convibra
www.convibra.org

Previsão do Valor de Mercado das Empresas com Algoritmos de *Machine Learning*

CÁSSIA DE OLIVEIRA FERREIRA
Universidade Federal de Minas Gerais - UFMG
LUA SYRMA ZANIAH SANTOS
Universidade Federal de Minas Gerais - UFMG
ROBERT ALDO IQUIAPAZA COAGUILA
Universidade Federal de Minas Gerais - UFMG
EWERTON ALEX AVELAR
Universidade Federal de Minas Gerais – UFMG
VALÉRIA GAMA FULLY BRESSAN
Universidade Federal de Minas Gerais – UFMG

Resumo

A pesquisa apresentada neste artigo buscou analisar o desempenho de diferentes algoritmos de *Machine Learning* (ML) na previsão do valor de mercado das empresas listadas em bolsa na Brasil, Bolsa, Balcão (B3). Foram considerados três algoritmos, a saber: *Random Forest* (RF), Redes Neurais Artificiais (RNA) e *k-Nearest Neighbors* (KNN). Foi empregada uma amostra composta por 194 empresas, no período de 2010 a 2019. Os dados foram extraídos da base Thomson Reuters Refinitiv Eikon® e utilizados para compor as variáveis aplicadas como preditoras do valor de mercado das empresas, a saber: Estrutura de capital, Retorno sobre o Ativo, Retorno sobre o Patrimônio Líquido e Crescimento. Dentre os resultados da pesquisa, cabe ressaltar o melhor desempenho dos modelos KNN. Contudo, a técnica de regressão quando comparada ao uso dos algoritmos RF e RNA, apresentou um maior desempenho preditivo. Os resultados levantam questionamentos sobre a superioridade de pelo menos parte dos algoritmos de ML em relação a técnicas estatísticas tradicionais para previsão do valor de mercado, muito relevante para a tomada de decisões acerca de suas aplicabilidades por parte de gestores e investidores das empresas listadas na bolsa brasileira, em virtude dos potenciais benefícios decorrentes dessa previsão. A pesquisa apresenta como contribuição a constatação de resultados que contrapõem ao apresentado usualmente na literatura visto que estudos têm evidenciado a superioridade no desempenho de algoritmos de ML na previsão de valores de mercado em relação a técnicas estatísticas tradicionais. Porém, os resultados encontrados, pelo menos em parte, levantam questionamentos sobre a superioridade sobre isso.

Palavras-Chave: Valor de mercado; *Machine Learning*; Indicadores econômico-financeiros.

1 Introdução

Prever o valor de mercado das empresas é um desafio complexo em virtude das indeterminações envolvidas e as diversas variáveis que influenciam tal valor, tais como condições macroeconômicas, sentimentos dos investidores e eventos políticos (Khaidem, Saha, & Dey, 2016). Há de se mencionar, ainda, que o valor da empresa é influenciado por outras variáveis relacionadas a sua estrutura econômico-financeira específica, tais como estrutura de capital, crescimento do ativo e lucratividade (Ihlenffeldt, Cunha Júnior; Silva, & Silva, 2017; Danise & Robert, 2009).

Em virtude disso, a análise e a previsão do valor de mercado é uma área relevante de estudos, na qual esforços têm sido direcionados para o desenvolvimento de métodos com melhor desempenho de previsão (Khaidem *et al.*, 2016). Há, na literatura, diversas abordagens cujo objetivo é prever o valor de mercado, tais como técnicas estatísticas de séries temporais (Cavdar & Aydin, 2020) e recentemente, pode-se mencionar o avanço da tecnologia computacional que possibilitou o emprego em larga escala de algoritmos ligados à Inteligência Artificial (IA), em especial de Aprendizado de Máquina (*Machine Learning* – ML).

Conforme apresentado por Mateus e Mendonça (2020), o ML é uma das áreas que envolvem a IA que, com uso de algoritmos, treina os sistemas computacionais, tornando-os, capazes de identificar padrões e tomar decisões com o mínimo de intervenção humana, além de fornecer resultados mais rápidos e precisos para as organizações. Assim, em geral, quando se fala em ML e seus diversos algoritmos de aprendizado, o que se objetiva é a identificação de padrões. Os algoritmos de ML são classificados em preditivos ou descritivos, sendo divididos em: (i) aprendizagem supervisionada; (ii) aprendizagem não supervisionada; e (iii) aprendizagem por reforço. Dentre os principais algoritmos voltados à aprendizagem supervisionada (foco da pesquisa apresentada neste trabalho), podem ser destacados: *Random Forest* (RF), Redes Neurais Artificiais (RNA) e *k-Nearest Neighbors* (KNN).

Diante do exposto e tendo em vista a relevância do assunto, o estudo apresentado neste artigo visou analisar o desempenho de diferentes algoritmos de ML (RF, RNA e KNN) na previsão do valor de mercado das empresas listadas na Brasil, Bolsa, Balcão (B3) entre os anos de 2010 e 2019. A escolha desse período deu-se por ser o primeiro ano da aplicação das *International Financial Reporting Standards* (IFRS) no mercado nacional e último ano antes

dos efeitos da pandemia de COVID-19. Como objetivos específicos, foram propostos e cumpridos: (i) comparar o desempenho dos algoritmos supracitados na previsão do valor de mercado das empresas a partir de dados econômico-financeiros; (ii) comparar o desempenho no uso dos algoritmos em relação a outras técnicas de previsão tradicionais como modelos de regressão com dados em painel; e (iii) discutir as implicações do emprego de algoritmos de ML para previsão em finanças corporativas.

Este estudo se justifica tendo em vista que se trata de uma área de pesquisa ainda em seus passos iniciais, fazendo-se necessária a realização de experimentos e testes para avaliar a confiabilidade de tais métodos computacionais para a previsão de valor. Esses tipos de modelos, quando implementados, testados e comprovados, podem resultar em uma ferramenta para a gestão de ativos, podendo ser aplicada tanto por pessoas comuns quanto por grandes gestoras de capitais visando gerar mais renda e também proteção de portfólio (Davenport & Bean, 2021; Natarajan, 2021). Ademais, ressalta-se que o estudo apresentado é relevante, considerando-se a importância de ferramentas de previsão para os agentes do mercado financeiro (Wu, Wang, Su, Tang, & Wu (2020); e a contribuição à literatura sobre o crescente emprego de IA em finanças (Ferreira, Gandomi, & Cardoso, 2021).

