

APLICAÇÃO DO ALGORITMO *K-NEAREST NEIGHBORS* (K-NN) PARA PREVISÃO DE ÍNDICES EM MERCADOS FINANCEIROS¹

Ewerton Alex Avelar²
Simone Evangelista Fonseca
Antônio Artur de Souza
Luís Henrique Marques Duarte
Vanderson Ribeiro Gonçalves

RESUMO

Esta pesquisa analisa se os preços históricos de negociação dos principais índices das maiores bolsas de valores do mundo explicam o retorno futuro dos ativos, por meio da aplicação do algoritmo de Aprendizado de Máquina (*Machine Learning* – ML) *K-Nearest Neighbors* (k-NN). Para tal, utilizou-se uma amostra composta pelas cotações diárias de 37 índices de tais bolsas no período de 2001 a 2019. Os modelos foram estimados por índice, conforme os preços de fechamento e de máximo, assim como o período completo e a divisão em subperíodos, cujos resultados se apresentaram superiores aos do desempenho médio do mercado. Com base na cotação máxima dos índices, os modelos obtiveram desempenho superior aos dos preços de fechamento, e os de subperíodos conseguiram melhores desempenhos. A eficiência dos mercados sob a forma fraca foi questionada no contexto contemporâneo de forte ascensão de algoritmos de ML para a previsão em finanças. Nesse contexto, os principais índices das maiores bolsas de valores do mundo foram analisados, com a obtenção de subsídios gerais que podem auxiliar na orientação de pesquisas futuras na área, em que o tema é relevante devido a contribuições sobre o uso de algoritmos de ML na previsão de preços de ativos de investimento.

Palavras-chave: *k-Nearest Neighbors*. *Machine Learning*. Previsão de Preços. Hipótese de Mercados Eficientes. Bolsa de Valores.

APPLICATION OF THE ALGORITHM *K-NEAREST NEIGHBORS* (K-NN) FOR FORECASTING INDEXES IN FINANCIAL MARKETS

ABSTRACT

This research analyzes whether the historical trading prices of the main indexes of the largest stock exchanges in the world explain the future return of assets, through the application of Machine Learning (ML) algorithm *K-Nearest Neighbors* (k-NN). To this end, a sample

¹ **Como citar este trabalho:** AVELAR, Ewerton Alex; FONSECA, Simone Evangelista; SOUZA, Antônio Artur de; DUARTE, Luís Henrique Marques; GONÇALVES, Vanderson Ribeiro. Aplicação do algoritmo *k-Nearest Neighbors* (k-NN) para previsão de índices em mercados financeiros. *Forscience*, Formiga, v. 12, n. 1, e01189, jan./jun. 2024. DOI: [10.29069/forscience.2024v12n1.e1189](https://doi.org/10.29069/forscience.2024v12n1.e1189).

² **Autor correspondente:** Ewerton Alex Avelar, e-mail: ewertonaavelar@gmail.com

composed of the daily quotations of 37 indexes of such exchanges in the period from 2001 to 2019 was used. The models were estimated by index, according to the closing and maximum prices, as well as the full period and the division into subperiods, whose results were superior to the average market performance. Based on the maximum quotation of the indexes, the models obtained superior performance than the closing prices, and those ones of subperiods achieved better performance. The efficiency of markets in the weak form has been questioned in the contemporary context of strong rise of ML algorithms for forecasting in finance. In this context, the main indexes of the largest stock exchanges in the world were analyzed, obtaining general subsidies that can help guide future research in the area, in which the topic is relevant due to contributions on the use of ML algorithms in predicting the prices of investment assets.

Keywords: k-Nearest Neighbors. Machine Learning. Price Forecasting. Efficiency Market Hypothesis. Stock Exchanges.

1 INTRODUÇÃO

O mercado de ações consiste em um mecanismo relevante para o levantamento de fundos por empresas de capital aberto, ao oferecer oportunidades para investidores e favorecer o desenvolvimento econômico-financeiro do país. Com os avanços da computação, o uso de algoritmos computacionais em finanças tem sido recorrente, com a respectiva produção de conhecimentos teóricos e práticos acerca desse mercado, sobretudo em termos da previsibilidade de preços dos ativos (WANG *et al.*, 2021; PUJITHA; MAHESH, 2021; KUMAR *et al.*, 2018). Demirel, Çam e Ünlü (2021) apontam que esse tipo de mercado é volátil, não linear, complicado, adaptativo, não paramétrico e imprevisível por natureza; logo, a previsão dos preços é um desafio para os agentes econômicos (especuladores, investidores, empresas, dentre outros), que buscam previsões acuradas para obter maior lucratividade.

No que tange ao comportamento dos preços dos ativos de investimento, é relevante destacar a vertente teórica da Hipótese de Mercados Eficientes (HME), na qual não é possível prever os preços de ações de forma a obter retornos acima da média considerados anormais, pois refletem informações dos ativos, conforme o nível de eficiência do mercado (FAMA, 1970; 1991). Shynkevich *et al.* (2017) ressaltam que, em sua forma fraca, por exemplo, a HME considera que os preços de mercado não poderiam ser previstos e permitir retornos anormais com base em dados históricos, pois estariam refletidos nos preços dos ativos.

Shynkevich *et al.* (2017) enfatizam o emprego de ferramentas sofisticadas por parte dos agentes econômicos para fins de previsão de preços de ativos no mercado financeiro, em especial, algoritmos de aprendizagem de máquina (*Machine Learning* – ML), a exemplo do k-

NN (*K-Nearest Neighbors* ou “k-vizinhos mais próximos”). Nesse sentido, Gite *et al.* (2021) ressaltam que o escopo da análise de preços das ações depende da capacidade em reconhecer os movimentos (aumentos ou reduções) com base em fundamentos técnicos e na compreensão das tendências ocultas do mercado. Convém salientar que os algoritmos de ML podem indicar previsões mais precisas no mercado financeiro, em que é preciso compreender a diversidade entre os mercados (SHYNKEVICH *et al.*, 2017).

Diante do exposto, este artigo objetivou analisar se preços históricos de negociação elucidam o retorno futuro dos principais índices das maiores bolsas de valores do mundo, por meio do uso do algoritmo k-NN. Para tanto, foram propostos e cumpridos os seguintes objetivos específicos: (i) verificar se diferentes tipos de preços históricos de negociação podem explicar o retorno de mercado do dia seguinte; (ii) mensurar o desempenho do algoritmo k-NN em diferentes contextos; e (iii) identificar se diferentes períodos de análise podem ser associados ao desempenho do algoritmo.

O k-NN consiste em um algoritmo de ML não paramétrico utilizado para a previsão de categorias de conjuntos de dados (FUKUNAGA; NARENDRA, 1975); identificar tendências, volatilidades e impulsos do mercado de ações (FACELI *et al.*, 2021; KUMAR *et al.*, 2018; CAVALCANTE *et al.*, 2016; ZHAO; SUN; WANG, 2014); e compreender os mercados, devido a inferências futuras de preços e retornos dos ativos de investimento (PUJITHA; MAHESH, 2021; WANG *et al.*, 2021). Para atingir o objetivo da pesquisa, tal algoritmo foi aplicado em uma amostra composta pelas cotações diárias de 37 índices das maiores bolsas de valores do mundo, no período de 2001 a 2019. Os modelos foram estimados por índice, em consonância aos preços de fechamento e de máximo, assim como ao período completo e à sua divisão em subperíodos.

Nesses termos, o presente artigo se justifica por aspectos acadêmicos e práticos, além das seguintes contribuições: (i) importância de ferramentas de previsão para os agentes do mercado de ações (DEMIREL; ÇAM; ÜNLÜ, 2021); (ii) foco do estudo em mercados distintos no mundo, e não apenas os individuais, algo realizado pela maioria dos estudos prévios; (iii) contribuição à literatura sobre a crescente automatização de processos em finanças (RUNDO *et al.*, 2019); e (iv) teste da HME (FAMA, 1970; 1991), em sua forma fraca e em contextos variados. Assim, contribui-se para esse campo de pesquisa que pode ser compreendido de modo consistentemente dinâmico (GITE *et al.*, 2021).

