

# **Aplicação de algoritmo de aprendizado de máquina (AM) para previsão do comportamento das ações negociadas no mercado brasileiro**

**Gabriel Donadio Costa**

Universidade Federal de Santa Catarina - UFSC, Brasil

**Rogério João Lunkes**

Universidade Federal de Santa Catarina - UFSC, Brasil

## **Aplicación de algoritmo de aprendizaje automático para predecir el comportamiento de las acciones negociadas en el mercado brasileño**

La economía moderna ofrece diversas opciones de inversión, lo que hace que la asignación de capital sea un proceso complejo, lento y arriesgado. Para ayudar en la toma de decisiones de los inversionistas, las herramientas de inteligencia artificial buscan descubrir patrones ocultos en los datos, y proporcionar informaciones útiles, oportunas y precisas. Este estudio evalúa la aplicación de algoritmos de aprendizaje automático (AA) en la selección de carteras en el mercado brasileño. Investigaciones recientes han buscado predecir el comportamiento del mercado bursátil implementando AA junto con métodos convencionales, como análisis técnicos o fundamentales (Anghel, 2021; Kamara et al., 2022; Nti et al., 2020b); sin embargo, pocos combinan los análisis en mercados emergentes y volátiles como Brasil. Por lo tanto, se entrenaron tres modelos de AA utilizando variables provenientes del análisis técnico y/o fundamental. La muestra incluyó 40.562 observaciones pertenecientes a seis empresas que cotizan en la Bolsa de Valores de Brasil (B3), durante el período de agosto de 1994 a diciembre de 2021. Los modelos entrenados únicamente con variables fundamentales o técnicas mostraron baja precisión, lo que indica baja capacidad de aprendizaje y generalización del algoritmo. En contraste, el modelo compuesto



<https://doi.org/10.18800/contabilidad.2025ESP.002>

**Contabilidad y Negocios 20 (esp.) 2025, pp. 48-70 / e-ISSN 2221-724X**

por la combinación de variables técnicas y fundamentales mostró una precisión promedio del 70,7 % en los 5 *folds*, coincidiendo con la literatura que sugiere que los modelos híbridos pueden proporcionar mayor precisión y menor volatilidad. Además, los resultados superan la precisión de estudios previos (Emir et al., 2012; Kim, 2003; Zhang & Zhao, 2009), lo que indica que el *support vector machine* (SVM) también puede aplicarse a mercados emergentes, incluso en escenarios de crisis, como la pandemia de COVID-19.

**Palabras clave:** aprendizaje automático, selección de cartera, análisis técnico, análisis fundamental

### **Application of machine learning algorithm to predict the behavior of stocks marketed in Brazil**

Modern economy offers several investment options, making capital assignments complex, slow and risky. In order to assist investors in the decision-making process, artificial intelligence tools aim at finding patterns hidden in data and providing useful, timely and accurate information. This work analyzes the application of machine learning algorithms in the selection of portfolios in the Brazilian market. Recent research intended to predict the stock market behavior implementing machine learning with conventional methods such as technical or fundamental analysis (Anghel, 2021; Kamara et al., 2022; Nti et al., 2020b); while few combine analyses in emerging and volatile markets like Brazil. Thus, three machine learning models were trained using variables from the technical and/or fundamental analysis. The sample included 40,562 observations from six companies listed on B3, from August 1994 to December 2021. Models trained only with fundamental or technical variables evidenced low accuracy, which was translated into low learning and generalization capacity of the algorithm. In contrast, the model including the combination of technical and fundamental variables revealed an average accuracy of 70,7 % on 5 folds, which was supported by the literature that indicates that hybrid models can provide greater accuracy and lower volatility. In addition, results exceed the accuracy of previous studies (e.g. Emir et al., 2012; Kim, 2003; Zhang & Zhao, 2009), which indicates that the Support Vector Machine - SVM can also be applied to emerging markets, even in crisis times, such as COVID-19 pandemic.

**Keywords:** machine learning, portfolio selection, technical analysis, fundamental analysis

### **Aplicação de algoritmo de aprendizado de máquina (AM) para previsão do comportamento das ações negociadas no mercado brasileiro**

A economia moderna é repleta opções de investimentos, o que torna a alocação de capital um processo complexo, demorado e arriscado. Para auxiliar a tomada de decisão dos investidores, ferramentas de inteligência artificial buscam a descoberta de padrões ocultos nos dados e fornecem informações úteis, tempestivas e precisas. Este estudo avalia a aplicação de algoritmos de aprendizado de máquina (AM) na seleção de portfólio no mercado brasileiro. Investigações

recentes têm buscado prever o comportamento do mercado acionário implementando AM em conjunto com métodos convencionais, como análises técnicas ou fundamentalistas (Anghel, 2021; Kamara et al., 2022; Nti et al., 2020b), entretanto, poucos combinam as análises em mercados emergentes e voláteis como o Brasil. Portanto, três modelos de AM foram treinados utilizando variáveis advindas das análises técnicas e/ou fundamentalistas. A amostra incluiu 40.562 observações pertencentes a seis empresas negociadas na Brasil, Bolsa, Balcão (B3), durante o período de agosto de 1994 a dezembro de 2021. Os modelos treinados somente com variáveis fundamentalistas ou técnicas apresentaram acurácia baixa, indicando baixa capacidade de aprendizagem e generalização do algoritmo. Em contraste, o modelo composto pela combinação de variáveis técnicas e fundamentalistas, apresentou acurácia média de 70,7% nos 5 *folds*, corroborando com a literatura que sugere que modelos híbridos podem fornecer maior acurácia e menor volatilidade. Além disso, os resultados superam a acurácia de estudos prévios (Emir et al., 2012; Kim, 2003; Zhang & Zhao, 2009), indicando que o *support vector machine* (SVM) também pode ser aplicável à mercados emergentes, inclusive em cenários de crise, como a pandemia da covid-19.

**Palavras-chave:** aprendizado de máquina, seleção portfólio, análise técnica, análise fundamentalista

## 1. INTRODUÇÃO

A alocação de capital tem sido um grande desafio para os investidores da economia moderna devido à grande quantidade, variedade e volatilidade dos investimentos. O mercado atual possui desde opções conservadoras como, as poupanças e títulos do governo, até as mais arriscadas como, bolsa de valores e cripto moedas. Visando diversificar os investimentos, aumentar a rentabilidade e encontrar alternativas para enfrentar fatores macroeconômicos adversos, a busca pelo mercado de ações têm apresentado um crescimento contínuo (Feijó et al., 2022). Dados da Brasil, Bolsa, Balcão (B3) evidenciam um aumento significativo no número de investidores pessoa física nos últimos anos, com mudanças na faixa etária e aumento da participação feminina. De 2020 a 2024, o número de investidores cresceu 80%, totalizando 5,1 milhões de investidores em renda variável e 19,4 milhões no total (B3, 2024).

O mercado de ações é uma alternativa de investimento que envolve um alto nível de volatilidade e complexidade, onde há intrínseco um certo grau de incerteza. Muitas vezes, estas incertezas estão associadas a eventos macroeconômicos, como decisões econômicas e políticas, já outras vezes podem tratar-se de ruídos, não-estacionariedade e caos determinístico (Marcelino et al., 2015). No âmbito das finanças, diversos estudos consideram que estas incertezas podem ser, em parte, previsíveis (Fama &

French, 1988; Huerta et al., 2013; Zhang et al., 2018). Para Fan e Palaniswami (2001), as séries históricas das cotações do mercado de ações possuem padrões que, se descobertos, possibilitam previsões a respeito do retorno futuro dos ativos.