2 Revisão da Literatura

2.1 Preditores do valor de mercado das empresas

A predição do valor de mercado das empresas a partir de indicadores econômico-financeiros tem sido objeto de estudo de diversos pesquisadores (*e.g.*, Caixe & Krauter, 2013; Medrado, Cella, Pereira, & Dantas, 2016; Zanon, Araújo, & Nunes, 2017; Vasconcelos & Forte, 2022). Dentre tais variáveis empregadas podem ser destacadas: Retorno sobre o Ativo (ROA), Retorno sobre o Patrimônio Líquido (ROE), tamanho das organizações, crescimento dos ativos, lucro antes dos juros (*earnings before interest taxes depreciation and amortization* – EBITDA) e estrutura de capital. Ressalta-se que Medrado *et al.* (2016) afirmam que o ROE consiste em um indicador econômico-financeiro que reflete a capacidade da empresa de agregar valor a si mesma, por meio de seus próprios recursos e nesse sentido, espera-se que maiores taxas influenciem positivamente no valor de mercado das empresas, mostrando que a organização está gerando rentabilidade para os acionistas.

Ademais, a rentabilidade da empresa pode influenciar o seu valor de mercado, porque os investidores podem dar prioridade aos títulos de empresas que obtiveram, relativamente às

demais, lucros contábeis mais satisfatórios ao enfrentarem o *trade-off* existente entre vender, comprar ou manter suas ações (Cho & Pucik, 2005; Caixe & Krauter, 2013). Vasconcelos e Forte (2022) consideram que o ROA se apresenta como um direcionador, tal como o ROE, e pode contribuir para o aumento do valor de mercado das empresas, apresentando a relação entre o luco líquido e os ativos totais, pressupondo que o mercado inclina-se a valorizar organizações mais rentáveis (Zanon *et al.*, 2017).

Outra variável utilizada nas pesquisas que analisam direcionadores de valor de mercado, é o tamanho das organizações. Nunes, Teixeira, Nossa e Galdi (2010) esclarecem que grandes empresas seriam mais intensamente pressionadas por diversos *stakeholders* em suas atividades, o que as levaria a incorrer em maiores custos políticos e possivelmente que poderiam afetar seus resultados e, consequentemente, em seus valores de mercado (Medrado *et al.*, 2016). No mesmo sentido, Silveira, Lanzana, Barros e Famá (2004) reforçam que grandes corporações poderiam estar sujeitas a maiores problemas de agência e, consequentemente, em maiores custos de monitoramento, o que resultaria em um menor valor de mercado e uma maior concentração acionária para tentar contornar esses problemas.

Adicionalmente, estudos incorporam a variável crescimento dos ativos como preditora do valor da empresa. Tais como Silva e Silva (2017) que constataram que as sociedades por ações tenderiam a investir de forma subótima a fim de expropriarem a riqueza dos detentores de títulos da empresa. Assim, o custo relacionado a esse problema de agência provavelmente seria maior para empresas em crescimento, que possuem maior flexibilidade na escolha de seus investimentos futuros. De acordo com Myers (1977), quanto mais as opções de crescimento na oportunidade de investimento forem estabelecidas, maior será o conflito entre os acionistas e os credores acerca do exercício dessas opções, uma vez que existem diferentes reivindicações sobre a empresa, o que poderia ser resolvido incluindo-se menos dívida na estrutura de capital, encurtando-se o vencimento de sua dívida, conforme argumentam Silva e Silva (2017).

O EBITDA por sua vez, é amplamente utilizado nesse tipo de estudo e representa quanto a empresa gera de recursos considerando exclusivamente suas atividades operacionais, excluindo-se os efeitos financeiros e de impostos (Ihlenffeldt *et al.*, 2018; Vasconcelos & Forte, 2022). É utilizado primordialmente como uma variável de desempenho das organizações, refletindo-se em uma *proxy* da capacidade operacional de geração de caixa,

devendo ser precificado de maneira positiva pelo mercado. Portanto, possui o potencial de demonstrar que os investidores valorizam as empresas que podem gerar caixa operacional, uma vez que pode apontar a capacidade da empresa de honrar suas obrigações contratuais, de gerar resultados e em distribuir dividendos (Medrado *et al.*, 2016).

Por fim, quanto à estrutura de capital, conforme pesquisa seminal de Modigliani e Miller (1958) a forma de financiamento seria indiferente, isto é, não seria determinante do valor da empresa. Contudo, a verificação do contrário por estudos posteriores motivou o surgimento de teorias que se utilizaram de elementos da realidade empresarial para analisar os fatores que explicariam o financiamento das empresas (Ihlenffeldt *et al.*, 2018; Silva & Silva, 2017). As indagações sobre como a estrutura de capital afeta o valor da organizações continuam sendo uma das preocupações mais relevantes no financiamento corporativo, conforme destacam Silva e Silva (2017) que afirmam não existir um consenso sobre o tema, apesar do grande volume de pesquisas acerca da relação entre esses fatores.

Os diversos estudos supracitados têm empregado tais variáveis em modelos econométricos para previsão do valor das empresas. Porém, recentemente, a aplicação de algoritmos relacionados à IA em finanças, levanta questionamentos sobre a superioridade dessas técnicas sobre os modelos tradicionais. Ferreira, Gandomi e Cardoso (2021) complementam que a partir da introdução de métodos computacionais em finanças, por volta de 1990, diversas pesquisas se concentraram na aplicação de Inteligência Artificial contextualizada ao ambiente do mercado de ações. Dentre as principais vantagens do uso de abordagens computacionais nesse contexto, os referidos autores citam a possibilidade de eliminação da "irracionalidade momentânea" ou das decisões que são tomadas sendo fundamentadas em emoções, além da utilização de informações em tempo tempo real.