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Hipótese de Mercados Eficientes (HME) e o preço dos ativos de investimento

A HME está ligada à concepção de que os preços dos ativos no mercado de capitais reagem às informações sobre eles (FAMA, 1970). Com isso, as formas (níveis) de eficiência dos mercados se classificam como fraca, na qual o preço corrente de um ativo reflete a informação passada sobre ele; semiforte, em que o valor incorporaria a tal informação e, ainda, a informações públicas sobre ele, disponíveis de modo igualitário a todos os agentes de mercado; e forte, com as informações passadas, públicas e privadas sobre o ativo e refletidas no preço (FAMA, 1970; 1991).

Diante disso, a literatura tem discutido acerca da HME em termos teóricos e empíricos. Geralmente, os economistas contestam tal vertente sob a perspectiva de que há certa previsibilidade nos preços de ações, proveniente de fatores psicológicos e comportamentais, embora os padrões de oscilações não sejam estáveis ao longo do tempo e rompidos, por exemplo, pelas crises financeiras (LIM; BROOKS, 2011; BALL, 2009; LO, 2005; MALKIEL, 2003; 2005; TIMMERMANN; GRANGER, 2004).

De acordo com Borges (2010), testes sobre a forma de eficiência fraca apontam que, dentre os índices de mercado de ações, em países da Europa, não há resultados suficientes para rejeitar a HME na Alemanha e Espanha (inclusive, o mercado espanhol se mostrou o mais eficiente). Por sua vez, nos mercados do Reino Unido e da França, existe forte tendência de reversão de preços à média, e Grécia e Portugal, após 2003, se aproximam à aleatoriedade. Para o caso do Brasil, os estudos tendem a apontar que o mercado reage sob o viés semiforte de eficiência (BATISTA; MAIA; ROMERO, 2018; GABRIEL; RIBEIRO; RIBEIRO, 2013; FREIRE; MAIA, 2012; CAMARGOS; ROMERO, 2006).

Apesar de previsões contrárias derivadas da HME, os agentes de mercado e pesquisadores tentam sistematicamente propor modelos para antecipar os movimentos dos preços de ativos nos mercados de capitais, mediante o avanço das técnicas computacionais de aprendizado (CAVALCANTE *et al.*, 2016; ZHAO; SUN; WANG, 2014). Destaca-se que os modelos para previsão de movimentos desses mercados consistem em um desafio em termos de aplicação prática e desenvolvimento de pesquisas; por conseguinte, aqueles fundamentos em ML são propostos em conformidade aos dados estruturados e não estruturados que, respectivamente, possuem um padrão fixo de disposição e/ou qualquer forma prescrita de

organização (AWAN *et al.*, 2021; KIESEL; MANZ; SCHIERECK, 2020; GANDHMAL; KUMAR, 2019; JEON; HONG; CHANG, 2018).

2.2 Emprego de *Machine Learning* (ML) e *k-Nearest Neighbors* (k-NN) em finanças

Estudos que abordam o uso de algoritmos de ML em finanças são relevantes para o avanço do conhecimento científico do campo em termos teóricos e práticos, por favorecerem o aprendizado e pelo fato de tal emprego ainda ser considerado incipiente na área financeira (DIXON; HALPERIN; BILOKON, 2020; GHODDUSI; CREAMER; RAFIZADEH, 2019). Algoritmos favorecem o tratamento de dados complexos, com o intuito de reforçar a percepção dos agentes de mercado e construir modelagens de mercados dinâmicos (DIXON; HALPERIN; BILOKON, 2020).

Dentre os algoritmos de classificação utilizados para prever movimentos dos mercados de capitais se destaca o k-NN associado a análises de eficiência (PUJITHA; MAHESH, 2021; KUMAR *et al.*, 2018; CHEN; HAO, 2017; ZHAO; SUN; WANG, 2014; ALKHATIB *et al.*, 2013; WANG *et al.*, 2013). Esse algoritmo de ML é não paramétrico e com previsão de categorias, em que classifica as observações conforme um conjunto conhecido de dados (FUKUNAGA; NARENDRA, 1975) e pode avaliar e prever os movimentos dos preços dos ativos (WANG *et al.*, 2013). Em finanças, o k-NN pode direcionar as escolhas dos investidores, pois favorece a compreensão da estrutura dinâmica e complexa das bolsas de valores e contribui para a generalização de resultados (PUJITHA; MAHESH, 2021; WANG *et al.*, 2021).

O k-NN tem sido utilizado com frequência na previsão de preços de ações, cujos resultados também são comparáveis aos de demais algoritmos de ML e aprendizado supervisionado, tais como Máquina de Vetores de Suporte (*Support Vector Machine – SVM*), *Random Forest* e *Naive Bayes* (sem traduções para o português) (KUMAR *et al.*, 2018; LI; LI; LI, 2018). O k-NN possibilita conhecer o espaço gráfico formado por padrões de comportamento dos preços de ativos; obter uma percepção geral dos padrões formados; e desenvolver modelos de previsão de preços (CAVALCANTE *et al.*, 2016; LI; LI; LI, 2018).

Por um lado, Faceli *et al.* (2021) apresentam algumas vantagens do k-NN em relação a outros algoritmos de ML: (a) o treinamento é simples (memoriza apenas objetos); (b) há aproximações locais da função-objetivo diferentes para cada dado a ser classificado; (c) aplicável mesmo em problemas complexos; e (d) naturalmente incremental, ou seja, quando

novos exemplos de treinamento estão disponíveis, é necessário apenas os armazenar na memória. Por outro lado, o referido algoritmo indica alguns aspectos negativos, tais como: (i) falta de representação compacta dos objetos; (ii) presença de atributos redundantes e irrelevantes; e (iii) dimensionalidade dos exemplos. Para fins de classificação, tal algoritmo é normalmente mensurado por meio da acurácia obtida conforme a Equação 1 adaptada de Cao *et al.* (2019):

$$\text{Acurácia} = \frac{N^{\circ} \text{ de previsões corretas}}{N^{\circ} \text{ de observações}} \times 100 \quad (1)$$

Na sequência, será abordado o k-NN nos investimentos realizados nas maiores bolsas de valores do mundo.

2.3 Uso do k-Nearest Neighbors (k-NN) na previsão de retornos de ativos de investimento

Pesquisas recentes sobre aplicações do k-NN para prever retornos de índices de bolsas de valores, apenas a partir do uso de dados históricos de negociação, apontam níveis de acurácias/acertos em torno de 60%, considerado desempenho satisfatório para tal tarefa e similar ao de outros algoritmos de Inteligência Artificial (IA) (QIAN; RASHEED, 2007; CARTA *et al.*, 2021). Wu, Coleman e Bawuah (2020), por exemplo, a partir da aplicação do k-NN em diferentes cenários, obteve valores de acurácia que variaram entre 49,9% (valor mínimo) a 70,9% (valor máximo). Por sua vez, Moon, Jun e Kim (2018) verificaram uma acurácia pouco superior a 60,0% em sua aplicação do k-NN. Por fim, destaca-se ainda que Cao *et al.* (2019), avaliaram a aplicação do modelo com diferentes índices de ações e observaram acurácias entre 70,59% (valor mínimo) e 74,83% (valor máximo).

Diante do exposto, o presente estudo analisou se preços históricos de negociação de índices podem explicar retornos futuros, em conformidade aos índices médios de desempenho das maiores bolsas de valores do mundo, por meio do algoritmo k-NN. Por se tratarem de dados históricos (passados) e relativos à forma fraca da HME, foi proposta a Hipótese 1 (H1): o algoritmo k-NN para prever o retorno futuro dos principais índices de mercados do mundo a partir de dados históricos gera resultados estatisticamente significantes e superiores à média.