Devido à diversidade e complexidade do mercado financeiro, os investidores precisam de informações tempestivas, úteis e precisas que auxiliem na tomada de decisão da forma mais rentável e segura de alocar seu capital. Chenhall e Moers (2015) destacam que, com o passar dos anos, o processo de alocação de capital vem se adaptando e evoluindo com o desenvolvimento de técnicas econométricas modernas, avanços computacionais e aperfeiçoamento dos algoritmos de aprendizado de máquina (AM). Estas novas tecnologias buscam proporcionar a redução de custos, aumento de eficiência e implementação de novas práticas, onde muitas vezes processos disruptivos são necessários.

Recentemente, a indústria financeira tem intensificado o uso de tecnologias relacionadas a inteligência artificial por demonstrarem serem ferramentas úteis para geração de valor, principalmente na detecção de fraudes corporativas, em *e-commerces*, cartões de crédito, previsões de curva de juros, retorno de títulos e seleção de portfólios (Bianchi et al., 2021; Buckmann et al., 2021; Chen et al., 2019; Dang et al., 2022; Huerta et al., 2013; Mun & Soong, 2021; Nti et al., 2020b; Shatshat & Ahmed, 2019). A complexidade destas tarefas faz com que os gestores lancem mão de ferramentas que auxiliam o processo de análise e tomada de decisão, das quais destacam-se os AM como, redes neurais, métodos *ensemble*, *random forest*, e máquina de vetores de suporte (SMV).

Na literatura, há cada vez mais indícios que sugerem que os métodos de AM podem superar os convencionais em problemas de predição econômica (Buckmann et al., 2021; Bianchi et al., 2021; Chen et al., 2019; Emir et al., 2012; Fan & Palaniswami, 2001; Mun & Soong, 2021). Estudos recentes têm buscado prever o comportamento do mercado acionário implementando o uso de AM em conjunto com análises técnicas, fundamentalistas e híbridas (Anghel, 2021; Kamara et al., 2022; Nti et al., 2020b). A maioria destes estudos concentram-se em países desenvolvidos e em períodos de normalidade econômica, onde há a aproximação com a teoria do mercado eficiente (Fan & Palaniswami, 2001; Huerta et al., 2012; Kim, 2003). Poucas são as pesquisas que analisam períodos de crises (econômicas e sanitárias), combinados com mercados voláteis e imprevisíveis como de países emergentes.

Portanto, este estudo investiga a aplicação de algoritmos de aprendizagem de máquina na previsão do comportamento de ações no mercado brasileiro. Para isto, foram testados três modelos distintos de AM, buscando prever o comportamento de seis ações negociadas no mercado brasileiro, baseados em análises técnicas, funda-

mentalistas e híbridas. Os dados foram coletados da base de dados Econômica, para o período de 18 de agosto de 1994 a 16 de dezembro de 2021, com periodicidade diária.

Ao propor um modelo de AM, destinado a apoiar investidores na previsão do comportamento de ações negociadas em mercados emergentes como o brasileiro, este estudo contribui para auxiliar gestores de fundos, *traders*, analistas financeiros e investidores individuais na promoção e otimização de estratégias de investimento, fornecendo subsídios técnicos para a tomada de decisões financeiras embasadas em dados. A pesquisa também contribui no âmbito teórico, ao integrar o uso de AM e variáveis advindas das análises técnicas e fundamentalistas, em mercados emergentes, durante um período relativamente longo (27 anos) e volátil. Essa integração de dados longitudinais permite a descoberta de padrões ocultos em dados financeiros complexos e uma melhor generalização dos resultados.

Além desta introdução, o presente trabalho está dividido nos seguintes tópicos: a seção 2 apresenta a referencial teórico do trabalho, a seção 3 descreve a metodologia da pesquisa, seção 4 traz os resultados das aplicações empíricas dos modelos e discussões e, por fim, a seção 5 conclui o artigo.

## 2. REFERENCIAL TEÓRICO

### 2.1. Análises para seleção de portfólio

O processo de seleção de portfólio no mercado de ações é uma tarefa complexa devido à vasta variedade e volatilidade das opções de investimentos e ao grande número de fatores externos que podem influenciar no processo de decisão (Marcelino et al., 2015). Reportagens jornalísticas, postagens viralizadas nas redes sociais ou eventos macroeconômicos podem acarretar na fuga ou atração de investidores para empresa, elevando ou diminuindo o preço das ações. O mercado de ações é uma alternativa de investimento que envolve um alto nível de complexidade, onde há intrínseco um certo grau de incerteza (Marcelino et al., 2015).

Devido à alta complexidade, os tomadores de decisões utilizam técnicas para prever o comportamento do mercado de ações. Duas ferramentas tradicionais e importantes para previsão do comportamento da ação são as análises técnicas e fundamentalistas (Renu & Christie, 2018).

A análise fundamentalista presume que o preço da ação, tanto no momento atual, quanto no futuro, depende do valor intrínseco da empresa e do retorno esperado. Esta análise utiliza informações públicas disponíveis, para analisar as ações com base

nos aspectos gerais da economia (inflação, taxas de juros e PIB), do setor em que atua (níveis de competição, ameaça de novos concorrentes e políticas governamentais) e nos fundamentos da própria empresa (características dos gestores de topo, finanças da organização, notícias sobre a empresa, participação no mercado e capacidade de endividamento, entre outros), e é indicada para investimentos de médio e longo prazo (Oliveira et al., 2013; Renu & Christie, 2018). Modelos de precificação de ativos com base na análise fundamentalista são amplamente estudados em finanças, como o *capital asset pricing model* (CAPM) que é aplicado para estimar o custo de capital das empresas e avaliar o desempenho de portfólios (Fama & French, 2004).

São exemplos de variáveis advindas da análise fundamentalista as advindas do resultado da organização, como lucro/preço, alavancagem, *book-to-market equity* (Fama & French, 1992), valor econômico adicionado (EVA), fluxo de caixa descontado (DCF), *debt-to-equity ratio* (D/E), *price-to-book* (P/B), *return-on-equity* (RoE), *return-on-asset* (RoA) (Cavalcante et al., 2005), índices de inflação, desemprego, informações setoriais (Silva et al., 2015), boato ou notícia vinculada a venda ou compra de ações, *status* financeiro de uma organização (Adebisi et al., 2012).

A análise técnica envolve o emprego de ferramentas para prever o comportamento futuro do preço da ação, tendências e pontos de inflexão, com base no padrão histórico (Lui & Mole, 1998). Seu foco é analisar a evolução dos mercados a partir da representação gráfica do histórico do preço de venda, preço de compra, volume negociado, entre outros. Portanto, esta análise busca compreender o comportamento histórico dos preços e volumes, visando determinar a tendência e o preço futuro dos ativos, sendo adequada a investimentos de curto prazo (Oliveira et al., 2013; Renu & Christie, 2018). King (1964) e Levy (1967) ainda vão além, expondo a existência de uma intercorrelação entre o preço das ações disponíveis em um mercado, indicando que parte do movimento do preço de uma ação está relacionado às tendências de mercados e setores.