2.2 Algoritmos de *Machine Learning*

Diversos estudos que empregaram algoritmos de IA para previsão de preços de ativos financeiros vêm apresentando modelos com alto desempenho de capacidade preditiva, potencializando retornos anormais, tais como: Qian e Rasheed (2007) e Shynkevich, McGinnity, Coleman, Belatreche e Li (2017). Destaca-se que tais estudos têm usado diferentes algoritmos de IA. Faceli, Lorena, Gama, & Carvalho (2021) reforçam que a crescente complexidade dos problemas a serem tratados, por ferramentas computacionais, bem como a velocidade e o volume de dados gerados por diversos setores, propiciou o

desenvolvimento de ferramentas computacionais mais autônomas e independentes da intervenção humana, para a aquisição de conhecimento. A grande parte dessas ferramentas é baseada em ML.

Os algoritmos de ML são categorizados principalmente em abordagens supervisionadas e não supervisionadas. No primeiro tipo, os dados de entrada nomeados e a saída desejada são fornecidos aos algoritmos de aprendizado. Já na abordagem de aprendizado não supervisionado, dados de entrada não rotulados são fornecidos ao algoritmo de aprendizado, e o algoritmo identifica os padrões e gera as saídas (Rouf *et al.*, 2021).

Nesse contexto, os algoritmos preditivos são aplicados a um conjunto de dados de treinamento rotulado para induzir um modelo preditivo capaz de prever, para um novo objeto, representado pelo valor dos atributos preditivos, o valor do atributo alvo. Em outras palavras, "um algoritmo de ML preditivo é uma função que, dado um conjunto de dados rotulados, constrói um estimador" (Faceli *et al.*, 2021, p. 50). Os modelos podem ser classificados como de regressão ou classificação, onde os modelos regressivos buscam encontrar valores reais (métricos) e os modelos de classificação buscam distingui-los em uma classificação binária (Faceli *et al.*, 2021).

Existem diferentes tipos de algoritmos de ML que podem ser aplicados na previsão de valor das empresas, tais como: RF, RNA e KNN. Sobre o primeiro, Sadorsky (2021) considera que são algoritmos fáceis de entender e tendem a ter bom desempenho para prever preços de ações (Basak, Kar, Saha, & Khaidem (2019); Khan *et al.*, 2022; Lohrmann & Luukka, 2019). Conforme Breiman (2001), RFs são constituídas por um grande número de árvores de decisão individuais que funcionam como um conjunto. Já no método de KNN, a previsão é mapeada em uma classificação baseada na proximidade. Por meio de estimações que utilizam a distância euclidiana, o KNN classifica os "k" vizinhos mais próximos no conjunto de treinamento. Esse método consiste em uma estrutura computacional não linear (Rouf *et al.*, 2021).

Por fim, as RNA "são compostas de unidades de processamento simples, que implementam funções matemáticas que simulam funções desempenhadas por neurônios. Essas unidades podem se conectar a várias outras conexões que simulam as sinapses, permitindo que as RNAs possam resolver problemas complexos" (Faceli *et al.*, 2021, p. 101). Portanto, trata-se de uma técnica inspirada no cérebro biológico, na qual um grande número

de neurônios artificiais são intensamente interconectados para resolver problemas complexos. Esses modelos entendem o contexto de um problema gerando múltiplas transformações no espaço de recursos, seguidas de não linearidade, no intuito de criar suas representações simplificadas (Rouf *et al.*, 2021; Sharma *et al.*, 2021). A partir desse ponto, passa-se a exposição de estudos relacionados ao tema proposto na presente pesquisa, trazendo suas principais contribuições e reflexões.

2.3 Estudos anteriores

Com objetivo de corroborar com a discussão proposta no presente artigo, buscou-se trazer nesta seção os estudos anteriores que contribuíram para a temática que se refere à utilização dos algoritmos de ML na previsão de valor das empresas, em especial RF, RNA e KNN, de maneira que as pesquisas apresentadas a seguir foram selecionadas considerando-se sua aderência e contribuições ao assunto. Vaisla e Bhatt (2010) empregaram redes neurais e técnicas estatísticas para modelar e prever os preços diários do mercado de ações e, em seguida, compararam os resultados desses dois modelos. Os resultados mostraram que as RNAs, quando treinadas com dados suficientes, entradas adequadas e com arquitetura adequada, previam muito bem os preços do mercado de ações. Os autores destacaram que a técnica estatística, embora bem construída, possui a capacidade de previsão reduzida à medida que as séries se tornam complexas. Dessa forma, Vaisla e Bhatt (2010) concluíram que as RNAs podem ser usadas como uma técnica alternativa melhor para prever os preços diários do mercado de ações. Na mesma direção, a proposta de Silva, Muntaser, Penedo e Pereira (2018) foi realizar uma comparação entre ferramentas que possibilitassem melhores previsões do valor de mercado: as RNAs e o modelo de regressão linear múltipla com dados em painel. Os autores coletaram dados de 384 empresas listadas na bolsa brasileira entre 2011 e 2015 e constataram que utilizar tais redes minimizava o erro da previsão.

Por sua vez, Santos, Barboza e Veiga (2020) colocaram o foco de suas análises em três empresas listadas na B3, com o intuito de comparar as predições em relação a tendência de suas respectivas ações por meio dos algoritmos RNA, RF e *Support Vector Machines* e concluíram que esses dois últimos obtiveram desempenhos melhores comparativamente às RNAs, com níveis de acurácia no intervalo de 91 a 97%. Outro trabalho que testou o mesmo tipo de algoritmo foi o de Sobreiro (2008) que avaliou a aplicação de RNAs na previsão do valor da *commodity* de açúcar, demonstrando que a aplicação obteve uma aproximação

significante comparada às cotações reais, evidenciando a importância da RNAs como um método alternativo na estimação de valores para as commodities a todos os atores envolvidos no setor sucroalcooleiro.