Vale ressaltar o emprego de diferentes períodos dos modelos de k-NN elucidados na literatura, o que gera mais e/ou menos observações, por variarem desde alguns meses (HUANG; YANG; CHUANG, 2008) a mais de uma década (SHYNKEVICH *et al.*, 2017).

Por se tratar de um modelo de aprendizado que explora exemplos/históricos para classificação, espera-se ainda que diferentes números de observação influenciem o seu desempenho. Chen e Hao (2017) ratificam tal expectativa ao preverem índices apenas com dados históricos a partir de outros modelos de ML, o que levou à elaboração da Hipótese 2 (H2): modelos do algoritmo k-NN que utilizam um número maior de observações apresentam desempenho estatisticamente superior em relação àqueles com menor quantidade.

Como a aplicação de algoritmos de ML é recorrente em pesquisas sobre mercados de capitais (CAVALCANTE *et al.*, 2016; ZHAO; SUN; WANG, 2014), espera-se que esse uso tenha sido, em partes, incorporado ao preço dos ativos segundo a HME, com a respectiva formulação da Hipótese 3 (H3): modelos do algoritmo k-NN estimados para a primeira década do século XXI têm desempenhos estatisticamente superiores em relação àqueles estimados para a segunda década do mesmo século.

Por fim, a maioria dos estudos acerca da previsão dos retornos diários de índices ou outros ativos financeiros utiliza o preço de fechamento como base, a exemplo de Qian e Rasheed (2007), Shynkevich *et al.* (2017) e Cao *et al.* (2019). Todavia, Novak e Velušček (2016) propõem os preços máximos diários para isso, por haver menor volatilidade desses valores em comparação aos do fechamento – ao final do pregão, os resultados tendem a apresentar maior variação. Assim, avaliou-se a superioridade do desempenho de modelos estimados com preços máximos em relação àqueles do fechamento na Hipótese 4 (H4): modelos do algoritmo k-NN que empregam preços máximos para cálculo do retorno têm desempenhos estatisticamente superiores em relação aos do fechamento.

Para estimar os modelos e realizar as análises da pesquisa, foram empregados os procedimentos metodológicos esclarecidos na próxima seção deste trabalho.

3 METODOLOGIA

Esta pesquisa possui caráter quantitativo quanto à abordagem, caracterizada com base na coleta e no tratamento de dados por meio da estatística (RICHARDSON, 1999; GIL, 2008), devido aos testes estatísticos para verificar as hipóteses desenvolvidas de acordo com o desempenho dos modelos estimados; e descritivo quanto aos objetivos, que, segundo Gil (2008), pretende estabelecer relações entre variáveis na descrição de fenômenos, o que, neste estudo, coaduna os desempenhos de diferentes modelos baseados em ML, para fins de previsão do retorno de índices consoantes a informações passadas.

Índice	Ticker	País / Região	Continente
DAX	^GDAXI	Alemanha	Europa
Euro Stoxx 50	^STOXX50E	Alemanha-Zurique	Europa
S&P Merval	M.BA	Argentina	Américas
S&P/ASX 200	^AXJO	Austrália	Oceania
ATX	^ATX	Áustria	Europa
BEL 20	^BFX	Bélgica	Europa
Bovespa	^BVSP	Brasil	Américas
S&P/TSX	^GSPTSE	Canadá	Américas
S&P CLX IPSA	^IPSA	Chile	Américas
Shanghai (SSE)	000001.SS	China	Ásia
Shenzhen Component (SZSE)	399001.SZ	China	Ásia
Hang Seng	^HSI	China	Ásia
KOSPI	^KS11	Coréia do Sul	Ásia
IBEX 35	^IBEX	Espanha	Europa
Down Jones	^DJI	Estados Unidos	Américas
Nasdaq 100	^NDX	Estados Unidos	Américas
Nasdaq	^IXIC	Estados Unidos	Américas
S&P 500	^GSPC	Estados Unidos	Américas
PSEI Composite	PSEI.PS	Filipinas	Ásia
OMX Helsinki 25	^OMXH25	Finlândia	Europa
CAC 40	^FCHI	França	Europa
AEX	^AEX	Holanda	Europa
BSE Sensex	^BSESN	Índia	Ásia
Nifty 50	^NSEI	Índia	Ásia
IDX Composite	^JKSE	Indonésia	Ásia
ISEQ Overall	^ISEQ	Irlanda	Europa
TA 35	TA35.TA	Israel	Ásia
Nikkei 225	^N225	Japão	Ásia
KLCI	^KLSE	Malásia	Ásia
S&P/BMV IPC	^MXX	México	Américas
NZX 50	^NZ50	Nova Zelândia	Oceania
PSI 20	PSI20.LS	Portugal	Europa
FTSE 100	^FTSE	Reino Unido	Europa
MOEX	IMOEX.ME	Rússia	Europa
SMI	^SSMI	Suíça	Europa
Taiwan Weighted	^TWII	Taiwan	Ásia
BIST 100	XU100.IS	Turquia	Europa

Quadro 1 - Informações sobre os índices componentes da amostra.

Fonte: Elaboração dos autores (2023).

Para a coleta de dados, foram identificadas primeiramente as maiores bolsas de valores em nível mundial, conforme a classificação do *website* Investing (2022). Após essa identificação, foi realizado o levantamento dos principais índices de cada bolsa e coletado o *ticker* (código) no *website* Yahoo!Finance (2022), importante fonte de dados para a previsão de preços de ativos com base em algoritmos de ML (AWAN *et al.*, 2021). Tais informações são fundamentais para a coleta de dados referentes às cotações diárias de cada índice a partir

do *software* R e das funções do pacote da Estrutura de Modelagem Financeira Quantitativa (*Quantitative Financial Modelling Framework* – *quantmod*), o qual auxilia agentes de mercado no teste e desenvolvimento dos modelos aplicados em finanças (RYAN *et al.*, 2020).

Na composição da amostra, consideraram-se 37 índices das maiores bolsas de valores a nível mundial (Quadro 1), cujas cotações foram coletadas em frequência diária de 2001 a 2019. Optou-se pelo encerramento da coleta no referido ano devido à emergência da pandemia da Doença do Novo Coronavírus 2019 (*Coronavirus Disease 2019* – COVID-19) em 2020, que afetou drasticamente os preços dos ativos em diferentes mercados (AVELAR *et al.*, 2022), e os modelos foram estimados para cada índice da amostra.

O k-NN foi empregado para fins de classificação, com o propósito de prever se o preço do ativo iria subir (1) ou descer (0) no pregão seguinte, a partir de informações passadas, como propõe Shynkevich *et al.* (2017). Para o treinamento do modelo, utilizaram-se 80% dos dados da amostra (MOON; JUN; KIM, 2018), parcela selecionada de forma aleatória. Com relação ao parâmetro *k* do agrupamento do método e para cada modelo de cada índice, se testaram valores entre um e vinte, de forma a atingir o melhor desempenho, em abordagem similar à de Cao *et al.* (2019). Com vistas ao teste das hipóteses H2 e H3, os modelos foram estimados para o período completo (2001-2019) e dois subperíodos da primeira e segunda década do século XXI (2001 a 2010 e 2011 a 2019).