Elder (1993) classifica os indicadores da análise técnica em três grupos, são eles: rastreadores de tendência, osciladores e mistos. Os rastreadores de tendência funcionam quando o mercado apresenta uma direção (tendência) definida. O uso em situações onde o mercado não apresenta uma tendência definida (mercado na horizontal) pode acarretar em resultados pouco fidedignos. São exemplos destes indicadores rastreadores as médias móveis simples (SMA) e média móvel exponencial (EMA), médias móveis convergente e divergente (MACD) (Elder, 1993), que suavizam os movimentos dos preços temporariamente exagerados (Levy, 1967).

O grupo dos osciladores são recomendados quando não há tendência de mercado definida, sendo o índice de força relativa (RSI) e o K-estocástico os indicadores mais

utilizados. Por fim, os indicadores mistos são utilizados para um grupo de ações, não sendo possível aplicá-lo a uma única ação. Seu objetivo é sinalizar um consenso dos investidores em relação a um grupo de ações, mercado como um todo ou a um setor específico da economia (Elder, 1993).

As análises técnicas e fundamentalistas ainda podem se complementar, possuindo foco em períodos de análise diferentes (Lui e Mole, 1998). Uma pesquisa realizada por Lui e Mole (1998) com corretores da bolsa de Hong Kong, afirmou que ambas análises são importantes para o processo de tomada de decisão e previsões dos movimentos da taxa de câmbio, sendo a análise técnica utilizada indicada para horizontes de tempo curtos, enquanto a fundamentalista é apropriada para horizontes de médio e longo prazo.

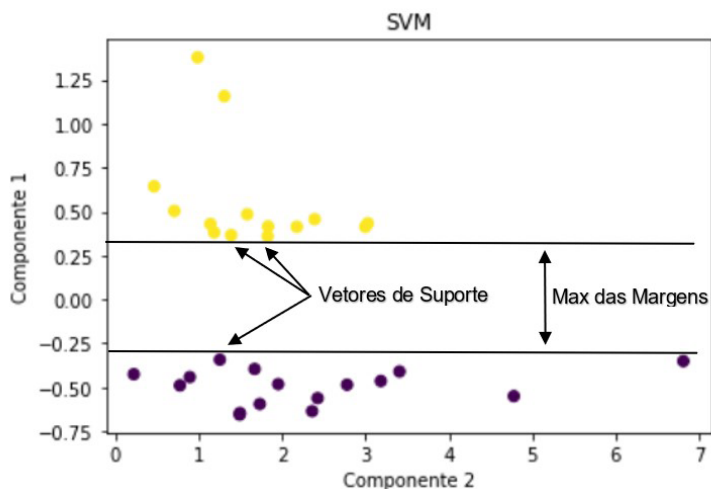
## 2.2. Aprendizado de máquina

Aprendizado de máquina é uma subárea da inteligência artificial que permite os computadores aprenderem com os dados (Géron, 2019). Em uma definição clássica, o AM é o campo de estudo que permite que os computadores aprendam sem ser explicitamente programados (Samuel, 1959). No campo da ciência de dados existem diversos algoritmos que objetivam o AM através de exemplos. Os algoritmos são modelos estatísticos e computacionais que têm o papel de reconhecer padrões nos dados, extrair informações e prever novas observações (Marcelino et al., 2015), estando amplamente inseridos nas atividades do cotidiano como, nos sistemas de recomendação de filmes da Netflix ou Youtube, no reconhecimento de *spams*, na direção de carros autônomos e em sistemas de seleção de alocação de ativos e seleção de portfólios.

Visando o funcionamento de forma satisfatória dos algoritmos de AM é necessário um grande conjunto de dados, que serão utilizados parte para treinamento do modelo e parte para o teste das previsões. Os dados também precisam ser representativos do problema a ser estudado, de forma que o algoritmo possa reconhecer padrões, aprender e ser capaz de generalizar os resultados. Portanto, é imprescindível analisar as estatísticas descritivas e, por vezes, retirar do conjunto de treinamento os dados errôneos, ruídos e *outliers* (Fan & Palaniswami, 2001).

## 2.3. Support vector machine (SMV) e finanças

Um dos algoritmos mais populares no âmbito do aprendizado de máquina, a SVM foi desenvolvido na década de 90, por meio dos trabalhos de Boser et al. (1992), e Vapnik (1995). Seu principal objetivo é determinar um hiperplano de separação que visa distinguir os dados em duas ou mais classes, visando atingir a separação máxima entre elas, conforme pode ser visto na figura 1 (Emir et al., 2012; Marcelino et al., 2015).

**Figura 1.** Maximização das margens pelo SVM

O SVM logo destacou-se pela capacidade de generalização, eficiência, versatilidade, facilidade em lidar com alta dimensionalidade dos dados e evitar *overfitting*, sendo utilizado para realizar tarefas de classificação não linear e linear, regressão, detecção de outliers, entre outros (Géron, 2019).

O trabalho de Fan e Palaniswami (2001) foi um dos pioneiros a testar a eficácia do SVM na seleção de portfólio, utilizando dados de ações negociadas na bolsa de valores australiana, durante o período de 1992 à 2000. O algoritmo foi treinado com indicadores financeiros de sete categorias: retorno de capital, lucratividade, alavancagem, crescimento, liquidez de curto prazo, retorno do investimento e risco. O portfólio selecionado pelo modelo obteve um retorno de 208%, enquanto o *benchmark* produziu um retorno acumulado no mesmo período de 71%, indicando utilidade do SVM para o problema de seleção de portfólio.

A pesquisa de Kim (2003) utilizou o SVM para previsão de séries temporais financeiras, comparando os resultados com os modelos de raciocínio baseado em casos e redes neurais *back-propagation*. O estudo utilizou 12 indicadores provenientes da análise técnica e procurou prever a direção das variações diárias no Korea Composite Stock Price Index (KOSPI). O período da amostra foi de janeiro de 1989 a dezembro 1998, totalizando 2.928 observações. Os resultados demonstraram que o SVM acertou 57,83% das previsões da direção das variações diárias do índice KOSPI, valor superior a acurácia apresentada pelos outros métodos, tendo a rede neural *back-propagation* uma acurácia de 54,73% e o algoritmo raciocínio baseado em casos uma acurácia de 51,97%. O estudo ainda aponta que o SVM utiliza o princípio da minimização de risco



estrutural, ocasionando uma melhor generalização dos dados temporais do que técnicas tradicionais.

O estudo de Emir et al. (2012) procurou construir um modelo financeiro ideal de seleção de portfólios utilizando redes neurais artificiais (RNA) e SVM. Os autores combinaram parâmetros da análise fundamentalista e da análise técnica, durante o período de 2002 a 2010, para as empresas da Bolsa de Valores de Istambul (ISE 30). Visando selecionar as variáveis que realmente são essenciais ao modelo, foi realizado o procedimento de redução de dimensionalidade na fase de pré-processamento dos dados. O procedimento resultou na seleção de 26 indicadores, sendo 14 provenientes da análise fundamentalista e 12 indicadores técnicos. Os resultados demonstraram que o modelo RNA apresentou uma acurácia média de 50,35%, já com o SVM a acurácia média foi de 66,19%, durante os anos de 2004 a 2010. A acurácia do modelo SVM também apresentou menor variabilidade (acurácia mínima de 63,33% e máxima de 70%), se comparado com o modelo RNA (acurácia mínima 32,5% e máxima de 67,5%), indicando que a máquina de suporte de vetor parece ser o modelo ideal para seleção de portfólios.