Já no trabalho de Khaidem *et al.* (2016) o algoritmo RF foi utilizado para prever as tendências futuras para um, dois e três meses das ações de empresas listadas na NASDAQ e na Bolsa de Valores da Coréia, no período 2008 a 2013. Dentre os resultados encontrados pelos autores, verificou-se que a utilização do algoritmo RF apresentou acurácia entre 84,5% e 94,5%, apresentando resultados distintos em decorrência do período analisado. Corrobora a esses achados a pesquisa de Khan et. al. (2022) que usaram 12 métodos de aprendizado de máquina aplicados a mídias sociais e dados financeiros para prever preços de ações e concluíram que o método RF foi consistentemente classificado como um dos melhores métodos.

Basak *et al.* (2019) utilizaram o mesmo algoritmo (RF) para preverem o valor de mercado das empresas *Apple* e *Facebook* utilizando indicadores técnicos e encontraram acuracidades entre 65% e 95%. Os autores afirmam que os resultados encontrados superam os resultados obtidos anteriormente por estudos que utilizaram abordagem tradicionais de previsão. Já Avelar, Leocádio, Campos, Ferreira e Orefici (2022a) obtiveram resultados semelhantes ao analisarem o desempenho do algoritmo RF e RNA na previsão do retorno futuro dos principais índices das maiores bolsas de valores do mundo no período 2001 a 2019, por meio de preços históricos de negociação. Os resultados apontaram que os desempenhos dos modelos estimados foram superiores à média de mercado, sendo que o RF apresentou os melhores resultados.

Cervelló-Royo e Guijarro (2020) compararam o desempenho de quatro algoritmos de aprendizado de máquina para validar a capacidade de previsão de indicadores técnicos populares no índice NASDAQ tecnológico. Os resultados apontaram que a metodologia de RF supera os outros algoritmos de aprendizado de máquina considerados na pesquisa, tendo sido capaz de prever a tendência do mercado 10 dias à frente, com uma precisão média de 80%. O algoritmo de RF também obteve a taxa de sucesso mais precisa (94,12%) no estudo de Daori *et al.* (2022), que buscaram prever os setores mais lucrativos do mercado de ações, considerando as previsões acerca do preço da ação, preço de fechamento e preço de abertura. Complementarmente, Alkhatib, Najadat, Hmeidi e Shatnawi (2013) utilizaram o algoritmo

KNN na previsão dos preços das ações em cinco empresas listadas na bolsa de valores da Jordânia e constaram que tal algoritmo é robusto e produz pequenas taxas de erro. Os autores destacam, no entanto, que existe um longo caminho para utilizar modelos avançados de previsão devido à falta de conhecimento em técnicas de IA na Jordânia.

Por fim, é importante destacar que, apesar de poucos, alguns estudos têm apresentado resultados semelhantes entre técnicas estatísticas e algoritmos de ML (Parray *et al.*, 2020; Jaggi *et al.*, 2021) ou até mesmo destacando a superioridade daquelas sobre estes últimos (Pyo *et al.*, 2017; Jang & Lee, 2019). A partir do arcabouço proporcionado pela literatura levantada até esse ponto, passa-se então, para a apresentação dos aspectos metodológicos relacionados ao desenvolvimento da pesquisa.

3. Metodologia

O estudo apresentado neste artigo pode ser classificado como descritivo, quantitativo e documental, a partir de dados extraídos das demonstrações financeiras obtidas por meio da base de dados *Thomson Reuters Refinitiv Eikon*®.

A população do estudo foi composta por empresas brasileiras de capital aberto, com ações negociadas na B3 no período de 2010 a 2019. A delimitação do período é justificada tendo em vista que, a partir de 2010, efetivou-se a adoção obrigatória do *International Financial Reporting Standard* (IFRS) no Brasil, o que implica na modificação significante das informações divulgadas pelas companhias e, portanto, pode afetar os dados econômico-financeiros de interesse da investigação, bem como as análises realizadas. Além disso, a amostra foi limitada ao ano de 2019 em decorrência da pandemia de COVID-19 ter se iniciado em 2020, alterando de forma significante os indicadores econômico-financeiros e o valor de empresas de diversos setores, conforme Avelar, Orefici, Borges, Campos e Ferreira (2022b). A amostra final foi constituída por por 194 empresas, totalizando 1.150 observações, conforme Tabela 1:

Tabela 1
Procedimento de seleção da amostra

	Empresas
Total de empresas de capital aberto brasileiras	576
(-) Companhias com PL negativo para cálculo das variáveis ROE e EC	79
(-) Companhias dados incompletos para cálculo das variáveis	303
Amostra Final	194

Fonte: Dados da pesquisa

Os dados extraídos da base *Thomson Reuters Refinitiv Eikon®* foram utilizados para compor os indicadores econômicos-financeiros utilizados para previsão do Valor de Mercado (VM) das empresas no estudo, os mesmos são destacados na Tabela 2. Ressalta-se que as variáveis utilizadas para a previsão do VM foram escolhidas a partir da revisão de literatura que apontou quais as variáveis mais utilizadas em estudos semelhantes.

Tabela 2
Descrição dos indicadores economômicos-financeiros

Indicador	Descrição	Operacionalização	Embasamento
EC	Estrutura de Capital	$\frac{Passivo\ Total}{Patrimônio\ Líquido}$	Danise e Robert (2009), Silva e Silva (2017), Ihlenffeldt *et al.* (2018)
ROA	Retorno sobre o Ativo	$\frac{Lucro\ Operacional}{Ativo}$	Ihlenffeldt *et al.* (2018), Vasconcelos e Forte (2022)
ROE	Retorno sobre o Patrimônio Líquido	$\frac{Lucro\ Operacional}{Patrimônio\ Líquido}$	Caixe & Krauter (2013), Medrado *et al.* (2016), Zanon *et al.* (2017), Ihlenffeldt *et al.* (2018), Vasconcelos e Forte (2022)
TAM	Ativo	*ln*(Ativo Total)	Nunes *et al.* (2010), Caixe e Krauter (2013), Medrado *et al.* (2016), Zanon *et al.* (2017)
CRES	Crescimento do Ativo	$\frac{Ativo\ t - (Ativo\ t - 1)}{Ativo\ t}$	Cho e Pucik (2005), Silva e Silva (2017)
EBITDA	EBITDA	$ln(Resultado\ Líquido + Juros + Impostos + Depreciação + Amortização)$	Medrado *et al.* (2016), Ihlenffeldt *et al.* (2018), Vasconcelos e Forte(2022)