Duas modelagens foram empregadas nesta pesquisa para cada índice da amostra, com vistas ao teste de H4: (i) modelo I – retornos baseados nos preços históricos de fechamento; e (ii) modelo II – preços históricos máximos diários. Em ambos foram usados retornos diários de uma janela que variou de um a trinta dias para prever a evolução do ativo no dia seguinte, como preconizam Cao *et al.* (2019) de modo similar. Para a estimativa do desempenho do modelo, optou-se pelo cálculo da acurácia (Equação 1), calculada a partir dos dados de teste (20% da amostra, previamente selecionada de forma aleatória). A Figura 1 ilustra a forma de treinamento e o teste do modelo básico de Ferreira, Gandomi e Cardoso (2021), para previsão de preços de ativos financeiros com base em IA:

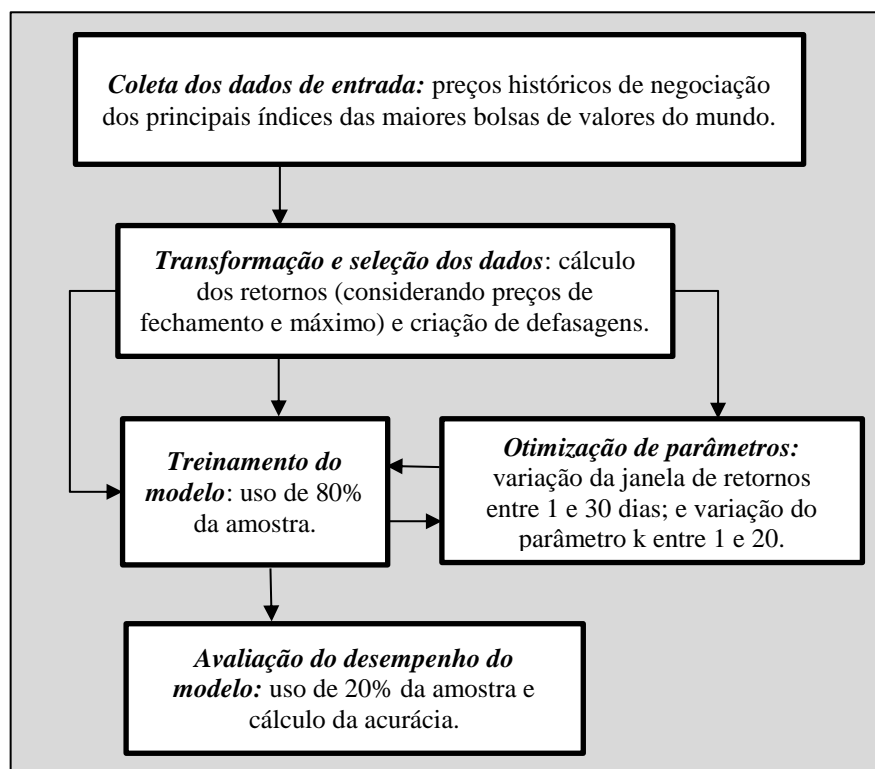


Figura 1 - Fluxograma do processo de uso de algoritmos de IA para previsão dos preços de ativos financeiros. Fonte: Elaboração dos autores (2023), com base em Ferreira, Gandomi e Cardoso (2021).

A análise dos resultados foi realizada em conformidade à estatística descritiva, para representar os resultados obtidos pelos modelos estimados; ao teste de Kolmogorov-Smirnov, para analisar a normalidade da distribuição; e ao teste t de Student, para testar as hipóteses propostas no estudo. Utilizaram-se os níveis de significância de 1%, 5% e 10% nos testes, e os dados foram tratados e analisados a partir do *software* R, com pacotes *Discrete Goodness-of-Fit Tests* (dgof); *quantmod*; *Functions for Classification* (class); *A Grammar of Data Manipulation* (dplyr); e *eXtensible Time Series* (xts).

4 RESULTADOS

Nesta seção são apresentados os resultados obtidos na pesquisa. Inicialmente, na subseção 4.1 se citam os modelos estimados com base nos retornos calculados a partir dos preços de fechamento diários (modelo I) e nos retornos mensurados com os preços máximos diários (modelo II) para cada índice de cada bolsa de valores considerado na amostra, cujos resultados são discutidos na subseção 4.2.

4.1 Análise dos resultados dos modelos

Diante dos diferentes períodos observados nesta subseção, a Tabela 1 apresenta os resultados gerais de acurácia para cada índice, durante os diferentes horizontes temporais, o teste de Kolmogorov-Smirnov e as estatísticas descritivas para ambos os modelos. Salienta-se que o referido teste apresentou normalidade na distribuição de todos os dados.

Tabela 1 - Resultados gerais de acurácia dos modelos estimados

Índices	Modelo I			Modelo II		
	Geral	2001-2010	2011-2019	Geral	2001-2010	2011-2019
AEX	0,5443	0,5601	0,5714	0,5524	0,5697	0,5848
ATX	0,5740	0,5909	0,5930	0,5816	0,6224	0,5903
BEL 20	0,5337	0,5584	0,5601	0,5463	0,6121	0,5714
BIST 100	0,5597	0,5692	0,5862	0,5748	0,5850	0,6158
Bovespa	0,5504	0,5957	0,5547	0,5708	0,5812	0,5945
BSE Sensex	0,5723	0,6007	0,5690	0,5934	0,6139	0,5972
CAC 40	0,5601	0,5688	0,5578	0,5431	0,5894	0,5624
DAX	0,5490	0,5847	0,5819	0,5629	0,5800	0,5653
Down Jones	0,5526	0,5895	0,5705	0,5516	0,5755	0,5638
Euro Stoxx 50	0,5832	0,5918	0,5788	0,5707	0,5918	0,5640
FTSE 100	0,5507	0,5595	0,5633	0,5739	0,5730	0,5950
Hang Seng	0,5499	0,5559	0,5926	0,5553	0,6073	0,5753
IBEX 35	0,5638	0,5667	0,5601	0,5452	0,5861	0,5918
IDX Composite	0,5663	0,5918	0,6039	0,5806	0,6041	0,6136
ISEQ Overall	0,5463	0,5622	0,5811	0,5561	0,5892	0,5923
KLCI	0,5429	0,5942	0,5565	0,5873	0,6196	0,6017
KOSPI	0,5750	0,6070	0,5825	0,5706	0,5992	0,5778
MOEX	0,5847		0,5847	0,5766		0,5766
Nasdaq	0,5600	0,5473	0,5817	0,5674	0,5775	0,5884
Nasdaq 100	0,5516	0,5694	0,5749	0,5611	0,5493	0,5615
Nifty 50	0,5469	0,6183	0,5933	0,5881	0,6489	0,5967
Nikkei 225	0,5581	0,5902	0,5568	0,5597	0,5820	0,5730
NZX 50	0,5595	0,6037	0,6102	0,5750	0,6481	0,6075
OMX Helsinki 25	0,5957		0,5957	0,5745		0,5745
PSEI Composite	0,5598	0,5630	0,5932	0,5769	0,5848	0,6271
PSI 20	0,6097		0,6097	0,6097		0,6097
S&P 500	0,5516	0,5936	0,5660	0,5411	0,5614	0,5593
S&P CLX IPSA	0,5707	0,6166	0,5747	0,6014	0,6632	0,6418
S&P Merval	0,5602	0,6108	0,5877	0,5940	0,6270	0,6082
S&P/ASX 200	0,5389	0,5950	0,5866	0,5637	0,5875	0,5891
S&P/BMV IPC	0,5571	0,5952	0,5522	0,5763	0,6228	0,5870
S&P/TSX	0,5546	0,5983	0,5775	0,5609	0,6040	0,6067
Shanghai (SSE)	0,5615	0,5807	0,5624	0,5962	0,6062	0,6141
Shenzhen Component	0,5501	0,5710	0,5637	0,5817	0,5989	0,5731
SMI	0,5565	0,5618	0,5663	0,5614	0,5645	0,5708
TA 35	0,5604	0,5813	0,5477	0,5770	0,5744	0,5777
Taiwan Weighted	0,5599	0,5699	0,5824	0,5763	0,5944	0,5745
Kolmogorov-Smirnov	1,177	0,840	0,530	0,569	0,667	0,830
Estatísticas descritivas						
Média	0,5600	0,5827	0,5765	0,5712	0,5969	0,5879
Mediana	0,5595	0,5871	0,5775	0,5739	0,5906	0,5877
Desvio-padrão	0,0153	0,0189	0,0162	0,0166	0,0258	0,0199
Coefficiente de variação	0,0273	0,0325	0,0281	0,0290	0,0432	0,0339

Índices	Modelo I			Modelo II		
	Geral	2001-2010	2011-2019	Geral	2001-2010	2011-2019
Mínimo	0,5337	0,5473	0,5477	0,5411	0,5493	0,5593
Máximo	0,6097	0,6183	0,6102	0,6097	0,6632	0,6418

Nota: ***, ** e * indicam que a variável é estatisticamente significativa a 1%, 5% e 10%, respectivamente.