Huerta et al. (2013) procuraram testar se o SVM seria capaz de auxiliar na previsão do preço das ações, utilizando como *input* as informações contábeis e o histórico de cotações. Para isto, em vez de utilizarem toda a amostra, foi utilizado uma parcela inicial dos dados. Os autores selecionaram 20% das ações com maior e 20% das ações com pior retorno para treinar o modelo, durante o período de 1981 a 2010. Os resultados do estudo evidenciaram que o retorno anual do portfólio nos oito setores excedeu 15% (desconsiderando os custos de transação), já a volatilidade ficou abaixo dos 8%.

Em uma abordagem mais recente, Nti et al. (2020b) testaram a utilização da SMV aprimorada com o algoritmo genético (GASVM) para seleção de *features*, otimização de parâmetros e previsão do preço das ações da bolsa de valores de Gana. Posteriormente, os autores compararam a acurácia do modelo GASVM com as técnicas clássicas de AM. Os resultados demonstraram que o modelo SVM apresentou a maior acurácia entre os modelos estudados, acertando 93,7% das predições, contra 80,1% das redes neurais, 75,3% das árvores de decisão e 82,3% do *random forest*.

No cenário nacional, Monteiro e Ribeiro (2024) testaram algoritmos de AM para previsão do comportamento de três ações negociadas no mercado de ações brasileiro, nomeadamente VALE3, PTR4 e ITUB4. Os algoritmos de SVM, redes neurais e um modelo híbrido, utilizando os dois algoritmos alcançaram acurácia máxima de 54,5%, 51,8% e 54,2%, indicando um valor próximo ao aleatório e corroborando a hipótese do mercado eficiente. Um estudo abrangente de Henrique et al. (2018) propôs um modelo de regressão para cálculo do retorno de dezoito ações brasileiras, norte-a-

americanas e chinesas, com período de treinamento de 15 anos. Algoritmo de *support vector regression*, utilizando Kernel linear e atualizações periódicas resultou em erros menores que o modelo de caminhada aleatória para quase todos os preços diários de ações. O trabalho de Santos (2020) analisou o comportamento das ações VALE3, PETR4 e ITUB5 durante o período de cinco anos (2015 a 2020) por meio do algoritmo SVM, utilizando validação cruzada de 5 *folds*. Os resultados indicam que a acurácia média para as ações VALE3, PETR4 e ITUB5 foram de 62,33%, 63,66% e 61,33%, respectivamente.

Portanto, a literatura indica a eficácia do algoritmo SVM para tarefas de predição do comportamento e preço das ações e seleção de portfólios. O presente trabalho diferencia-se dos demais uma vez que busca testar a eficácia do SVM na predição do comportamento de ações negociadas em um mercado emergente e volátil, durante um período relativamente longo (27 anos), no qual incluí uma crise sanitária e econômica.

### 3. METODOLOGIA

#### 3.1. Fonte de dados, amostra e identificação das variáveis

Os dados da pesquisa foram obtidos na base Economática e o período de análise compreendeu de 18 de agosto de 1994 a 16 de dezembro de 2021, com periodicidade diária. A população deste estudo compreendeu todas as ações da bolsa de valores B3 (antiga BM&F BOVESPA). Destas, foram excluídas as ações que apresentavam dados faltantes ou incompletos durante o período de análise, resultando em 40.562 observações, pertencentes a seis empresas, conforme tabela 1.

**Tabela 1. Amostra do estudo**

Ação	Tipo	Empresa	Setor	Valor de mercado <sup>a</sup>	Volume médio <sup>b</sup>
<b>CMIG4</b>	PN	Companhia Energética de Minas Gerais	Energia e Saneamento	27,60	13,04
<b>LAME4</b>	PN	Lojas Americanas	Varejo	1,14	3,64
<b>PETR4</b>	PN	Petróleo Brasileiro S.A.	Petróleo e Gás	506,18	30,69
<b>UNIP6</b>	PN	Unipar Carbocloro	Petroquímico	6,09	0,12
<b>VALE3</b>	ON	Vale S.A.	Mineração	242,04	30,09
<b>VIVT3</b>	ON	Telefônica Brasil	Telecomunicações	81,06	2,19

Nota. De StatusInvest, de Statusinvest, 2025 (<https://statusinvest.com.br/acoes>).

<sup>a</sup> Valor em bilhões de reais (R\$)

<sup>b</sup> Volume médio representa a média em milhões de negociações diárias de cada ação, nos meses de janeiro, fevereiro e março de 2025.

Como é possível observar na tabela 1, as empresas que compõem a amostra deste estudo apresentam diferentes tamanhos e pertencem a setores de análise distintos. Estes fatores constituem uma dificuldade adicional no treinamento dos modelos, podendo prejudicar a generalização dos resultados, uma vez que variáveis setoriais e empresariais têm o poder de influenciar no comportamento da ação.

**3.1.1. Mensuração das variáveis.** As variáveis independentes provenientes das análises fundamentalistas foram calculadas com as informações das demonstrações financeiras (Balanço Patrimonial e Demonstração do Resultado do Exercício) e do histórico de emissão de ações. Na tabela 2 são apresentados os indicadores/siglas e classes de fatores (Haugen & Baker, 1996) e a fórmula utilizada para o cálculo.

Tabela 2. Variáveis da análise fundamentalista

Indicador (sigla)	Classes de fatores (Haugen & Baker, 1996)	Fórmula	
D/E	Risco	$D / E_t = \frac{Passivo\ Circulante_t + Passivo\ não\ circulante_t}{Patrimônio\ Líquido_t}$	(1)
Cobertura de juros (TIE)	Risco	$TIE_t = \frac{EBITDA_t}{Despesas\ Financeiras_t}$	(2)
Market capitalization (MCAP)	Liquidez	$MCAP_t = Núm\ ações\ Outstanding_t \times Preço\ Fechamento_t$	(3)
Trading volume (TV)	Liquidez	$TV_t = \frac{Volume_t}{Market\ Capitalization_t}$	(4)
Earning-to-price (E/P)	Nível de preço	$E/P_t = \frac{EBITDA_t}{Peço\ Fechamento_t}$	(5)
Sales-to-price (S/P)	Nível de preço	$S/P_t = \frac{Receita\ Operacional_t}{Preço\ Fechamento_t}$	(6)
P/B	Nível de preço	$P/B_t = \frac{Market\ Capitalization_t}{Patrimônio\ Líquido_t}$	(7)
Margem de lucro bruto (GPM)	Potencial de crescimento	$P/B_t = \frac{Market\ Capitalization_t}{Patrimônio\ Líquido_t}$	(8)
Capital turnover (CT)	Potencial de crescimento	$CT_t = \frac{Receita\ Operacional_t}{Ativo\ Total_t}$	(9)
RoA	Potencial de crescimento	$RoA_t = \frac{EBITDA_t}{Ativo\ Total_t}$	(10)
RoE	Potencial de crescimento	$RoE_t = \frac{EBITDA_t}{Patrimônio\ Líquido_t}$	(11)

As variáveis advindas da análise técnica foram compostas por nove indicadores, calculados utilizando a cotação diária (preço de fechamento ajustado a splits e dividendos) e o volume diário negociado das ações. A tabela 3 apresenta os indicadores/siglas, classes de fatores de acordo com Haugen e Baker (1996) e a fórmula utilizada para o cálculo de cada um dos indicadores.

