado de ações, é uma tarefa desafiadora que tem atraído investidores e pesquisadores de diferentes áreas nos últimos anos. A dificuldade é intensificada pela Hipótese do Mercado Eficiente (HME), que afirma que não é possível prever uma série temporal do mercado financeiro, uma vez que os preços do mercado sempre refletem todas as informações disponíveis. Entretanto, no início do século XXI, estudos mostraram que os preços das ações são parcialmente previsíveis com base nos padrões de preços anteriores e evidenciaram a influência de fatores psicológicos e comportamentais nos preços dos ativos. Diante disso, a teoria de finanças comportamentais surgiu contrapondo-se à HME para explicar os comportamentos não racionais dos agentes do mercado em determinados momentos, levando a períodos de ineficiência do mercado. Essas janelas de tempo permitem alguma previsibilidade do comportamento do mercado considerando o uso de indicadores técnicos. A análise técnica baseia suas previsões em indicadores matemáticos construídos sobre o preço das ações e auxilia os investidores a definirem qual a tendência geral dos preços dos ativos. Trabalhos recentes têm obtido resultados promissores na predição de tendência do mercado. Neste trabalho, é proposto um modelo de aprendizado de máquina para predição de reações de curto prazo nos preços dos ativos do mercado brasileiro. Como dados de entrada são utilizados indicadores técnicos e, como saída, duas possíveis sugestões para o investidor: “Compra” ou “Neutro”. De modo a reduzir eventuais perdas causadas por erros de classificação do modelo, foram propostas 5 estratégias de negociação, utilizando-se das ferramentas *stop loss* e *stop gain*. Os resultados das estratégias foram comparados à estratégia *buy and hold*, comumente utilizada no mercado. O modelo proposto apresentou precisão de até 70% nas indicações de “Compra”, e ao ser associado com uma estratégia de negociação, o resultado superou a estratégia *buy and hold* em 96% das vezes.

Palavras-chaves: Aprendizado de máquina; Classificação; Mercado de ações; Análise técnica; Estratégias de negociação.

Abstract

The financial market forecast behavior, specifically the stock market, is a challenging task that has attracted investors and researchers from different areas in the last years. The difficulty is intensified by the Efficient Market Hypothesis (EMH), which states that it is not possible to predict a financial market timeseries, since that market prices always reflect all available information. Although, at the beginning of the 21st century, studies have shown that stock prices are partially predictable based on previous price patterns and shown the influence of psychological and behavioral factors in stock prices. So, the theory of financial changes emerged to counter the EMH to explain the non-rational behaviors of market agents at certain times, leading to periods of market inefficiency. These time windows allow some predictability of the market behavior considering the use of technical indicators. The technical analysis bases its predictions on mathematical indicators built on the stock price and helps investors to define the general trend of asset prices. Recent studies have obtained promising results in predicting market trends. In this work, a machine learning model is proposed for short-term changes prediction in asset prices in the Brazilian market. As input data technical indicators are used and, as an output, two possible suggestions for investors: “Buy” or “Neutral”. In order to reduce any losses caused by errors in the model’s classification, 5 negotiation strategies were proposed, using the stop loss and stop gain tools. The results of the strategies were compared to the buy and hold strategy, commonly used in the stock market. The proposed model presented an accuracy of up to 70% in “Buy” suggestions, and when associated with a strategy of trading, the result beat the buy and hold strategy 96% of the times.

Keywords: Machine Learning; Classification; Stock Market; Technical Analysis; Trading strategies.

Lista de ilustrações

Figura 2.1 – Representação de um <i>candlestick</i> de alta (verde) e um de baixa (vermelho)	23
Figura 2.2 – Representação de um gráfico diário de <i>candlestick</i> .	23
Figura 2.3 – Representação Média Móvel	24
Figura 2.4 – Representação MACD	26
Figura 2.5 – Representação RSI	27
Figura 2.6 – Representação da árvore de decisão	29
Figura 2.7 – Representação da rede neural MLP	32
Figura 2.8 – Matriz de confusão	35
Figura 2.9 – Métodos de Balanceamento dos Dados	37
Figura 3.1 – Estrutura Classificador Binário	42
Figura 3.2 – Coleta de dados	42
Figura 3.3 – Cálculo dos indicadores	43
Figura 3.4 – Definição dos parâmetros	43
Figura 3.5 – Treino e teste do modelo	44
Figura 3.6 – Análise dos resultados	44
Figura 3.7 – Fluxograma das etapas de desenvolvimento do modelo	45
Figura 4.1 – Histograma Dados de Treino - LREN3	48
Figura 4.2 – Matriz de confusão - Árvore de Decisão - LREN3	48
Figura 4.3 – Matriz de confusão - Rede Neural - LREN3	49
Figura 4.4 – Matriz de confusão - <i>XGBoost</i> - LREN3	49
Figura 4.5 – Histograma de variação Rede neural MLP - Classificação + Alvo	51
Figura 4.6 – Histograma de variação ativos - Classificação + Alvo	53
Figura 4.7 – Histograma de variação ativos - Classificação + Alvo	54
Figura 5.1 – Estratégia 0 - Saída ao final do período 1	58
Figura 5.2 – Estratégia 0 - Saída ao final do período 2	58
Figura 5.3 – Estratégia 1 - Saída ao final do período	59
Figura 5.4 – Estratégia 1 - Saída no <i>stop loss</i>	60
Figura 5.5 – Estratégia 2 - Saída no alvo	61
Figura 5.6 – Estratégia 2 - Saída ao final do período	61
Figura 5.7 – Estratégia 3 - Saída no alvo	62

Figura 5.8 – Estratégia 3 - Saída ao atingir o <i>stop loss</i>	62
Figura 5.9 – Estratégia 4 - Saída no alvo	63
Figura 5.10–Estratégia 4 - Saída ao atingir o <i>stop loss</i>	64
Figura 5.11–Retorno médio esperado por ativo	67
Figura 5.12–Retorno total esperado por ativo	68
Figura 5.13–Retorno total por estratégia	72
Figura 5.14–Retorno total por método	72
Figura 5.15–Gráfico de Efeitos Principais	73
Figura 5.16–Gráfico de Dispersão: Retorno total X Precisão do Classificador	74

Lista de tabelas

Tabela 2.1 – Classificação das operações no mercado financeiro	22
Tabela 4.1 – Parâmetros e hiperparâmetros dos Classificadores	47
Tabela 4.2 – Métricas de desempenho	50
Tabela 4.3 – Papéis avaliados	52
Tabela 4.4 – Métricas de desempenho dos classificadores	54
Tabela 4.5 – Métricas de desempenho dos classificadores x classificador aleatório . .	55
Tabela 5.1 – Papéis avaliados	65
Tabela 5.2 – Parâmetros utilizados para Classificação	66
Tabela 5.3 – Número de Operações de acordo com o método e estratégia	69
Tabela 5.4 – Retorno esperado por estratégia	70
Tabela 5.5 – Comparação com estratégia <i>buy and hold</i>	71

Sumário

1	Introdução	14
1.1	Contexto do problema	14
1.2	Problema de pesquisa	16
1.3	Objetivo	17
1.4	Justificativa	17
2	Referencial teórico	18
2.1	Hipóteses de mercado	18
2.2	Indicadores de análise técnica para o mercado de ações	20
2.2.1	Média móvel simples e exponencial	24
2.2.2	Média móvel convergente e divergente (MACD)	25
2.2.3	Índice de força relativa (RSI)	26
2.3	Aprendizado de máquina	27
2.3.1	Árvore de decisão	28
2.3.2	Rede neural	31
2.3.3	<i>Extreme gradient boosting</i> (XGBoost)	33
2.3.4	Métricas de desempenho	34
2.3.5	Balanceamento dos dados	36
2.4	Gestão de carteiras	37
2.4.1	Investimento	37
2.4.2	Gestão de risco e retorno	37
2.4.3	Teoria de portfólio	38
2.4.4	Gestão ativa x Gestão passiva	39
2.4.5	<i>Buy and hold</i>	39
2.4.6	<i>Benchmarking</i>	40
3	Modelo Proposto	41
4	Experimentos computacionais e resultados empíricos	46
5	Resultados: Estratégias de negociação	57
6	Conclusão	75

Referências 77

1 Introdução

1.1 Contexto do problema

Ao longo dos anos, o interesse de investidores e pesquisadores no desenvolvimento de modelos para análise do comportamento do preço de ações tem aumentado de maneira significativa (SHAH *et al.*, 2019). Entretanto, essa é uma tarefa desafiadora, uma vez que o mercado de ações é afetado por fatores altamente inter-relacionados, abrangendo aspectos econômicos, políticos, psicológicos e ainda fatores específicos da empresa (ZHONG; ENKE, 2017).

Devido à complexidade do mercado de ações, diversas teorias surgiram para tentar explicar seu comportamento, entre elas estão a Hipótese do Mercado Eficiente e Hipótese do Mercado Adaptável. A Hipótese do Mercado Eficiente (HME) afirma que o mercado financeiro é informacionalmente eficiente, ou seja, os preços do mercado sempre refletem todas as informações existentes (FAMA, 1970). Dessa forma, desenvolver um sistema preditor baseado nas informações disponíveis e obter retornos acima da média do mercado no longo prazo não seria possível, pois, estas informações já estariam refletidas nos preços atuais das ações.

Entretanto, no início do século XXI, o domínio da HME tornou-se menos universal. Economistas e estatísticos passaram a acreditar que os preços das ações são parcialmente previsíveis com base nos padrões de preços anteriores e enfatizaram a influência de elementos psicológicos e comportamentais na determinação do preço dos ativos (MALKIEL, 2003). Nesse contexto, estudiosos da área de finanças comportamentais atribuíram a ineficiência do mercado a teorias psicológicas e heurísticas cognitivas como julgamentos pessoais, excesso de confiança, reações exageradas, expectativas em relação ao futuro, otimismo/pessimismo, ego e autoestima (NASEER; TARIQ, 2015).

Na tentativa de reconciliar as teorias econômicas com a economia comportamental surgiu a Hipótese do Mercado Adaptável (HMA) (LO, 2019). De acordo com a HMA, é possível explorar fragilidades na eficiência do mercado de modo a obter retornos positivos de uma carteira de ações. A literatura apresenta diversos estudos relacionados com as tentativas de previsão e análise do mercado financeiro. Neste contexto, duas abor-

dagens tiveram destaque: análise técnica e análise fundamentalista. Analistas técnicos acreditam que o preço é capaz de explicar exhaustivamente os movimentos do mercado, portanto, suas estratégias são baseadas em indicadores matemáticos calculados sobre os preços passados das ações. Por outro lado, a análise fundamentalista considera dados provenientes de notícias, rentabilidade e fatores macroeconômicos (PICASSO *et al.*, 2019).

De forma complementar, a teoria de Dow, um dos principais conceitos aplicados na análise técnica, surgiu com informações que permitiram um maior entendimento do comportamento do mercado de ações. Esta teoria busca compreender as tendências e oscilações de preços de diversos tipos de ativos, com base em dados do passado. Segundo a teoria de Dow, o mercado de ações se movimenta em ondas, ou seja, em tendências, sendo possível prever as próximas oscilações através de desenhos e padrões identificados.

O desenvolvimento da análise financeira foi favorecido pela disponibilidade crescente de recursos computacionais e o avanço da ciência dos dados (PICASSO *et al.*, 2019). Inicialmente, foram aplicados métodos estatísticos sobre dados passados com o objetivo de prever valores futuros do mercado, tais como Médias Móveis. Em seguida, técnicas computacionais baseadas em Inteligência Artificial foram desenvolvidas, com destaque para as técnicas de Aprendizado de Máquina.

Associado ao desenvolvimento de técnicas para avaliação do mercado financeiro, os avanços nas tecnologias computacionais também levaram a uma maior disponibilidade de informações que passaram a ser exploradas pelos investidores. Um importante marco do avanço *Web* ocorreu em meados de 1999, com o lançamento do sistema de *home broker*, que tornou viável a participação de pequenos e médios investidores no mercado de ações. Este sistema permite que o investidor, através de uma corretora, envie ordens de compra e/ou venda diretamente ao sistema de negociações, facilitando o acesso aos dados e informações em tempo real e contribuindo para a popularização desse tipo de atividade.

Diante disso, as críticas a HME deram origem a estudos que utilizaram abordagens novas e bem-sucedidas que combinam indicadores de análise técnica e padrões gráficos com metodologias de econometria, estatística, mineração de dados e inteligência artificial (ARÉVALO *et al.*, 2017).

Trabalhos focados no comportamento financeiro, como (PICASSO *et al.*, 2019), revelam que as emoções influenciam nas decisões no mercado de ações. Estudiosos da área

defendem que os sentimentos podem levar a flutuações de curto prazo no mercado, mas no longo prazo os fundamentos da empresa fazem com que o valor e o preço de mercado convirjam. Enquanto o preço é definido pela cotação do ativo, isto é, o preço negociado na bolsa, o valor de um ativo é um conceito mais subjetivo e está relacionado com o que a empresa proporciona aos seus acionistas e o potencial no mercado. Um exemplo disto origina-se do Prêmio Nobel Robert Shiller, que demonstrou que os preços das ações são extremamente voláteis no curto prazo, porém, relativamente previsíveis em longos períodos (SHAH *et al.*, 2019).

1.2 Problema de pesquisa

A previsão do mercado de ações é um problema desafiador, cuja complexidade está relacionada a múltiplos fatores que podem afetar o preço de uma ação no curto e médio prazo. Em função dessa complexidade, há um crescente interesse no uso de técnicas de inteligência computacional para aplicações em finanças (PICASSO *et al.*, 2019).

Diante das dificuldades na previsão do valor dos ativos, visto que o mercado de ações pode apresentar elevada volatilidade no curto e médio prazo, os investidores utilizam-se muitas vezes de estratégias para reduzir as perdas e obter um desempenho superior à média do mercado. Essas estratégias fazem parte da gestão ativa de carteira que consiste, principalmente, na compra e venda de ativos com base em previsões, análises e experiência do investidor.

Para que a gestão de riscos apresente resultados positivos, o investidor deve ser capaz de identificar os comportamentos que distanciem o preço dos ativos de seu valor real e que sejam alvo de correções, para executar as ações de compra e/ou venda dos ativos.

Considerando esse contexto, este trabalho tem como motivação responder às seguintes perguntas: algoritmos de aprendizado de máquina (AM) podem ser treinados para sugerir investimentos (compras e vendas) de ações com base em indicadores técnicos? É possível utilizar AM para prever reações de curto e médio prazo ¹ no mercado de ações de modo a obter retornos lucrativos? Técnicas de gestão de riscos podem ser associadas com AM para aumentar os retornos nas operações?

¹ No contexto do trabalho, serão consideradas operações de *swing trade*, conforme descrito na Tabela 2.1.

1.3 Objetivo

Este trabalho tem como objetivo investigar a efetividade de modelos de aprendizado de máquina na predição de reações de curto prazo nos preços de ativos do mercado brasileiro com base na combinação de indicadores técnicos e estratégias de negociação.

1.4 Justificativa

Apesar da existência de diversos indicadores técnicos utilizados como ferramentas de apoio nas decisões de investimento, o uso desses indicadores de maneira isolada não assegura que as diversas oscilações do mercado sejam captadas de forma a garantir resultados lucrativos.

Nos últimos anos, técnicas de aprendizado de máquina têm sido amplamente estudadas por seu perfil orientado a dados, isto é, a capacidade de aprender automaticamente a partir de grandes volumes de dados (LUDERMIR, 2021). Diante disso, utilizar mais de um indicador técnico de forma simultânea para treinar modelos, se mostra como uma oportunidade interessante para o desenvolvimento de uma ferramenta para análise do mercado de ações. Ademais, a possibilidade de aprimorar o modelo continuamente é um diferencial importante devido à volatilidade do mercado e as particularidades de cada empresa/setor analisado.

2 Referencial teórico

2.1 Hipóteses de mercado

Nesta Seção abordaremos as principais hipóteses de mercado existentes, bem como as motivações de cada uma delas. Estas hipóteses surgiram para explicar os comportamentos do mercado e são utilizadas até hoje para o desenvolvimento de técnicas que visam o aumento dos retornos nas operações no mercado financeiro.

De acordo com a hipótese do mercado eficiente (HME), um mercado é considerado eficiente quando reflete completamente todas as informações disponíveis na determinação dos preços dos ativos. Sendo assim, a eficiência de um mercado implica na impossibilidade de obter retornos acima da média do mercado no longo prazo. A eficiência do mercado é comumente classificada em 3 níveis, de acordo com o conjunto de informações (MALKIEL, 1989):

- **Eficiência fraca:** a eficiência fraca da HME afirma que os preços refletem integralmente as informações contidas na sequência histórica de preços. Dessa forma, a análise de padrões de preços passados não ajudam na previsão dos preços futuros. A forma fraca de eficiência está relacionada à teoria do passeio aleatório, isto é, as variações nos valores das ações independem uma da outra.
- **Eficiência semi-forte:** a forma semi-forte da HME afirma que os preços atuais refletem não apenas as informações históricas de preços, mas também todas as informações publicamente disponíveis. Sendo assim, análises de balanços, demonstrações de resultados ou qualquer outra informação pública de uma empresa (análise fundamental), não levarão a lucros acima do mercado.
- **Eficiência forte:** a forma forte da HME afirma que todas as informações conhecidas por qualquer participante do mercado sobre uma empresa, são totalmente refletidas nos preços de mercado. Portanto, mesmo aqueles que possuem informações privilegiadas do mercado não conseguem utilizá-las para obter maiores lucros.

No contexto geral da HME, se os mercados são eficientes, a análise de padrões de preços passados para prever os preços futuros (conhecida como análise técnica), não seria possível, pois qualquer informação resultante desta análise já estaria refletida nos preços dos ativos. Deste modo, nenhuma análise permitiria que o investidor superasse a estratégia de comprar e manter o ativo em carteira (*buy and hold*¹) (MALKIEL, 1989).

Em resumo, a HME defende que os mercados incorporam de forma completa, precisa e instantânea todas as informações disponíveis aos preços dos ativos. Implícito neste contexto está a ideia de que os agentes do mercado são seres econômicos racionais que decidem de forma ótima considerando os riscos e probabilidades. As suposições de racionalidade, no entanto, geraram questionamentos em estudiosos da área, principalmente em psicólogos e economistas comportamentais.

A Hipótese do Mercado Adaptativo (HMA) foi proposta por (LO, 2019), que defendia algumas ideias provenientes da intersecção entre a psicologia e a economia. Dessa forma, a HMA é formada por uma associação da HME com questões comportamentais. A HMA pode ser vista com uma nova versão da HME, derivada de princípios evolucionários.

Os componentes primários da HMA consistem nas seguintes ideias:

1. Os indivíduos agem em seu próprio interesse;
2. Indivíduos cometem erros;
3. Indivíduos aprendem e se adaptam;
4. A competição impulsiona a adaptação e a inovação;
5. A seleção natural molda a ecologia do mercado;
6. A evolução determina a dinâmica do mercado.

A HME e HMA apresentam um ponto de partida comum (1), porém, distinguem-se nos pontos (2) e (3), uma vez que, em mercados eficientes, os investidores não cometem erros e não há aprendizado e adaptação. Em (4) afirma-se que a adaptação dos agentes do mercado não ocorre de forma independente das forças do mercado, mas é impulsionada pela competição. As interações entre os participantes do mercado são regidos pela seleção

¹ A estratégia *buy and hold* está definida na Seção 2, no item “Gestão de Carteiras”.

natural e em (6) conclui-se que a dinâmica do mercado é formada pela combinação de todos esses fatores - interesses dos indivíduos, competição, adaptação, seleção natural e condições ambientais.

Dentro do contexto da HMA, os agentes têm a capacidade de adaptação para atingir maiores retornos, porém, não de forma 100% racional, devido às limitações cognitivas e preferências. Os agentes com maior capacidade de adaptação são aqueles com maior chance de sobreviver mais tempo dentro do mercado. Por outro lado, àqueles que não conseguem se adaptar tendem a acumular perdas percebidas ao longo do tempo e desaparecerem do mercado em função à aversão a perda.

O processo de adaptação e aprendizado ocorre a medida que as estratégias adotadas são recompensadas (retornos lucrativos) e as que não apresentam sucesso são inibidas (prejuízos). Através da observação do mercado, os agentes tendem a aprender quais estratégias funcionam. Nos casos em que o ambiente externo é estável, essa tendência leva em direção a um ponto ótimo.