Fonte: Elaboração dos autores (2023).

De acordo com a Tabela 1, verificou-se que a média de acurácia foi próxima a 57% em todos os modelos, tendo havido baixa dispersão em torno da média, em conformidade ao desvio-padrão. A maior acurácia foi obtida com o índice Nifty 50 da *National Stock Exchange of India* (61,83%) no primeiro subperíodo, e a menor, em BEL20 da *Brussels Stock Exchange* (53,37%), ao se considerar todo o período de análise.

Na sequência, a Tabela 2 apresenta os resultados dos testes t referentes ao modelo I. Por um lado, constatou-se que, em todos os horizontes de tempo analisados, os desempenhos dos modelos foram estatisticamente significantes e superiores ao retorno médio esperado aos mercados, o que ratifica H1; por outro lado, os desempenhos dos modelos com período fracionado superaram os dos modelos com todas as observações, aspecto contrário à H2. Ademais, não foram observadas diferenças estatísticas significantes entre os modelos estimados com cada subperíodo, com a impossibilidade de confirmar H3, pois, na última hipótese, era esperado desempenho superior aos modelos estimados com dados da primeira década do século XXI, em relação àqueles estimados para a segunda.

Tabela 2 - Teste t para as acurácias do modelo I

Subamostras comparadas			T	Sig.	Subamostras comparadas			T	Sig.
Geral	Mercado	20,0	***	Geral	2001-2010	-5,5	***		
2001-2010	Mercado	23,0	***	Geral	2011-2019	-4,5	***		
2011-2019	Mercado	24,0	***	2001-2010	2011-2019	1,5			

Nota: ***, ** e * indicam que a variável é estatisticamente significativa a 1%, 5% e 10%, respectivamente.

Fonte: Elaboração dos autores (2023).

No que concerne ao modelo II e com base em retornos de preços máximos diários, a média de acurácia se aproxima a 58,5% e houve baixa dispersão em torno da média. A maior acurácia foi obtida pelo modelo estimado para o índice S&P CLX IPSA, que enfoca ações da Bolsa de Santiago (66,32%) na primeira década do século XXI, e a menor acurácia foi observada para o índice S&P 500 (54,11%), considerando todo o período de análise.

A Tabela 3 apresenta os resultados obtidos a partir dos testes t referentes ao modelo II. Assim como no modelo anterior, verificou-se que, em todos os horizontes temporais

analisados, os desempenhos dos modelos estimados foram estatisticamente significantes e superiores ao retorno esperado de mercado, ao ratificar H1. Todavia, os desempenhos dos modelos estimados com os subperíodos superaram os resultados notados no período completo, o que vai ao encontro do esperado em H2. Novamente, não foram observadas diferenças estatisticamente significantes entre os modelos estimados com os subperíodos, com a impossibilidade de corroborar H3.

Tabela 3 - Teste t para as acurácias do modelo II

Subamostras comparadas		T	Sig.	Subamostras comparadas		T	Sig.
Geral	Mercado	24,0	***	Geral	2001-2010	-4,5	***
2001-2010	Mercado	20,0	***	Geral	2011-2019	-3,9	***
2011-2019	Mercado	24,0	***	2001-2010	2011-2019	1,6	

Nota: ***, ** e * indicam que a variável é estatisticamente significativa a 1%, 5% e 10%, respectivamente.
 Fonte: Elaboração dos autores (2023).

Finalmente, a Tabela 4 indica os resultados dos testes t, com a comparação dos modelos I e II. Nesse caso, os desempenhos do modelo II, estimado com cotações máximas, superaram aqueles baseados nos preços de fechamento, o que confirmou os apontamentos de Novak e Velušček (2016) e ratificou H4.

Tabela 4 - Teste t para as acurácias dos modelos I e II

Modelo I	Modelo II	T	Sig.
Geral	Geral	-3,0	***
2001-2010	2001-2010	-2,6	**
2011-2019	2011-2019	-2,7	***

Nota: ***, ** e * indicam que a variável é estatisticamente significativa a 1%, 5% e 10%, respectivamente.
 Fonte: Elaboração dos autores (2023).

O subtópico a seguir se destina a debater os resultados obtidos nesta investigação.

4.2 Discussão dos resultados

Diversas considerações sobre a aplicação do algoritmo k-NN para explicar o retorno futuro dos principais índices das maiores de bolsas em nível mundial a partir de dados históricos são ressaltadas nesta pesquisa. Todos os modelos desenvolvidos com base no algoritmo obtiveram desempenho superior à média de mercado; logo, nem todas as informações passadas foram precificadas pelo mercado, o que vai ao encontro da forma fraca de eficiência de mercado sublinhada por Fama (1970; 1991).

Nesse prisma, resultados semelhantes foram reportados com frequência por estudos que empregaram algoritmos de IA no mercado financeiro, como Shynkevich *et al.* (2017), Moon, Jun e Kim (2018), Cao *et al.* (2019) e Wu *et al.* (2020). Porém, tais pesquisas geralmente foram desenvolvidas com menor abrangência de mercados de capitais, em comparação à amostra da presente investigação. Os resultados manifestaram desempenho satisfatório do algoritmo, como apontado por Qian e Rasheed (2007) e Carta *et al.* (2021), o que reforça os achados da literatura e a importância das discussões acerca da HME no contexto de automatização de decisões de investimento baseadas em IA; logo, ratificou-se H1.

Quando empregados os subperíodos em ambos os modelos, o desempenho do algoritmo k-NN foi superior ao do período completo. Esse resultado não corrobora o esperado para o modelo (FACELI *et al.*, 2021), tampouco os aspectos observados em outros algoritmos (CHEN; HAO, 2017). Desse modo, não foi possível ratificar H2, o que pode ser explicado pelo fato de o período completo englobar uma série de eventos heterogêneos refletidos em cada índice ao longo do século XXI, como crises e *booms* econômicos que prejudicam o desempenho dos modelos fundamentados em todos os dados no treinamento. Em contrapartida, os subperíodos tenderiam a conter dados mais homogêneos, o que contribui para o desempenho dos modelos estimados nesses contextos.

Além disso, em ambos os modelos com os subperíodos, não foram verificadas diferenças estatisticamente significantes, o que impossibilitou atestar H3. De modo similar à HME, tais achados poderiam ser relacionados à falta de precificação de todas as informações pelos mercados estudados, inclusive com o uso de algoritmos de ML para previsão. Outra possível explanação poderia ser relacionada ao efeito dos algoritmos ML nos preços dos mercados de capital até 2019 (último ano do estudo), o que teria impossibilitado a percepção de modelos baseados em tais algoritmos nos índices gerais.

Como argumentam Novak e Velušček (2016), algoritmos treinados de acordo com os preços máximos apresentaram um desempenho superior àqueles com valores de fechamento. Apesar de os últimos serem empregados de forma recorrente no cálculo de retorno de ações, sua volatilidade afeta negativamente a *performance* das previsões; por conseguinte, a constatação de que os preços máximos para cálculo dos retornos possibilitam acurácia maior dos modelos ratifica H4. Menos recorrente, essa medida orienta as decisões dos agentes de mercado do mesmo modo que os preços de fechamento e pode ser implementada em estudos anteriores que utilizaram a cotação de fechamento, o que possivelmente otimiza o

desempenho dos modelos estimados, como mencionam Qian e Rasheed (2007), Shynkevich *et al.* (2017) e Cao *et al.* (2019).