**Tabela 3. Variáveis da análise técnica**

Indicador (sigla)	Classes de fatores (Haugen & Baker, 1996)	Fórmula	
<b>SMA</b> * Calculado para 3, 7, 15, e 30 dias	Histórico de preço	$MMSt = \frac{(P_{(t-1)} + P_{(t-2)} + \dots + P_n)}{n}$	(12)
<b>EMA</b> * Calculado para 5, 9, 12, 21, 26	Histórico de preço	$MME_t = P_{(t-1)} K + MME_{(t-1)} \times (1-K)$ Sendo $K = \frac{2}{n+1}$	(13)
<b>Força relativa (FR)</b>	Histórico de preço	$FR = \frac{U}{D}$ Sendo: U média das cotações dos últimos 14 dias em que o preço da cotação subiu e a média das cotações dos últimos 14 dias em que o preço da ação desceu	(14)
<b>RSI</b>	Histórico de preço	$IFR = 100 \frac{100}{1 + FR}$	(15)
<b>Dummies (RSI)</b>	Histórico de preço	Dummies indicando RSI > 70 ou RSI < 30	
<b>MACD</b>	Histórico de preço	$MACD_t = EMA12_t - EMA26_t$	(16)
<b>Dummy sinal tendência alta (MACD)</b>	Histórico de preço	Dummy indicando se o $MACD_t > EMA9_t$	
<b>On balance volume (OBV)</b>	Histórico de preço	$OBV_t = OBV_{t-1} + \{volume_t, \text{ se } P_{(t-1)} > P_{(t-2)}\} 0, \text{ se } P_t = P_{(t-1)} - volume_t, \text{ se } P_{(t-1)} < P_{(t-2)}$	(17)
<b>OBV/Volume</b>	Histórico de preço	$OBV / Volume_t = \frac{OBV_t}{Volume_t}$	(18)

Este estudo utilizou o comportamento do preço das ações como variável dependente dos modelos de classificação. Para o cálculo do comportamento do preço das ações foi utilizado dados diários do preço de abertura, diminuído do preço de fechamento ajustado a splits e dividendos. A variável dependente assume o valor 1 quando o preço de fechamento for maior ou igual ao preço de aberto e 0 quando o preço de fechado ajustado for menor que o preço de abertura.

$$\text{Comportamento da ação}_t = \text{Preço fechamento ajustado}_t - \text{Preço abertura}_t$$

$$\begin{cases} 1, \text{ se Preço de fechamento ajustado}_t \geq \text{Preço de abertura}_t \\ 0, \text{ se Preço de fechamento ajustado}_t < \text{Preço de abertura}_t \end{cases} \quad (19)$$

3.2. Modelos empíricos e medidas de desempenho

Foram elaborados três modelos empíricos, que utilizaram o algoritmo SMV na previsão do comportamento e retorno de ações negociadas no mercado brasileiro. Para facilitar o entendimento, um resumo das informações dos modelos foi apresentado na tabela 4.

Tabela 4. Resumo dos Modelos Empíricos

Modelo	Tarefa	Variável dependente	Variável independente
A	Classificação	Comportamento das ações	Variáveis fundamentalistas
B	Classificação	Comportamento das ações	Variáveis técnicas
C	Classificação	Comportamento das ações	Híbrido (fundamentalistas e técnicas)

A busca dos parâmetros ideias desempenham um papel crucial na construção de modelos de previsão de ações (Lee, 2009), portanto, a ferramenta *grid search* auxilia na definição dos valores ótimos para os parâmetros C, gamma além de testar funções *kernel* diferentes para o SVM. A otimização dos hiper parâmetros do modelo foram definidos testando os *kernels linear, rbf e sigmoid*, parâmetro C com os valores 0,1; 1; 2; 5; 10; 100; 1000; 10000; e 50000; e o parâmetro *gamma* com os valores 0,001; 0,01; 0,1; 0,5; 1; 1,5; 2; “auto” e “scale” (Lee, 2009; Li & Sun, 2020; Monteiro & Ribeiro, 2024). Tendo em vista que a variável dependente do modelo de classificação não apresenta desbalanceamento entre as classes, optou-se pela aplicação da validação cruzada com 5 *folds* (Lee, 2009).

O presente estudo utilizará como métrica principal a acurácia e métricas adicionais a precisão (relacionada ao erro tipo I – falsos positivos), revocação (erro tipo II – falsos negativos) e F1-score, devido ao balanceamento entre as classes da variável dependente do modelo de classificação, que podem ser calculadas conforme equações apresentadas na sequência:

$$\text{Acurácia} = \frac{\text{Verdadeiro Positivo} + \text{Verdadeiro negativo}}{\text{Verdadeiro Positivo} + \text{Verdadeiro Negativo} + \text{Falso Positivo} + \text{Falso Negativo}}$$

(20)

$$\text{F1 score} = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

(21)

$$\text{Precisão} = \frac{\text{Verdadeiro Positivo}}{\text{Verdadeiro Positivo} + \text{Falso Positivo}}$$

(22)

$$\text{Revocação (recall)} = \frac{\text{Verdadeiro Positivo}}{\text{Verdadeiro Positivo} + \text{Falso Negativo}}$$

(23)

### 3.3. Pré-processamento dos dados

Antes das etapas de treinamento e teste, os dados foram submetidos a etapa de pré-processamento, onde realizou-se a limpeza, compilação e normalização dos dados. Foram utilizados os *softwares* Excel e Python, para emprego das seguintes etapas: 1) Eliminação dos dados faltantes ou incompletos; 2) Separação da amostra em conjunto de treinamento do algoritmo (80%), contendo 32.450 observações, pertencentes ao período de agosto de 1994 a junho de 2016, e conjunto de teste (20%), contendo 8.112 observações relativamente ao período de junho de 1996 a dezembro de 2021; e 3) Normalização, utilizando o método *StandardScaler*, que transforma a distribuição dos dados de forma que a média fique igual a 0 e o desvio padrão igual a 1, conforme equação a seguir:

$$z = \frac{(x - \mu)}{s} \quad (24)$$

Onde  $\mu$  representa a média da amostra de treinamento e  $s$  o desvio padrão da mesma amostra.

## 4. RESULTADOS DA PESQUISA

A estatística descritiva das principais variáveis e indicadores, apresentada na tabela 5, demonstra a heterogeneidade da amostra, que é formada por empresas de diferentes tamanhos, volumes e preços das ações, níveis de endividamento e retornos das ações. Também é possível observar o balanceamento entre as classes da variável dependente.