Porém, de acordo com (LO, 2019), com as mudanças do ambiente externo, como variação de recursos ou número de agentes, a tendência de adaptação nem sempre é direta e estratégias que funcionavam bem em determinados contextos, podem não mais apresentarem retornos favoráveis após a mudança de fatores externos, gerando assim momento de ineficiência informacional.

2.2 Indicadores de análise técnica para o mercado de ações

Nesta Seção iremos expor brevemente as duas principais abordagens utilizadas para analisar o mercado de ações e, em seguida, apresentar os indicadores técnicos que serão utilizados neste trabalho.

Análise técnica e análise fundamentalista são as duas principais abordagens utilizadas para analisar o mercado de ações e justificar os seus movimentos. Os analistas técnicos defendem que o preço é capaz de explicar de maneira exaustiva as oscilações do mercado, portanto, suas estratégias são baseadas nos preços das ações e indicadores matemáticos calculados sobre eles (SERAFINI; PEREIRA, 2010).

Por outro lado, a análise fundamentalista busca avaliar se há um valor real ou intrínseco que se relaciona diretamente com o desempenho da empresa, ou seja, tem

como objetivo determinar o real valor de uma ação. Os analistas fundamentalistas avaliam as empresas com base em suas expectativas de crescimento e rendimento, na análise de seus riscos e em outros indicadores. De forma resumida, o analista fundamentalista busca focar seus estudos no comportamento econômico e financeiro das empresas e através das comparações entre empresas de um mesmo setor, avaliando a conjuntura econômica e política (WAGNER, 2007).

Os indicadores técnicos são calculados a partir dos dados referentes aos preços das ações e utilizados para prever a direção de movimento, ou seja, se o ativo apresenta tendência de alta ou de baixa (DEY *et al.*, 2016). Os indicadores técnicos são comumente classificados em 4 categorias: tendência, *momentum*, volume e volatilidade.

Os indicadores de tendência auxiliam o investidor a definir qual a tendência geral dos preços dos ativos. As tendências podem ser de alta, de baixa ou lateral (quando os preços movimentam-se sem uma direção definida). As tendências podem ser classificadas também em relação ao horizonte de análise: curto, médio ou longo prazo. Um mesmo ativo pode apresentar diferentes tendências de acordo com o prazo analisado (DEBASTIANI, 2008). A maior parte dos indicadores de tendência baseiam-se no uso de médias móveis, que serão detalhadas na sequência.

Os indicadores de *momentum* indicam a velocidade da mudança de preço da ação, utilizada para identificar uma possível reversão de tendência. Os indicadores de volume refletem o entusiasmo dos investidores na compra ou venda dos ativos e, por fim, os indicadores de volatilidade apresentam a faixa de flutuação dos preços das ações para avaliar o risco e identificar o nível de suporte e resistência. Suportes são níveis de preço em que o interesse de compra supera o interesse de venda e, por isso, dificultam a queda dos preços a partir daquele ponto e resistências agem da maneira inversa, de modo que consistem em níveis de preço em que o interesse de venda supera o de compra e, por isso, dificultam o aumento dos preços. (DEBASTIANI, 2008).

A bolsa de valores permite a utilização de diferentes estratégias. É possível realizar investimentos de longo prazo ou optar pela especulação com diferentes tipos de *trade*. As classificações dessas operações estão relacionadas aos prazos de atuação, conforme descrito na Tabela 2.1.

Tabela 2.1 – Classificação das operações no mercado financeiro

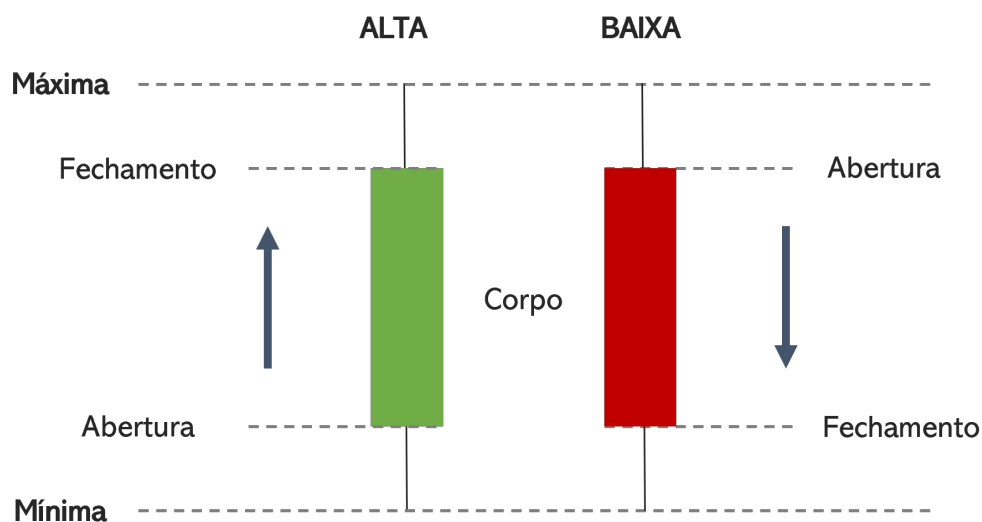
Operação	Prazo	Período	Intervalo análise gráfica
Scalper	Ultra curto prazo	Minutos	Menos de 1 min até 15 min
Day Trader	Curto prazo	Horas	De 5 min até 4 horas
Swing Trader	Médio prazo	Dias ou semanas	De 1 hora até 1 dia
Position Trader	Longo prazo	Meses ou anos	1 dia, 1 semana, 1 mês

Fonte: Elaborado pela autora

Entre as ferramentas disponíveis para análise do mercado financeiro encontra-se o gráfico de *candlestick* (também conhecido como gráfico de velas), que representa as variações de valores dos ativos em um determinado período. Cada barra de *candlestick* representa um período, sendo composto por quatro preços (DEBASTIANI, 2007).

1. Preço de abertura: representa o preço pelo qual foi fechado o primeiro negócio dentro do período;
2. Preço de fechamento: representa o preço pelo qual foi fechado o último negócio dentro do período;
3. Preço máximo: representa o preço mais alto pelo qual foi fechado um negócio dentro do período;
4. Preço mínimo: representa o preço mais baixo pelo qual foi fechado um negócio dentro do período.

A cor em que se representa o corpo do *candle* é definida pelo movimento dos preços dentro do período. Na Figura 2.1, o *candle* verde representa um período de alta (preço de abertura menor do que o preço de fechamento), e o *candle* vermelho representa um período de baixa (preço de abertura maior do que o preço de fechamento) (DEBASTIANI, 2007).

Figura 2.1 – Representação de um *candlestick* de alta (verde) e um de baixa (vermelho)

Fonte: Elaborado pela autora

O *candlestick* pode representar diferentes períodos (horas, dias, meses, semanas, etc). A Figura 2.2 apresenta um gráfico de *candlestick* diário.

Figura 2.2 – Representação de um gráfico diário de *candlestick*.

Fonte: Adaptado pela autora de (TRADING VIEW).

2.2.1 Média móvel simples e exponencial

A média móvel é um dos principais indicadores da categoria de rastreadores de tendência e é amplamente utilizada com o objetivo de entender o comportamento atual do preço do ativo e quais são as tendências demonstradas em um curto período. Os investidores enxergam oportunidades de compra e venda com médias móveis, quando ocorrem cruzamentos entre médias de diferentes prazos ou entre médias e preços, conforme ilustrado na Figura 2.3.

Figura 2.3 – Representação Média Móvel



Fonte: Adaptado pela autora de (TRADING VIEW).

A média móvel simples (MMS) indica o preço médio de um ativo em um período N com base no preço de fechamento. Seja $\{p_1, p_2, \dots, p_N\}$ uma série temporal com os preços de fechamento de um determinado ativo para N períodos de tempo. A média móvel simples dos k últimos preços é calculada de acordo com a equação (2.1)

(DEBASTIANI, 2008).

$$MMS(k) = \frac{1}{k} \sum_{i=N-k+1}^N p_i, \quad (2.1)$$

Além da MMS, outro indicador muito utilizado é a média móvel exponencial (MME), que se caracteriza por atribuir um peso aos valores da série temporal, conforme exibido na equação (2.2), onde , $0 < \alpha < 1$ (DEBASTIANI, 2008).

$$MME(k) = \frac{p_k + (1 - \alpha)p_{k-1} + (1 - \alpha)^2 p_{k-2} + \dots + (1 - \alpha)^k p_0}{1 + (1 - \alpha) + (1 - \alpha)^2 + \dots + (1 - \alpha)^k}, \quad (2.2)$$

2.2.2 Média móvel convergente e divergente (MACD)

A Média Móvel Convergente e Divergente (do inglês, *Moving Average Convergence Divergence* - MACD) foi desenvolvida em 1979 por Gerald Appel, analista e administrador de recursos em Nova York, e tem como característica principal a identificação de sinais de mudança no equilíbrio das forças entre compradores e vendedores. O MACD é um indicador técnico que mostra a diferença entre dois sinais, um rápido e um lento, gerados a partir do movimento das médias móveis exponenciais (APPEL; DOBSON, 2007).

A representação gráfica do indicador é feita através de uma linha denominada MACD e uma linha de sinal. A linha MACD é composta pela subtração de duas médias exponenciais (usualmente 12 e 26 dias) e a linha de sinal é composta por uma média móvel exponencial do próprio valor do MACD (9 dias). Na Figura 2.4, as linhas de sinal e MACD estão representadas pelas cores laranja e azul, respectivamente.

Os sinais de venda ocorrem quando a linha MACD cruza de cima para baixo a linha de sinal e os sinais de compra ocorrem quando a linha MACD cruza de baixo para cima a linha de sinal (APPEL; DOBSON, 2007). Os sinais de venda e compra estão indicados na Figura 2.4 pelos números 1 e 2, respectivamente.

O MACD é calculado pela subtração da média móvel exponencial longa na média móvel exponencial curta, conforme mostra a equação (2.3) (DEBASTIANI, 2008).

$$MACD = MME_{curta} - MME_{longa}, \quad (2.3)$$

Figura 2.4 – Representação MACD



Fonte: Adaptado pela autora de (TRADING VIEW).

2.2.3 Índice de força relativa (RSI)

O Índice de Força Relativa (do inglês, *Relative Strength Index* - RSI) mede a velocidade e a magnitude dos movimentos dos preços de determinado ativo. O RSI é medido em uma escala de 0 a 100. O RSI marca sobrecompra quando ultrapassa a região dos 70 pontos e sobrevenda quando está abaixo dos 30 pontos. As regiões de sobrevenda e sobrecompra são sinais de quando o mercado tende a fazer movimentos para correção ou reversão de tendência. A Figura 2.5 ilustra o gráfico do RSI e os momentos de sobrecompra e sobrevenda podem ser identificados pelos números 1 e 2, respectivamente..

O RSI é comumente utilizado em um período de 14 dias e é calculado conforme mostra a equação (2.4) (DEBASTIANI, 2008).

Figura 2.5 – Representação RSI



Fonte: Adaptado pela autora de (TRADING VIEW).

$$RSI = 100 + \frac{100}{1 + RS}, \quad (2.4)$$

$$RS = \frac{\text{Média de N períodos de alta}}{\text{Média de N períodos de baixa}}, \quad (2.5)$$

2.3 Aprendizado de máquina

Nesta Seção abordaremos conceitos de aprendizado de máquina e métodos utilizados em problemas de classificação, com foco nos métodos que serão utilizados neste trabalho: árvore de decisão, rede neural e *XGBoost*.

O aprendizado de máquina é uma subárea da inteligência artificial (IA) que utiliza métodos computacionais para automatizar processos decisórios com base em dados. Os métodos de aprendizado de máquina podem ser classificados em três diferentes tipos, conforme detalhado a seguir (LANGLEY; SIMON, 1995).

- Aprendizado supervisionado: quando os dados de treino têm rótulos conhecidos e o aprendizado é utilizado para prever dados com rótulos desconhecidos;
- Aprendizado não supervisionado: o objetivo do aprendizado não supervisionado é identificar padrões e relações para realizar a classificação de novos dados em agrupamentos;
- Aprendizado semi-supervisionado: utilizado quando há uma grande quantidade de dados de treino, no entanto, apenas alguns são supervisionados.

No aprendizado supervisionado existem basicamente duas categorias de algoritmos: classificação e regressão. A classificação tem como objetivo identificar a qual classe pertence cada amostra e permite a separação dos dados com base em características comuns, sem a necessidade de assumir que todos os indivíduos de um grupo têm o mesmo comportamento. Por outro lado, a regressão tem como objetivo prever um valor numérico, sendo possível extrapolar os resultados das medidas disponíveis para estimar valores de experimentos não realizados (BATISTA; MONARD, 2021).

Técnicas de aprendizado de máquina têm sido exploradas como ferramenta de predição de tendências do mercado de ações em trabalhos como (PICASSO *et al.*, 2019) e (CAROSIA *et al.*, 2021). Nesses estudos, dados históricos são analisados para identificação de padrões que poderão, posteriormente, ser utilizados para prever tendências futuras. Vários algoritmos têm sido utilizados na previsão do preço de ações como redes neurais, máquina de vetores de suporte (SVM), árvores de decisão e *Extreme Gradient Boosting* (XGBoost) (SHAH *et al.*, 2019).

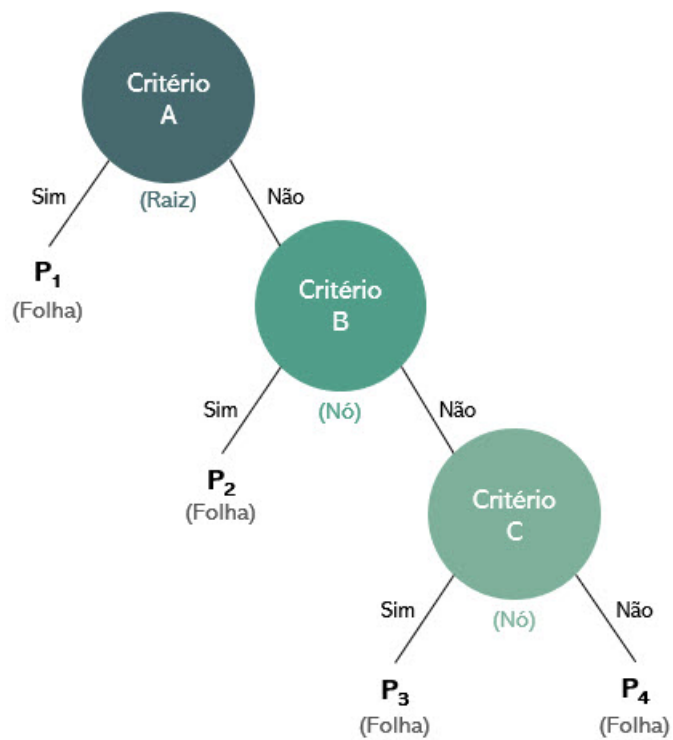
A seguir, apresentamos uma breve descrição desses algoritmos.

2.3.1 Árvore de decisão

A árvore de decisão é um algoritmo de aprendizado de máquina cuja estrutura tem formato de árvore e consiste em um determinado número de nós que representam os

atributos e os ramos representam os resultados. O nó de maior hierarquia é denominado nó raiz. A partir da raiz existem conexões (ramos) para os outros nós. Estes ramos representam as saídas dos testes. Os nós que não apresentam conexões com outros, são chamados de folhas e representam o rótulo de uma classe. A Figura 2.6 apresenta a estrutura de uma árvore de decisão.

Figura 2.6 – Representação da árvore de decisão



Fonte: Elaborado pela autora

A construção de uma árvore de decisão inicia-se com um conjunto de treinamento, que é dividido a partir de condições sobre uma das variáveis independentes, representada pelo nó raiz. Este processo é repetido recursivamente por cada nó subsequente até que seja possível atribuir um único valor para a variável dependente. De maneira geral, para a construção de uma árvore, as seguintes etapas são seguidas:

1. Definição das variáveis que serão utilizadas nos nós (desde a raiz até o último nó);
2. Definição das regras que serão utilizadas para dividir os dados de um determinado nó, com base no valor da variável representada pelo nó;
3. Definição das regras para decidir se um ramo deve ser encerrado, formando uma folha, ou se deve dividir-se em novos nós;
4. Definição do valor predito para a variável em cada folha.

Para a escolha da variável que será utilizada no passo 1, pode-se utilizar o índice de Gini (também chamado de impureza de Gini) ou outras medidas que buscam computar o ganho de informação ou a entropia em cada nó. O nó raiz é definido como o atributo que apresenta o menor índice. Essa medida é um índice de dispersão estatística que mede a heterogeneidade dos dados, ou seja, mede o grau de probabilidade de uma determinada variável ser classificada erroneamente quando escolhida aleatoriamente. Caso todos os elementos de um conjunto de dados pertençam a uma única classe, o conjunto é classificado como puro e o índice de Gini é igual a 0. Um valor igual a 1 representa que os elementos estão randomicamente distribuídos entre mais de uma classe e um valor igual a 0,5 significa que os elementos estão igualmente distribuídos entre as classes. O índice de Gini é definido pela equação (2.6).

$$Gini = 1 - \sum_{i=1}^N (p_i)^2, \quad (2.6)$$

em que p_i representa a probabilidade de um item com rótulo i ser escolhido dentre as amostras.

Durante a construção de árvores de decisão, dificuldades podem ser encontradas, como o *overfitting*, principalmente nos casos em que o número de folhas na árvore é alto. Isso ocorre quando o modelo “memoriza” os dados previamente conhecidos e perde

a habilidade de prever novos valores. Para evitar isso pode-se utilizar técnicas de “poda” da árvore, limitando o número de folhas ou níveis (LAST *et al.*, 2002).

De maneira geral, as árvores de decisão são representações simples do conhecimento e um meio eficiente de construir classificadores que predizem classes baseadas nos valores de atributos de um conjunto de dados (GARCIA, 2003).

2.3.2 Rede neural

As redes neurais artificiais (RNA) consistem em modelos matemáticos baseados em características do cérebro humano, cuja funcionalidade é obtida por meio de aprendizado e generalização. As redes neurais reconhecem e classificam padrões para posteriormente generalizar o conhecimento adquirido (HAYKIN, 2009).

Uma RNA é composta por diversas unidades interconectadas denominadas neurônios. Essas unidades realizam processamentos matemáticos a partir de sinais de entradas, simulando os aspectos de neurônios biológicos. Para uma maior compreensão do funcionamento das RNAs, é pertinente que se faça uma analogia com os elementos biológicos do cérebro humano.

De acordo com (PANDYA; MACY, 1995), o sistema nervoso de um ser humano consiste de três estágios básicos: o estágio inicial, no qual estímulos de entrada externos são captados pelas unidades “sensoras”, o estágio de processamento destes sinais pelo sistema nervoso central (o cérebro) e o estágio no qual as respostas aos estímulos de entrada são obtidas.

