



DESEMPENHO FINANCEIRO E ESG: UMA ABORDAGEM COMPARATIVA POR MEIO DE APRENDIZADO DE MÁQUINA

Maria Clara de Oliveira Ramos¹

Allisson Silva dos Santos²

RESUMO

Este estudo almeja compreender como os indicadores ESG (ambiental, social e de governança) são capazes de contribuir para a predição do desempenho financeiro das empresas, utilizando-os como variáveis preditoras, em diferentes modelos de aprendizado de máquina. Metodologicamente, a pesquisa utilizou sete algoritmos de classificação (Random Forest, K-Nearest Neighbors (KNN), Naïve Bayes, Regressão Logística (Logit), Support Vector Machine (SVM) linear, Bagging e Gradient Boosting Machine (GBM)), avaliados por métricas de acurácia e curvas ROC/AUC. Os resultados destacaram o modelo Gradient Boosting como o mais eficaz, atingindo 67,5% de acurácia no conjunto de teste. A análise revelou que a variável ambiental foi o preditor mais relevante para o ROA, seguida pela dimensão social, enquanto a governança apresentou menor impacto. Conclui-se que ações ambientais devem ser encaradas pelo mercado como investimentos capazes de gerar retorno financeiro, e não apenas custos. O trabalho contribui para a literatura ao integrar ESG, finanças e machine learning, incentivando a transparência e a adoção de políticas corporativas sustentáveis.

Palavras-chave: ESG. Desempenho financeiro. Aprendizado de máquina.

FINANCIAL PERFORMANCE AND ESG: A COMPARATIVE APPROACH THROUGH MACHINE LEARNING

ABSTRACT

This study aims to understand how ESG (environmental, social, and governance) indicators can contribute to predicting the financial performance of companies, using them as predictor variables in different machine learning models. Methodologically, the research used seven classification algorithms (Random Forest, K-Nearest Neighbors (KNN), Naïve Bayes, Logistic Regression (Logit), Linear Support Vector Machine (SVM), Bagging, and Gradient Boosting Machine (GBM)), evaluated by accuracy metrics and ROC/AUC curves.

Keywords: ESG. Financial performance. Machine learning.

¹ Graduada em Análise e Desenvolvimento de Sistemas (UEPB) e graduanda em Letras – Inglês (UEPB).

² Doutorando e Mestre em Administração pelo Programa de Pós-Graduação em Administração - PPGA da Universidade Federal da Paraíba - UFPB

The results highlighted the Gradient Boosting model as the most effective, achieving 67.5% accuracy in the test set. The analysis revealed that the environmental variable was the most relevant predictor for ROA, followed by the social dimension, while governance had the least impact. It is concluded that environmental actions should be viewed by the market as investments capable of generating financial returns, and not just costs. This work contributes to the literature by integrating ESG, finance, and machine learning, encouraging transparency and the adoption of sustainable corporate policies.

1 INTRODUÇÃO

Os princípios ESG (*Environmental, Social and Governance*) podem ser entendidos como um guia de políticas para as empresas se tornarem mais responsáveis sobre esses três eixos da sigla. Assim, empresas se comprometem em realizar mudanças que tenham impactos positivos na governança, na sociedade e no meio ambiente (Amaral; Willerding; Lapolli, 2023). A preocupação com o meio ambiente e o meio social tem aumentado a cada dia, principalmente diante dos problemas observados na sociedade, o que faz com que a ajuda de empresas, adotando os princípios ESG, seja crucial para resolver esses impasses, ou pelo menos, diminuir os impactos deles.

Além dos impactos ambientais e sociais, a implementação do ESG pode influenciar diretamente o desempenho financeiro das empresas. Um dos indicadores mais utilizados para medir essa performance é o retorno sobre ativos (*Return on Assets* – ROA), que representa a eficiência da empresa em gerar lucro a partir de seus ativos (Ou *et al.*, 2023). Dessa forma, avaliar a relação entre a adoção do ESG e o ROA se torna fundamental para entender se essas práticas agregam valor econômico às empresas.

Nos últimos anos, o mundo vem sofrendo grandes mudanças ambientais, sociais e governamentais. Segundo dados da Organização Internacional de Sustentabilidade, o uso de recursos naturais está ocorrendo 1,7 vezes mais rápido do que a natureza consegue se regenerar (CNN Brasil, 2024). Na área social, segundo a Oxfam (2024), a riqueza dos cinco maiores bilionários do mundo dobrou desde 2020, enquanto a riqueza de 5 bilhões de pessoas, cerca de 60% da população global, diminuiu nesse mesmo período. Esse contexto é preocupante, porque com a falta de acesso a recursos básicos, como educação, saúde e com a escassez de recursos naturais há a diminuição de oportunidades de crescimento da sociedade como um todo. Esses motivos mostram a importância para que as empresas adotem práticas sustentáveis e responsáveis, que estão conectadas aos princípios do ESG, auxiliando o Estado a enfrentar essas adversidades tanto no eixo ambiental, quanto no social e governamental.

Segundo a pesquisa Global Trends de 2023, quase 70% dos brasileiros afirmam procurar comprar produtos de marcas socialmente responsáveis, mesmo que isso signifique um custo mais elevado (EXAME, 2024). Junto a isso, Giese *et al.* (2019) afirmam que empresas que adotam esses princípios apresentam menores riscos sistemáticos, resultando em um custo de capital mais baixo, o que gera avaliações mais altas no mercado financeiro para essas empresas. Assim, empresas comprometidas com os princípios ESG tendem a ser mais interessantes para investidores e consumidores.

Diante disso, com o auxílio da evolução da tecnologia, especialmente os avanços em inteligência artificial e suas subáreas, o aprendizado de máquina pode desempenhar um papel importante na análise das empresas, podendo assim contribuir para a análise de dados financeiros e corporativos de acordo com os princípios do ESG. Esses sistemas de aprendizado de máquina funcionam por meio de treinamentos realizados com conjuntos de dados que são definidos previamente, e, por meio desses dados, aprendem a identificar padrões (Zhou, 2021).