**Tabela 5. Estatística descritiva**

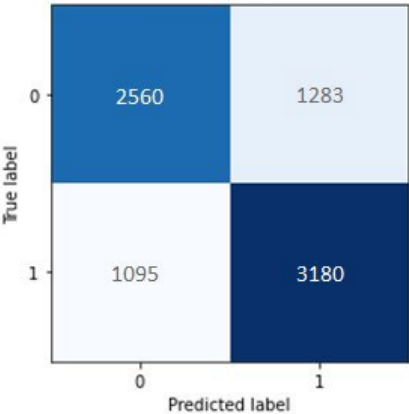
Variável/Indicador	Média	Desvio padrão	Valor mínimo	Valor máximo
Comportamento da ação	0,509	0,496	0	1
Valor da ação	9,49	13,02	0,01	106,26
Volume (em milhares)	151.920,38	424.734,80	0,35	10.989.848
Ativo (em milhares)	99.879,37	207.897,75	341,01	1.239.448,00
Debt-to-equity	1,41	2,05	0,10	12,51
Return on assets	0,02	0,03	-0,58	0,21

A primeira aplicação (modelo A) utilizou somente as variáveis fundamentalistas para treinamento do modelo, conforme tabela 1. Os resultados do modelo foram:

acurácia média 56,46%, medida F1 score igual a 68,95%, precisão média de 59,07% e revocação média de 82,78%. O modelo B fez uso de variáveis advindas da análise técnica para treinamento do algoritmo, conforme tabela 2. Este modelo apresentou os seguintes resultados: acurácia média 55,72%, medida F1-score igual a 67,98%, precisão média de 58,83% e revocação média de 80,50%.

O modelo empírico C aplicou o algoritmo SVM na previsão do comportamento de ações do mercado brasileiro, utilizando a combinação da análise técnica e análise fundamentalista. O modelo de classificação apresentou a média de acurácia nos cinco *folds* de 70,7%, medida F1-score igual a 72,74%, precisão média de 71,25% e revocação média de 74,33. A figura 2 ilustra a matriz de confusão do modelo de classificação.

**Figura 2.** Matriz de confusão do modelo de classificação



Dentre as 40.562 observações que constituem a amostra, foi utilizado 80% para treinamento do algoritmo e 20% para teste. Pode-se observar que dos 8.118 dados de teste, o modelo conseguiu classificar corretamente 5.760, isto é, o algoritmo SMV treinado com variáveis técnicas e fundamentalistas, conseguiu prever se o preço da ação iria subir, se manter ou descer em 70,7% das vezes.

5. DISCUSSÃO

Nos resultados dos modelos empíricos foi possível verificar que o treinamento com variáveis provenientes somente da análise fundamentalista (modelo A) teve um resultado ligeiramente superior, se comparado com o modelo que utilizou somente variáveis da análise técnica (acurácia do modelo A = 0,564; acurácia do modelo B = 0,557). Apesar da pequena diferença no poder preditivo dos modelos, os resultados foram na contramão do que é indicado na literatura. Lui e Mole, (1998), Oliveira et al. (2013),

e Renu e Christie (2018) destacam que a análise fundamentalista é ideal para investimentos a médio e longo prazo e que a análise técnica conseguiria captar melhor as alterações de curto prazo, como é o caso das variações diárias do preço das ações objeto deste estudo.

Cabe ressaltar também que o desempenho dos dois primeiros modelos foi mais baixo do que indicado pela literatura (Emir et al., 2012; Kim, 2003; Nti et al., 2020b; Zhang & Zhao, 2009), indicando que o algoritmo conseguiu aprender pouco com os dados e generalizar pouco os resultados, tendo sua acurácia próxima ao aleatório. Possíveis explicações para estes resultados, reside na composição heterogênea de empresas que constituem a amostra na qual as diferenças setoriais e organizacionais podem ter prejudicado a generalização dos resultados e na inclusão de períodos de crise na etapa de teste do modelo. Parte da literatura tentou isolar os efeitos das variáveis externas, por exemplo, Huerta et al. (2013) construíram um modelo diferente para cada um dos setores do mercado americano, Nti et al. (2020b) separaram a análise de empresas dos setores de petróleo e bancos. Já Emir et al. (2012) excluiu da sua análise os anos 2000 a 2001, devido ao período de crise financeira ocorrida na Turquia. No entanto, em comparação ao cenário nacional, os modelos A e B tiveram acurácia similar ao estudo de Monteiro e Ribeiro (2024), o que pode indicar a dificuldade de previsão do comportamento das ações em mercados emergentes, voláteis e instáveis, como o brasileiro.

Quando combinados os dados das análises técnica e fundamentalista (modelo C), o modelo ganha poder de predição e capacidade de generalização dos resultados, atingindo uma acurácia média de 70,7% para os 5 *folds*. Este resultado vem ao encontro da revisão sistemática de Nti et al. (2020a), na qual aponta que modelos híbridos podem fornecer uma melhor acurácia na previsão das ações. O modelo C também apresentou resultados melhores que a maioria dos estudos da revisão da literatura (Emir et al., 2012; Kim, 2003; Zhang & Zhao, 2009), o que pode indicar a superioridade de modelos híbridos para previsão do comportamento das ações.

Em relação aos erros de classificação, a medida de precisão, relacionada ao erro tipo I foi de 59% para o modelo A, 58,8% para o B, e 71,25% para o modelo C, indicando que o modelo C possui a maior capacidade de evitar falsos positivos. Já a medida de revocação (*recall*) relacionada ao erro tipo II, foi de 82,7%, 80,5% e 74,3% para os modelos A, B e C, respectivamente. Estes resultados demonstram que apesar de apresentar a melhor acurácia e precisão, o modelo C possui a menor medida de identificação correta dos verdadeiros positivos. No entanto, como os modelos de classificação desta pesquisa visam prever o comportamento das ações (aumento ou diminuição



do preço da ação), as medidas de falsos positivos e falsos negativos são igualmente importantes, portanto, consideramos também a medida F1-score. Esta medida é uma média ponderada da precisão e revocação e leva em consideração tanto os falsos positivos quanto os falsos negativos. Os valores de F1-score para os modelos A, B e C foram de 68,9%, 67,9% e 72,7%, respectivamente. Novamente, o modelo C (híbrido) apresentou as melhores medidas de erro.

Portanto, considerando as medidas de desempenho para a tarefa de classificação (A, B e C), verifica-se que o modelo C, treinado para a tarefa de classificação, com variáveis híbridas (advindas da combinação das análises técnica e fundamentalista) apresenta a melhor acurácia, precisão e medida F1-score, podendo assim, ser uma ferramenta útil para auxiliar os investidores no processo de previsão do comportamento das ações e seleção de portfólio.

## 6. CONCLUSÕES, LIMITAÇÕES E RECOMENDAÇÕES

Este trabalho teve como objetivo investigar a aplicação de algoritmos de aprendizagem de máquina na previsão do comportamento de ações no mercado brasileiro. Para isto, foram testados três modelos distintos de AM, buscando prever o comportamento de seis ações negociadas no mercado brasileiro, baseados em análises técnicas, fundamentalistas e híbridas. A amostra compreendeu seis empresas, durante o período de 27 anos (de 18 de agosto de 1994 a 16 de dezembro de 2021), totalizando 40.562 observações.