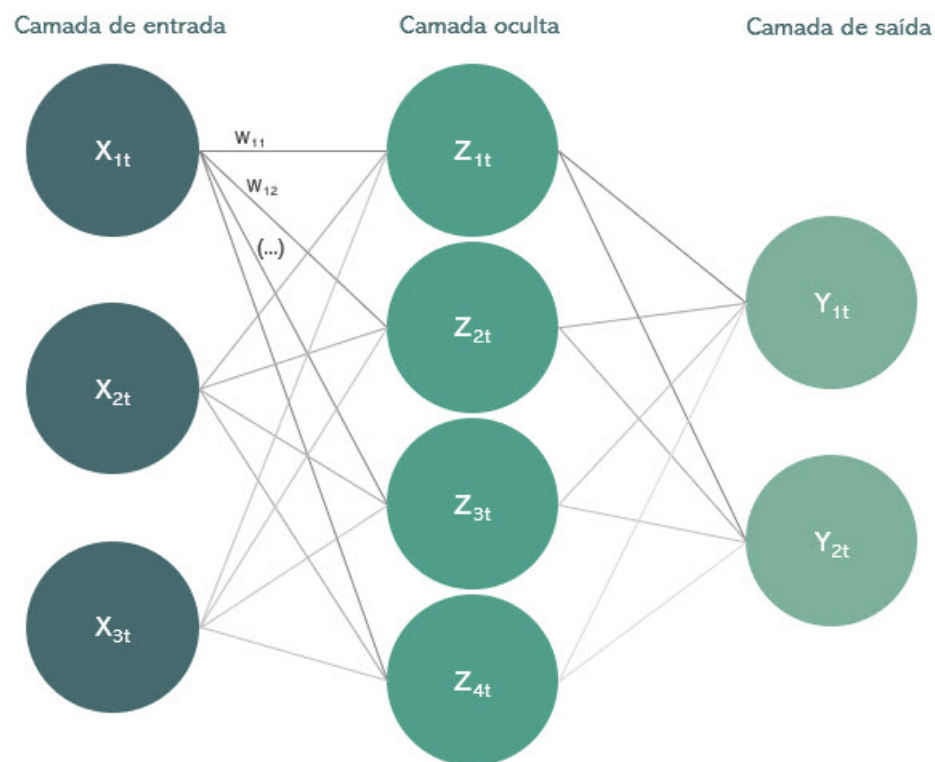
Os neurônios biológicos responsáveis pelo processamento das informações no cérebro humano são representados, nas RNAs, como as unidades básicas de processamento, os “neurônios artificiais”. Estes neurônios são interligados por conexões, assim como as conexões biológicas, denominadas “conexões sinápticas”. Do mesmo modo como ocorre no cérebro humano, todos os dados recebidos pela RNA, são passados entre os neurônios através destas conexões sob a forma de impulsos.

Para representar o aprendizado do cérebro humano, são definidos pesos para as conexões entre os neurônios artificiais e representam o conhecimento armazenado na RNA. Através das etapas de treinamento (análogas ao processo de aprendizado), os pesos são ajustados e tornam a rede capaz de “decidir” com maior ou menor eficiência.

Existem diversos modelos de arquiteturas de RNAs. Vários modelos foram propostos com o intuito de aprimorar os resultados alcançados em determinadas aplicações. Alguns exemplos de arquiteturas são: MLP (*Multilayer Perceptron*), RNN (*Recurrent Neural Network*), LSTM (*Long Short-Term Memory*). Neste trabalho foi utilizado o perceptron multicamadas (MLP).

A rede MLP é composta por 3 partes principais: camada de entrada, camadas ocultas e camada de saída. Os nós de cada camada são interligados por um grande número de conexões. As conexões estão associadas a pesos, os quais armazenam o conhecimento e o generalizam para um conjunto de dados desconhecidos, conforme ilustra a Figura 2.7 (HAYKIN, 2009).

Figura 2.7 – Representação da rede neural MLP



Fonte: Elaborado pela autora

O número de neurônios e camadas de uma rede neural é consequência do problema a ser resolvido. A camada de saída em um problema de classificação representa as classes para as quais os padrões podem ser associados.

O funcionamento das redes neurais MLP ocorrem em dois momentos, um para

frente (*feedforward*) e um para trás (*backpropagation*). Na etapa *feedforward*, a camada de entrada recebe os dados fornecidos à rede, aplica-os à camada oculta ponderando-os através dos pesos sinápticos das ligações entre os neurônios artificiais e aplicando a função de ativação que transfere o resultado para a segunda camada oculta. Este processo é repetido nas seguintes camadas até que a saída seja produzida (BOMBARDELLI, 2021).

Na etapa de *backpropagation* é onde ocorre o aprendizado. Os neurônios na MLP são treinados com algoritmos de aprendizado baseados em retropropagação, mecanismos que permitem com que o modelo ajuste iterativamente os pesos na rede com objetivo de minimizar uma função custo. Em cada iteração, o gradiente do erro quadrático médio é calculado em todos os pares de entrada e saída. Em seguida, os pesos da primeira camada oculta são atualizados com o valor do gradiente e propagados de volta por toda a rede. Esse processo é repetido até que o gradiente para cada par entrada-saída tenha convergido, ou seja, o gradiente atualizado não variou mais do que um limite de convergência pré-definido, em comparação com o valor anterior (BOMBARDELLI, 2021).

2.3.3 *Extreme gradient boosting (XGBoost)*

O *Extreme Gradient Boosting (XGBoost)* é um algoritmo de classificação, o qual é considerado uma melhoria do algoritmo *gradient boosting*. O *XGBoost* utiliza-se do princípio de árvore de decisão com aumento de gradiente. Isto significa que o método utiliza o algoritmo gradiente descendente para minimizar a perda enquanto novos modelos são adicionados.

O modelo é construído em estágios, do mesmo modo que os outros métodos de boosting, e os generaliza permitindo a otimização de uma função de perda diferenciável arbitrária. De acordo com o autor do algoritmo, Tianqi Chen, o que diferencia o *XGBoost* dos demais algoritmos de *gradient boosting* é que ele utiliza uma formalização de modelo mais regularizada para controle do sobreajuste, implicando em um melhor desempenho. O *XGBoost* utiliza a pontuação de similaridade (*similarity score*) e o ganho (*gain*) para determinar as melhores divisões dos nós.

$$Pont.similaridade = \frac{(\sum_{i=1}^N Residual_i)^2}{\sum_{i=1}^N [Prob\ anterior_i \cdot (1 - Prob\ anterior_i)] + \lambda}, \quad (2.7)$$

Em que, o valor residual é a diferença entre o valor real (observado) e o valor previsto, a probabilidade anterior é a probabilidade de um evento calculado em uma etapa anterior. A probabilidade inicial é assumida como 0.5 para cada observação e é utilizada para a construção da primeira árvore. Para as árvores seguintes, a probabilidade é recalculada com base na previsão inicial e nas previsões de todas as árvores seguintes. Por fim, λ é um parâmetro de regularização. O aumento de λ reduz desproporcionalmente a influência das folhas pequenas (com poucas observações), embora tenha um impacto mínimo nas folhas maiores (com muitas observações).

Após o cálculo da pontuação de similaridade em cada folha, é calculado o ganho utilizando a equação (2.8).

$$Ganho = Folha_{esquerda} \text{ similaridade} + Folha_{direita} \text{ similaridade} - Raiz \text{ similaridade}, \quad (2.8)$$

O nó dividido com maior ganho é escolhido como a melhor divisão para a árvore (DOBILAS, 2021).

2.3.4 Métricas de desempenho

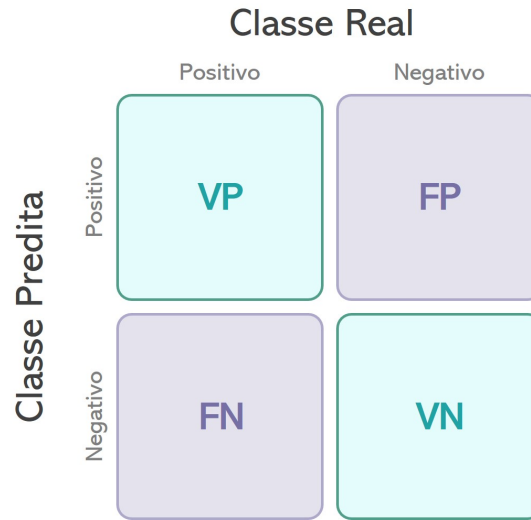
Nesta Seção apresentamos algumas métricas de desempenho que são comumente utilizadas para avaliar modelos de classificação e serão utilizadas para analisar os resultados do modelo proposto neste trabalho no Capítulo 3.

As métricas de desempenho são importantes ferramentas para avaliar a performance dos classificadores. As métricas de classificação são calculadas a partir dos verdadeiros positivos (VP), falsos positivos (FP), verdadeiros negativos (VN) e falsos negativos (FN), que são tabulados na matriz de confusão, conforme apresentado na Figura 2.8 (LEVER *et al.*, 2016).

A relevância de cada um dos 4 fatores depende do propósito do classificador. Para uma ferramenta de apoio para investidores, falsos negativos têm menor criticidade em relação aos falsos positivos, ou seja, é menos prejudicial para o investidor perder uma oportunidade de compra em comparação com a realização de uma compra que acarretará prejuízo. Diante disso, 3 importantes métricas podem ser utilizadas: acurácia, precisão e *recall* (LEVER *et al.*, 2016).

A acurácia indica a fração das predições que estão corretas, conforme apresentado na equação (2.9).

Figura 2.8 – Matriz de confusão



Fonte: Elaborado pela autora

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}, \quad (2.9)$$

Apesar de ser uma métrica muito utilizada, uma acurácia elevada não necessariamente representa um bom classificador. É importante entender como os FPs e FNs se dividem. Uma métrica útil para entender os falsos negativos é o *recall*, que representa a fração de amostras positivas que foram corretamente classificadas, conforme exibido na equação (2.10) (LEVER *et al.*, 2016).

$$\text{Recall} = \frac{VP}{VP + FN}, \quad (2.10)$$

Entretanto, o *recall* não considera os verdadeiros negativos e os falsos positivos. Um classificador que prediz todos os valores como positivos apresenta elevado *recall*. Dessa forma, para complementar a análise do classificador, utiliza-se a precisão, que indica a fração dos valores preditos como positivos que são corretos. A equação (2.11) apresenta o cálculo da precisão (LEVER *et al.*, 2016).

$$\text{Precisão} = \frac{VP}{VP + FP}, \quad (2.11)$$

Com a combinação dessas métricas é possível avaliar o desempenho do classificador.

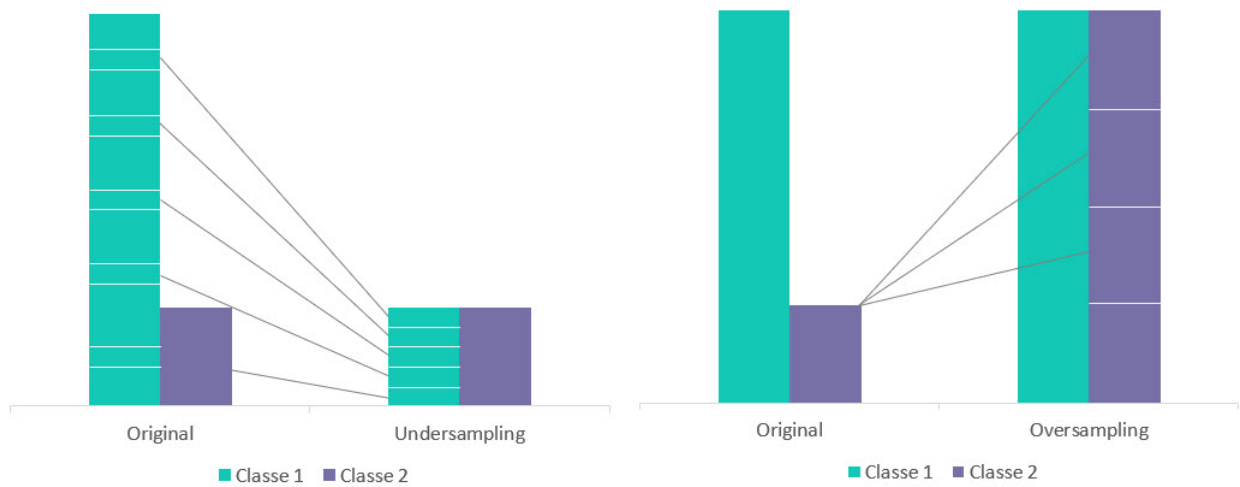
2.3.5 Balanceamento dos dados

Ao trabalhar com problemas de classificação de dados, é possível deparar-se com conjuntos de dados desbalanceados, isto é, uma quantidade significativamente maior de dados de uma classe em relação às demais. A maioria dos algoritmos de AM apresentam dificuldades em criar um modelo que classifique com precisão os dados da classe minoritária. Isso ocorre, pois o processo de aprendizado leva em consideração as classes majoritárias dos dados para generalização do aprendizado e, assim, produz uma classificação que consiga explicar a maior parte dos dados (HASANIN; KHOSHGOFTAAR, 2018), mas não necessariamente os eventos raros.

Para lidar com dados desbalanceados é possível utilizar métodos de *sampling* ou *resampling*. Estes métodos consistem na modificação da estrutura do conjunto de dados desbalanceados, de modo a deixá-lo com quantidades equivalentes de amostras para cada classe. As duas principais formas de balancear os dados consistem na remoção (*undersampling*) ou na adição (*oversampling*) de amostras. Estes métodos são classificados como abordagens ao nível de dados, pois, o processo de balanceamento dos dados ocorre de maneira isolada no conjunto de dados, antes de iniciar o processo de treinamento do classificador.

O processo de *undersampling* consiste na remoção aleatória de amostras da classe majoritária, enquanto o *oversampling* consiste na replicação de amostras da classe minoritária. Ambos os métodos apresentam vantagens e desvantagens. De um lado, a remoção de amostras leva a uma perda no volume dos dados e pode privar o modelo de informações importantes acerca da classe majoritária, reduzindo a capacidade do classificador de discriminar corretamente as classes. Por outro lado, a adição de amostras pode levar o modelo ao *overfitting*, uma vez que o classificador possivelmente criará diversas regras de classificação para uma mesma amostra, tornando-a muito específica.

Figura 2.9 – Métodos de Balanceamento dos Dados



Fonte: Elaborado pela autora

2.4 Gestão de carteiras

Nesta Seção introduzimos alguns conceitos relacionados à gestão de carteiras de renda variável, com objetivo de subsidiar as estratégias de negociação que serão apresentadas no Capítulo 5. Vale ressaltar que a gestão de carteiras, ou de forma mais geral, gestão de portfólio é um tema complexo e multifacetado. Uma referência mais completa sobre esse tema pode ser encontrada em (ELTON *et al.*, 2012).

2.4.1 Investimento

De forma simplificada, um investimento consiste na alocação de capital com a expectativa de um benefício futuro. Em outra perspectiva, um investimento é uma operação que através de sua análise demonstra segurança e um retorno adequado. De acordo com (REILLY; BROWN, 2011), a incerteza dos pagamentos futuros, juntamente com o tempo em que os investidores abdicam do capital investido, é compensada pelo retorno esperado por eles.

2.4.2 Gestão de risco e retorno

Risco e retorno são variáveis fundamentais na tomada de decisão de investimentos e permitem avaliar, de modo quantitativo, se o investimento está de acordo com

as expectativas do investidor. O retorno consiste no valor esperado da distribuição de probabilidade da taxa de rentabilidade do investimento, enquanto o risco é medido pela variância² da distribuição dessa mesma taxa de rentabilidade. Em suma, o investidor racional busca maximizar a rentabilidade dos investimentos (retorno) de forma simultânea com que busca minimizar os riscos dos mesmos.

Entre as técnicas de gestão de risco, pode-se citar duas estratégias comumente utilizadas pelos investidores: *stop loss* e *stop gain*. O *stop loss* é uma estratégia utilizada pelos investidores para gerenciar riscos e limitar perdas de uma determinada operação no mercado financeiro. O *stop loss* baseia-se na ideia que deve existir um limite de perdas para determinado ativo. Atualmente, com a digitalização dos processos de negociação, as plataformas de negociação oferecem a estratégia de *stop loss* através de uma ordem de venda programada para ser disparada automaticamente, caso o valor do ativo atinja a porcentagem de perda determinada pelo investidor.

A ordem de *stop gain* é uma estratégia análoga ao *stop loss*. A diferença entre elas é que o *stop gain* ajuda o investidor a vender um ativo quando o valor atingir o patamar esperado, evitando que uma possível desvalorização prejudique seus lucros. Em suma, caso o ativo atinja o valor definido pelo *stop gain*, uma ordem de venda programada é disparada automaticamente para garantir o ganho esperado pelo investidor.

2.4.3 Teoria de portfólio

A teoria do portfólio foi desenvolvida por Harry Markowitz nos anos 1950, e se tornou o fundamento matemático das estratégias de construção de carteira diversificada. A teoria defende a ideia de que “não se deve colocar todos os ovos no mesmo cesto”, isto é, investir todo o capital em um único ativo pode levar a perdas drásticas. Sendo assim, para maximizar os ganhos, o investidor deve buscar ativos com diferentes relações entre risco e retorno, equilibrando investimento com perspectivas maiores com investimentos mais seguros (menor risco).

A escolha dos ativos onde investir e a repartição do capital pelos vários ativos, pode ser formulada de modo a obter um nível de risco adequado a cada investidor, reconhecendo-se diversas formas de obter as porcentagens adequadas para alocar a cada ativo.

² Medida de dispersão em torno de um valor esperado.

Markowitz (1952, 1959) construiu um modelo de diversificação, com base num algoritmo matemático, segundo o qual é possível ao investidor obter melhores retornos para um determinado nível de risco, ou para um determinado nível de retorno, o mais baixo nível de risco. Sendo possível construir um número infinito de portfólios através da modificação do peso relativo de cada ativo no portfólio, considerando previamente definido quais os ativos onde investir. Os portfólios constituídos com base neste modelo de alocação são denominados de eficientes, constituindo a fronteira de eficiência (MARKOWITZ; FABOZZI, 2002).

2.4.4 Gestão ativa x Gestão passiva

Após a definição do portfólio, estratégias podem ser utilizadas com o objetivo de minimizar os riscos e aumentar o retorno. Existem diversas estratégias que podem ser adotadas pelos investidores. De maneira geral, elas são agrupadas em duas classes: gestão ativa e gestão passiva.

A gestão ativa da carteira consiste na observação frequente do mercado e das informações disponíveis e na utilização de indicadores e estimativas para uma frequente compra e venda de ativos, com base nas novas informações disponíveis. Este tipo de estratégia é compatível com mercados que não são perfeitamente eficientes, em que os investidores podem obter retornos lucrativos.

Por outro lado, a gestão passiva assume que os preços dos ativos refletem todas as informações disponíveis, portanto, não seria possível obter retornos acima da média do mercado. Dentro da gestão passiva, duas estratégias são comumente utilizadas: *buy and hold* e *benchmark* (MARKOWITZ; FABOZZI, 2002).

2.4.5 *Buy and hold*

A estratégia de *buy and hold* consiste em comprar ativos e mantê-los por um determinado período de tempo na carteira visando rendimentos no longo prazo. Essa estratégia é fundamentada, em boa parte, pela HME, uma vez que, se o mercado for eficiente o tempo todo, não existirá um momento certo para investir e obter retornos lucrativos no curto prazo. Dessa forma, investir em uma boa empresa por um considerável período de tempo, oferece bons retornos, desconsiderando a volatilidade do curto prazo.

2.4.6 *Benchmarking*

O *Benchmarking* é conhecido como uma ferramenta estratégica na qual as empresas realizam um estudo da concorrência e das melhores práticas das empresas de um mesmo setor que podem ser replicadas. No mercado financeiro, o *benchmark* adota um índice como referência para avaliar o desempenho do investimento. É uma ferramenta utilizada para comparar os investimentos com alguma métrica do mercado. Considera-se que um investimento tem uma boa rentabilidade quando tem um desempenho acima da média de um determinado setor ou mercado.

3 Modelo Proposto

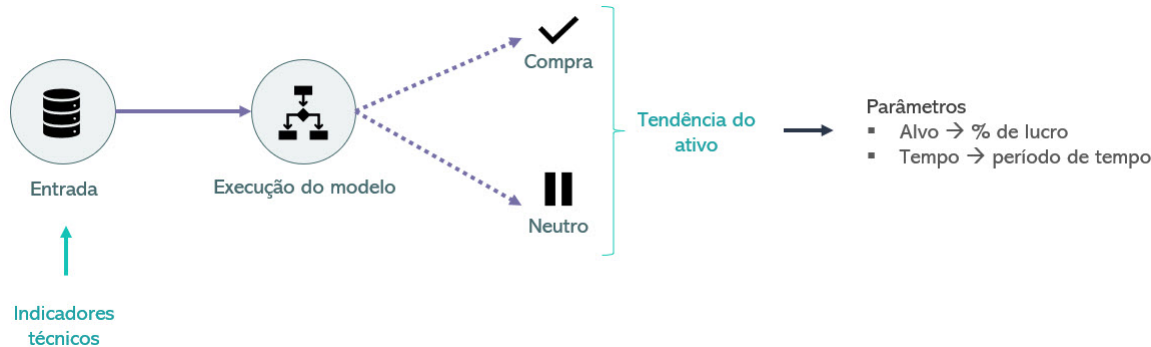
Este trabalho busca avaliar modelos que auxiliem o investidor em suas operações no mercado brasileiro de ações através da previsão da movimentação de cada ativo. O modelo proposto consiste em um classificador binário, que utiliza como parâmetros de entrada os indicadores técnicos e tem como saída a previsão de movimento do ativo. Foram selecionados indicadores técnicos amplamente utilizados por investidores no mercado de ações e que coincidem com os indicadores utilizados em (PICASSO *et al.*, 2019).