A tecnologia está evoluindo rapidamente e trazendo novas possibilidades em diversas áreas. Isso inclui o uso de aprendizado de máquina ligado aos princípios do ESG (Del Vitto; Marazzina; Stocco, 2023). A predição de classificação do ESG pode ser um grande contribuinte, pois oferece uma visão mais clara e rápida acerca das empresas e seu comprometimento com os princípios ESG (Lee *et al.*, 2022).

Adicionalmente, saber em que empresas os *stakeholders* estão se concentrando e como essas empresas lidam com o ESG pode ser uma maneira de analisar seus futuros rendimentos e se estão de acordo com seus ideais (De Lucia; Pazienza; Bartlett, 2020). Ademais, empresas que adotam princípios alinhados com o ESG tendem a ser mais resilientes e rentáveis a longo prazo, uma vez que estão mais preparadas para lidar com os desafios econômicos, ambientais e sociais (McKinsey & Company, 2023)

Apesar da adoção crescente de indicadores ESG pelas empresas, não está claro se essas práticas possuem, de fato, uma grande influência preditiva em torno do desempenho financeiro. Diante dessa lacuna e da crescente demanda por uma mudança nas práticas empresariais, impulsionada pelos inúmeros problemas, bem como da busca por maior transparência corporativa, compreender essa relação sob uma perspectiva tecnológica pode gerar insights relevantes para investidores e gestores.

Nesse contexto, este estudo busca compreender como os indicadores ESG (ambiental, social e de governança) são capazes de contribuir para a predição do desempenho financeiro (ROA) das empresas, utilizando-os como variáveis preditoras, em diferentes modelos de aprendizado de máquina. Embora o objetivo principal não seja classificar o nível ESG em si, os modelos treinados aprendem a identificar padrões entre os scores ESG e a performance financeira, o que oferece uma nova perspectiva sobre como esses indicadores podem ser utilizados para análise estratégica. Com base nos resultados obtidos, é possível entender se existe alguma associação entre os pilares ESG e o desempenho financeiro, e se modelos de aprendizado de máquina conseguem capturar esse padrão de forma eficaz.

2 REVISÃO DA LITERATURA

2.1 ENVIRONMENTAL, SOCIAL AND GOVERNANCE

A preocupação com o meio ambiente existe há muito tempo, impulsionada pela degradação causada por inúmeros fatores, especialmente pelos danos resultantes da Revolução Industrial. No entanto, até o início dos anos 2000, não havia ações concretas por parte das empresas para enfrentar esses problemas. Em 2004, uma parceria entre a ONU e instituições financeiras de nove países deu origem a um marco importante: a criação da carta *Who Cares Wins*, que tinha como objetivo incentivar o mercado financeiro a adotar práticas positivas e conscientes que englobassem governança, responsabilidade social e ambiental (Pires; Silva, 2023).

O ESG orienta práticas de sustentabilidade e responsabilidade no mercado, incentivando as empresas a se tornarem mais comprometidas com a sociedade de maneira eficiente. Esses

princípios permitem avaliar e classificar empresas com base em ações que impactam o meio ambiente, a sociedade e a governança (Amaral; Willerding; Lapolli, 2023). Segundo Bergman, Deckelbaum e Karp (2020), o campo ambiental do ESG incentiva a redução das emissões de carbono e a adoção de práticas mais sustentáveis, como o uso de energia renovável, o emprego de materiais recicláveis e o incentivo ao transporte público. No campo social, muitas empresas adotam políticas de inclusão e diversidade, promovendo um ambiente mais igualitário em termos de gênero, raça e orientação sexual, além de melhorias nas condições de trabalho. Por fim, na governança destaca-se a transparência, incentivando as empresas a trabalharem com códigos de ética, a divulgarem relatórios periódicos de suas ações e a implementarem práticas que assegurem uma gestão ética e responsável.

Atualmente, o ESG se tornou um fator essencial no mercado, sendo uma das principais preocupações dos investidores. Segundo uma reportagem da revista Exame (2024), 99% dos investidores consideram as práticas de ESG ao tomar decisões de investimento, demonstrando disposição em apoiar empresas que adotam práticas sustentáveis, mesmo sem um retorno imediato de grande impacto. Entretanto, a implementação do ESG traz alguns desafios. A mesma reportagem destaca a preocupação dos investidores com a falta de transparência das empresas, especialmente em relação às suas ações ambientais, sociais e de governança. Essa falta de clareza pode resultar na prática conhecida como *greenwashing*, onde as empresas divulgam informações enganosas sobre suas iniciativas sustentáveis, criando uma imagem distorcida de suas ações (Lovato, 2013).

Apesar da crescente demanda por empresas que adotem práticas responsáveis, especialmente por parte da população, ainda há uma mentalidade predominante de se importar mais com o lucro. Para combater o *greenwashing* e promover maior transparência, a União Europeia implementou regulamentos que exigem a divulgação de informações financeiras sustentáveis, categorizadas em três níveis, garantindo maior confiança nas práticas ambientais das empresas europeias (Pires; Silva, 2023).

Diante disso, muitas empresas compartilham publicamente seus relatórios, detalhando metas e resultados alcançados. A Nike, por exemplo, reportou em seu relatório de 2023 uma redução de 69% nas emissões de gases de efeito estufa e o alcance de 96% de energia renovável globalmente em suas instalações, além de um investimento de US\$ 142,7 milhões em comunidades, com ênfase em equidade e acesso ao esporte, incluindo um compromisso de US\$ 40,8 milhões para apoiar a comunidade negra e iniciativas inclusivas (Nike, 2023).

Outro exemplo é a Unilever, que em seu relatório de 2023 foi reconhecida com a avaliação AAA- do CDP (Carbon Disclosure Project), organização global sem fins lucrativos que avalia o impacto ambiental das empresas, sendo uma alta pontuação da empresa nas avaliações ambientais. A Unilever destacou-se em ações de combate às mudanças climáticas, alcançando uma redução de 74% nas emissões de gases de efeito estufa desde 2015 (Unilever, 2023).