Diante do exposto, verificou-se que a HME, em sua forma clássica, necessita de debates no âmbito da ascensão de algoritmos de IA. Os modelos podem ser amplamente empregados pelos agentes de mercado no processo de avaliação de ativos e tomada de decisão de investimento. Como frisam Rundo *et al.* (2019), a fase contemporânea de automatização da avaliação e análise de ativos de investimento em finanças impõe desafios aos pesquisadores para compreenderem a eficiência de mercado e comportamento dos preços, dada a intensificação do uso da técnica.

5 CONSIDERAÇÕES FINAIS

Este artigo analisou se preços históricos de negociação explicam o retorno futuro dos principais índices das maiores bolsas de valores do mundo, por meio do uso do algoritmo k-NN. Para tanto, a pesquisa descritiva e quantitativa utilizou dados diários dos principais índices de tais bolsas e foram estimados dois tipos de modelos distintos baseados no algoritmo – retornos baseados nos preços históricos de fechamento dos índices; e nos valores históricos máximos diários –, cuja estimativa contemplou o período completo de 2001 a 2019 e a respectiva divisão em dois subperíodos: 2001 a 2010 e 2011 a 2019.

Os resultados evidenciaram que todos os modelos desenvolvidos a partir do algoritmo k-NN obtiveram um desempenho estatisticamente superior à média de mercado. Por isso, o emprego desse algoritmo de ML supera o retorno médio esperado com base em dados passados, ao questionar a eficiência de tais mercados na forma fraca da HME (FAMA, 1970; 1991). Aqui, ressalta-se a importância de considerar tal hipótese em um contexto contemporâneo de automatização, no qual os agentes de mercado têm amplo acesso ao uso de modelos de IA para auxiliar na avaliação de ativos e na tomada de decisão de investimentos.

Esperava-se que um maior número de observações para treinamento do algoritmo possibilitasse desempenhos estatisticamente superiores dos modelos estimados, mas, contrariamente, aqueles com melhor *performance* indicaram menores quantidades de observação. Parcialmente, isso se deve ao fato de fenômenos específicos de um período terem sido compreendidos de maneira adequada em seu contexto, o que contribuiu para o ajuste dos modelos com subperíodos. Destaca-se, ainda, que não se notaram diferenças estatisticamente

significantes entre os desempenhos dos modelos com cada subperíodo, provavelmente pela dificuldade do mercado em precificar o uso de algoritmos de ML na última década.

Nos três períodos estudados, o modelo II apresentou os melhores desempenhos, com preços máximos em detrimento aos de fechamento. Apesar de ser uma abordagem pouco frequente de análise e por ser menos influenciada pela volatilidade diária do fim de pregão, ela possibilitou dados adequados à estimação dos modelos, em que pode ser aplicada na orientação de decisões dos agentes de mercado sobre compra e venda, como no preço de fechamento. Tal constatação orienta o desenvolvimento de novas investigações e possibilita o aprimoramento de pesquisas existentes com uso de algoritmos de ML no contexto dos mercados de capitais.

Salienta-se que a pesquisa apresenta limitações, como: (i) uso das próprias informações de preços para previsão, sem considerar outras variáveis no treinamento do algoritmo (indicadores técnicos, por exemplo); (ii) emprego de apenas um algoritmo (k-NN), que utiliza um método peculiar de previsão (distância entre as observações) em relação aos demais algoritmos de ML; (iii) desconsideração explícita da característica temporal dos preços analisados em cada índice; e (iv) falta de relevância de outras variáveis externas públicas que poderiam afetar os retornos dos índices em cada país.

Apesar das limitações supracitadas, os resultados contribuíram com a literatura sobre o emprego de algoritmos de ML (especialmente, o k-NN), em se tratando da previsão dos preços de ativos listados no mercado financeiro. *A priori*, foram abordados os principais índices das maiores bolsas mundiais, enquanto a maioria dos estudos aborda um ou poucos mercados. *A posteriori*, evidenciou-se que o número de observações utilizado para o treinamento do algoritmo gerou diferenças relevantes no seu desempenho, com o respectivo questionamento acerca da eficiência dos mercados investigados em sua forma fraca (de acordo com a HME) em um contexto de ampla automatização por IA.

Uma vez que os modelos estimados para alguns dos índices apresentaram desempenho discrepante, as características específicas de cada mercado também podem ser abordadas com a análise das influências individuais das especificações dos mercados em pesquisas posteriores. Igualmente, estudos com outros algoritmos de ML (como *Random Forest*, *Naive Bayes* e SVM, por exemplo) podem replicar a metodologia apresentada neste trabalho para comparação de desempenho. Além disso, é possível aplicar os modelos propostos com ações de empresas em detrimento a índices, e as crises globais e específicas de cada mercado conseguem ser exploradas com base em novas modelagens baseadas em algoritmos de ML.

REFERÊNCIAS

- AWAN, M. J.; RAHIM, M. S. M.; NOBANEE, H.; MUNAWAR, A.; YASIN, A.; ZAIN, A. M. Social media and stock market prediction: a big data approach. **Computers, Materials & Continua**, [s.l.], v. 67, n. 2, p. 2569-2583, 2021. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3827106. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.32604/cmc.2021.014253>
- ALKHATIB, K.; NAJADAT, H.; HMEIDI, I.; SHATNAWI, M. K. A. Stock price prediction using k-Nearest Neighbor (kNN) algorithm. **International Journal of Business, Humanities and Technology**, [s.l.], v. 3, n. 3, p. 32-44, 2013. Disponível em: https://www.ijbhtnet.com/journals/Vol_3_No_3_March_2013/4.pdf. Acesso em: 1º ago. 2024.
- AVELAR, E. A.; OREFICI, J. B. P.; BORGES, S. L.; CAMPOS, O. V.; FERREIRA, P. O. The COVID-19 pandemic effects on the financial performance of Brazilian listed companies. **Revista de Administração da UFSM**, Santa Maria, v. 15, n. 3, p. 389-410, 2022. Disponível em: <https://periodicos.ufsm.br/reaufsm/article/view/67372>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.5902/1983465967372>
- BALL, R. The global financial crisis and the efficient market hypothesis: what have we learned? **Journal of Applied Corporate Finance**, Chicago, v. 21, n. 4, p. 8-16, 2009. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1502815. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1111/j.1745-6622.2009.00246.x>
- BATISTA, A. R. A.; MAIA, U.; ROMERO, A. Stock market under the 2016 Brazilian presidential impeachment: a test in the semi-strong form of the efficient market hypothesis,. **Contabilidade & Finanças**, São Paulo, v. 29, n. 78, 405-417, 2018. Disponível em: <https://www.scielo.br/j/rcf/a/qsDPvq9R349gn8GqstnXgSr/?lang=en>. Acesso em: 1º ago. 2024. DOI: <http://dx.doi.org/10.1590/1808-057x201805560>
- BORGES, M. R. Efficient market hypothesis in European stock markets. **The European Journal of Finance**, Lisboa, v. 16, n. 7, p. 711-726, 2010. Disponível em: <https://www.repository.utl.pt/bitstream/10400.5/26311/1/MRBORGES.2010.pdf>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1080/1351847X.2010.495477>
- CAMARGOS, M. A.; ROMERO, J. A. R. Reaction of Sao Paulo stock exchange corporative events: a test of the Efficient Market Hypothesis. **Revista de Gestão**, São Paulo, v. 13, n. 3, p. 57-74, 2006. Disponível em: <https://www.revistas.usp.br/rege/article/view/36566>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.5700/issn.2177-8736.rege.2006.36566>
- CAO, H.; LIN, T.; LI, Y.; ZHANG, H. Stock price pattern prediction based on complex network and Machine Learning. **Complexity**, [s.l.], v. 1, p. 1-12, 2019. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1155/2019/4132485>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1155/2019/4132485>
- CAVALCANTE, R. C.; BRASILEIRO, R. C.; SOUZA, V. L. F.; NÓBREGA, J. P.;

OLIVEIRA, L. I. Computational intelligence and financial markets: a survey and future directions. **Expert Systems with Applications**, [s.l.], v. 55, p. 194-211, 2016. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S095741741630029X>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1016/j.eswa.2016.02.006>