O modelo será treinado de modo a sugerir duas possíveis ações ao investidor:

- Compra: sugere ao investidor que, com base nos dados históricos e nos valores atuais dos indicadores, o ativo apresenta um movimento de alta;
- Neutro: sugere ao investidor que, com base nos dados históricos e nos valores atuais dos indicadores, o ativo apresenta um movimento lateral ou de baixa.

É importante ressaltar que as classes “Compra” e “Neutro” estão associadas a dois parâmetros (alvo e tempo) que devem ser definidos pelo investidor. O parâmetro alvo define a porcentagem que o investidor visa realizar de lucro dentro de um determinado período de tempo. Por exemplo, caso o usuário defina o valor alvo como 3% e um período t igual a 5, o modelo retornará a classe “Compra” apenas nos momentos em que os indicadores técnicos demonstrarem uma movimentação de alta superior a 3% no intervalo de 5 períodos, em relação ao valor atual do ativo. A Figura 3.1 apresenta a estrutura do classificador.

Figura 3.1 – Estrutura Classificador Binário



Fonte: Elaborado pela autora

O desenvolvimento do modelo consiste em 6 etapas principais:

1. Coleta de dados: coleta do histórico dos ativos (valor de abertura, valor de fechamento, valor mínimo e valor máximo) considerando a periodicidade de interesse. Estes dados podem ser obtidos facilmente em plataformas de negociação ou através de APIs¹. Os dados apresentados a seguir foram obtidos na plataforma Yahoo Finance;

Figura 3.2 – Coleta de dados

Data	Abertura	Máxima	Mínima	Fechamento
01/03/2010	R\$ 4.52	R\$ 4.64	R\$ 4.43	R\$ 4.62
02/03/2010	R\$ 4.66	R\$ 4.82	R\$ 4.57	R\$ 4.80
03/03/2010	R\$ 4.80	R\$ 4.81	R\$ 4.63	R\$ 4.63
04/03/2010	R\$ 4.63	R\$ 4.74	R\$ 4.60	R\$ 4.68
05/03/2010	R\$ 4.74	R\$ 4.77	R\$ 4.62	R\$ 4.71
08/03/2010	R\$ 4.70	R\$ 4.84	R\$ 4.61	R\$ 4.82
09/03/2010	R\$ 4.81	R\$ 4.91	R\$ 4.81	R\$ 4.82
10/03/2010	R\$ 4.86	R\$ 4.90	R\$ 4.78	R\$ 4.80

Periodicidade:
Parâmetro definido pelo investidor
(Ex. diária)

Fonte: Elaborado pela autora

2. Cálculo dos indicadores técnicos: cálculo dos indicadores técnicos apresentados na Seção 2.2 para a base de dados coletada (considerando a definição dos intervalos

¹ Uma API - do inglês, *Application Programming Interface*, é uma interface de programação que possibilita a comunicação entre plataformas através de uma série de padrões e protocolos.

necessários para cada indicador);

Figura 3.3 – Cálculo dos indicadores

Data	Abertura	Máxima	Mínima	Fechamento	Média Móvel	RSI	MACD
01/03/2010	R\$ 4.52	R\$ 4.64	R\$ 4.43	R\$ 4.62	I ₁₁	I ₂₁	I ₃₁
02/03/2010	R\$ 4.66	R\$ 4.82	R\$ 4.57	R\$ 4.80	I ₁₂	I ₂₂	I ₃₂
03/03/2010	R\$ 4.80	R\$ 4.81	R\$ 4.63	R\$ 4.63	I ₁₃	I ₂₃	I ₃₃
04/03/2010	R\$ 4.63	R\$ 4.74	R\$ 4.60	R\$ 4.68	I ₁₄	I ₂₄	I ₃₄
05/03/2010	R\$ 4.74	R\$ 4.77	R\$ 4.62	R\$ 4.71	I ₁₅	I ₂₅	I ₃₅
08/03/2010	R\$ 4.70	R\$ 4.84	R\$ 4.61	R\$ 4.82	I ₁₆	I ₂₆	I ₃₆
09/03/2010	R\$ 4.81	R\$ 4.91	R\$ 4.81	R\$ 4.82	I ₁₇	I ₂₇	I ₃₇
10/03/2010	R\$ 4.86	R\$ 4.90	R\$ 4.78	R\$ 4.80	I ₁₈	I ₂₈	I ₃₈

Intervalos para cálculo dos indicadores:
Parâmetro definido pelo investidor
(Ex. período curto e longo média móvel)

Fonte: Elaborado pela autora

- Definição dos parâmetros: escolha do valor alvo (% variação) que será utilizado como parâmetro de classificação, e do período t para atingimento do alvo (o período t está associado à periodicidade dos dados coletados no item 1, ou seja, se a frequência dos dados coletados for diária, o período t corresponde a t dias). A Figura 3.4 detalha o processo de construção da base de dados de acordo com os parâmetros definidos. Inicialmente, desloca-se a variação t períodos (no exemplo, $t = 1$ dia). Em seguida, classifica-se o alvo de acordo com o parâmetro alvo definido. Caso a variação seja superior ao valor definido, a amostra é classificada como 1. Caso contrário, a classificação é definida como 0;

Figura 3.4 – Definição dos parâmetros

Data	Abertura	Máxima	Mínima	Fechamento	Média Móvel	RSI	MACD	Variação (Fechamento)	Alvo	Classe
01/03/2010	R\$ 4.52	R\$ 4.64	R\$ 4.43	R\$ 4.62	I ₁₁	I ₂₁	I ₃₁	3.7%	3.8%	1
02/03/2010	R\$ 4.66	R\$ 4.82	R\$ 4.57	R\$ 4.80	I ₁₂	I ₂₂	I ₃₂	3.8%	-3.6%	0
03/03/2010	R\$ 4.80	R\$ 4.81	R\$ 4.63	R\$ 4.63	I ₁₃	I ₂₃	I ₃₃	-3.6%	1.0%	0
04/03/2010	R\$ 4.63	R\$ 4.74	R\$ 4.60	R\$ 4.68	I ₁₄	I ₂₄	I ₃₄	1.0%	0.7%	0
05/03/2010	R\$ 4.74	R\$ 4.77	R\$ 4.62	R\$ 4.71	I ₁₅	I ₂₅	I ₃₅	0.7%	2.4%	1
08/03/2010	R\$ 4.70	R\$ 4.84	R\$ 4.61	R\$ 4.82	I ₁₆	I ₂₆	I ₃₆	2.4%	-0.1%	0
09/03/2010	R\$ 4.81	R\$ 4.91	R\$ 4.81	R\$ 4.82	I ₁₇	I ₂₇	I ₃₇	-0.1%	-0.4%	0
10/03/2010	R\$ 4.86	R\$ 4.90	R\$ 4.78	R\$ 4.80	I ₁₈	I ₂₈	I ₃₈	-0.4%	?	?

Parâmetros:
Período = 1
Alvo = 2%

Fonte: Elaborado pela autora

- Treino do modelo: após a criação da base de dados composta pelos indicadores técnicos, define-se uma porcentagem de divisão do *dataset* e uma parte dos dados é

utilizada para treino no modelo. O treino consiste em um aprendizado supervisionado em que o rótulo de cada amostra corresponde à classificação em “Compra” ou “Neutro” com base na variação do ativo no período t ;

5. Teste do modelo: após o treino, utiliza-se a outra parte do *dataset* para teste. O modelo classifica cada uma das amostras nas classes “Compra” ou “Neutro”, com base nos indicadores técnicos calculados;

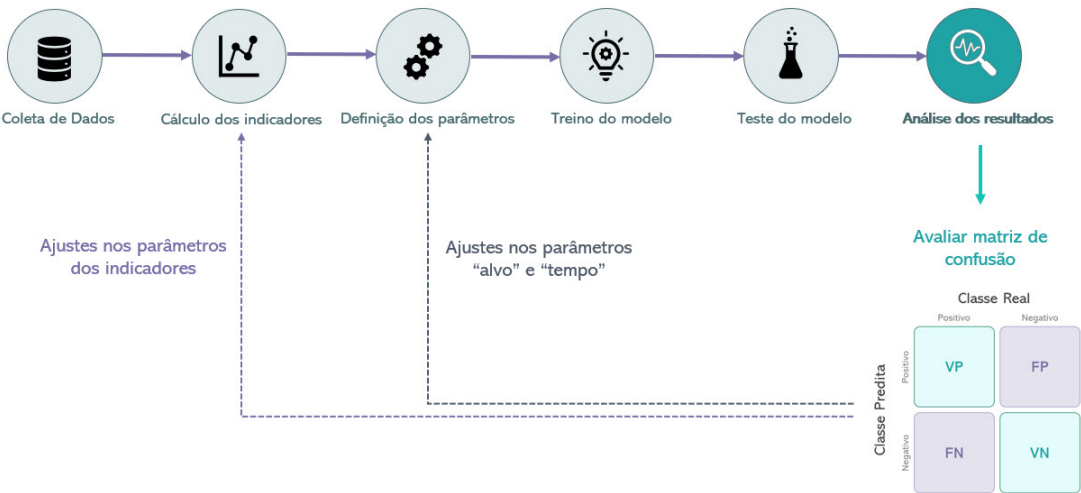
Figura 3.5 – Treino e teste do modelo

Data	Abertura	Máxima	Mínima	Fechamento	Média Móvel	RSI	MACD	Variação (Fechamento)	Alvo	Classe
01/03/2010	R\$ 4.52	R\$ 4.64	R\$ 4.43	R\$ 4.62	I_{11}	I_{21}	I_{31}	3.7%	3.8%	1
02/03/2010	R\$ 4.66	R\$ 4.82	R\$ 4.57	R\$ 4.80	I_{12}	I_{22}	I_{32}	3.8%	-3.6%	0
03/03/2010	R\$ 4.80	R\$ 4.81	R\$ 4.63	R\$ 4.63	I_{13}	I_{23}	I_{33}	-3.6%	1.0%	0
04/03/2010	R\$ 4.63	R\$ 4.74	R\$ 4.60	R\$ 4.68	I_{14}	I_{24}	I_{34}	1.0%	0.7%	0
05/03/2010	R\$ 4.74	R\$ 4.77	R\$ 4.62	R\$ 4.71	I_{15}	I_{25}	I_{35}	0.7%	2.4%	1
08/03/2010	R\$ 4.70	R\$ 4.84	R\$ 4.61	R\$ 4.82	I_{16}	I_{26}	I_{36}	2.4%	-0.1%	0
09/03/2010	R\$ 4.81	R\$ 4.91	R\$ 4.81	R\$ 4.82	I_{17}	I_{27}	I_{37}	-0.1%	-0.4%	0
10/03/2010	R\$ 4.86	R\$ 4.90	R\$ 4.78	R\$ 4.80	I_{18}	I_{28}	I_{38}	-0.4%	?	?

Fonte: Elaborado pela autora

6. Análise dos resultados: por fim, para validação do modelo são utilizadas as métricas de validação apresentadas no item 2.4.

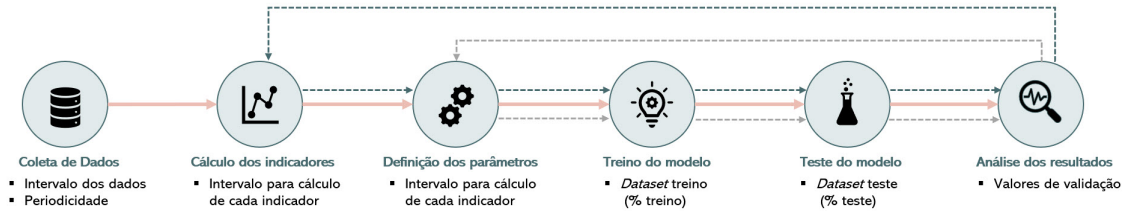
Figura 3.6 – Análise dos resultados



Fonte: Elaborado pela autora

As etapas descritas acima estão sintetizadas no Fluxograma da Figura 3.7

Figura 3.7 – Fluxograma das etapas de desenvolvimento do modelo



Fonte: Elaborado pela autora

Caso os resultados obtidos no teste do modelo sejam satisfatórios, o usuário pode aplicar o modelo para predição dos movimentos futuros utilizando como base de dados os valores atuais dos ativos de interesse e definindo o alvo e o período que deseja analisar. Após a execução, o usuário terá como resposta a classe correspondente ao ativo analisado, isto é, “Compra” ou “Neutro”.

Com o apoio da classificação obtida pelo modelo, o usuário poderá definir algumas estratégias de negociação para investir de acordo com as respostas dadas pelo classificador.

4 Experimentos computacionais e resultados empíricos

Neste Capítulo apresentamos resultados de experimentos computacionais realizados com base em dados reais extraídos da Bolsa de Valores de São Paulo (B3). Os experimentos têm como objetivo investigar a viabilidade empírica do modelo de classificação supervisionado proposto no Capítulo anterior e comparar as métricas de desempenho dos modelos baseados em árvore de decisão, rede neural e *XGBoost*.

Para a implementação dos modelos de classificação, inicialmente, foi selecionado um conjunto de papéis de alta volatilidade, listados na B3, de diferentes setores. Para cada ativo selecionado, coletaram-se os dados referentes a abertura, fechamento, mínima e máxima do período entre 01 de janeiro de 2015 e 31 de dezembro de 2019, com uma periodicidade diária. Posteriormente, foram calculados os indicadores técnicos para cada amostra. Os indicadores utilizados foram média móvel simples e exponencial, MACD e RSI. Estes indicadores foram adotados por serem comumente utilizados por analistas técnicas e por apresentarem bons resultados em trabalhos como (PICASSO *et al.*, 2019).

Em seguida, os modelos de árvore de decisão, rede neural e *XGBoost* foram implementados em *python*, utilizando as bibliotecas *scikit learn* e *xgboost*, que apresentam uma série de ferramentas simples e eficientes comumente utilizadas nas mineração e análise de dados. Os parâmetros e hiperparâmetros adotados estão apresentados na Tabela 4.1. Para os parâmetros não presentes na tabela, foram adotados os valores padrões de cada modelo.

Os modelos foram executados em um computador com um processador Intel Core i9 2.40GHz e 64GB de memória RAM, sob plataforma Windows 11. O tempo de execução dos modelos não é objeto de trabalho deste estudo, porém, para fins de registro foram coletados. O tempo de execução para coleta e tratamento da base de dados (5 anos, com periodicidade diária, para 7 ativos) foi de 3.7s e a execução dos classificadores, incluindo treino e teste, foi de 26.8s.

Tabela 4.1 – Parâmetros e hiperparâmetros dos Classificadores

Modelo	Descrição	Valor
Árvore de Decisão	Observações	2409
	Treinamento	1806
	Teste	603
	Profundidade máxima	5
Rede neural MLP	Observações	2409
	Treinamento	1806
	Teste	603
	Nº de camadas ocultas	3
	Nº de neurônios ocultos	(100,50,25)
	Nº máx. de iterações	2000
XGBoost	Observações	2409
	Treinamento	1806
	Teste	603
	Profundidade máxima	6

Fonte: Elaborado pela autora

A seguir estão detalhados os resultados obtidos com o ativo LREN3 (Lojas Renner), empresa do setor de consumo e varejo.

Os seguintes parâmetros foram utilizados:

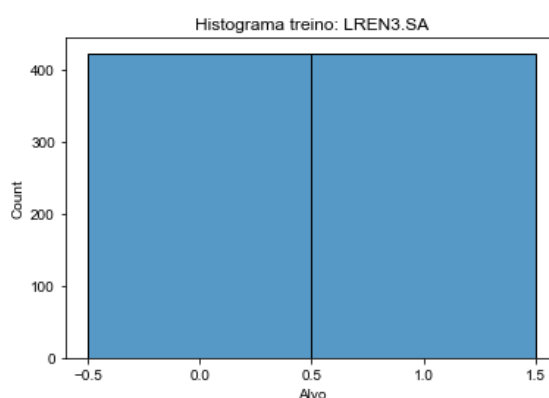
- Alvo: 2% $\rightarrow$ meta da variação do preço do ativo (o modelo indica compra caso o ativo apresente um retorno esperado maior ou igual a 2%);
- Período: 5 dias $\rightarrow$ prazo para atingir o valor alvo buscado;
- % de divisão dos dados: 75% $\rightarrow$ divisão dos dados em treino e teste (75% para treino e 25% para teste).

Antes da execução dos classificadores, os dados foram avaliados para definição de qual método de balanceamento seria utilizado. Para isso, foram realizados testes com 08 ativos, alternando o método de balanceamento. Avaliou-se a precisão e o *recall* dos classificadores utilizando *undersampling* e *oversampling* para balanceamento dos dados de treino. Para todos os papéis avaliados, as métricas dos classificadores foram superiores ao utilizar o *undersampling*. Diante disso, este foi o método adotado para as análises

apresentadas a seguir. A Figura 4.1 apresenta o histograma de distribuição dos dados de treino após o balanceamento dos dados.

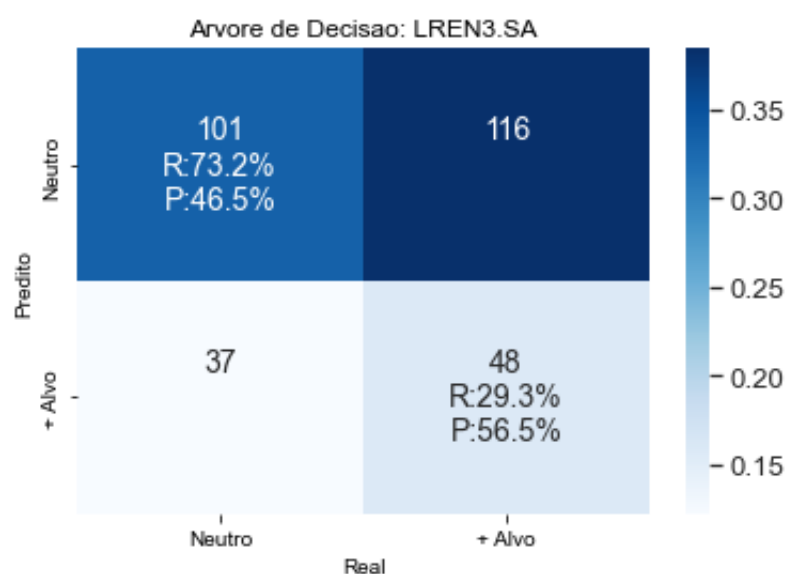
Após o pré-processamento dos dados, os métodos de classificação foram executados e as matrizes de confusão calculadas, conforme exibido nas Figuras 4.2, 4.3 e 4.4.