2.2 DESEMPENHO FINANCEIRO

Para entender o impacto de diferentes fatores no desempenho financeiro, é fundamental primeiro definir esse conceito. Em termos gerais, o desempenho financeiro é o resultado da maneira como grandes ou pequenas empresas, projetos ou até mesmo indivíduos gerenciam seus investimentos de forma eficiente, buscando maximizar seus lucros e garantir a sustentabilidade e

lucratividade econômica. O objetivo é obter resultados positivos ao longo de toda a sua atuação. Diante disso, para avaliar como uma empresa está se firmando no mercado, é fundamental mensurar e analisar seu desempenho financeiro. Essa análise é baseada em indicadores, que fornecem informações essenciais para interpretar os resultados das ações tomadas por essas instituições ao longo do período analisado, e diante do resultado tomar as próximas decisões que vão ser implementadas nesse negócio (Alexandrino, 2020).

Existem indicadores que apoiam a análise do desempenho e dos intuitos organizacionais, mas é importante compreender como eles serão interpretados, já que podem ser categorizados de maneira diferente (ex: liquidez, rentabilidade e estrutura de capital). Entre os mais relevantes estão giro do ativo (GA), retorno sobre o ativo (ROA), retorno sobre o patrimônio líquido (ROE), margem líquida (ML), EBITDA e margem EBITDA (Oliveira *et al.*, 2017).

Embora existam diversos indicadores disponíveis, este estudo foca no ROA, pois ele é essencial para avaliar a eficiência financeira da empresa e será utilizado para fundamentar análises realizadas com aprendizado de máquina no contexto do ESG. O ROA mede o retorno obtido sobre os ativos da empresa, refletindo sua capacidade de gerar lucro a partir dos recursos investidos (Oliveira *et al.*, 2017). Esse indicador é calculado conforme a equação 1.

$$ROA = \frac{\text{Lucro Operacional}}{\text{Ativo Total}} \quad (1)$$

Ao analisar o ROA, é possível obter *insights* importantes sobre a rentabilidade da empresa, auxiliando na comparação do desempenho entre diferentes períodos ou em relação a outras empresas. Nos últimos anos, fatores ambientais, sociais e de governança passaram a ter um papel cada vez mais relevante na análise de desempenho financeiro. Segundo Alexandrino (2020), o mercado vem crescendo de forma expressiva, o que impulsiona o surgimento de novas estratégias e torna o ambiente de negócios cada vez mais competitivo. Dessa forma, as empresas precisam se adaptar a esse cenário, investindo no socioambiental.

De acordo com o estudo de Khan, Serafeim e Yoon (2016), foi observado que instituições que possuem um maior interesse em investir em índices de sustentabilidade tendem a se destacar perante os acionistas, gerando uma visibilidade maior para a empresa. Além disso, o estudo sugere que a adoção de práticas sustentáveis pode impactar positivamente os resultados financeiros das empresas. Um exemplo prático disso é o relatório divulgado pela Natura, que demonstra os níveis de comprometimento da empresa com o ESG e como tais iniciativas impactaram de forma positiva a sociedade, além de impactos positivos na contribuição para a redução de riscos, o fortalecimento do relacionamento com os *stakeholders* e a melhoria na eficiência operacional (Natura e Co, 2023).

Por fim, no estudo proposto por Eccles, Ioannou e Serafeim (2012), foram analisadas diversas empresas com diferentes níveis de comprometimento com as responsabilidades ambientais e os investimentos correspondentes, ao longo de um extenso período. Embora não seja possível afirmar com certeza a existência de uma relação causal, os dados demonstram que empresas com maior engajamento com o meio ambiente tendem a apresentar um desempenho financeiro superior em comparação com aquelas com menor envolvimento em temas de sustentabilidade, especialmente a longo prazo. Esse efeito é mais evidente em setores que dependem intensamente de capital humano ou que operam com recursos naturais.

2.3 MODELOS DE APRENDIZADO DE MÁQUINA

O aprendizado de máquina, também conhecido pelo termo em inglês *machine learning*, é um dos principais ramos da inteligência artificial. O objetivo principal dessa área é treinar sistemas para que consigam tomar decisões e realizar tarefas específicas, utilizando dados para identificar padrões e poder fazer previsões. Assim, há o interesse em produzir uma máquina que consiga, de forma contínua, aprender através de informações, dados pré-estabelecidos pela equipe para, por fim, ter a capacidade de dar respostas corretas, seja de classificação, predição ou detecção (Paixão *et al.*, 2022).

De maneira geral, no aprendizado de máquina, o sistema consegue aprender baseado em dados obtidos através de banco de dados. Esses dados servem para que seja possível ensinar a máquina de acordo com o objetivo principal do modelo. Existem diversas aplicações onde pode-se diferenciar animais, realizar cálculos para previsões, construir modelos analíticos e resolver tarefas complexas (Zhou, 2021).

Por meio desses dados, fornecidos por bancos de dados, o modelo consegue consumir essas informações, formando padrões através das características dos dados recebidos e ligando com seus rótulos. Durante toda a sua utilização, o modelo vai melhorando cada vez mais com novos dados fornecidos. Esse processo lembra bastante o aprendizado de uma criança, que, como observado por Piaget (1986), na sua teoria do desenvolvimento cognitivo, propõe que, a partir da interação da criança com o ambiente, ela consegue desenvolver e expandir conhecimentos já adquiridos, através da assimilação de novos ambientes, refinando esses conhecimentos.

Adicionalmente, é importante destacar que existem diferentes formas de ensinar um modelo, também conhecidos como modelos de aprendizagem, conforme apresentado por Bishop (2006). Os três principais modelos são: aprendizado supervisionado, não supervisionado e semi-supervisionado. A respeito do aprendizado supervisionado, Goodfellow, Bengio e Courville (2015) introduziram os modelos em questão, como sendo os que aprendem através de rótulos e características. O modelo desta tipologia é treinado com dados de entrada (*inputs*), em conjunto com seus rótulos (*labels*).

Diferentemente do aprendizado supervisionado, o não supervisionado não conta com rótulos predefinidos. O modelo trabalha apenas com as características dos dados, buscando identificar padrões e agrupamentos de maneira autônoma (Fernandes, 2021). Sobre o aprendizado semi-supervisionado, Zhu (2005) propôs que seria um método que combina características do aprendizado supervisionado e não supervisionado. Nesse caso, o modelo é treinado utilizando tanto dados rotulados quanto não rotulados.