Os modelos treinados somente com variáveis fundamentalistas (A) e somente com variáveis técnicas (B) apresentaram acurácia próximo ao aleatório, respectivamente de 56,46% e 55,72%. Estes resultados podem indicar que a heterogeneidade da amostra, que compreendeu empresas de diversos tamanhos e setores, pode ter contribuído para a falta de aprendizado do algoritmo, prejudicando a generalização dos resultados. Conforme indicado pela literatura, esperava-se que o modelo com variáveis provenientes da análise técnica (B) produzisse melhores resultados, tendo em vista que esta análise consegue captar melhor as alterações de curto prazo (Lui & Mole, 1998), como é o caso das variações diárias do preço das ações, objeto deste estudo (Oliveira et al., 2013; Renu & Christie, 2018).

Com a combinação das variáveis provenientes das análises fundamentalistas e técnicas, foi possível notar uma melhora significativa no desempenho do modelo C. Este resultado vem ao encontro da literatura (Nti et al., 2020a), ao apontar que modelos híbridos podem fornecer uma melhor acurácia na previsão das ações. O modelo C também apresentou resultados melhores que a maioria dos estudos da revisão da

literatura (Emir et al., 2012; Kim, 2003; Zhang & Zhao, 2009), podendo indicar que o algoritmo SMV também pode ser utilizado para a realidade brasileira e para períodos de crises, como a pandemia da covid-19.

Considerando as medidas de desempenho dos modelos de classificação, concluiu-se que o modelo C, treinado para a tarefa de classificação, com variáveis híbridas (advindas da combinação das análises técnica e fundamentalista) apresenta a melhor acurácia, F1-score e precisão (medida relacionada ao erro tipo I). O que indica que este modelo pode ser uma ferramenta útil para auxiliar os investidores no processo de previsão do comportamento das ações e seleção de portfólio.

Ao identificar padrões ocultos nos dados e aumentar a precisão das previsões, o presente trabalho auxilia no processo de seleção de portfólio dos investidores, ao propor um modelo de aprendizado de máquina para previsão do comportamento das ações negociadas no mercado brasileiro. Desta forma, contribuí para a prática de diversos profissional da área financeira. Gestores de fundos e *traders* podem utilizar o modelo para maximizar o retorno e ajustar o risco da carteira de investimentos, desta forma, otimiza a composição do portfólio. Os analistas financeiros podem utilizar modelos de AM para gerar *insights* baseados nas variáveis advindas da análise técnica e fundamentalista para complementar suas análises e aumentar a eficiência das suas recomendações de investimentos. Modelos de AM para previsão do comportamento das ações também podem democratizar o acesso a estratégias de investimento, considerando os diversos perfis de investimentos. Enquanto indivíduos conservadores podem utilizar este trabalho para evitar ativos voláteis e buscar reduzir os riscos, os agressivos podem explorar oportunidades identificadas pelo modelo de previsão, maximizando o retorno da sua carteira.

Além das contribuições práticas o presente trabalho apresenta outros pontos que se destacam na literatura, como é o caso da utilização de técnicas híbridas para a previsão do comportamento do mercado de ações e do amplo período de análise dos dados (27 anos, n=40.562), que auxiliaram no treinamento e generalização do modelo.

Apesar dos cuidados na elaboração deste estudo, ele não está isento de limitações. A seleção de uma amostra heterogênea, contendo empresas de diferentes tamanhos e setores, pode ter prejudicado o treinamento do modelo. Outro fator externo que pode ter atrapalhado a aprendizagem do algoritmo foi a inclusão de períodos de crise na amostra (Emir et al., 2012), como a pandemia da covid-19, que pode ter alterado o comportamento das ações e diminuído acurácia dos modelos de predição. A utilização de uma única base de dados para definição da amostra e mensuração das variáveis também constituem uma limitação, uma vez que foram analisadas somente

empresas que possuíam os dados completos no banco de dados. O período de análise dos dados (vinte e sete anos) também pode ser interpretado como uma limitação do estudo, uma vez que pode diminuir o número de empresas analisadas e acarretar em um viés de sobrevivência na amostra.

Como recomendações para trabalhos futuros sugere-se a ampliação do escopo da pesquisa, de forma a utilizar o algoritmo de AM para composição de portfólios de investimento considerando as diferentes estratégias de investidores (*buy and hold*, *swing trade*, *value investment*) e perfis de investidores (conservadores/agressivos). Também sugere-se a diversificação das variáveis que possam influenciar o preço da ação como, por exemplo, a análise de sentimento das notícias vinculadas na mídia e nas redes sociais. Por fim, recomenda-se também a utilização de outros AM como, redes neurais e métodos *ensembles*.

#### **Contribuição de autoria:**

**Costa, G. D.:** Conceituação, Metodologia, Software, Pesquisa, Curadoria de dados, Escrita-rascunho original. **Lunkes, R. J.:** Conceituação, Metodologia, Validação, Análise Formal, Pesquisa, Escrita-rascunho original, Escrita, revisado e edição, Supervisão

Gabriel Donadio Costa (Costa, G. D.)

Rogério João Lunkes (Lunkes, R. J.)

#### **Declaração de conflito de interesse**

Os autores declaram que durante o processo de pesquisa, não existiu nenhum tipo de interesse pessoal, profissional ou econômico que tenha podido influir no julgamento e/ou ações dos pesquisadores no momento de elaborar e publicar o artigo.

## **REFERÊNCIAS**

- Adebiyi, A. A., Ayo, C. K., Adebiyi, M., & Otokiti, S. O. (2012). Stock price prediction using neural network with hybridized market indicators. *Journal of Emerging Trends in Computing and Information Sciences*, 3(1), 1-9.
- Anghel, D. G. (2021). A reality check on trading rule performance in the cryptocurrency market: Machine learning vs. technical analysis. *Finance Research Letters*, 39, 1-8. <https://doi.org/10.1016/j.frl.2020.101655>
- Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2), 1046-1089. <https://doi.org/10.1093/rfs/hhaa062>

- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In L. Valiant and M. Warmuth (Eds.), *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144-152). Association for Computing Machinery. <https://doi.org/10.1145/130385.13040>
- Brasil, Bolsa, Balcão. (2024). Perfil pessoas físicas: Nossos investidores. [https://www.b3.com.br/pt\\_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/](https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/).
- Buckmann, M., Joseph, A., & Robertson, H. (2021, 15-16 de junho). *An interpretable machine learning workflow with an application to economic forecasting* [Apresentação em conferência]. 11th ECB Conference on Forecasting Techniques, virtual. <https://doi.org/10.2139/ssrn.4130517>
- Cavalcante, F., Misumi, J. Y., & Rudge, L. F. (2005). *Mercado de capitais: O que é, como funciona*. Elsevier.
- Chen, J. C., Dunn, A., Hood, K., Driessen, A., & Batch, A. (2019). Off to the races: A comparison of machine learning and alternative data for predicting economic indicators. In K. G. Abraham, R. S. Jarmin, B. Moyer, & M. D. Shapiro (Eds.), *Big data for 21st century economic statistics* (pp. 373 – 402). University of Chicago Press.
- Chenhall, R. H., & Moers, F. (2015). The role of innovation in the evolution of management accounting and its integration into management control. *Accounting, Organizations and Society*, 47, 1-13. <https://doi.org/10.1016/j.aos.2015.10.002>
- Dang, N. H., Van Vu, T. T., & Le Dao, T. N. (2022). Accounting information and stock returns in Vietnam securities market: Machine learning approach. *Contabilidad y Negocios*, 17(33), 94-118. <https://doi.org/10.18800/contabilidad.202201.004>
- Elder, A. (1993). *Trading for a living: Psychology, trading tactics, money management* (Vol. 31). John Wiley & Sons.
- Emir, S., Dinçer, H., & Timor, M. (2012). A stock selection model based on fundamental and technical analysis variables by using artificial neural networks and support vector machines. *Review of Economics & Finance*, 2(3), 106-122. <https://doi.org/10.18800/contabilidad.202201.004>
- Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1), 3-25. [https://doi.org/10.1016/0304-405X\(88\)90020-7](https://doi.org/10.1016/0304-405X(88)90020-7)
- Fama, E. F., & French, K. R. (1992). The cross section of expected stock returns. *The Journal of Finance*, 47(2), 427-465. <https://doi.org/10.1111/j.1540-6261.1992.tb04398.x>
- Fama, E. F., & French, K. R. (2004). The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives*, 18(3), 25-46. <https://doi.org/10.1257/0895330042162430>