Figura 4.1 – Histograma Dados de Treino - LREN3



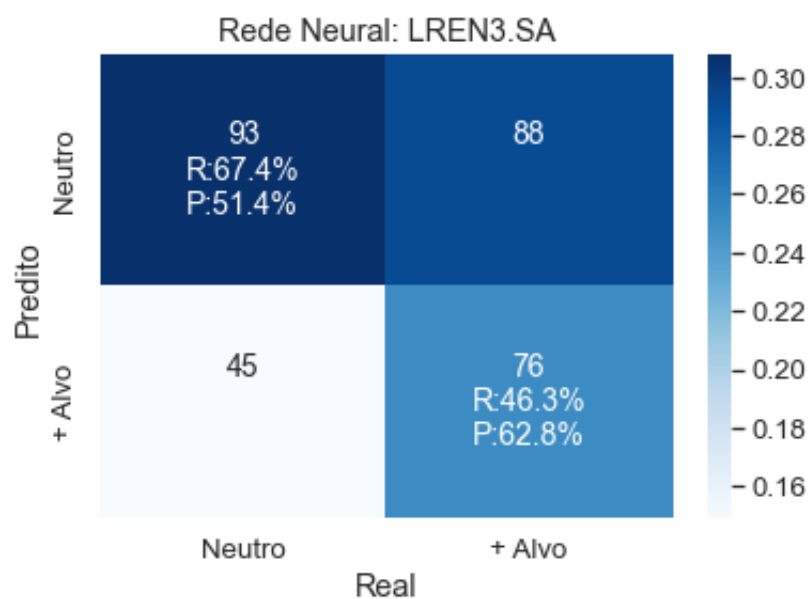
Fonte: Elaborado pela autora (*python*)

Figura 4.2 – Matriz de confusão - Árvore de Decisão - LREN3

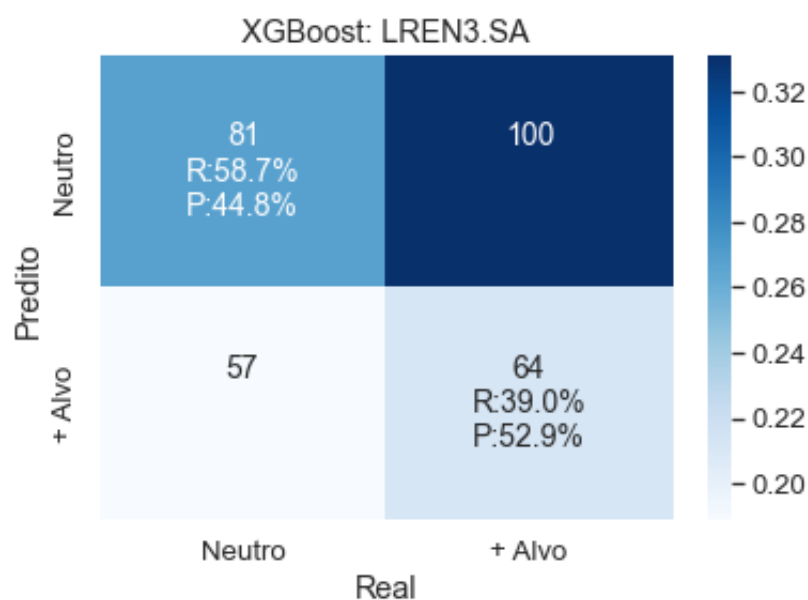


Fonte: Elaborado pela autora (*python*)

Figura 4.3 – Matriz de confusão - Rede Neural - LREN3



Fonte: Elaborado pela autora (*python*)

Figura 4.4 – Matriz de confusão - *XGBoost* - LREN3

Fonte: Elaborado pela autora (*python*)

A Tabela 4.2 apresenta os resultados obtidos para os três métodos e as principais métricas avaliadas.

Tabela 4.2 – Métricas de desempenho

Métrica	Árvore de Decisão	Rede Neural MLP	<i>XGBoost</i>
Acurácia	49.0%	56.0%	48.0%
Precisão (+ Alvo)	56.5%	62.8%	52.9%
<i>Recall</i> (+ Alvo)	29.3%	46.3%	39.0%
Precisão (Neutro)	46.5%	51.4%	44.8%
<i>Recall</i> (Neutro)	73.2%	67.4%	58.7%

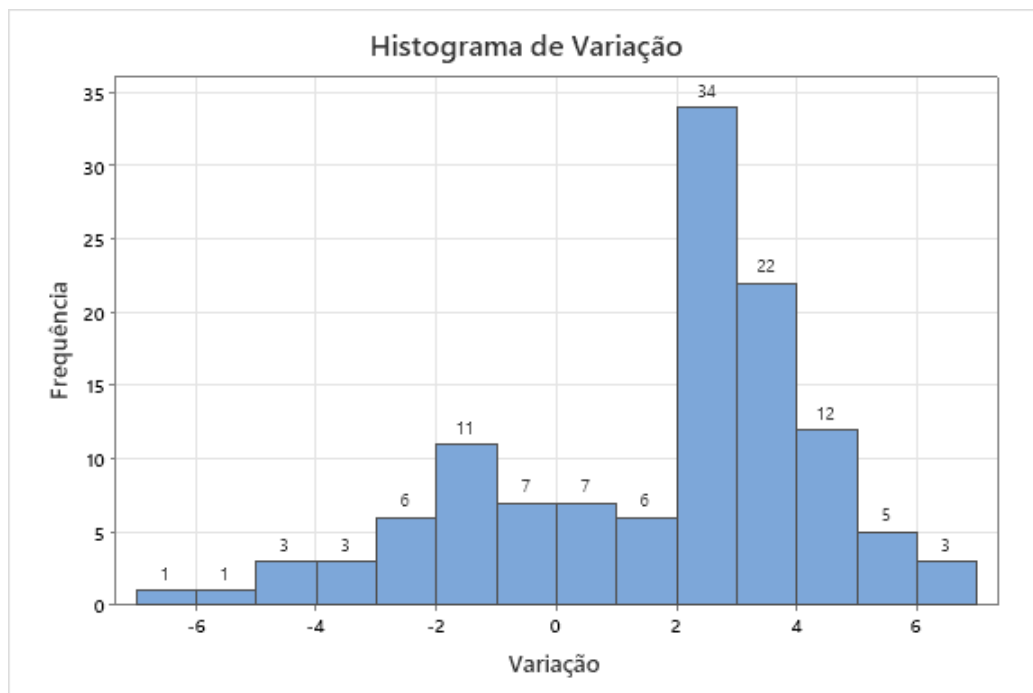
Fonte: Elaborado pela autora

Ao analisar os resultados apresentados na Tabela 4.2, observa-se que a rede neural apresentou a maior acurácia, com 56% dos dados classificados corretamente. Entretanto, conforme explicado na Seção 2.3.4, para o caso de métodos de classificação de ativos, busca-se reduzir a quantidade de falsos negativos, ou seja, reduzir a quantidade de vezes em que o modelo sugere a compra de um determinado ativo que não resulta em alta no preço. Esse comportamento é explicado pela precisão (+ Alvo).

De acordo com os resultados obtidos, a rede neural foi o modelo que apresentou maior precisão ao indicar a compra de um ativo. Das 121 vezes que o modelo sugeriu a compra, 76 vezes resultaram em uma variação maior ou igual ao alvo buscado, representando aproximadamente 63% de acerto, conforme mostra a Figura 4.3.

É importante ressaltar que ao utilizar um alvo de 2%, variações abaixo deste valor não são classificadas como “Compra”, porém, não indicam necessariamente prejuízo em caso de investimento. Ou seja, das 45 classificações incorretas do método (falsos positivos), existem variações no intervalo entre 0 e 2% e, portanto, apenas variações inferiores a 0% resultam em prejuízo. Para identificar esse comportamento de forma mais detalhada foi gerado um histograma com a variação de todas as amostras que foram classificadas como “Compra”, conforme mostra a Figura 4.5.

Figura 4.5 – Histograma de variação Rede neural MLP - Classificação + Alvo



Fonte: Elaborado pela autora (*python*)

Ao analisar os resultados do histograma, verifica-se que das 45 classificações incorretas realizadas pelo método, 13 amostras estão no intervalo entre 0 e 2%, ou seja, não refletem em prejuízo caso o investidor siga a recomendação do modelo e compre o ativo. Diante disso, apenas 32 amostras classificadas como “+ Alvo” resultaram em prejuízo. Por outro lado, caso essas 13 amostras positivas sejam consideradas como “acertos” do modelo, a precisão da rede neural passa de 63% para aproximadamente 74%.

Outro *insight* importante obtido através da análise do histograma da Figura 4.5 diz respeito aos retornos obtidos para as amostras classificadas como “Compra”. Apesar de o alvo de variação ser de 2%, verificam-se retornos superiores no período de 5 dias. Sendo que 76 amostras apresentaram retorno superior a 2%, podendo chegar a 7% de variação.

Após a análise dos resultados do ativo LREN3, um conjunto de papéis de alta volatilidade, listados na B3, de diferentes setores foi selecionado para avaliação dos modelos de classificação propostos e comparação dos resultados, conforme apresentado na Tabela 4.3.

Tabela 4.3 – Papéis avaliados

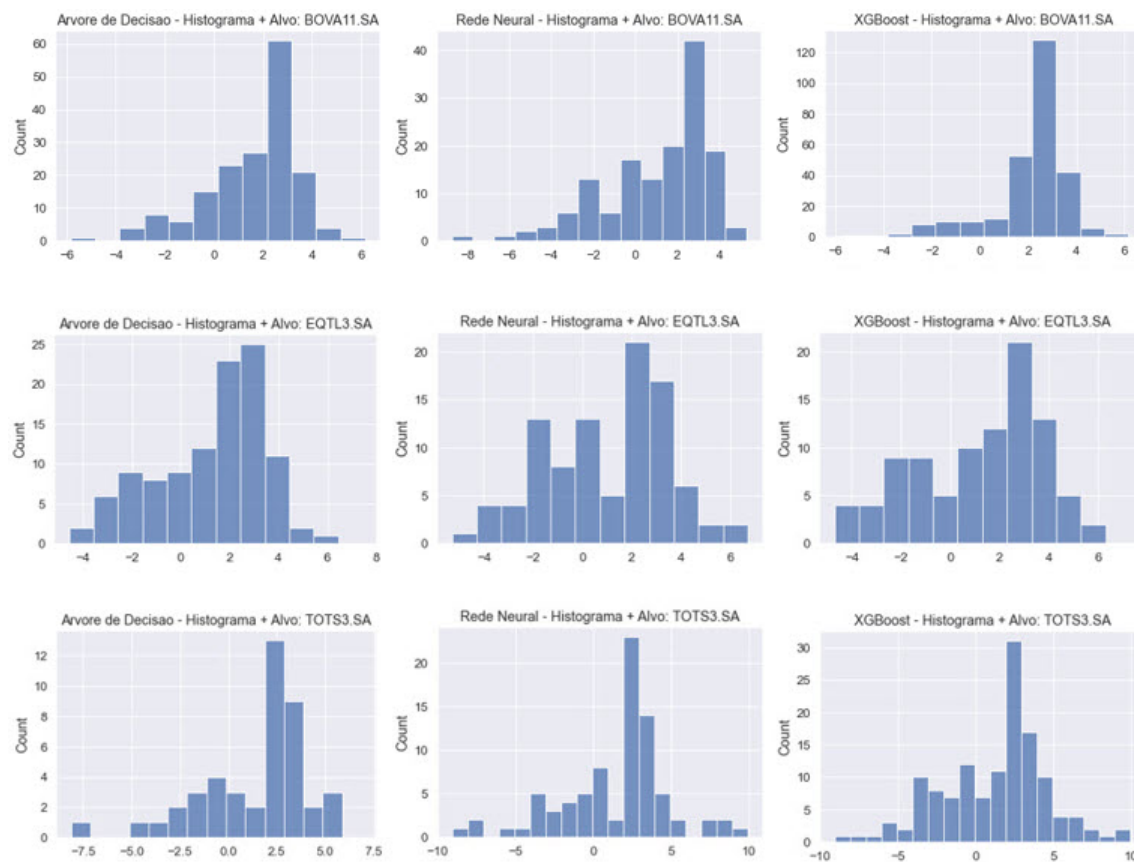
Ativo	Empresa/Índice	Setor
ABEV3	Ambev	Consumo e Varejo
BOVA11	Ishares Ibovespa	ETF (Ref. BOVESPA)
EQTL3	Equatorial Energia	Energia e Saneamento
ITSA4	Itaúsa	Holding
PETR4	Petrobras	Petróleo E Gás
TOTS3	Totvs	Tecnologia e Internet

Fonte: Elaborado pela autora

De forma análoga à análise do ativo LREN3, foram gerados histogramas que apresentam a porcentagem de variação para as classificações “+ Alvo” realizadas pelos 3 modelos: árvore de decisão, rede neural MLP e *XGBoost*, conforme mostram as Figuras 4.6 e 4.7.

Através da análise dos histogramas, pode-se perceber que, para todos os ativos, existe uma concentração de variação em torno de 2%, parâmetro adotado como “alvo” neste trabalho. Conforme explicado anteriormente, retornos entre 0% e 2% são considerados erros de classificação segundo as métricas de avaliação de desempenho dos modelos, porém, na perspectiva de investimentos podem ser considerados resultados positivos para o investidor.

Figura 4.6 – Histograma de variação ativos - Classificação + Alvo



Fonte: Elaborado pela autora (*python*)

A Tabela 4.4 resume os resultados obtidos pelos classificadores.

Figura 4.7 – Histograma de variação ativos - Classificação + Alvo

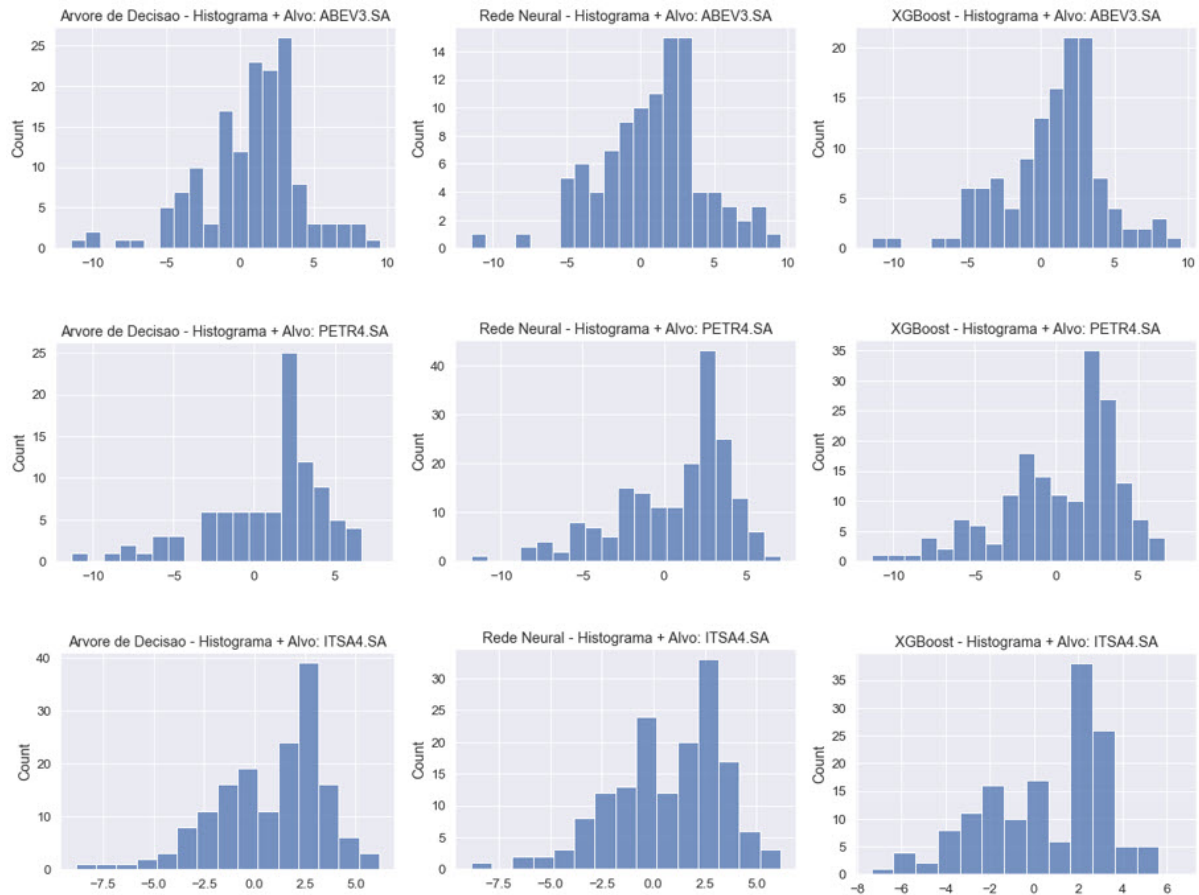
Fonte: Elaborado pela autora (*python*)

Tabela 4.4 – Métricas de desempenho dos classificadores

Ativo	Métrica	Árvore de Decisão	Rede neural MLP	XGBoost
ABEV3	Precisão (+Alvo)	40.6%	39.4%	43.5%
	Recall (+Alvo)	53.3%	45.9%	44.3%
BOVA11	Precisão (+Alvo)	45.2%	53.4%	80.0%
	Recall (+Alvo)	60.2%	34.2%	96.5%
EQTL3	Precisão (+Alvo)	54.6%	47.9%	53.2%
	Recall (+Alvo)	41.8%	32.6%	35.5%
ITSA4	Precisão (+Alvo)	50.9%	45.5%	48.3%
	Recall (+Alvo)	57.7%	50.0%	50.7%
PETR4	Precisão (+Alvo)	55.1%	52.2%	48.0%
	Recall (+Alvo)	36.2%	63.1%	55.0%
TOTS3	Precisão (+Alvo)	61.4%	60.5%	53.0%
	Recall (+Alvo)	15.9%	28.8%	41.8%

Fonte: Elaborado pela autora

Ao avaliar os resultados dos modelos para os ativos em questão, verificou-se que a precisão média dos classificadores foi de 52% e o *recall* médio foi 47%. Estes resultados nos motivaram a testar um novo método de classificação: um modelo aleatório.

Este modelo funciona da seguinte forma: para cada amostra de teste é definido um rótulo aleatório. Este rótulo pode ser 0 ou 1, representando as classes "Neutro" e "Compra", respectivamente.

As probabilidades de cada classe são definidas com base nos dados históricos, ou seja, a probabilidade de uma amostra ser rotulada com "Compra" é dada pela proporção de rótulos "Compra" nos dados históricos. Por exemplo, em um histórico de 100 amostras de um ativo (100 dias), caso em 20 dias a variação seja superior ao "Alvo" definido, considera-se uma probabilidade de 20% para este ativo. Sendo assim, o classificador aleatório irá rotular os dados de teste com uma probabilidade de 20% para "Compra" e 80% para "Neutro".

Este classificador foi testado para os mesmos ativos definidos na Tabela 5.1 e os resultados obtidos estão apresentados na Tabela 4.5.

Tabela 4.5 – Métricas de desempenho dos classificadores x classificador aleatório

Ativo	Métrica	Árvore de Decisão	Rede neural MLP	<i>XGBoost</i>	Aleatório
ABEV3	Precisão (+Alvo)	40.6%	39.4%	43.5%	27.0%
	Recall (+Alvo)	53.3%	45.9%	44.3%	34.8%
BOVA11	Precisão (+Alvo)	45.2%	53.4%	80.0%	25.4%
	Recall (+Alvo)	60.2%	34.2%	96.5%	27.3%
EQTL3	Precisão (+Alvo)	54.6%	47.9%	53.2%	25.9%
	Recall (+Alvo)	41.8%	32.6%	35.5%	30.9%
ITSA4	Precisão (+Alvo)	50.9%	45.5%	48.3%	42.1%
	Recall (+Alvo)	57.7%	50.0%	50.7%	39.8%
PETR4	Precisão (+Alvo)	55.1%	52.2%	48.0%	43.0%
	Recall (+Alvo)	36.2%	63.1%	55.0%	38.0%
TOTS3	Precisão (+Alvo)	61.4%	60.5%	53.0%	34.6%
	Recall (+Alvo)	15.9%	28.8%	41.8%	43.9%

Fonte: Elaborado pela autora

Os resultados da Tabela 4.5 mostram que o classificador aleatório apresentou precisão média de 33.9% e o *recall* médio de 36.3%. Sendo assim, o classificador aleatório não superou os modelos propostos (árvore de decisão, rede neural MLP e *XGBoost*) nas métricas em questão para os ativos testados neste trabalho.

Diante destes resultados, a Seção a seguir tem como objetivo explorar possibilidades para aumentar os retornos do investidor unindo estratégias de negociação aos resultados dos modelos de classificação apresentados nesta Seção.

5 Resultados: Estratégias de negociação

Neste Capítulo apresentamos a segunda contribuição original desta dissertação, que consiste em combinar os resultados obtidos com o modelo de classificação - apresentados nos Capítulos precedentes - com estratégias de negociação de ativos. Essa combinação mostra que é possível compensar um baixo desempenho do classificador adotando estratégias adequadas.

A aplicação prática no mercado financeiro do modelo apresentado na Seção 3 pode ser feita com base na classificação dos ativos. O investidor define os papéis que deseja avaliar, os parâmetros alvo e período e executa o modelo utilizando os dados de fechamento atuais.