De maneira mais detalhada, um processo de aprendizagem simples, utilizando aprendizado supervisionado, se basearia em duas etapas: treinamento e teste. Para iniciar o treinamento da máquina, que é a primeira etapa, é necessário já ter um banco com os dados que serão utilizados. Caso esses dados não estejam formatados de maneira adequada, será preciso realizar um pré-processamento. Supondo que os dados já estejam prontos para uso, eles são divididos em duas partes: uma para o treinamento e outra para os testes das máquinas (Géron, 2019).

Após o treinamento, o modelo passa pela fase de teste. Nessa etapa, ele recebe novos *inputs*, previamente separados, mas sem os rótulos visíveis. O objetivo é avaliar se o modelo consegue fazer previsões corretas, baseando-se apenas no conhecimento adquirido. Em outras

palavras, verifica-se se a máquina aprendeu a classificar corretamente os dados sem precisar das respostas previamente fornecidas (St-Aubin, 2024).

A partir da fase de teste, é possível medir a eficácia do modelo, analisando se ele consegue gerar previsões confiáveis para novos dados. Uma métrica comumente utilizada para essa avaliação é a acurácia, que indica o percentual de acertos do modelo em relação aos testes realizados (NG, 2018).

3 METODOLOGIA

Este trabalho tem o intuito de compreender como os indicadores ESG são capazes de contribuir para a predição do desempenho financeiro das empresas, utilizando-os como variáveis preditoras, em diferentes modelos de aprendizado de máquina. A pesquisa foi desenvolvida em cinco etapas principais: escolha do banco de dados, separação dos dados, padronização dos dados, treinamento do modelo e, finalmente, a avaliação dos resultados. Os dados foram coletados na plataforma Refinitiv Eikon. Foi considerado o valor do ROA por empresa e por ano, abrangendo registros de 1998 a 2024, e organizações das seguintes nacionalidades: Chile, Estados Unidos, Peru, México, Brasil, Argentina, Colômbia e Reino Unido.

Para a análise dos dados, foram escolhidos sete algoritmos de classificação passíveis de implementação em aprendizado de máquina: Random Forest, K-Nearest Neighbors (KNN), Naïve Bayes, Regressão Logística (Logit), Support Vector Machine (SVM) linear, Bagging e Gradient Boosting Machine (GBM). Um dos motivos que justificam a escolha do Random Forest é sua capacidade de lidar com variáveis correlacionadas, conforme demonstrado por De Lucia, Pazienza e Bartlett (2020). Além disso, os classificadores Naïve Bayes são naturalmente capazes de lidar com múltiplas classes (Géron, 2019). Já o Support Vector Machine (SVM) é considerado um modelo de aprendizado de máquina poderoso e versátil, aplicável à classificação linear e não linear, à regressão e também à detecção de *outliers* (Géron, 2019). Quanto ao K-Nearest Neighbors (KNN), a escolha se deve à sua natureza simples e interpretável (Domingos, 2012).

Os demais algoritmos foram escolhidos por representarem diferentes abordagens de classificação, permitindo uma comparação mais completa dos resultados. A ideia central é analisar quais desses algoritmos oferecem um resultado mais satisfatório de acordo com o objetivo da pesquisa, identificando qual se comporta melhor e qual apresenta a melhor avaliação para o contexto em questão. Além disso, esses modelos foram escolhidos pela facilidade de aplicação e pela eficácia em problemas de classificação.

A avaliação dos modelos foi realizada em etapas, considerando o valor da acurácia (mensuração da taxa de previsões corretas feitas pelos modelos) e a curva ROC e o valor de AUC (mensuração da capacidade do modelo de distinguir entre as classes em todos os limiares de decisão possíveis). Por fim, foram apresentados dados mais detalhados sobre o processo de treinamento, incluindo gráficos e métricas obtidas durante as fases de treinamento e avaliação. A divisão do conjunto de dados foi realizada de forma aleatória, alocando 70% para o treinamento do modelo e os 30% restantes para a etapa de teste.

Em uma primeira amostra de dados, havia cerca de 21441 observações de empresas com suas respectivas variáveis. Todavia, nem todos os dados estavam completos, sendo necessária uma padronização e limpeza dessas informações para eliminar dados faltantes, o que resultou em 1968 dados para serem utilizados. O processo de winsorização foi aplicado ao nível de 1% inferior e 99% superior para minimizar a influência de valores extremos nas análises.

O estudo se concentrou nas variáveis que seriam utilizadas: ROA, *Social Score*, *Environmental Score* e *Governance Score*. Os dados foram padronizados, transformados em painel, e o ROA foi convertido em variável binária (0 e 1) através de clusterização, uma técnica de aprendizado não supervisionado que agrupa dados em clusters naturais (Bishop, 2006), utilizando o algoritmo k-means com k=2 centros.

A clusterização foi aplicada diretamente à variável ROA após o processo de winsorização. Depois da formação dos dois clusters, foi realizada uma reclassificação na qual o cluster que tinha o valor 2 foi convertido para o valor 0, enquanto o cluster 1 permaneceu com o mesmo valor, resultando em 1299 observações com valores 0 (menor retorno) e 669 com classificação 1 (maior retorno). Além disso, foram criadas variáveis defasadas (t-1) dos indicadores ESG para utilização nos modelos.

4 ANÁLISE DOS RESULTADOS

Após o tratamento dos dados, foram analisados os resultados da estatística descritiva das variáveis ESG e ROA com dados contínuos, conforme apresentado na Tabela 1. O ROA em sua forma primária apresentou média de 7,74%, com desvio-padrão de 8,5%, variando de -25,08% até 38,25%. Esta característica mostra uma considerável variação no desempenho financeiro das empresas analisadas.

Para os indicadores ESG, o pilar Ambiental (ENV) registrou média de 40,22 pontos, com valores entre 0,00 e 92,07, mostrando uma grande variação nas práticas ambientais adotadas. O pilar de Governança (GOV) apresentou a maior média (49,71) entre os três componentes, com pontuações entre 4,31 e 93,08. Já o pilar Social (SOC) mostrou média de 47,53, com valores entre 0,60 e 93,01. As medianas próximas às médias em todos os indicadores sugerem distribuições relativamente equilibradas, embora o pilar Social apresenta mediana ligeiramente superior à média (52,08 e 47,53), indicando a possibilidade de alguns valores mais baixos influenciarem a média negativamente.