Fonte: Dados da pesquisa

Ressalta-se que se utilizou a regressão com dados em painel conforme modelo apresentado na Equação 1. Os dados obtidos foram divididos entre dados de treino e dados de teste, sendo que o período 2010 a 2016 foi utilizado como dados de treino e o período 2017 a 2019 para dados de teste. Salienta-se que, no modelo apresentado, o *i* e o *t* representam, respectivamente, a organização e o período, enquanto μ equivale ao erro.

$$VM = \beta 0 + \beta 1EC_{it} + \beta 2ROA_{it} + \beta 3RO_{it} + \beta 4TAM_{it} + \beta 5CRES_{it} + \beta 6EBITDA_{it} + \mu_{it} \quad (1)$$

Em que:
VM = Valor de Mercado

EC = Estrutura de capital
ROA =Retorno sobre o Ativo
ROE = Retorno sobre o Patrimônio Líquido
TAM = Tamanho
CRES = Crescimento
EBITDA = *Earnings Before Interest, Taxes, Depreciation and Amortization*

A escolha do melhor modelo com dados em painel foi realizada a partir dos testes que comparam os modelos *Pooled*, efeitos fixos e efeitos aleatórios: (i) teste F de Chow – comparação entre *Pooled* e Efeitos Fixos; (ii) teste de Breush-Pagan – comparação entre *Pooled* e Efeitos Aleatórios; e (iii) teste de Hausmann – comparação entre Efeitos Aleatórios e Efeitos Fixos (Fávero, 2013). Para verificar a existência de heteroscedasticidade e autocorrelação do modelo proposto, foram empregados os testes de Wald Modificado e Wooldridge, respectivamente.

Alguns parâmetros específicos foram ajustados para determinados algoritmos: (a) RF – possibilitou-se uma variação entre 5 (cinco) e 50 (cinquenta) árvores estimadas, para que o algoritmo selecionasse a quantidade que forneceria o melhor desempenho para cada índice, mitigando problemas de *overfitting*; (b) KNN – possibilitou-se uma variação entre 1 (um) e 50 (cinquenta) para o parâmetro "k", de modo que o algoritmo selecionasse o melhor "k-vizinhos mais próximos" para cada índice; e (c) RNA – optou-se por uma arquitetura com 2 neurônios de entrada e 2 camadas, visto ser a configuração com melhor desempenho. Para mensurar o desempenho dos modelos, utilizou-se o *Root Mean squared error* (RMSE) que corresponde ao desvio-padrão das diferenças entre o previsto e o real. Conforme indicado por Faceli *et al.* (2021), trata-se de uma mensuração adequada para problemas de regressão. A fórmula para cálculo do RMSE é dada pela Equação 2.

$$RMSE = \frac{\sum \sqrt{(y_i - x_i)^2}}{n} \qquad (2)$$

Na Figura 1, apresenta-se a forma de treinamento e teste dos modelos, com base no fluxograma básico de Ferreira *et al.* (2021) que utilizou algoritmos para previsão do movimento de preços de ativos financeiros com base em IA. Destaca-se que todos os dados foram tratados e analisados a partir do software livre RStudio, um ambiente de desenvolvimento integrado para R, que é uma linguagem de programação para gráficos e cálculos estatísticos, e recursos de bibliotecas de manipulação e visualização dos dados, assim

como implementação das instruções necessárias para o aprendizado, utilizando os seguintes pacotes: *Training of Neural Networks* (neuralnet), *Breiman and Cutler's Random Forests for Classification and Regression* (randomForest); *Classification and Regression Training* (caret)*; Evaluation Metrics for Machine Learning* (Metrics).

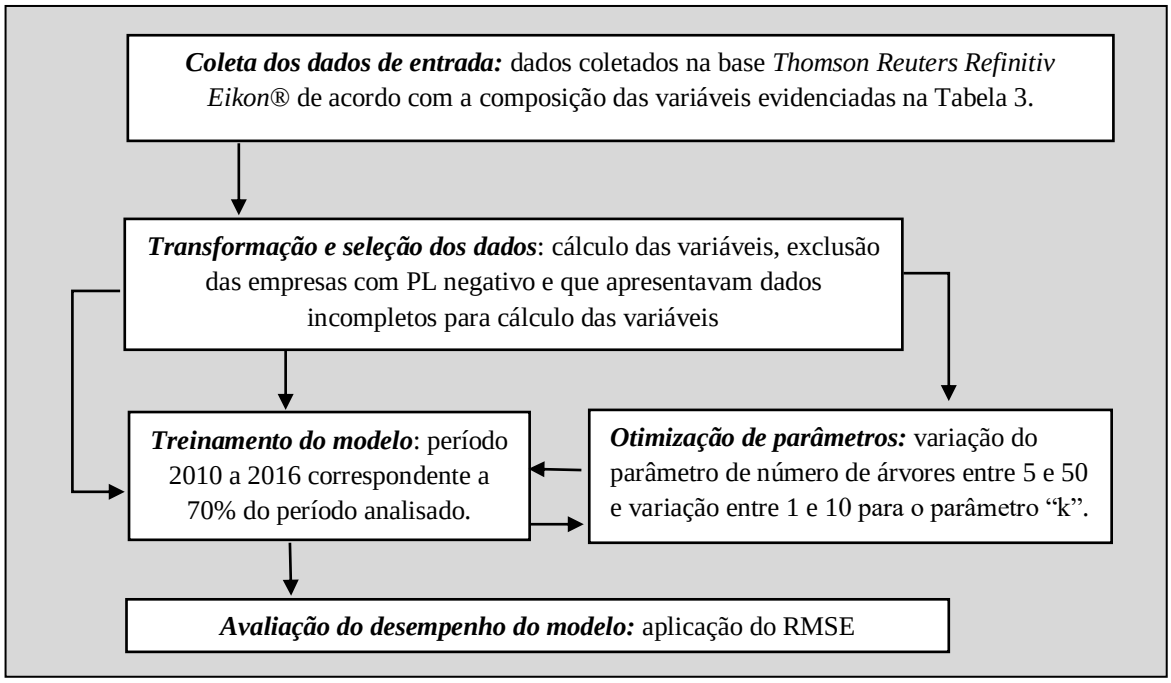

Figura 1. Fluxograma do processo de uso de algoritmos aplicado neste estudo
Fonte: Desenvolvido pelos autores com base em Ferreira *et al.* (2021)