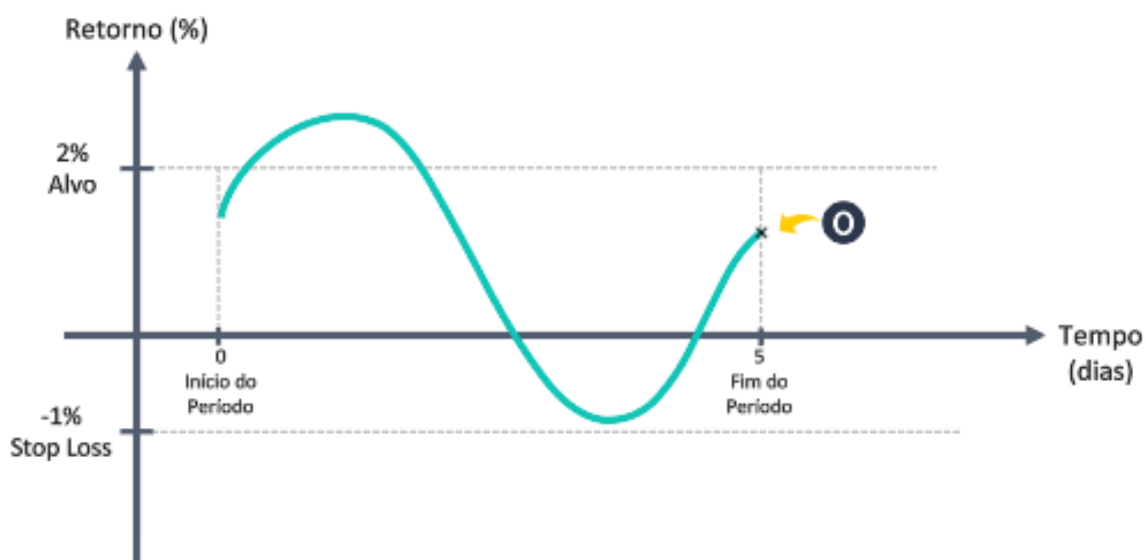
Após a execução do modelo, todos os ativos que apresentarem o rótulo “Compra” (+ Alvo) indicam uma movimentação de alta. Conforme definido na Seção 2, a precisão indica a fração dos valores preditos como positivos que são corretos. Diante disso, precisões abaixo de 100% refletem classificações incorretas realizadas pelo modelo. Com o intuito de atenuar as perdas para o investidor nos casos de classificação incorreta, foram propostas 5 estratégias de negociação.

Antes de detalhar cada uma das estratégias de negociação, é importante ressaltar que os conceitos de *stop loss* e *stop gain* apresentados na Seção 2 serão utilizados. A seguir estão descritas as cinco estratégias propostas neste trabalho. Para exemplificar as estratégias, será considerado um período igual a 5 dias, um alvo igual a 2% e *stop loss* de 1%, sobre o valor de compra.

1. **Estratégia 0:** Saída no final do período - sem *stop loss*

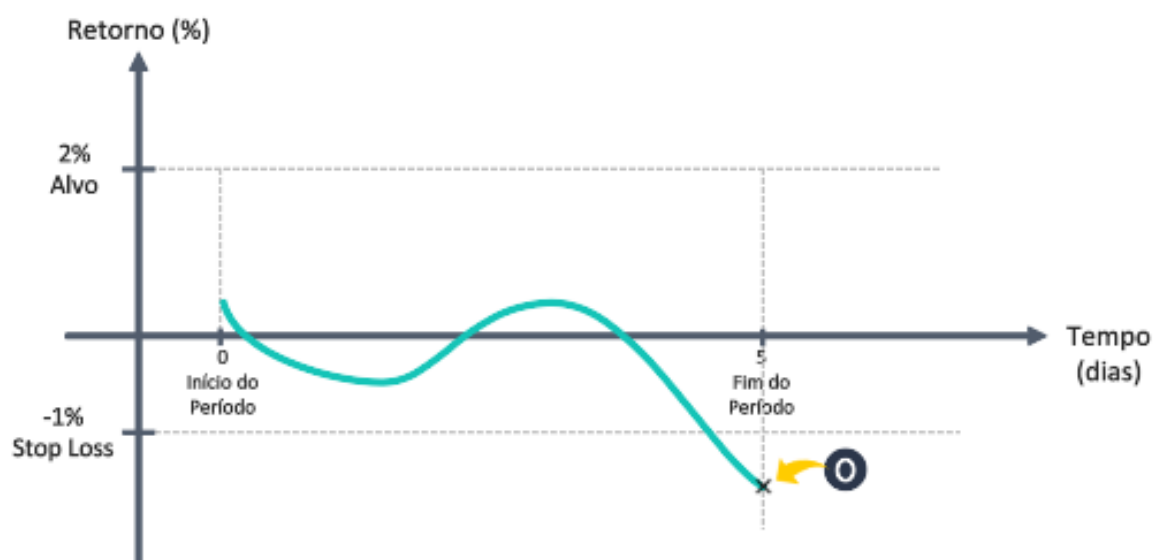
Nessa estratégia o investidor compra o ativo caso o rótulo seja “Compra” e vende após o final do período, ou seja, após o fechamento do 5º dia. Na estratégia 0 existe apenas uma saída possível e o retorno é calculado com base na variação no preço do ativo no dia da compra e no dia da venda. As Figuras 5.1 e 5.2 apresentam dois exemplos de retorno com a estratégia 0.

Figura 5.1 – Estratégia 0 - Saída ao final do período 1



Fonte: Elaborado pela autora

Figura 5.2 – Estratégia 0 - Saída ao final do período 2

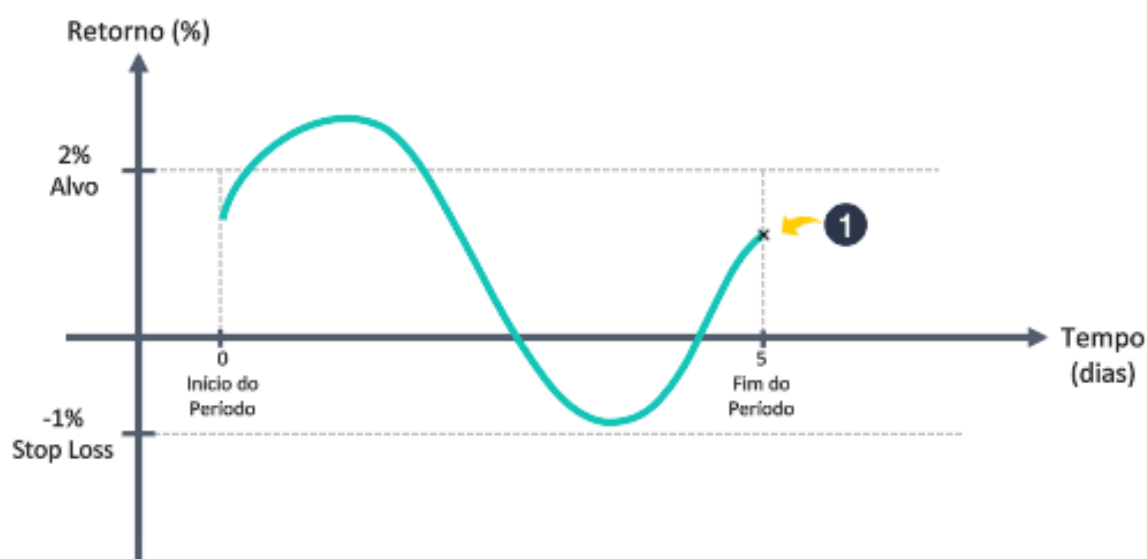


Fonte: Elaborado pela autora

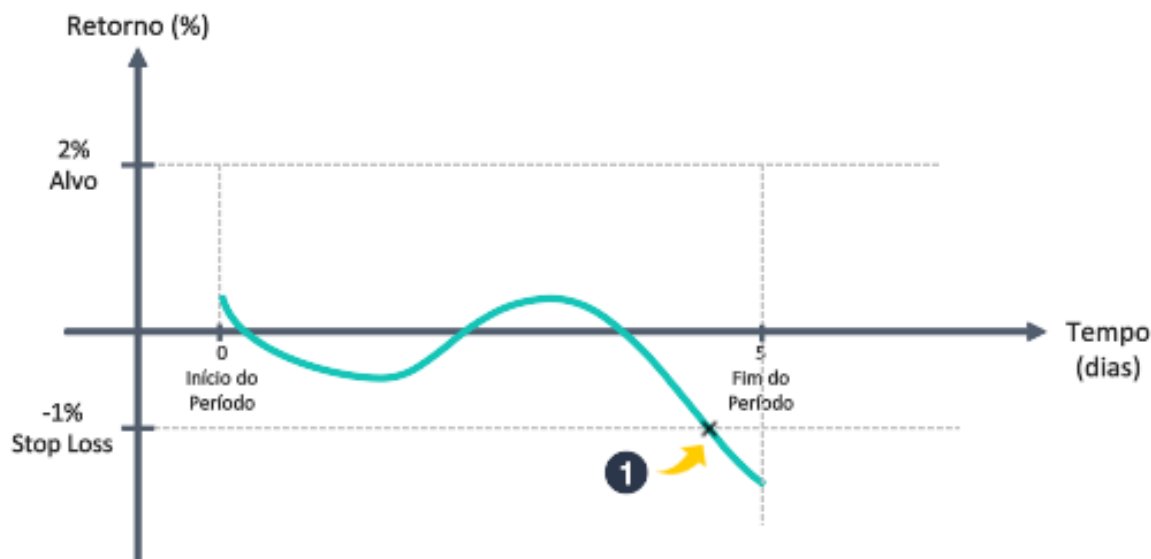
2. Estratégia 1: Saída no final do período - com *stop loss*

A estratégia 1 é semelhante a anterior, porém, utiliza-se do recurso de *stop loss* para minimizar as perdas. Isto é, o investidor compra o ativo caso o rótulo seja “Compra” e vende após o final do período ou caso o preço do ativo atinja o valor definido como limite de perda. Dessa forma, existem duas saídas possíveis, a primeira saída é ao final do período e o retorno é calculado com base na variação no preço do ativo no dia da compra e no dia da venda, e a segunda saída é através do *stop loss* e o retorno coincide com o limite definido pelo investidor. Portanto, o retorno mínimo da estratégia 1 é o valor do *stop loss*. As Figuras 5.3 e 5.4 apresentam as duas saídas possíveis da estratégia 1, ao final do período e ao atingir o *stop loss*, respectivamente.

Figura 5.3 – Estratégia 1 - Saída ao final do período



Fonte: Elaborado pela autora

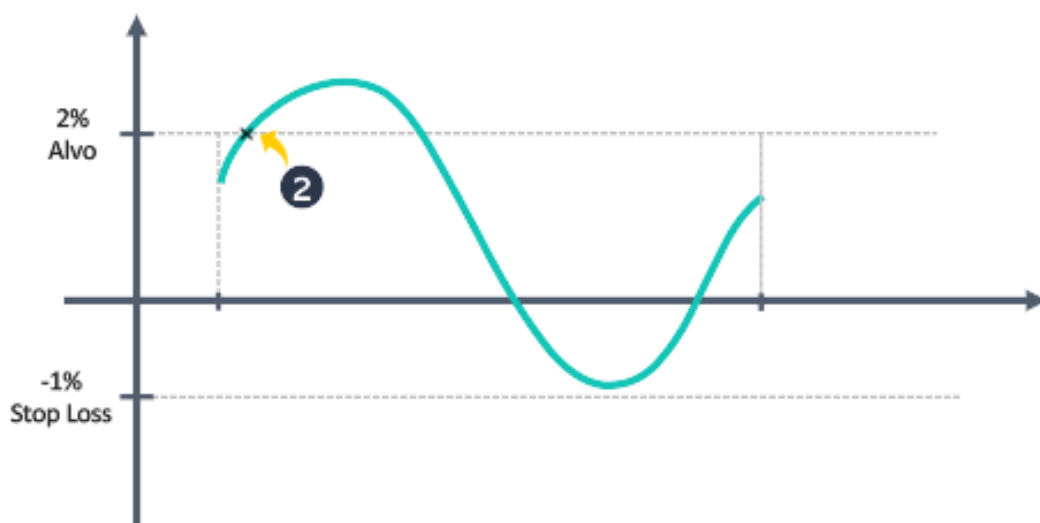
Figura 5.4 – Estratégia 1 - Saída no *stop loss*

Fonte: Elaborado pela autora

3. Estratégia 2: Saída no alvo - sem *stop loss*

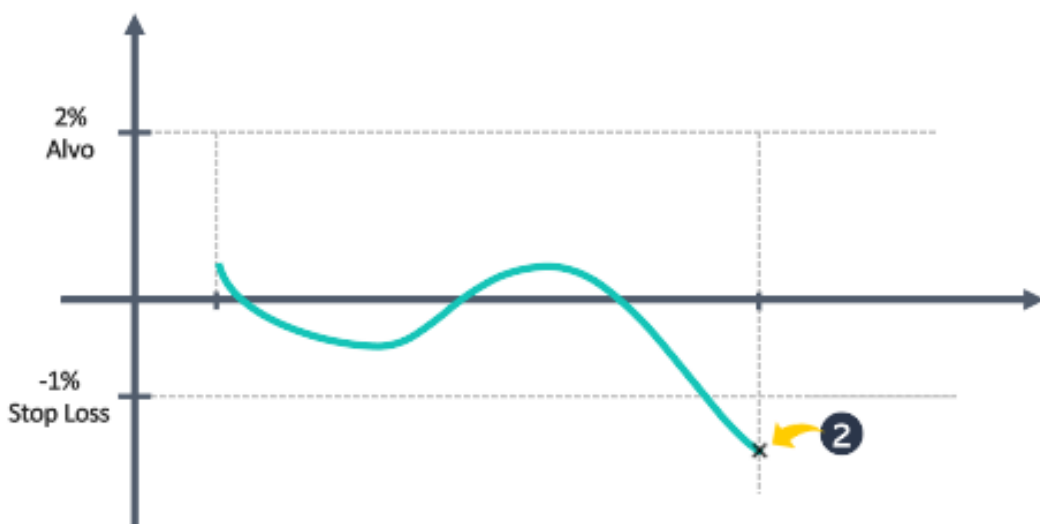
Na estratégia 2, o critério de venda do ativo consiste no atingimento do alvo, ou seja, caso o preço do ativo atinja o valor do parâmetro alvo, o investidor vende o papel. Caso o alvo não seja atingido dentro do período definido, considera-se a saída ao final do período, de forma semelhante à estratégia 0. Dessa forma, existem duas saídas possíveis nessa estratégia, a primeira saída ocorre caso o preço atinja o alvo, sendo este o retorno do investimento e, a segunda saída possível corresponde ao final do período e o retorno é calculado com base na variação no preço do ativo no dia da compra e no dia da venda. Nesse caso, sabe-se que o retorno obtido será inferior ao alvo, uma vez que este seria o primeiro critério de saída. Conforme os parâmetros de exemplo, o investidor poderia ter dois retornos possíveis. No primeiro cenário o retorno seria igual a 2% (valor alvo) e no segundo, o retorno seria a variação após o 5º dia, sendo que este valor é inferior aos 2% do alvo. As Figuras 5.5 e 5.6 apresentam as duas saídas possíveis da estratégia 2, ao atingir o alvo e ao final do período, respectivamente.

Figura 5.5 – Estratégia 2 - Saída no alvo



Fonte: Elaborado pela autora

Figura 5.6 – Estratégia 2 - Saída ao final do período



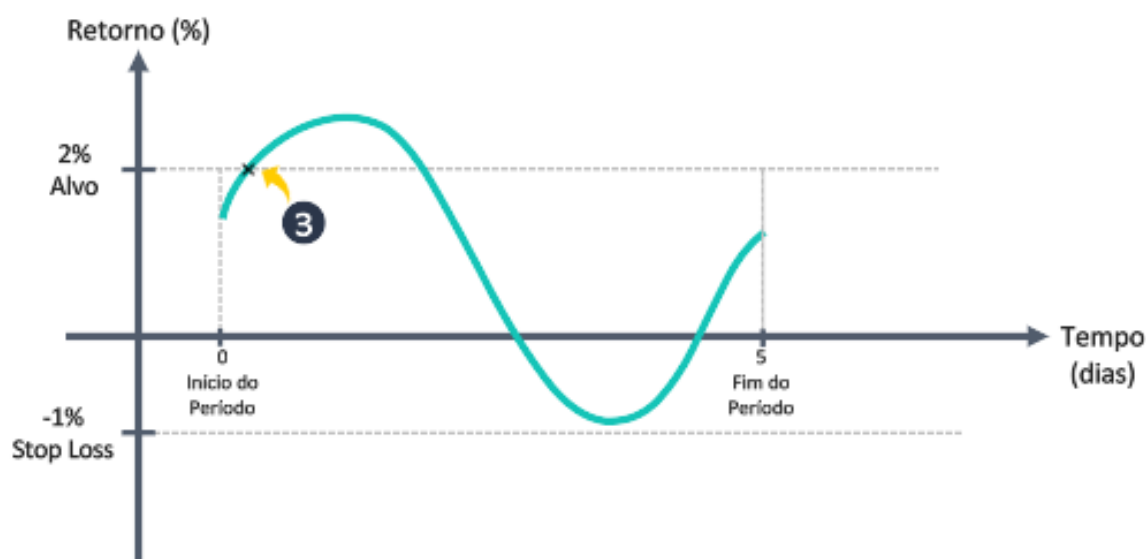
Fonte: Elaborado pela autora

4. Estratégia 3: Saída no alvo - com *stop loss*

A estratégia 3 segue os mesmos critérios da estratégia 2, com a inclusão do recurso *stop loss*. Dessa forma, além das duas saídas possíveis explicadas anteriormente (no alvo ou ao final do período), existe uma terceira saída possível que ocorre se o preço

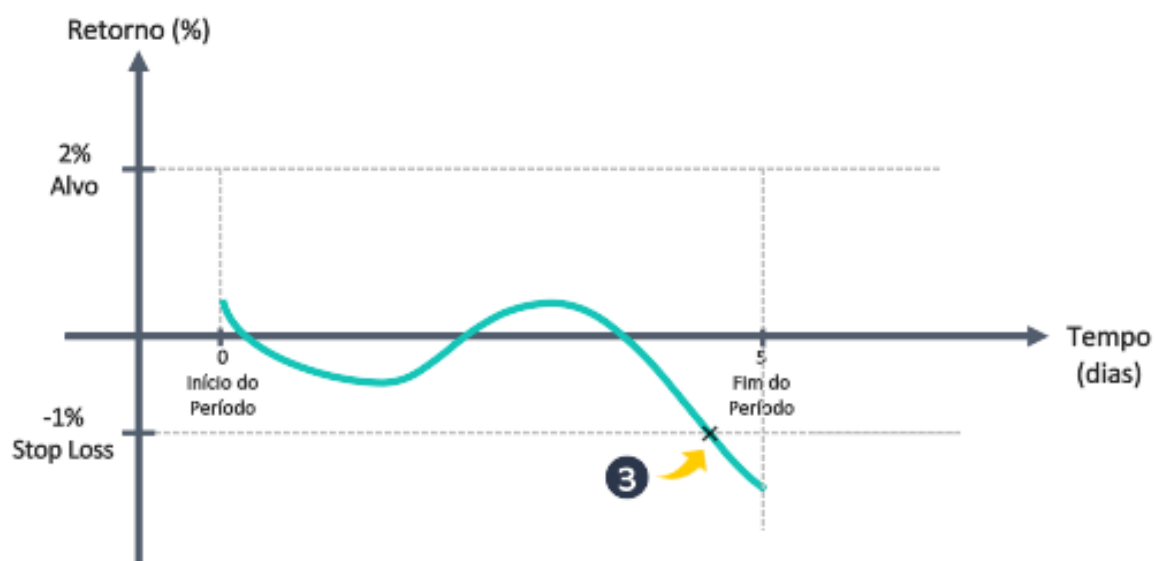
do ativo reduzir abaixo do *stop loss*. Nesse caso, as saídas possíveis são: retorno = 2% (saída no alvo), retorno = -1% (saída no *stop loss*) e retorno entre -1% e 2% (saída ao final do período). As Figuras 5.7 e 5.8 apresentam as duas saídas possíveis da estratégia 3, ao atingir o alvo e ao atingir o *stop loss*, respectivamente.