Tabela 1 – Estatística descritiva do ROA e dos escores ESG

Estatística	ROA	ENV	GOV	SOC
Mínimo	-0,25082	0,00	4,31	0,60
1.º Quartil	0,03922	16,67	31,24	25,79
Mediana	0,07310	41,55	49,95	52,08
Média	0,07746	40,22	49,71	47,53
3.º Quartil	0,11559	61,77	68,48	69,08
Máximo	0,38253	92,07	93,08	93,01
Desvio padrão	0,08585	26,42	23,15	26,21

Fonte: dados da pesquisa.

Posteriormente à aplicação da clusterização, os dados foram reorganizados com base nos dois grupos formados a partir do ROA depois da função de winsorização. O algoritmo k-means com dois centros ($k=2$), sobre os valores winsorizados do ROA, segmentou os dados em duas classes. O cluster 0, associado ao menor retorno sobre ativos, abrangeu 1.299 observações (66,01 %), enquanto o cluster 1, relacionado ao maior retorno, agrupou 669 observações (33,99 %), como é possível observar na Tabela 2.

Tabela 2 – Distribuição dos clusters obtidos via k-means sobre o ROA

Classe (Cluster)	Nº de Observações	Percentual (%)	Interpretação
0	1.299	66,01 %	Baixo retorno sobre ativos
1	669	33,99 %	Alto retorno sobre ativos

Fonte: dados da pesquisa.

A análise comparativa dos sete modelos de aprendizado de máquina aplicados à classificação de ROA com base em dados ESG revela limitações significativas na capacidade preditiva destes modelos para o contexto estudado. Os resultados estão presentes nas Tabelas 3 e 4. A sensibilidade indica o quanto o modelo acerta os casos positivos, enquanto especificidade mostra sua capacidade de identificar corretamente os casos negativos, conforme Géron (2019).

Tabela 3 – Aplicação dos modelos de aprendizado de máquina (dados de treino)

Modelo	Acurácia	Kappa	Sensibilidade	Especificidade	Balanced Accuracy
Random Forest	1	1	1	1	1
KNN	0,6761	0,1035	0,12179	0,96150	0,54165
Naïve Bayes	0,6609	0,0371	0,05769	0,97140	0,51454
Logistic Regression	0,6601	0	0	1	0,5000
SVM Linear	0,6601	0	0	1	0,5000
Bagging	0,9949	0,9886	0,9872	0,9989	0,9930
Gradient Boosting	0,6972	0,1655	0,16026	0,97360	0,56693

Fonte: dados da pesquisa.

Tabela 4 – Aplicação dos modelos de aprendizado de máquina (dados de teste)

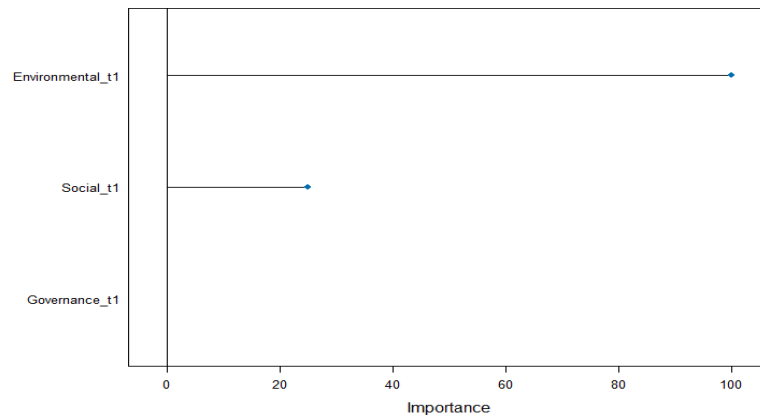
Modelo	Acurácia	Kappa	Sensibilidade	Especificidade	Balanced Accuracy
Random Forest	0,6447	0,1239	0,27861	0,8333	0,55597
KNN	0,6413	0,0099	0,07463	0,93333	0,50398
Naïve Bayes	0,6684	0,065	0,07960	0,97179	0,52570
Logistic Regression	0,6599	0	0	1,0000	0,5000
SVM Linear	0,6599	0	0	1,0000	0,5000
Bagging	0,6176	0,072	0,26866	0,79744	0,53305
Gradient Boosting	0,6751	0,1086	0,13433	0,95385	0,54409

Fonte: dados da pesquisa.

O Random Forest apresentou *overfitting*, com desempenho perfeito no treino (acurácia de 100%), mas queda drástica no teste (64,47%). Similarmente, o Bagging demonstrou comportamento parecido, com acurácia de 99,49% no treino reduzida para 61,76% no teste. Os modelos KNN, Naive Bayes e GBM, embora tenham apresentado menor discrepância entre treino e teste, evidenciaram baixa sensibilidade (7 - 28%) para a classe positiva, privilegiando a classe majoritária negativa. O modelo de Regressão Logística mostrou-se completamente generalista, classificando todos os exemplos como negativos, resultando em sensibilidade nula.

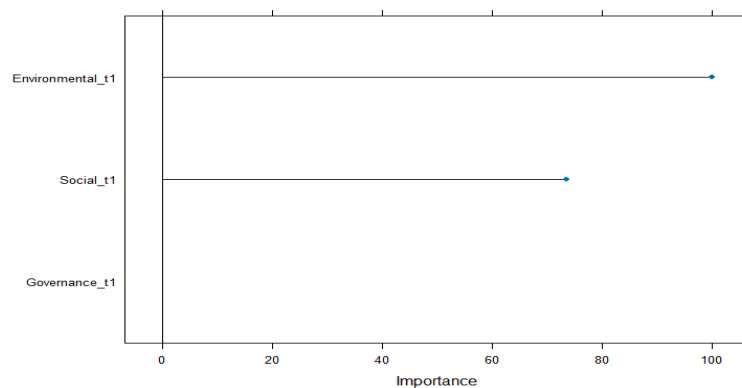
Um padrão consistente entre a maioria dos modelos foi a alta taxa da variável ambiental, com exceção do modelo bagging (Figura 6), como principal elemento na relevância do impacto no ROA, seguida pela variável social, enquanto a governança apresentou menor relevância. Esses resultados podem ser visualizados por meio das Figuras 1, 2, 3, 4, 5 e 7.