- Fan, A., & Palaniswami, M. (2001, July). Stock selection using support vector machines. In Institute of Electrical and Electronics Engineers (Ed.), *IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222)* (Vol. 3, pp. 1793-1798). IEEE. <https://doi.org/10.1109/IJCNN.2001.938434>
- Feijó, C., Araújo, E. C., & Bresser-Pereira, L. C. (2022). Política monetária no Brasil em tempos de pandemia. *Brazilian Journal of Political Economy*, 42, 150-171. <https://doi.org/10.1590/0101-31572022-3353>
- Géron, A. (2019). *Mãos à obra: Aprendizado de máquina com Scikit-Learn & TensorFlow*. Alta Books.
- Haugen, R. A., & Baker, N. L. (1996). Commonality in the determinants of expected stock returns. *Journal of Financial Economics*, 41(3), 401-439. [https://doi.org/10.1016/0304-405X\(95\)00868-F](https://doi.org/10.1016/0304-405X(95)00868-F)
- Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of Finance and Data Science*, 4(3), 183-201. <https://doi.org/10.1016/j.jfds.2018.04.003>
- Huerta, R., Corbacho, F., & Elkan, C. (2013). Nonlinear support vector machines can systematically identify stocks with high and low future returns. *Algorithmic Finance*, 2(1), 45-58. <https://doi.org/10.3233/AF-13016>
- Kamara, A. F., Chen, E., & Pan, Z. (2022). An ensemble of a boosted hybrid of deep learning models and technical analysis for forecasting stock prices. *Information Sciences*, 594, 1-19. <https://doi.org/10.1016/j.ins.2022.02.015>
- Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2), 307-319. [https://doi.org/10.1016/S0925-2312\(03\)00372-2](https://doi.org/10.1016/S0925-2312(03)00372-2)
- King, B. F. (1964). *The latent statistical structure of security price changes* [Unpublished PhD dissertation]. University of Chicago.
- Lee, M. C. (2009). Using support vector machine with a hybrid feature selection method to the stock trend prediction. *Expert Systems with Applications*, 36(8), 10896-10904. <https://doi.org/10.1016/j.eswa.2009.02.038>
- Levy, R. A. (1967). Relative strength as a criterion for investment selection. *The Journal of Finance*, 22(4), 595-610. <https://doi.org/10.2307/2326004>
- Li, X., & Sun, Y. (2020). Stock intelligent investment strategy based on support vector machine parameter optimization algorithm. *Neural Computing and Applications*, 32(6), 1765-1775. <https://doi.org/10.1007/s00521-019-04566-2>
- Lui, Y. H., & Mole, D. (1998). The use of fundamental and technical analyses by foreign exchange dealers: Hong Kong evidence. *Journal of International Money and Finance*, 17(3), 535-545. [https://doi.org/10.1016/S0261-5606\(98\)00011-4](https://doi.org/10.1016/S0261-5606(98)00011-4)

- Marcelino, S., Henrique, P. A., & Albuquerque, P. H. M. (2015). Portfolio selection with support vector machines in low economic perspectives in emerging markets. *Economic Computation & Economic Cybernetics Studies & Research*, 49(4), 261-278.
- Monteiro, C. A., & Ribeiro, M. H. (2024, July). Previsão da direção do preço de ação brasileira utilizando redes neurais e máquinas de vetores suporte. In Sociedade Brasileira de Computação (Ed.), *2024: Anais do XI Encontro Nacional de Computação dos Institutos Federais* (pp. 98-105). Sociedade Brasileira de Computação. <https://doi.org/10.5753/encompif.2024.1983>
- Mun, W. C. P. & Soong, V. (2021). Forecasting yield curve with macro-driven models: A comparison between machine learning and traditional statistical approaches. <https://rpubs.com/WendyChongPooiMun/YieldCurve>
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020a). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, 53(4), 3007-3057. <https://doi.org/10.1007/s10462-019-09754-z>
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020b). Efficient stock-market prediction using ensemble support vector machine. *Open Computer Science*, 10(1), 153-163. <https://doi.org/10.1515/comp-2020-0199>
- Oliveira, F. A., Nobre, C. N., & Zárte, L. E. (2013). Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index—Case study of PETR4, Petrobras, Brazil. *Expert Systems with Applications*, 40(18), 7596-7606. <https://doi.org/10.1016/j.eswa.2013.06.071>
- Renu, I. R. & Christie, P. (2018). Fundamental Analysis versus technical analysis. A comparative review. *International Journal of Recent Scientific Research*, 9(1), 23009-23013.
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229. <https://doi.org/10.1147/rd.33.0210>
- Santos, G. C. (2020). *Algoritmos de machine learning para previsão de ações da B3* [Dissertação de mestrado, Universidade Federal de Uberlândia]. Repositório Institucional - Universidade Federal de Uberlândia. <https://repositorio.ufu.br/handle/123456789/29897>
- Shatshat, M. A. H. I., & Ahmed, S. (2019). Information technology governance linkage to the financial report quality in Libyan commercial banks. *International Journal of Business Society*, 3(10), 13-17.

- Silva, A., Neves, R., & Horta, N. (2015). A hybrid approach to portfolio composition based on fundamental and technical indicators. *Expert Systems with Applications*, 42(4), 2036-2048. <https://doi.org/10.1016/j.eswa.2014.09.050>
- Statusinvest. (2025). *StatusInvest*. <https://statusinvest.com.br/acoes>
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer. <https://doi.org/10.1007/978-1-4757-2440-0>
- Zhang, Z., & Zhao, Q. (2009). The application of SVMs method on exchange rates fluctuation. *Discrete Dynamics in Nature and Society*, 2009(1), 1-8. <https://doi.org/10.1155/2009/250206>
- Zhang, X., Zhang, Y., Wang, S., Yao, Y., Fang, B., & Philip, S. Y. (2018). Improving stock market prediction via heterogeneous information fusion. *Knowledge-Based Systems*, 143, 236-247. <https://doi.org/10.1016/j.knosys.2017.12.025>

Data de recepção: 06/11/2024

Data de revisão: 27/11/2024

Data de aceitação: 24/04/2025

Contato: rogerio.lunkes@ufsc.br