Figura 5.7 – Estratégia 3 - Saída no alvo



Fonte: Elaborado pela autora

Figura 5.8 – Estratégia 3 - Saída ao atingir o *stop loss*

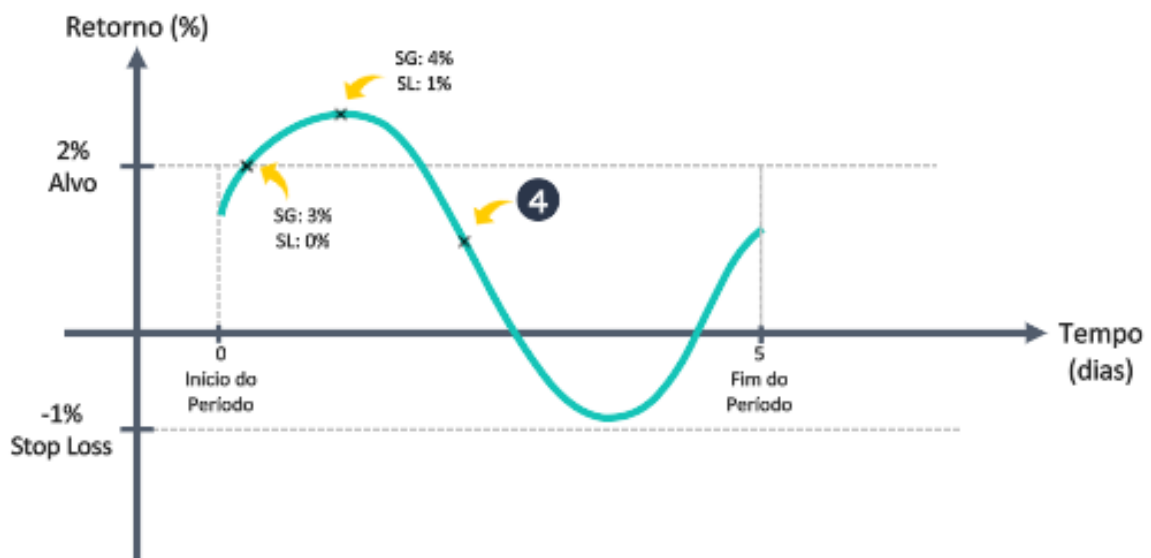


Fonte: Elaborado pela autora

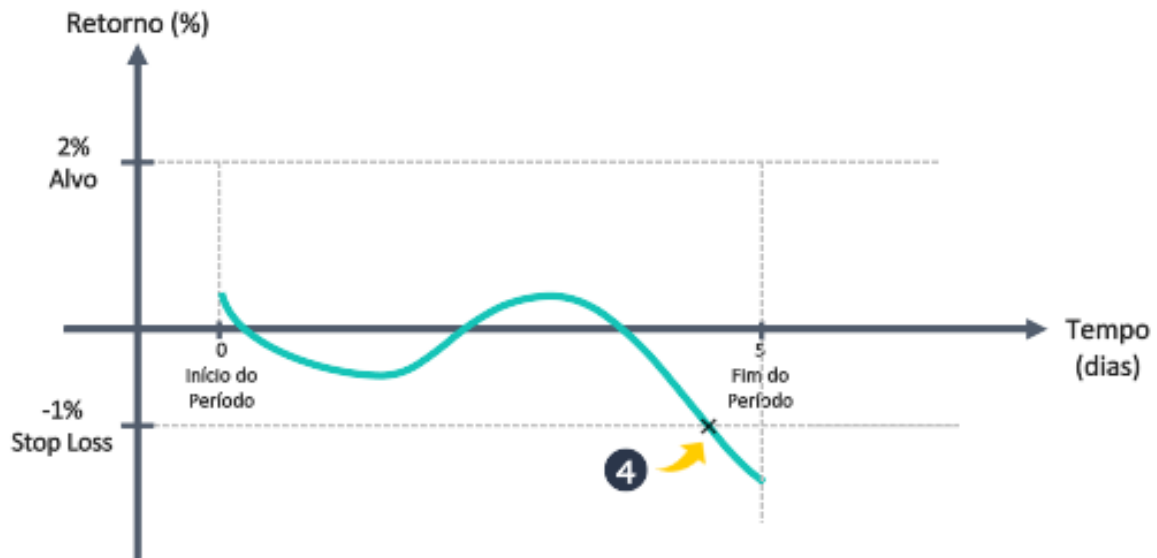
5. Estratégia 4: Saída no final do período - com ajuste do *stop loss*

Por fim, a estratégia 4 consiste no ajuste do *stop gain* e *stop loss* de modo a aumentar o retorno esperado. Para isso, são definidos 4 valores iniciais: *stop loss*, *stop gain* (alvo), incremento do *stop loss* e incremento do *stop gain*. A estratégia tem como objetivo minimizar a perda a cada operação. Dessa forma, caso o ativo atinja o valor do alvo dentro do período, o *stop loss* e o *stop gain* são ajustados, reduzindo as perdas caso o preço do ativo diminua. Por exemplo, se o investidor define o *stop gain* como 2%, o *stop loss* como -1% e o incremento como 1%, caso o valor do ativo alcance os 2% , ao invés de sair da operação, o *stop loss* e o *stop gain* são ajustados para 0% e 3%, respectivamente (incremento de 1%). Com isso, a cada vez que o ativo atinge o valor alvo, o investidor assegura um retorno maior para a operação. As Figuras 5.9 e 5.10 apresentam as duas saídas possíveis da estratégia 4, ao atingir o alvo (ajustado) e ao atingir o *stop loss*, respectivamente.

Figura 5.9 – Estratégia 4 - Saída no alvo



Fonte: Elaborado pela autora

Figura 5.10 – Estratégia 4 - Saída ao atingir o *stop loss*

Fonte: Elaborado pela autora

6. **Estratégia 4:** Saída no final do período - com ajuste do *stop loss*

As estratégias apresentadas acima foram aplicadas a um conjunto de ativos, utilizando os resultados do classificador definido na Seção 3. Para avaliar os resultados, as seguintes premissas foram adotadas:

1. **Premissa 1:** Considera-se que o investidor realiza uma operação de compra sempre que o classificador define o rótulo com “Compra” (+ Alvo);
2. **Premissa 2:** As operações são realizadas de forma independente. Isto é, não existe um limite de número de operações ativas simultaneamente. Cada ativo é tratado de forma independente.
3. **Premissa 3:** As operações não são exclusivas no período. Isto é, dentro de um mesmo período, são permitidas entradas múltiplas no mesmo papel. Por exemplo, se o classificador sugerir compra em 3 dias consecutivos (t , $t+1$, $t+2$), será contabilizado o retorno das 3 operações de forma independente.
4. **Premissa 4:** O tempo máximo de duração da operação é definido pelo parâmetro “Período”.

Tabela 5.1 – Papéis avaliados

Ativo	Empresa/Índice	Setor
ABEV3	Ambev	Consumo e Varejo
BOVA11	Ishares Ibovespa	ETF (Ref. BOVESPA)
EQTL3	Equatorial Energia	Energia e Saneamento
ITSA4	Itaúsa	Holding
LREN3	Lojas Renner	Consumo e Varejo
PETR4	Petrobras	Petróleo E Gás
TOTS3	Totvs	Tecnologia e Internet

Fonte: Elaborado pela autora

5. **Premissa 5:** O classificador utiliza como base para cálculo dos indicadores técnicos o preço de fechamento do ativo. Dessa forma, oscilações *intraday* não foram consideradas.

Para avaliar os retornos esperados com a utilização de cada estratégia foi definido um conjunto de papéis de alta volatilidade, listados na B3, de diferentes setores, conforme apresentado na Tabela 5.1.

Após a escolha dos ativos, foi realizada a coleta de dados (abertura, fechamento, mínima e máxima) do período entre 01 de janeiro de 2015 e 31 de dezembro de 2019, com uma periodicidade diária. Em seguida, foram definidos os parâmetros do classificador conforme descrito na Seção 3. A Tabela 5.2 apresenta os parâmetros adotados.

Tabela 5.2 – Parâmetros utilizados para Classificação

Parâmetro	Valor
Periodicidade dos dados	Diária
Alvo	2%
Período	5 dias
% Treino / % Teste	75% / 25%
Período Treino	01/01/2015 - 01/10/2018
Período Teste	02/10/2018 - 31/12/2019
Período Curto Média Móvel	5 dias
Período Longo Média Móvel	20 dias
Período RSI	14 dias
Período Curto MACD	12 dias
Período Longo MACD	26 dias
Período Sinal MACD	9 dias

Fonte: Elaborado pela autora

Com os dados coletados e os parâmetros definidos, o modelo de classificação foi executado para os ativos selecionados. De forma análoga à apresentada na Seção 4, as matrizes de confusão de cada método foram geradas para todos os papéis analisados.

Conforme descrito acima, as estratégias 0 e 1 são semelhantes, assim como as estratégias 2 e 3, sendo que a única diferença é o uso ou não de *stop loss*. As duas primeiras apresentam a saída ao final do período e as duas seguintes têm a saída ao atingir o alvo ou ao final do período, caso o alvo não seja atingido.

Após a execução dos modelos, o primeiro indicador analisado foi o retorno médio das operações para cada estratégia, conforme mostra a Figura 5.11. O retorno médio corresponde à porcentagem de variação esperada em cada operação realizada.

Figura 5.11 – Retorno médio esperado por ativo

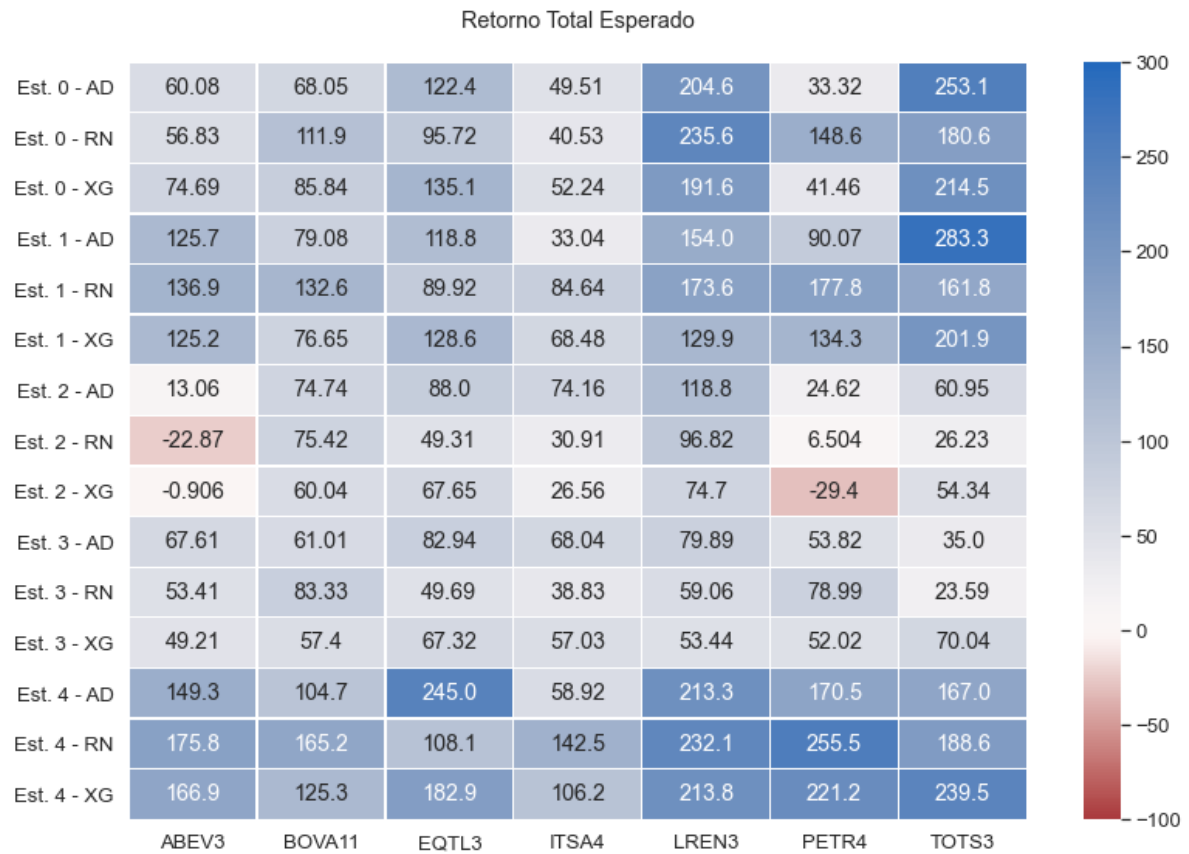


Fonte: Elaborado pela autora (*python*)

Em seguida, foi analisado o retorno total obtido durante o período de teste, ou seja, o valor acumulado após todas as operações. O número de operações varia de acordo com o ativo e a estratégia, conforme apresenta a Tabela 5.3. A Figura 5.12 apresenta o

retorno total para cada ativo.

Figura 5.12 – Retorno total esperado por ativo



Fonte: Elaborado pela autora (*python*)

Para avaliar o resultado das estratégias de negociação, o retorno total foi comparado com o valor do *buy and hold*. A estratégia de *buy and hold* consiste em comprar ativos e mantê-los por um determinado período de tempo. Neste trabalho, o período de tempo considera a data de compra como data inicial e a data de venda como a data final.

A Tabela 5.4 apresenta o retorno obtido por cada estratégia e o retorno do *buy and hold*. De forma análoga ao Capítulo 3, foi incluído também os resultados obtidos pelo classificador aleatório.

Tabela 5.3 – Número de Operações de acordo com o método e estratégia

Ativo	Método	Número de operações no período				
		Estrat. 0	Estrat. 1	Estrat. 2	Estrat. 3	Estrat. 4
LREN3	AD	122	122	139	139	103
	RN	148	148	133	133	130
	XG	143	143	112	112	156
	Aleatório	99	102	160	150	106
EQTL3	AD	126	126	147	149	163
	RN	73	73	75	75	70
	XG	92	92	93	93	87
	Aleatório	106	86	126	141	106
TOTS3	AD	138	167	74	76	114
	RN	65	65	71	71	74
	XG	106	106	137	137	110
	Aleatório	115	94	146	142	113
ABEV3	AD	93	91	145	147	88
	RN	124	124	150	150	124
	XG	123	123	116	116	123
	Aleatório	65	72	93	85	87
PETR4	AD	101	99	122	122	129
	RN	168	168	198	198	148
	XG	153	153	166	166	177
	Aleatório	10	123	160	165	119
ITSA4	AD	59	60	135	135	66
	RN	176	176	158	158	176
	XG	129	129	149	149	128
	Aleatório	94	90	155	156	114
BOVA11	AD	96	96	82	82	82
	RN	170	170	139	139	122
	XG	141	141	150	150	128
	Aleatório	175	178	220	243	160

Fonte: Elaborado pela autora

A Tabela 5.5 apresenta a porcentagem de vezes que cada estratégia superou o retorno obtido pela estratégia *buy and hold*. Através da análise da tabela, verifica-se que as Estratégias 2 e 3 apresentaram desempenhos inferiores em comparação com as demais. Isto pode ser explicado pelo perfil dessas estratégias. Conforme explicado anteriormente,

Tabela 5.4 – Retorno esperado por estratégia

Ativo	Método	Retorno esperado no período (%)					Buy and hold
		Estrat. 0	Estrat. 1	Estrat. 2	Estrat. 3	Estrat. 4	
LREN3	AD	204.6	154	118.8	79.9	213.3	91.4
	RN	235.6	173.6	96.8	59.1	232.1	
	XG	191.6	129.9	74.7	53.4	213.8	
	Aleatório	76.4	98.7	65.9	71.2	144	
EQTL3	AD	122.4	118.8	88	82.9	245	98.6
	RN	95.7	89.9	49.3	49.7	108.1	
	XG	135.1	128.6	67.6	67.3	182.9	
	Aleatório	77.2	124.3	48.4	33.4	157.1	
TOTS3	AD	253.1	283.3	60.9	35	167	171.7
	RN	180.6	161.8	26.2	23.6	188.6	
	XG	214.5	201.9	54.3	70	239.5	
	Aleatório	218	117.1	99.5	83	244	
ABEV3	AD	60.1	125.7	13.1	67.6	149.3	5.2
	RN	56.8	136.9	-22.9	53.4	175.8	
	XG	74.7	125.2	-0.9	49.2	166.9	
	Aleatório	-31.6	39.9	-53.7	0.1	78.6	
PETR4	AD	33.3	90.1	24.6	53.8	170.5	34.5
	RN	148.6	177.8	6.5	79	255.5	
	XG	41.5	134.3	-29.4	52	221.2	
	Aleatório	105.2	93.7	16.3	59.1	171.7	
ITSA4	AD	49.5	33	74.2	68	58.9	38.1
	RN	40.5	84.6	30.9	38.8	142.5	
	XG	52.2	68.5	26.6	57	106.2	
	Aleatório	60.7	54.9	-9	42.6	110.1	
BOVA11	AD	68	79.1	74.7	61	104.7	45
	RN	111.9	132.6	75.4	83.3	165.2	
	XG	85.8	76.6	60	57.4	125.3	
	Aleatório	91.4	134.6	42.4	85	212.7	

Fonte: Elaborado pela autora

essas estratégias têm a saída (venda do ativo) sempre que o alvo é atingido, dessa forma o retorno máximo possível fica limitado ao parâmetro alvo definido pelo usuário.

Tabela 5.5 – Comparação com estratégia *buy and hold*

	Retorno acima do <i>buy and hold</i> (#)	Retorno acima do <i>buy and hold</i> (%)
Estratégia 0	23	82.1%
Estratégia 1	24	85.7%
Estratégia 2	7	25.0%
Estratégia 3	15	53.6%
Estratégia 4	27	96.4%

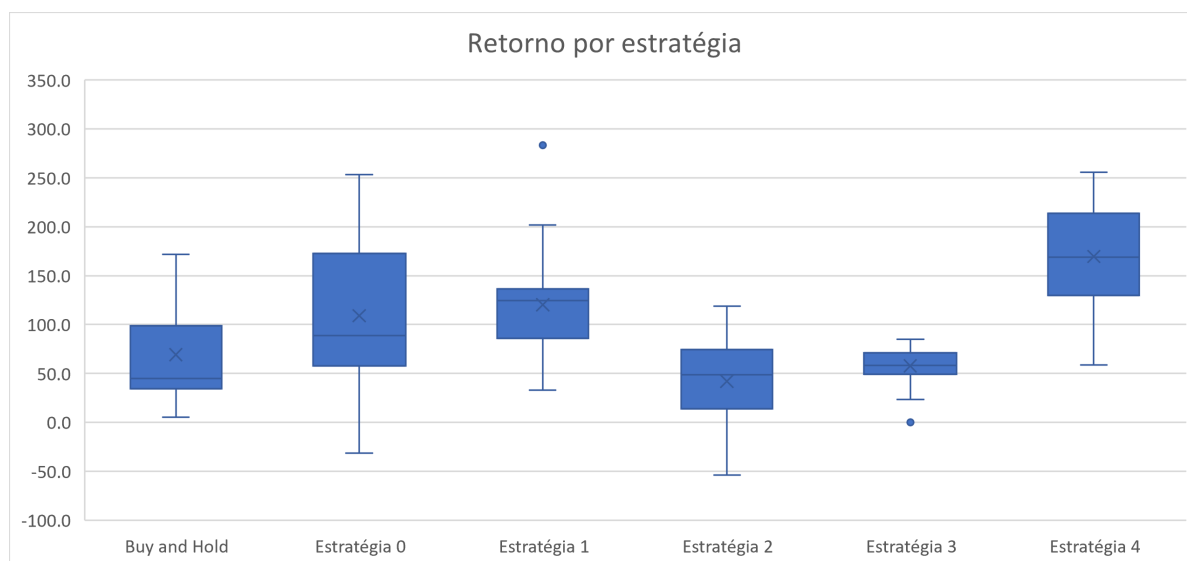
Fonte: Elaborado pela autora

Por outro lado, a Estratégia 4 apresentou o melhor desempenho, superando o retorno do *buy and hold* em 96.4%. Isto significa que dos 28 resultados (7 ativos e 4 métodos), apenas um foi inferior ao retorno caso o investidor tivesse comprado o ativo no início do período e vendido ao final.

Para comparar os retornos totais esperados das estratégias foi gerado um boxplot ¹, conforme mostra a Figura 5.13. Como evidenciado anteriormente, as estratégias 2 e 3 apresentam retornos totais menores em comparação com as outras estratégias. Além disso, é possível identificar o impacto da utilização do *stop loss*. As estratégias 1 e 3 apresentam uma menor amplitude nos retornos esperados em relação às estratégias 0 e 2, respectivamente. Isto ocorre devido à utilização do *stop loss*, que limita as perdas no valor definido.