Figura 1 - Impacto das variáveis ESG no ROA - Modelo Random Forest



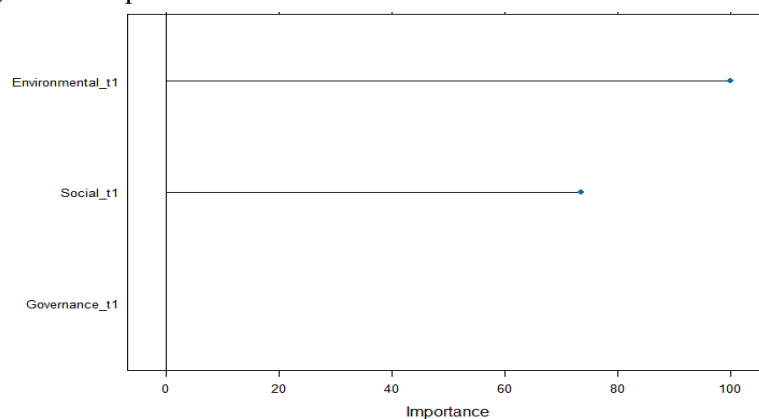
Fonte: dados da pesquisa.

Figura 2 - Impacto das variáveis ESG no ROA - Modelo KNN



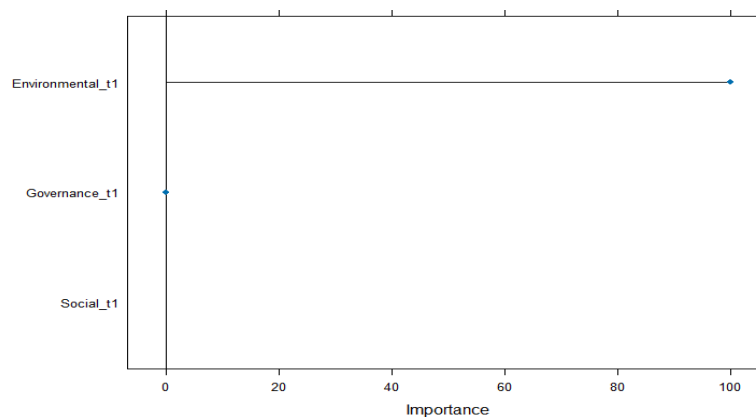
Fonte: dados da pesquisa.

Figura 3 - Impacto das variáveis ESG no ROA - Modelo Naive Bayes



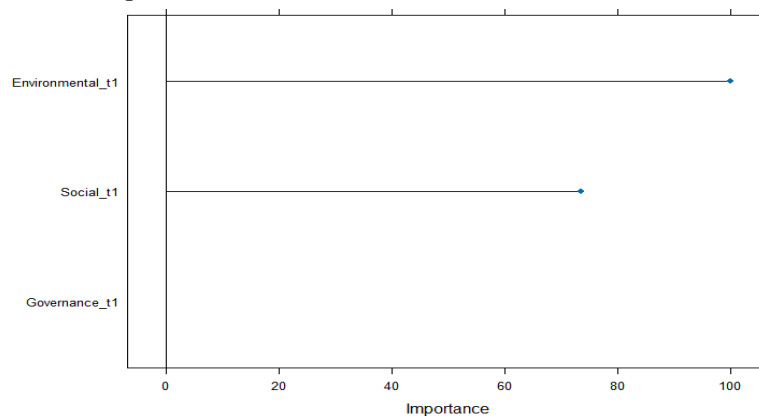
Fonte: dados da pesquisa.

Figura 4 - Impacto das variáveis ESG no ROA - Modelo Logit



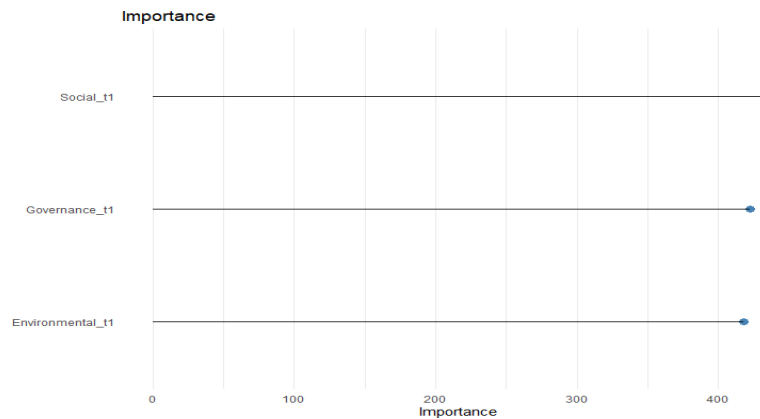
Fonte: dados da pesquisa.

Figura 5 - Impacto das variáveis ESG no ROA - Modelo SVM Linear



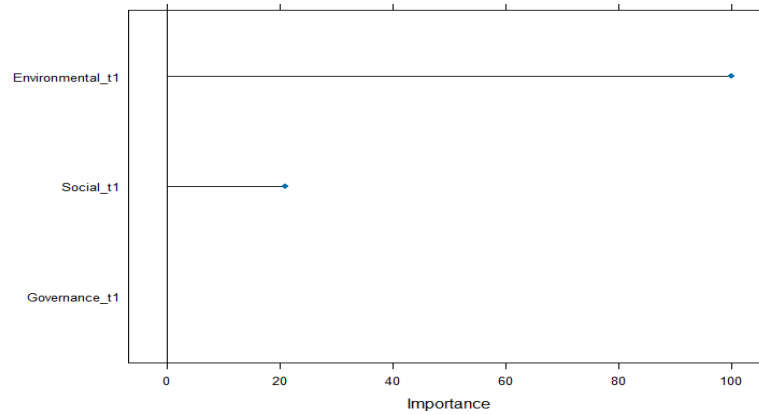
Fonte: dados da pesquisa.