CARTA, Salvatore et al. Explainable AI for financial forecasting. In: **International Conference on Machine Learning, Optimization, and Data Science**. Cham: Springer International Publishing, 2021. p. 51-69. Disponível em: https://www.researchgate.net/publication/353483959_Explainable_AI_for_Financial_Forecasting. Acesso em: 1º ago. 2024. DOI: https://doi.org/10.1007/978-3-030-95470-3_5

CHEN, Y.; HAO, Y. A feature weighted support vector machine and k-nearest neighbor algorithm for stock market indices prediction. **Expert Systems with Applications**, [s.l.], v. 80, p. 3400355, 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0957417417301367>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1016/j.eswa.2017.02.044>

DEMIREL, U.; ÇAM, H.; ÜNLÜ, R. Predicting stock prices using machine learning methods and deep learning algorithms: the sample of the Istanbul Stock Exchange. **Gazi University Journal of Science**, [s.l.], v. 34, n. 1, p. 63-82, 2021. Disponível em: <https://dergipark.org.tr/en/pub/gujs/issue/60264/679103>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.35378/gujs.679103>

DIXON, M. F.; HALPERIN, I.; BILOKON, P. **Machine learning in finance**. New York: Springer International, 2020. DOI: <https://doi.org/10.1007/978-3-030-41068-1>

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. **Inteligência Artificial: uma abordagem de Aprendizado de Máquina**. Rio de Janeiro: LTC, 2011.

FAMA, E. F. Efficient capital markets: a review of theory and empirical work. **The Journal of Finance**, New York, v. 25, n. 2, p. 383-417, 1970. Disponível em: <https://www.jstor.org/stable/2325486>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.2307/2325486>

FAMA, E. F. Efficient capital markets: II. **The Journal of Finance**, New York, v. 46, n. 5, p. 1575-1617, 1991. Disponível em: <https://www.jstor.org/stable/2328565>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.2307/2328565>

FERREIRA, F. G. D. C.; GANDOMI, A. H.; CARDOSO, R. T. N. Artificial Intelligence applied to stock market trading: a review. **IEEE Access**, [s.l.], v. 9, p. 30898-30917, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9350582>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1109/ACCESS.2021.3058133>

FREIRE, A. P. F.; MAIA, P. A. A informação contábil e a Hipótese do Mercado Eficiente: um estudo empírico utilizando o setor de petróleo, gás e biocombustível do Brasil no período 2006-2009. **Revista de Finanças Aplicadas**, São Paulo, v. 1, p. 1-20, 2012. Disponível em: <http://www.financasaplicadas.fia.com.br/index.php/financasaplicadas/article/view/37>. Acesso em: 1º ago. 2024.

FUKUNAGA, K.; NARENDRA, P. M. A branch and bound algorithm for computing k-Nearest Neighbors. **IEEE Transactions on Computers**, [s.l.], v. C-24, n. 7, p. 750-753, 1975. Disponível em: <https://ieeexplore.ieee.org/document/1672890>. Acesso em: 1º ago. 2024. <https://doi.org/10.1109/T-C.1975.224297>

GABRIEL, F. S.; RIBEIRO, R. B.; RIBEIRO, K. C. S. Hipóteses de mercado eficiente: um estudo de eventos a partir da redução do IPI. **Revista de Gestão, Finanças e Contabilidade**, Salvador, v. 3, n. 1, p. 36-52, 2013. Disponível em: <https://revistas.uneb.br/index.php/financ/article/view/95/199%26amp%3Bgt%3B%26lt%3B>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.18028/rgfc.v3i1.95>

GANDHMAL, D. P.; KUMAR, K. Systematic analysis and review of stock market prediction techniques. **Computer Science Review**, [s.l.], v. 34, p. 100190, 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S157401371930084X>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1016/j.cosrev.2019.08.001>

GHODDUSI, H.; CREAMER, G. G.; RAFIZADEH, N. Machine learning in energy economics and finance: a review. **Energy Economics**, [s.l.], v. 81, p. 709-727, 2019. Disponível em: <https://ideas.repec.org/a/eee/eneeco/v81y2019icp709-727.html>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1016/j.eneco.2019.05.006>

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GITE, S.; KHATAVKAR, H.; KOTECHA, K.; SRIVASTAVA, S.; MAHESHWARI, P.; PANDEY, N. Explainable stock prices prediction from financial news articles using sentiment analysis. **Peer Journal of Computer Science**, [s.l.], v. 7, e340, 2021. Disponível em: <https://peerj.com/articles/cs-340/>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.7717/2Fpeerj-cs.340>

HUANG, C.; YANG, D.; CHUANG, Y. Application of wrapper approach and composite classifier to the stock trend prediction. **Expert Systems with Applications**, [s.l.], v. 34, n. 4, 2870-2875, 2008. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0957417407001819>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1016/j.eswa.2007.05.035>

INVESTING. **Classificação das maiores bolsas mundiais**. São Paulo, 2022. Disponível em: <https://br.investing.com/>. Acesso em: 1º ago. 2024.

JEON, S.; HONG, B.; CHANG, V. Pattern graph tracking-based stock price prediction using big data. **Future Generation Computer Systems**, [s.l.], v. 80, p. 171-187, 2018. Disponível em: <https://sciencedirect.com/science/article/abs/pii/S0167739X17301991>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1016/j.future.2017.02.010>

KIESEL, F.; MANZ, F.; SCHIERECK, D. The conditional stock market response to banks' distressed asset sales on CDS availability. **Applied Economics**, [s.l.], v. 52, n. 56, p. 6123-6135, 2020. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/00036846.2020.1784388>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1080/00036846.2020.1784388>

KUMAR, I.; DOGRA, K.; UTREJA, C.; YADAV, P. A comparative study of supervised Machine Learning algorithms for stock market trend prediction. *In: INTERNATIONAL CONFERENCE ON INVENTIVE COMMUNICATION AND COMPUTATIONAL TECHNOLOGIES*, 2., 2018, [s.l.]. **Proceedings...** [s.l.]: ICICCT, 2018, p. 1003-1007. Disponível em: <https://www.semanticscholar.org/paper/A-Comparative-Study-of-Supervised-Machine-Learning-Kumar-Dogra/4ed67297d994da53b84332d98b0d02f3bccfa155>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1109/ICICCT.2018.8473214>

LI, D.; LI, Z.; LI, R. Automate the identification of technical patterns: a K-nearest-neighbour model approach. **Applied Economics**, [s.l.], v. 50, n. 17, p. 1978-1991, 2018. Disponível em: <https://ideas.repec.org/a/taf/applec/v50y2018i17p1978-1991.html>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1080/00036846.2017.1383596>

LIM, K. P.; BROOKS, R. The evolution of stock market efficiency over time: a survey of the empirical literature. **Journal of Economic Surveys**, [s.l.], v. 25, n. 1, p. 69-108, 2011. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6419.2009.00611.x>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1111/j.1467-6419.2009.00611.x>

LO, A. W. Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. **Journal of Investment Consulting**, [s.l.], v. 7, n. 2, p. 21-44, 2005. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1702447. Acesso em: 1º ago. 2024.