¹ Em estatística descritiva, o boxplot (diagrama de caixa) é um gráfico utilizado para avaliar a distribuição empírica dos dados e é formado pelo primeiro e terceiro quartil e pela mediana. O boxplot auxilia na visualização da dispersão e assimetria de um conjunto de dados, sendo ainda possível a identificação de *outliers*.

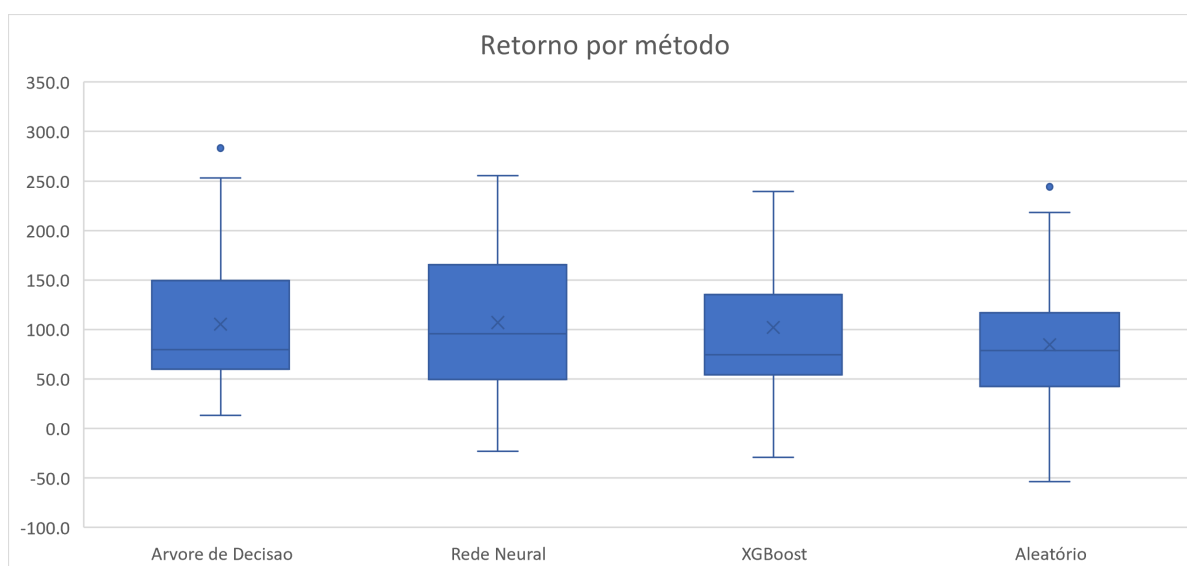
Figura 5.13 – Retorno total por estratégia



Fonte: Elaborado pela autora

Com intuito de analisar de forma mais detalhada os resultados, foi gerado um gráfico que apresenta o retorno de acordo com o método de classificação utilizado: árvore de decisão, rede neural MLP e XGBoost, conforme exibido na Figura 5.14.

Figura 5.14 – Retorno total por método

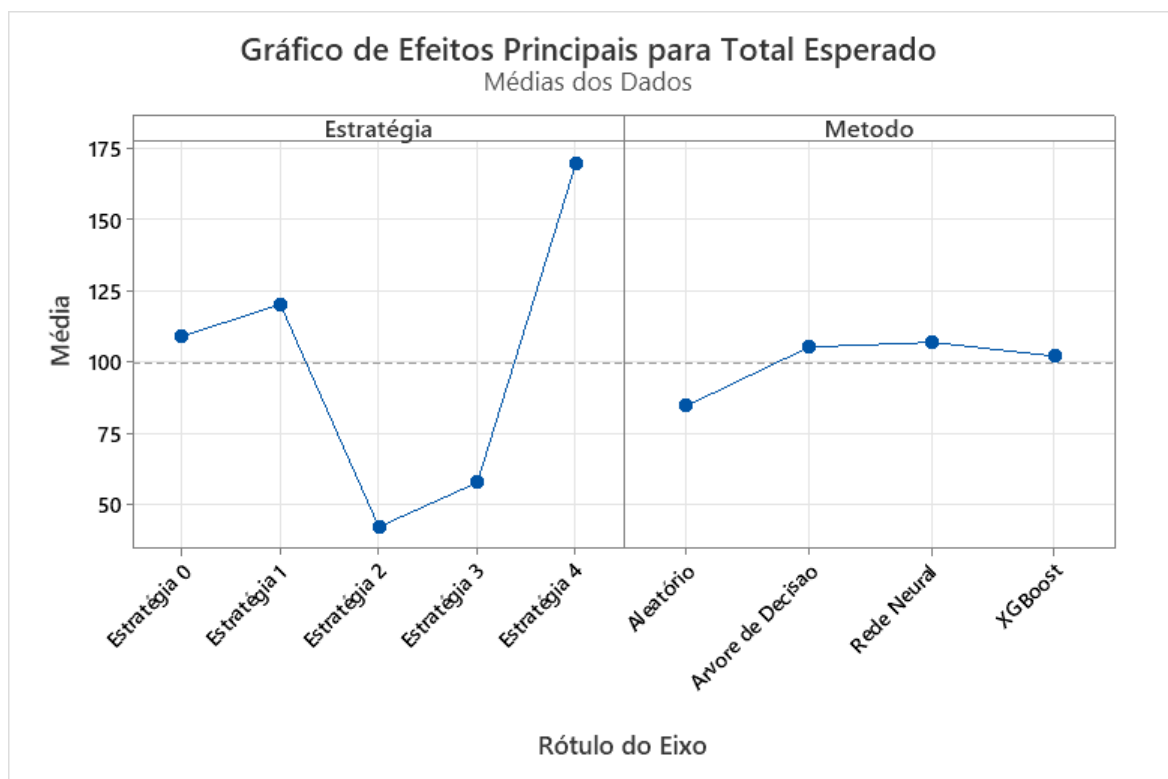


Fonte: Elaborado pela autora

A avaliação dos dois diagramas de caixa sugere uma maior relação entre a

estratégia adotada com o retorno obtido. Para validar esse comportamento foi utilizado um gráfico de efeitos principais. O gráfico de efeitos principais, examina as diferenças entre as médias de nível para um ou mais fatores. Observa-se um efeito principal quando diferentes níveis de um fator afetam a resposta de maneira diferente. Graficamente, quanto maior a diferença na inclinação entre as linhas, maior o grau da potencial interação, e linhas paralelas indicam que esta pode ser inexistente. O gráfico de efeitos principais foi gerado pelo pacote computacional Minitab 21®, conforme ilustra a Figura 5.15.

Figura 5.15 – Gráfico de Efeitos Principais

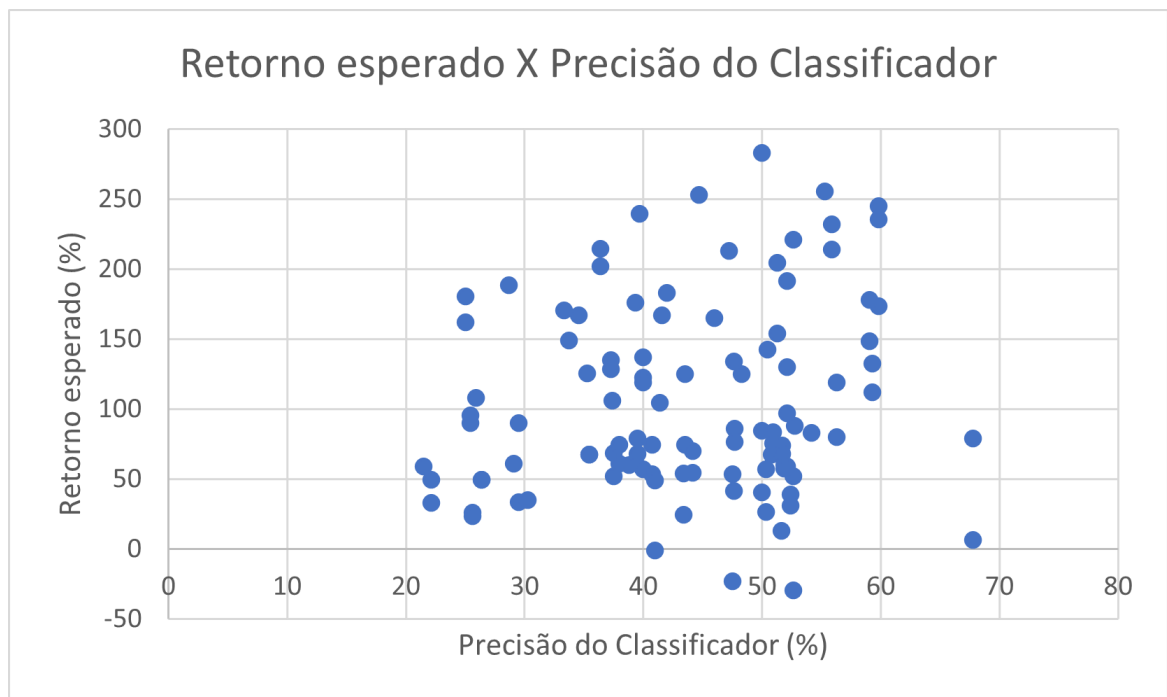


Fonte: Elaborado pela autora - Obtido pelo Software Minitab®

Conforme explicado anteriormente, quando a linha é horizontal (paralela ao eixo x), não existe nenhum efeito principal, ou seja, cada nível afeta a resposta do mesmo modo e a resposta média é a mesma em todos os níveis. Este comportamento pode ser observado ao avaliar o método de classificação. Os três métodos utilizados apresentaram retornos totais próximos. Por outro lado, quando a linha não é horizontal, significa que existe um efeito principal e, quanto maior a inclinação da linha, maior a magnitude do efeito principal. Isto é identificado ao avaliar as estratégias utilizadas. A estratégia 4 apresenta o maior retorno total médio, enquanto a estratégia 2 apresenta o menor.

Através da análise dos gráficos das Figuras 5.13, 5.14 e 5.15 pode se concluir que a estratégia tem uma influência significativa no retorno obtido pelo investidor, enquanto o modelo de classificação não gera o mesmo impacto. Isto pode ser evidenciado pelo gráfico da Figura 5.16 que exibe a relação entre o retorno total obtido com a precisão do classificador utilizado. No gráfico é possível observar que a amostra com maior retorno total apresentou uma precisão de 50% no momento da sua classificação.

Figura 5.16 – Gráfico de Dispersão: Retorno total X Precisão do Classificador



Fonte: Elaborado pela autora

6 Conclusão

Conforme apresentado neste trabalho e em estudos relacionados, os indicadores técnicos podem ser utilizados como ferramenta de auxílio para investimento no mercado de ações. Este trabalho propôs a utilização de indicadores técnicos como dados de entrada de modelos de classificação. Para a construção dos modelos foram utilizados três diferentes métodos de classificação: árvore de decisão, rede neural MLP e *XGBoost*.

O estudo avaliou um conjunto de sete papéis de alta volatilidade, listados na B3, de diferentes setores por um período de cinco anos, sendo os 45 primeiros meses utilizados para treino dos modelos e os 15 meses seguintes para teste.

Este trabalho tinha como motivação responder 3 perguntas:

1. Algoritmos de aprendizado de máquina (AM) podem ser treinados para sugerir investimentos (compras e vendas) de ações com base em indicadores técnicos?
2. É possível utilizar AM para prever reações de curto e médio prazo no mercado de ações de modo a obter retornos lucrativos?
3. Técnicas de gestão de riscos podem ser associadas com AM para aumentar os retornos nas operações?

As evidências que temos com os resultados obtidos neste trabalho, mostram que o modelo de classificação proposto permite ao investidor obter sugestões de investimentos com base nas classes “Compra” e “Neutro”, com precisão de até 70% nas indicações de compra. Além disso, ao incluir estratégias de negociação utilizando-se de ferramentas de *stop loss* e *stop gain*, os retornos obtidos pelo modelo superaram o retorno da estratégia de *buy and hold* em até 96.4%.

A suposição de que a associação de técnicas de gestão de carteiras com algoritmos de aprendizado de máquina pode aumentar os retornos obtidos, foi confirmada neste trabalho ao comparar os resultados obtidos pela estratégia 0 - resultados do classificador, com a estratégia 4 - resultados do classificador com uso de ferramentas de gestão de carteiras. Isto é, o retorno total médio passou de 117% na estratégia 0 para 173% na estratégia 4.

Por fim, vale a pena ressaltar que este trabalho apresentou resultados para alguns ativos da B3, em um período de tempo pré-estabelecido e sem variações nos parâmetros dos modelos. Na perspectiva de análise do classificador, existem oportunidades de melhorias nos resultados através de alterações nos parâmetros dos modelos de classificação e também nos parâmetros dos dados de entrada como periodicidade dos dados, períodos utilizados para cálculo dos indicadores técnicos e % de variação alvo.

Possibilidades adicionais incluem testes com outros classificadores, outros indicadores técnicos, ou ainda, outros dados de entrada, como notícias que podem ser incorporadas aos modelos através da análise de sentimentos.

Em relação às estratégias de negociação, os trabalhos futuros incluem a implementação do *backtesting* para avaliar os retornos obtidos pelos ativos. Com a utilização do *backtesting*, além da avaliação do retorno isolado de cada papel, será possível ainda a análise de carteiras considerando diversificação, eventuais taxas de corretagem e impostos relacionados com o número de operações realizadas. Além disso, destaca-se a possibilidade ao mudar os parâmetros de entrada, as estratégias que não apresentaram um bom desempenho poderiam apresentar melhores resultados.

Referências

APPEL, G.; DOBSON, E. *Understanding MACD*. [S.l.]: Traders Press, 2007. v. 34. Citado na página 25.

ARÉVALO, R.; GARCÍA, J.; GUIJARRO, F.; PERIS, A. A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting. *Expert Systems with Applications*, v. 81, p. 177–192, 2017. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417417301823>>. Citado na página 15.

BATISTA, G.; MONARD, M.-C. Sniffer: Um ambiente computacional para gerenciamento de experimentos de aprendizado de máquina supervisionado. 12 2021. Citado na página 28.

BOMBARDELLI, W. W. *Aprendizado de máquina em dados estruturados para modelagem de fenômenos hidráulicos em sistemas de microirrigação*. Tese (Doutorado) — Universidade de São Paulo, 2021. Citado na página 33.

CAROSIA, A. E. de O.; COELHO, G. P.; SILVA, A. E. A. da. Investment strategies applied to the brazilian stock market: A methodology based on sentiment analysis with deep learning. *Expert Systems with Applications*, v. 184, p. 115470, 2021. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417421008812>>. Citado na página 28.

DEBASTIANI, C. A. *Candlestick: um método para ampliar lucros na Bolsa de Valores*. [S.l.]: Novatec Editora, 2007. Citado na página 22.

DEBASTIANI, C. A. *Análise Técnica de Ações: Identificando Oportunidades de Compra e Venda*. Paperback. Novatec, 2008. ISBN 978-8575221532. Disponível em: <<https://lead.to/amazon/com/?op=bt&la=en&cu=usd&key=8575221531>>. Citado 3 vezes nas páginas 21, 25 e 26.

DEY, S.; KUMAR, Y.; SAHA, S.; BASAK, S. Forecasting to classification: Predicting the direction of stock market price using xtreme gradient boosting. 10 2016. Citado na página 21.

DOBILAS, S. *XGBoost: Extreme Gradient Boosting — How to Improve on Regular Gradient Boosting?* 2021. Disponível em: <<https://towardsdatascience.com/xgboost-extreme-gradient-boosting-how-to-improve-on-regular-gradient-boosting-5c6acf66c70a>>. Citado na página 34.

ELTON, E.; GRUBER, M.; BROWN, S. *Moderna teoria de carteiras e análise de investimentos*. [S.l.]: Elsevier Brasil, 2012. Citado na página 37.

FAMA, E. F. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, [American Finance Association, Wiley], v. 25, n. 2, p. 383–417, 1970. ISSN 00221082, 15406261. Disponível em: <<http://www.jstor.org/stable/2325486>>. Citado na página 14.

- GARCIA, S. C. O uso de árvores de decisão na descoberta de conhecimento na área de saúde. 2003. Citado na página 31.
- HASANIN, T.; KHOSHGOFTAAR, T. M. The effects of random undersampling with simulated class imbalance for big data. *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, p. 70–79, 2018. Citado na página 36.
- HAYKIN, S. *Neural Networks and Learning Machines 3rd Ed.* NY: NYL Pearson Prentice Hall. 2009. Citado 2 vezes nas páginas 31 e 32.
- LANGLEY, P.; SIMON, H. A. Applications of machine learning and rule induction. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 38, n. 11, p. 54–64, nov 1995. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/219717.219768>>. Citado na página 28.
- LAST, M.; MAIMON, O.; MINKOV, E. Improving stability of decision trees. *IJPRAI*, v. 16, p. 145–159, 03 2002. Citado na página 31.
- LEVER, J.; KRZYWINSKI, M.; ALTMAN, N. Classification evaluation. *Nature Methods*, v. 13, n. 8, p. 603–604, 2016. Citado 2 vezes nas páginas 34 e 35.
- LO, A. W. Chapter 6. the adaptive markets hypothesis. In: _____. *Adaptive Markets: Financial Evolution at the Speed of Thought*. Princeton University Press, 2019. p. 176–221. Disponível em: <<https://doi.org/10.1515/9780691196800-008>>. Citado 3 vezes nas páginas 14, 19 e 20.
- LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, v. 35, n. 101, p. 85–94, abr. 2021. Disponível em: <<https://www.revistas.usp.br/eav/article/view/185035>>. Citado na página 17.
- MALKIEL, B. G. Efficient market hypothesis. In: *Finance*. [S.l.]: Springer, 1989. p. 127–134. Citado 2 vezes nas páginas 18 e 19.
- MALKIEL, B. G. The efficient market hypothesis and its. *Journal of Economic Perspectives*, v. 17, n. 1, p. 59–82, 2003. Cited By :755. Disponível em: <www.scopus.com>. Citado na página 14.
- MARKOWITZ, H. M.; FABOZZI, F. J. *The theory and practice of investment management*. [S.l.]: John Wiley & Sons, 2002. Citado na página 39.
- NASEER, M.; TARIQ, D. Y. The efficient market hypothesis: A critical review of the literature. *The IUP Journal of Financial Risk Management*, 12 2015. Citado na página 14.
- PANDYA, A. S.; MACY, R. B. *Pattern recognition with neural networks in C++*. [S.l.]: CRC press, 1995. Citado na página 31.
- PICASSO, A.; MERELLO, S.; MA, Y.; ONETO, L.; CAMBRIA, E. Technical analysis and sentiment embeddings for market trend prediction. *Expert Systems with Applications*, v. 135, p. 60–70, 2019. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417419304142>>. Citado 5 vezes nas páginas 15, 16, 28, 41 e 46.

REILLY, F. K.; BROWN, K. C. *Investment Analysis and Portfolio Management (Text Only)*. [S.l.]: Cengage Learning, 2011. Citado na página 37.

SERAFINI, D. G.; PEREIRA, P. L. V. *Sistemas técnicos de trading no mercado de ações brasileiro: testando a hipótese de eficiência de mercado em sua forma fraca e avaliando se a análise técnica agrega valor*. [S.l.], 2010. Disponível em: <<https://ideas.repec.org/p/fgv/eesptd/260.html>>. Citado na página 20.

SHAH, D.; ISAH, H.; ZULKERNINE, F. Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*, v. 7, n. 2, 2019. ISSN 2227-7072. Disponível em: <<https://www.mdpi.com/2227-7072/7/2/26>>. Citado 3 vezes nas páginas 14, 16 e 28.

WAGNER, N. Stephen h. penman, financial statement analysis and security valuation (3rd ed), mcgraw-hill/irwin, international edition (2007), isbn-10: 007-125432-3. *The International Journal of Accounting*, v. 42, n. 4, p. 450–453, 2007. Disponível em: <<https://EconPapers.repec.org/RePEc:eee:accoun:v:42:y:2007:i:4:p:450-453>>. Citado na página 21.

ZHONG, X.; ENKE, D. Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications*, v. 67, p. 126–139, 2017. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417416305115>>. Citado na página 14.