Figura 6 - Impacto das variáveis ESG no ROA - Modelo Bagging



Fonte: dados da pesquisa.

Figura 7 - Impacto das variáveis ESG no ROA - Modelo Gradient Boosting Machine



Fonte: dados da pesquisa.

No estudo de Velte (2017), que analisou empresas alemãs, constatou-se uma relação positiva e significativa entre o desempenho do ESG e o ROA, mas sem impacto significativo sobre o Q de Tobin. Além disso, a governança corporativa foi identificada como o componente mais influente no desempenho financeiro, diferindo dos resultados da análise feita com aprendizado de máquina anteriormente, que apontaram o fator ambiental como o de maior significância. Em relação aos valores de AUC, eles foram próximos ao intervalo de 0,50 a 0,60 nos testes, como podem ser visualizados na Tabela 6, pouco superiores a classificadores aleatórios (0,50), mostrando a dificuldade de generalização para novos dados.

Todavia, dentre os modelos testados, o Gradient Boosting se destacou com a maior acurácia no conjunto de teste (67,5%), conforme Tabela 5. Isso significa que este modelo foi capaz de classificar corretamente 67,5% das observações entre os dois grupos de desempenho do ROA. Embora este resultado represente uma melhoria em relação aos outros modelos avaliados, ele ainda é considerado insuficiente para aplicações práticas que pedem uma maior confiabilidade na classificação.

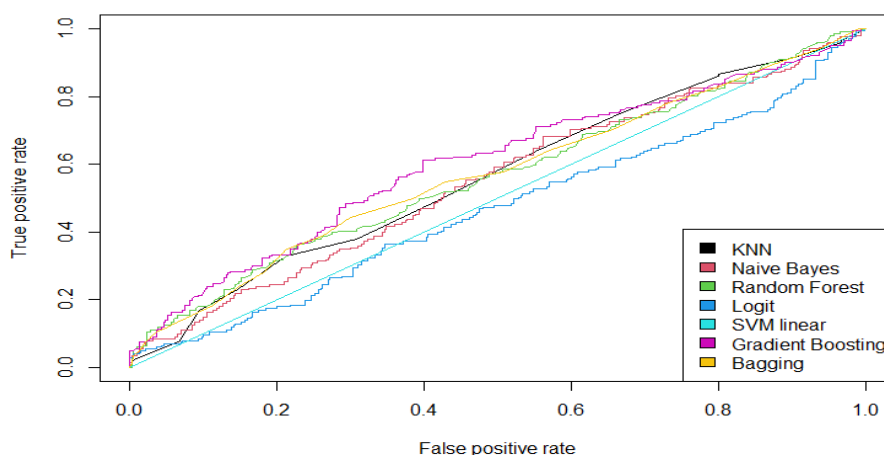
Tabela 6 – Aplicação dos modelos de aprendizado de máquina

Modelo	AUC (Treino)	AUC (Teste)	Δ (Treino – Teste)
Random Forest	1,0000	0,5653	0,4347
KNN	0,6453	0,5641	0,0812
Naive Bayes	0,6074	0,5508	0,0566
Logit	0,5271	0,5336	-0,0065
SVM Linear	0,5000	0,5000	0,0000
Bagging	1,0000	0,5603	0,4397
Gradient Boosting Machine	0,7077	0,6023	0,1054

Fonte: dados da pesquisa.

Na Figura 8, que apresenta a curva ROC, podemos observar que o modelo Gradient Boosting gera uma curva que se distancia mais da linha diagonal, que representa classificação aleatória, e se aproxima mais do canto superior esquerdo do gráfico, que seria a melhor classificação que um modelo poderia ter.

Figura 8 - Curvas ROC comparativas dos modelos de predição do ROA
ROC curves for test data



Fonte: dados da pesquisa.

Assim, estes resultados sugerem que, apesar da relevância teórica das variáveis ESG para o desempenho financeiro, os modelos testados não conseguiram capturar a completude para previsões confiáveis do ROA. Dentre eles, o Gradient Boosting se destacou como o modelo mais adequado, alcançando a maior acurácia no conjunto de teste (67,51%) e melhor balanceamento entre sensibilidade e especificidade, além de uma curva ROC mais próxima de um modelo adequado. Uma das maneiras de tentar melhorar esse desempenho é a utilização de outras variáveis e explorar outros modelos que possam eventualmente se adequar mais a esses dados.

De Lucia, Pazienza e Bartlett (2020) empregaram modelos de aprendizado de máquina — especificamente algoritmos de regressão baseados em árvores de decisão, como Random Forest e Gradient Boosting, além de regressão logística para análise inferencial — para investigar o impacto dos componentes ESG no desempenho financeiro de empresas europeias (medido, por exemplo, por ROA). Os modelos baseados em árvores demonstraram um elevado poder preditivo, superando abordagens tradicionais, o que evidencia que a incorporação de variáveis ESG melhora significativamente a previsão dos indicadores financeiros. Ademais, o estudo destaca que a

integração de práticas ESG pode aprimorar a tomada de decisões de investimento, promovendo maior rentabilidade e redução de riscos, sendo a governança corporativa identificada como a dimensão de ESG com maior influência nos lucros das empresas.

Em contraste com a pesquisa em desenvolvimento, o estudo de De Lucia, Pazienza e Bartlett (2020) utilizou diversas variáveis financeiras para complementar os resultados, destacando a governança como o componente ESG mais impactante. Já no estudo atual, a dimensão ambiental apresentou maior relevância para o desempenho financeiro. No entanto, tanto no presente trabalho quanto no referido artigo, o algoritmo Gradient Boosting se destaca em relação aos demais modelos empregados.

Por fim, as diferenças nos resultados entre os estudos podem ser explicadas pelas distintas variáveis utilizadas, pelos contextos regionais e setoriais analisados, e pelas metodologias aplicadas. Essas discrepâncias reforçam a importância de considerar fatores contextuais ao empregar modelos de aprendizado de máquina na previsão de pontuações ESG e seu impacto sobre o desempenho financeiro das empresas.