4 Resultados e Discussões

Destaca-se que o ajuste do modelo de regressão adotado foi realizado a partir da utilização de testes estatísticos, fundamentados na literatura, que objetivaram a sua validação e adequabilidade. O resultado do teste de multicolinearidade apresentou um VIF médio de 1,25, o que indica ausência do problema. Na Tabela 3, são evidenciados os resultados da modelagem econométrica proposta e as estatísticas de validação.

Tabela 3
Validação e estimação do modelo proposto.

Variável	Coeficiente
Ativo	0,89438 ***
ROA	0,09279 ***
CRES	-6,16273 ***
EC	-1,53531 ***
Constante	2,46888 ***
No. de observações	1150
No. de empresas	194

Chow	11,09***
Breusch-Pagan	39,73**
Hausman	61,50**
Wooldridge	41,06**
Wald	86,30***

Notas: As significâncias estatísticas dos testes são representadas por meio da seguinte simbologia: 10%; **5%; ***1%.

Na presença de heteroscedasticidade e/ou autocorrelação, utilizou-se o modelo estimado por *Generalized Least Squares* (GLS) com ajustes para heterocedasticidade e autocorrelação conforme indicado por Wooldridge (2010). Adicionalmente foram aplicados os testes de Chow, *Lagrange Multiplier*, de Breush-Pagan e Teste de Hausman, com a finalidade de estimar o modelo de dados em painel adequado para esta pesquisa. Verifica-se que todas as variáveis analisadas foram significantes ao nível de 1%. Evidenciando que são consideradas preditoras para o Valor de Mercado das empresas da amostra, assim como constatado por estudos anteriores.

A partir dos coeficientes gerados, estimou-se o RMSE do modelo de regressão comparativamente ao resultado dos algoritmos empregados na análise. Sabe-se que o RMSE representa a distância média entre as previsões e os valores reais. Logo, quanto maior o RMSE do modelo, pior é o seu desempenho, já que valores altos indicam maior quantidade de erros de predição. Os resultados obtidos são apresentados na Tabela 4. Pode-se observar que o algoritmo KNN apresentou menor erro de previsão quando comparado aos demais algoritmos e à regressão.

Os resultados apontaram que nem todos os modelos desenvolvidos com base em algoritmos de ML obtiveram desempenhos superiores ao apresentado pela técnica de regressão. Dessa forma, os resultados encontrados contrapõem ao esperado pelo estudo e ao apresentado usualmente na literatura visto que estudos tem evidenciado superioridade no desempenho de algoritmos de IA na previsão de valores de mercado. Os resultados obtidos são mais alinhados a Pyo *et al.* (2017) e Jang e Lee (2019). Ressalta-se que resultados reportados por estudos que empregaram algoritmos de IA no mercado de capitais, tais como Shynkevich *et al.* (2017) e Wu *et al.* (2020) evidenciaram que todos os modelos desenvolvidos com base em algoritmos de IA obtiveram desempenhos superiores ao esperado pelo mercado. Logo, a constatação realizada neste estudo evidencia a importância das

discussões acerca do emprego de algoritmos de ML na previsão, visto a divergência nos resultados encontrados quando comparado ao apresentado na literatura.

Tabela 4
Desempenho dos algoritmos

TÉCNICA	RMSE♣
K-Nearest Neighbors (KNN)	0,3687
Regressão com dados em painel	0,4128
Random Forest (RF)	0,6860
Redes Neurais Artificiais (RNA)	0,7088

Notas: ♣*Root Mean squared error*

A verificação de que o modelo baseado no algoritmo KNN apresentou resultado superior aos dos demais algoritmos de ML, quando se empregou o RMSE como medida de desempenho. Esses resultados corroboram com o estudo de Alkhatib *et al.* (2013), que constaram que o modelo KNN é robusto e produz pequenas taxas de erro. Verificou-se, ainda, que o desempenho do modelo baseado no algoritmo RF foi superior ao baseado em RNA. Assim, tal achado corrobora com Avelar *et al.* (2022a) que obtiveram desempenhos dos modelos estimados com RF superiores à média de mercado e ao desempenho do algoritmo RNA na previsão do retorno futuro dos principais índices das maiores bolsas de valores do mundo. O resultado do presente estudo corrobora parcialmente com os resultados de Basak *et al.* (2019), Cervelló-Royo e Guijarro (2020) e Daori *et al.* (2022), visto a precisão do algoritmo RF ser maior que RNA, contudo, as taxas de erros encontradas por estes estudos foram menores quando comparadas ao evidenciado na presente pesquisa. Observou-se que os modelos baseados no algoritmo RNA apresentaram resultados inferiores aos dos demais algoritmos ML. Uma possível explicação para esses achados se refere ao uso de algoritmos de IA mais adequados para aplicação em dados transversais.

5 Considerações Finais

Este artigo visou analisar o desempenho de diferentes algoritmos de *Machine Learning* na previsão do VM das empresas listadas na B3, quais sejam: RF, RNA e *KNN*. Os dados utilizados para análise foram extraídos da base *Thomson Reuters Refinitiv Eikon*®. As variáveis EC, ROA, ROE e CRES foram aplicadas como preditoras do VM das empresas. Ao final, o estudo utilizou-se de uma amostra composta por 194 empresas, no período de 2010 a 2019.