MALKIEL, B. G. The efficient market hypothesis and its critics. **Journal of Economic Perspectives**, [s.l.], v. 17, n. 1, p. 59-82, 2003. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/089533003321164958>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1257/089533003321164958>

MALKIEL, B. G. Reflections on the efficient market hypothesis: 30 years later. **The Financial Review**, [s.l.], v. 40, n. 1, p. 1-9, 2005. Disponível em: <http://www.e-m-h.org/Malkiel2005.pdf>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1111/j.0732-8516.2005.00090.x>

MOON, K. S.; JUN, S.; KIM, H. Speed up of the majority voting ensemble method for the prediction of stock price directions. **Economic Computation and Economic Cybernetics Studies and Research**, [s.l.], v. 52, n. 1, p. 215-228, 2018. Disponível em: https://www.researchgate.net/publication/323890209_Speed_up_of_the_Majority_Voting_Ensemble_Method_for_the_Prediction_of_Stock_Price_Directions. Acesso em: 1º ago. 2024. DOI: <http://dx.doi.org/10.24818/18423264/52.1.18.13>

NOVAK, M. G.; DEJAN, V. Prediction of stock price movement based on daily high prices, **Quantitative Finance**, v. 16, n. 5, 793-826, 2016. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/14697688.2015.1070960>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1080/14697688.2015.1070960>

PUJITHA, J.; MAHESH, G. Stock market price predictions using Python. *In: INTERNATIONAL CONFERENCE ON FORENSICS, ANALYTICS, BIG DATA, SECURITY*, 1., 2021, Bengaluru. **Proceedings...** Bengaluru: FABS, 2021, p. 1-5. Disponível em: <https://ieeexplore.ieee.org/document/9702683>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1109/FABS52071.2021.9702683>

QIAN, B.; RASHEED, K. Stock market prediction with multiple classifiers. **Applied Intelligence**, v. 26, p. 25-33, 2007. Disponível em: <https://link.springer.com/article/10.1007/s10489-006-0001-7>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1007/s10489-006-0001-7>

RICHARDSON, M. **Fundamentos da metodologia científica**. São Paulo: Atlas, 1999.

RYAN, J. A.; ULRICH, J. M.; SMITH, E. B.; THIELEN, W.; TEETOR, P.; BRONDER, S. **Package ‘quantmod’**. [s.l.], 2020. Disponível em: <https://cran.r-project.org/web/packages/quantmod/quantmod.pdf>. Acesso em: 2 fev. 2022.

RUNDO, F.; TRENTA, F.; STALLO, A. L.; BATTIATO, S. Machine Learning for quantitative finance applications: a survey. **Applied Sciences**, [s.l.], v. 9, n. 24, p. 5574, 2019. Disponível em: <https://www.mdpi.com/2076-3417/9/24/5574>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.3390/app9245574>

SHYNKEVICH, Y.; MCGINNITY, T. M.; COLEMAN, S. A.; BELATRECHE, A.; LI, Y. Forecasting price movements using technical indicators: investigating the impact of varying input window length. **Neurocomputing**, [s.l.], v. 264, p. 71-88, 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0925231217311074>. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1016/j.neucom.2016.11.095>

TIMMERMANN, A.; GRANGER, C. W. J. Efficient market hypothesis and forecasting. **International Journal of Forecasting**, [s.l.], v. 20, n. 1, p. 15-27, 2004. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0169207003000128>. Acesso em: 1º ago. 2024. DOI: [https://doi.org/10.1016/S0169-2070\(03\)00012-8](https://doi.org/10.1016/S0169-2070(03)00012-8)

WANG, D. D. S.; GUO, C.; YANG, Y.; ZHANG, Y.; WANG, Z. Stock prediction method based on integrated learning. **International Core Journal of Engineering**, [s.l.], v. 7, n. 12, p. 678-682, 2021. Disponível em: <http://www.icj-e.org/download/ICJE-7-12-678-682.pdf>. Acesso em: 1º ago. 2024. DOI: [https://doi.org/10.6919/ICJE.202112_7\(12\).0093](https://doi.org/10.6919/ICJE.202112_7(12).0093)

WANG, W.; HO, K.-Y.; LIU, W.-M.; WANG, K. The relation between news events and stock price jump: an analysis based on neural network. In: INTERNATIONAL CONGRESS ON MODELLING AND SIMULATION, 20., 2013, Adelaide. **Proceedings...** Adelaide, 2013, p. 1406-1411. Disponível em: <https://www.mssanz.org.au/modsim2013/F8/wang.pdf>. Acesso em: 1º ago. 2024.

WU, Mengyun; COLEMAN, Martha; BAWUAH, Jonas. The predictive power of K-nearest neighbor (KNN): The effect of corporate governance mechanisms on earnings management. **Sage Open**, v. 10, n. 3, p. 2158244020949537, 2020. Disponível em: <https://ideas.repec.org/a/sae/sagope/v10y2020i3p2158244020949537.html>. Acesso em: 1º ago. 2024. Acesso em: 1º ago. 2024. DOI: <https://doi.org/10.1177/2158244020949537>

YAHOO!FINANCE. **Tickers**. New York, 2022. Disponível em: <https://finance.yahoo.com/>. Acesso em: 1º ago. 2024.

ZHAO, W.; SUN, C.; WANG, J. The research on price prediction of second-hand houses based on KNN and stimulated annealing algorithm. **International Journal of Smart Home**,

[s.l.], v. 8, n. 2, p. 191-200, 2014. Disponível em:
https://gvpress.com/journals/IJSH/vol8_no2/19.pdf. Acesso em: 1º ago. 2024. DOI:
<http://dx.doi.org/10.14257/ijsh.2014.8.2.19>

AGRADECIMENTOS

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento da pesquisa.

DADOS DOS AUTORES:

Ewerton Alex Avelar

E-mail: ewertonalexavelar@gmail.com

Curriculum Lattes: <http://lattes.cnpq.br/6758338709153591>

Graduado em Ciências Contábeis pela Universidade Federal de Minas Gerais (UFMG), Especialista em Educação pela Universidade Federal Fluminense (UFF), Mestre em Administração pela Universidade Federal de Lavras (UFLA) e Doutor em Administração pela UFMG. Atualmente, é Professor Adjunto do Departamento de Ciências Contábeis da UFMG, Professor Permanente do Centro de Pós-graduação e Pesquisas em Controladoria e Contabilidade (CEPCON/UFMG) e Coordenador do curso de graduação em Ciências Contábeis da UFMG. É o Líder no Núcleo de Estudos sobre Aplicação de Algoritmos em Contabilidade e Finanças (NEACONF/UFMG).

Simone Evangelista Fonseca

E-mail: simone_fonseca16@hotmail.com

Curriculum Lattes: <http://lattes.cnpq.br/5220117639109190>

Graduada em Administração pela Universidade Federal de Minas Gerais (UFMG), Especialista em Gestão Estratégica de Negócios pelo Centro de Pós-Graduação e Pesquisas em Administração da UFMG (CEPEAD/UFMG). Mestre e Doutora em Administração, linha de pesquisa em Finanças pela UFMG. Integrante e pesquisadora dos núcleos NUFI, NECEU, NEGEC e NEACONF da UFMG. Atualmente, é Professora Adjunta na Universidade Federal de Ouro Preto (UFOP), área de Administração Financeira.

Antônio Artur de Souza

E-mail: antonioarturdesouza@gmail.com

Curriculum Lattes: <http://lattes.cnpq.br/0597505816212353>

Graduado em Ciências Contábeis pela Universidade Federal de Santa Catarina (UFSC), Bacharel em Letras pela Universidade Federal de Minas Gerais (UFMG), Especialista em Educação a Distância pela Universidade Federal Fluminense (UFF), Mestre em Engenharia de Produção pela UFSC, Ph.D. em *Management Science* pela Universidade de Lancaster, na Inglaterra, e Pós-doutor em Finanças pela Universidade de Grenoble, na França. Atualmente, é Professor Titular do Departamento de Ciências Administrativas da UFMG.

Luís Henrique Marques Duarte

E-mail: luishmduarte@gmail.com

Curriculum Lattes: <http://lattes.cnpq.br/1162586369203116>

Graduando em Ciências Contábeis pela Universidade Federal de Minas Gerais (UFMG).

Vanderson Ribeiro Gonçalves

E-mail: vribeirogoncalves@gmail.com

Curriculum Lattes: <http://lattes.cnpq.br/6856729928326346>

Graduado em Ciências Contábeis pelo Centro Universitário Estácio Ribeirão Preto.

Graduado em Gestão de Cooperativas pela Universidade Federal de Viçosa (UFV).

Mestre em Controladoria e Contabilidade pela Universidade Federal de Minas Gerais.