Os resultados da pesquisa evidenciaram a possibilidade de se obter melhor desempenho em previsões a partir de algoritmos com base nos modelos KNN, quando comparados a previsões geradas pelos algoritmos de RF, RNA e utilizando-se a técnica de regressão. Contudo, cabe destacar que a técnica de regressão, quando comparada ao uso dos algoritmos RF e RNA, apresentou com maior desempenho preditivo. Tal constatação levanta questionamentos sobre a superioridade de pelo menos parte dos algoritmos de ML em relação a técnicas estatísticas tradicionais para previsão do valor de mercado, o que contribui para a tomada de decisões acerca de suas aplicabilidades por parte de gestores e investidores das empresas listas na bolsa brasileira, em virtude dos potenciais benefícios monetários decorrentes dessa previsão

Ademais, ressalta-se que há espaço para discussão do emprego do uso de algoritmos de ML, no contexto de ascensão do uso de algoritmos de IA em previsão do valor de mercado das empresas. Os modelos baseados em tais algoritmos podem ser amplamente aperfeiçoados, ajustados a fim de que sejam empregados pelos agentes de mercado no processo de previsão do valor das empresas e tomada de decisão de investimentos.

Destaca-se que a principal limitação desta pesquisa, considera-se a falta de dados em algumas empresas que limitou o número de observações utilizadas nos modelos propostos. Ademais, a amostra limitada e não probabilística, impossibilita a generalização dos resultados. Entretanto, apesar das limitações, os achados da pesquisa contribuem para a maior compreensão do uso de algoritmos para previsão do valor de mercado das empresas, visto que se trata de uma área de pesquisa ainda em seus passos iniciais, fazendo-se necessária a realização de experimentos e testes para avaliar a confiabilidade de tais métodos para a previsão de valor. Ressalta-se ainda, que este estudo contribui para ampliar as discussões quanto a existência de superioridade dos algoritmos de ML em relação a técnicas estatísticas tradicionais para previsão do valor de mercado. Como estudos futuros, sugere-se o emprego de outros algoritmos de IA para previsão do valor de mercado, tais como *Support Vector Machine* (SVM) e/ou RNA como *long short-term memory* (LSTM), o emprego de outros dados de entrada diferentes dos indicadores econômicos-financeiros, como evolução de preços históricos das ações das empresas. Além disso, sugere-se, ainda, empregar diferentes métricas de avaliação de desempenhos dosa algoritmos para fins comparativos, também contribuirá para o avanço do tema.

Referências

Akinwande, M. , Dikko, H. e Samson, A. (2015) Fator de inflação de variância: como condição para a inclusão de variável(is) supressora(s) na análise de regressão. *Open Journal of Statistics* , 5 , 754-767. doi: 10.4236/ojs.2015.57075.

Alkhatib, K., Najadat, H., Hmeidi, I., & Shatnawi, M. K. A. (2013). Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm. *International Journal of Business, Humanities and Technology*, *3*(3).

Avelar, E. A., Leocádio, V. A., Campos, O. V., Ferreira, P. O., & Orefici, J. B. P. (2022a). Algoritmo Random Forest para Previsão de Comportamento de Preços de Ativos. *Revista FSA*, 19(10), 45–65. https://doi.org/10.12819/2022.19.10.3.

Avelar, E. A., Orefici, J. B. P., Borges, S. L., Campos, O. V., & Ferreira, P. O. (2022b). The Covid-19 pandemic effects on the financial performance of Brazilian listed companies. *Revista de Administração da UFSM*, 15, 389-410.

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552–567. https://doi.org/10.1016/j.najef.2018.06.013.

Caixe, D. F., & Krauter, E. (2013). A influência da estrutura de propriedade e controle sobre o valor de mercado corporativo no Brasil. *Revista Contabilidade & Finanças*, 24(62), 142–153. https://doi.org/10.1590/S1519-70772013000200005.

Cavdar, S. C., & Aydin, A. D. (2015). An Empirical Analysis for the Prediction of a Financial Crisis in Turkey through the Use of Forecast Error Measures ," JRFM , MDPI, vol. 8(3), páginas 1-18, agosto.

Cervelló-Royo, R., & Guijarro, F. (2020). Forecasting stock market trend: A comparison of machine learning algorithms. *Finance, Markets and Valuation*, 6(1), 37–49. https://doi.org/10.46503/NLUF8557.

Cho, H. J., & Pucik, V. (2005). Relationship between innovativeness, quality, growth, profitability, and market value. *Strategic management journal*, 26(6), 555-575.

Danise, D.; Robert, H. (2009). Revisiting The Relationship Between Insider Ownership and Performance. *Journal of Business and Economic Studies*, 15(2), 32-43.

Daori, H., Alharthi, M., Alanazi, A., Alzahrani, G., Aborokbah, M., & Aljehane, N. (2022). *Predicting Stock Prices Using the Random Forest Classifier* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-2266733/v1.

Davenport, T. H.; Bean, R. The Pursuit of AI-Driven Wealth Management. 2021. Disponível em: https://sloanreview.mit.edu/article/the-pursuit-of-aidriven-wealth-management/. Acesso em: 10/12/2022.

Faceli, K., Lorena, A. C., Gama, J., & Carvalho, A. C. P. D. L. F. D. (2011). Inteligência artificial: uma abordagem de aprendizado de máquina.

Ferreira, F. G. D. C., Gandomi, A. H., & Cardoso, R. T. N. (2021). Artificial Intelligence Applied to Stock Market Trading: A Review. *IEEE Access*, *9*, 30898–30917. https://doi.org/10.1109/ACCESS.2021.3058133.

Ihlenffeldt, E. L., Cunha júnior, C. A., Leal, F. S., & Scarpin, J. E. (2018). A Influência Da Estrutura De Capital No Valor De Mercado Das Empresas Brasileiras De Capital Aberto. *A Influência Da Estrutura De Capital No Valor De Mercado Das Empresas Brasileiras De Capital Aberto*, 12(2). Disponível em: https://revistagc.com.br/ojs/index.php/rgc/article/view/125.

Jaggi, M., Mandal, P., Narang, S., Naseem, U., & Khushi, M. (2021). Text Mining of Stocktwits Data for Predicting Stock Prices. *Applied System Innovation*, *4*(1), 13.

Jia, J., Zhao, A., & Guan, S. (2017). Forecasting based on high-order fuzzy-fluctuation trends and particle swarm optimization machine learning. *Symmetry*, *9*(7), 124.

Khaidem, L., Saha, S., & Dey, S. R. (2016). *Predicting the direction of stock market prices using random forest*. http://arxiv.org/abs/1605.00003.

Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 13(7), 3433–3456. https://doi.org/10.1007/s12652-020-01839-w.

Lohrmann, C., & Luukka, P. (2019). Classification of intraday S&P500 returns with a Random Forest. *International Journal of Forecasting*, 35(1), 390–407. https://doi.org/10.1016/j.ijforecast.2018.08.004.

Mateus, F. M. Q., Mendonça, M. C. (2020). Machine Learning na Melhoria de Processos Internos: Estudo de Caso na Indústria de Varejo Brasileira. 106 f. Universidade Federal do Rio de Janeiro. Disponível em: http://www.monografias.poli.ufrj.br/monografias/monopoli10031889.pdf.

Medrado, F., Cella, G., Pereira, J. V., & Dantas, J. A. (2016). Relação Entre O Nível De Intangibilidade Dos Ativos E O Valor De Mercado Das Empresas. *Revista de Contabilidade e Organizações*, 10(28), 32. https://doi.org/10.11606/rco.v10i28.119480.

Modigliani, F., & Miller, M. H. (1963). Corporate income taxes and the cost of capital: a correction. *The American economic review*, 53(3), 433-443.

Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of financial economics*, 5(2), 147-175.

Natarajan, G. (2021). AI in investing is about human empowerment, not displacement. Disponível em: https://economictimes.indiatimes.com/markets/stocks/news/ai-in-investing-is-about-humanempowerment-not displacement/articleshow/83685863.cms?from=mdr. Acesso em: 04 jan. 2022.

Nunes, J., Teixeira, A., Nossa, V., & Galdi, F. (2010). Análise das variáveis que influenciam a adesão das empresas ao índice BM&F Bovespa de sustentabilidade empresarial. *Base – Revista de Administração e Contabilidade da Unisinos*, 7(4), 328–340. https://doi.org/10.4013/base.2010.74.06.

Parray, I. R., Khurana, S. S., Kumar, M., & Altalbe, A. A. (2020). Time series data analysis of stock price movement using machine learning techniques. *Soft Computing*, 24, 16509–16517. https://doi.org/10.1007/s00500-020-04957-x

Pyo, S., Lee, J., Cha, M., & Jang, H. (2017). Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets. PLoS ONE, 12(11), 1-17. https://doi.org/10.1371/journal.pone.0188107

Qian, B. & Rasheed, K. (2007), 'Stock market prediction with multiple classifiers', *Applied Intelligence*, vol. 26, n. 1, pp. 25–33.

Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H.-C. (2021). Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. *Electronics*, 10(21), 2717. https://doi.org/10.3390/electronics10212717.

Sadorsky, P. (2021). A Random Forests Approach to Predicting Clean Energy Stock Prices. *Journal of Risk and Financial Management*, 14(2), 48. https://doi.org/10.3390/jrfm14020048

Santos, G. C., Barboza, F., & Veiga, A. C. P. (2020). Prevendo O Mercado De Ações Brasileiro Utilizando Machine Learning. XVII Simpep – Simpósio de Engenharia de Produção. Recuperado em 19 de janeiro de 2023, de https://www.researchgate.net/profile/Flavio-Barboza/publication/358054560_PREVENDO_O_MERCADO_DE_ACOES_BRASILEIRO_UTILIZANDO_MACHINE_LEARNING/links/61eea50dc5e3103375b9f87a/PREVENDO-O-MERCADO-DE-ACOES-BRASILEIRO-UTILIZANDO-MACHINE-LEARNING.pdf.

Sharma, S., Ahmed, S., Naseem, M., Alnumay, W. S., Singh, S., & Cho, G. H. (2021). A Survey on Applications of Artificial Intelligence for Pre-Parametric Project Cost and Soil Shear-Strength Estimation in Construction and Geotechnical Engineering. *Sensors*, 21(2), 463. https://doi.org/10.3390/s21020463.

Shynkevich, Y., McGinnity, T.M., Coleman, S.A., Belatreche, A. & Li, Y. (2017), 'Forecasting price movements using technical indicators: Investigating the impact of varying input window length', *Neurocomputing*, vol. 264, pp. 71–88.

Silva, J. P., & Silva, D. C. (2017). *Análise da relação entre estrutura de capital, crescimento, lucratividade e valor de mercado das companhias brasileiras de capital aberto*, 18(1), 15–25.

Silva, V. P., Muntaser, J. G. S., Penedo, A. S. T., & Pereira, V. S. (2018). Redes Neurais Artificiais X Regressão Com Dados Em Painel: Prevendo O Valor De Mercado Das Empresas. *Redes Neurais Artificiais X Regressão Com Dados Em Painel: Prevendo O Valor De Mercado Das Empresas*, 33(2), 133-146.

Silveira, A. de M., Lanzana, na P., Barros, L. A. B. de C., & Famá, R. (2004). *Efeito dos acionistas controladores no valor das companhias abertas brasileiras*. 39(4).

Vaisla, K. S., & Bhatt, D. A. K. (2010). *An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting*. 02(06).

Vasconcelos, T. de, & Forte, D. (2022). Analysis of factors that influence the companies' market value in England, France, and Germany. How can managers maximize such value? *Revista de Gestão dos Países de Língua Portuguesa*, 21(1), 43–66. https://doi.org/10.12660/rgplp.v21n1.2022.82338.

Wu, D., Wang X., Su, J., Tang, B., Wu S. (2020). "A Labeling Method for Financial Time Series Prediction Based on Trends" *Entropy* 22, no. 10: 1162. https://doi.org/10.3390/e22101162.

Zanon, A. R. M., de Araújo, C. G., & Nunes, A. (2017). Influence Of Dividend Policy On Market Value Of Brazilian Companies. *Revista De Gestão*, 7(3).

Agradecimento à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento do presente estudo.