5 CONSIDERAÇÕES FINAIS

Este estudo apresenta uma contribuição acadêmica ao investigar, por meio de técnicas de aprendizado de máquina, a influência dos critérios ESG no desempenho financeiro das empresas. Ao explorar a aplicação de modelos de *machine learning* nessas três dimensões, o trabalho preenche uma lacuna na literatura, que ainda conta com poucas pesquisas integrando ESG, performance financeira e aprendizado de máquina. Ademais, demonstra a viabilidade de análises de dados voltadas ao retorno financeiro, evidenciando que, com conjuntos de dados mais amplos e variados, será possível quantificar de forma precisa como as práticas ESG impactam os resultados corporativos. Para o mercado, esses achados sugerem que iniciativas ambientais não devem ser encaradas apenas como custos, mas sim como investimentos capazes de gerar retorno. Assim, o estudo pode incentivar as empresas a adotarem políticas alinhadas aos princípios ESG e reforçar a importância da coleta e divulgação transparentes de informações socioambientais.

Ao investigar a relação entre os princípios ESG e o desempenho financeiro das empresas, medido pelo ROA, este estudo utilizou diversos modelos de aprendizado de máquina para analisar se os indicadores ESG podem efetivamente contribuir para a previsão do desempenho financeiro corporativo.

Este achado sugere que investimentos em práticas ambientalmente responsáveis podem estar mais diretamente relacionados ao retorno financeiro das empresas. Ademais, os resultados mostraram que, embora alguns modelos apresentem alta acurácia durante o treinamento, sua performance no conjunto de teste indica problemas de *overfitting*, especialmente quando se considera a baixa sensibilidade para detectar a classe positiva. O fator ambiental mostrou maior relevância na explicação do ROA, enquanto os indicadores sociais e de governança tiveram menor impacto.

Em comparação com estudos anteriores, como os de Velte (2017) e De Lucia, Pazienza e Bartlett (2020), este trabalho identificou o fator ambiental como o mais influente no contexto analisado, divergindo de outras pesquisas que apontaram a governança como o elemento principal. Isso evidencia como fatores regionais, setoriais e temporais podem alterar a dinâmica da relação ESG e desempenho financeiro.

Apesar do potencial dos indicadores ESG, sua aplicação isolada na predição do desempenho financeiro é limitada. Ressalta-se a importância da integração de variáveis financeiras adicionais e da aplicação de técnicas adicionais, já que poucas variáveis e modelos de aprendizado de máquina foram utilizados neste estudo. Assim, para pesquisas futuras, recomenda-se a inclusão de variáveis financeiras adicionais, análise setorial mais detalhada, utilização de séries temporais mais longas, e exploração de técnicas avançadas de *feature engineering* e balanceamento de classes. Adicionalmente, seria valioso investigar outras métricas de desempenho financeiro além do ROA, como o Q de Tobin, EVA (*Economic Value Added*), entre outras.

REFERÊNCIAS

ALEXANDRINO, Thaynan Cavalcanti. *Fatores determinantes para o desempenho financeiro e contábil: um estudo com base em dados ESG e o uso de técnicas de machine learning*. 2023. 180 f. Dissertação (Mestrado em Ciências Contábeis) – Universidade Federal de Pernambuco, Recife, 2023. Disponível em:

<https://repositorio.ufpe.br/bitstream/123456789/38600/1/DISSERTA%C3%87%C3%83O%20Thaynan%20Cavalcanti%20Alexandrino.pdf>. Acesso em: 20 abr. 2025.

AMARAL, Melissa Ribeiro do; WILLERDING, Inara Antunes Vieira; LAPOLLI, Édis Mafra. Sustentabilidade organizacional: a sinergia entre ESG e organizações saudáveis. In: CONGRESSO INTERNACIONAL DE CONHECIMENTO E INOVAÇÃO, 1., 2023, Florianópolis. *Anais [...]*. Florianópolis: CIKI, 2023. Disponível em: <https://proceeding.ciki.ufsc.br/index.php/ciki/article/download/1428/848/5956>. Acesso em: 18 out. 2024.

BERGMAN, Mark S.; DECKELBAUM, Ariel J.; KARP, Brad S. *Introduction to ESG*. New York: Paul, Weiss, Rifkind, Wharton & Garrison LLP, 2020. Disponível em: <https://www.paulweiss.com/media/3980354/9july20-intro-to-esg.pdf>. Acesso em: 20 abr. 2025.

BISHOP, Christopher M. *Pattern recognition and machine learning*. New York: Springer, 2006. Disponível em: <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>. Acesso em: 15 nov. 2024.

CNN BRASIL. Humanidade esgotou recursos da Terra disponíveis para 2024. *CNN Brasil*, 1 ago. 2024. Disponível em: <https://www.cnnbrasil.com.br/internacional/humanidade-esgotou-recursos-da-terra-disponiveis-para-2024/>. Acesso em: 18 out. 2024.

DE LUCIA, Caterina; PAZIENZA, Pasquale; BARTLETT, Mark. Does good ESG lead to better financial performances by firms? Machine learning and logistic regression models of public enterprises in Europe. *Sustainability*, Basel, v. 12, n. 13, p. 5317, 2020. DOI: 10.3390/su12135317. Disponível em: <https://www.mdpi.com/2071-1050/12/13/5317>. Acesso em: 20 abr. 2025.

DEL VITTO, Alessandro; MARAZZINA, Daniele; STOCCO, Davide. ESG ratings explainability through machine learning techniques. *Annals of Operations Research*, v. 65, p. 1-25, 2023. Disponível em: https://www.researchgate.net/publication/372466858_ESG_ratings_explainability_through_machine_learning_techniques. Acesso em: 15 nov. 2024.

DOMINGOS, Pedro. A few useful things to know about machine learning. *Communications of the ACM*, New York, v. 55, n. 10, p. 78–87, 2012.

ECCLES, Robert G.; IOANNOU, Ioannis; SERAFEIM, George. The Impact of Corporate Sustainability on Organizational Processes and Performance. *Management Science*, v. 60, n. 11, p. 2835–2857, 2014. DOI: 10.1287/mnsc.2014.1984. Disponível em: https://www.nber.org/system/files/working_papers/w17950/revisions/w17950.rev0.pdf. Acesso em: 20 abr. 2025.

EXAME. Brasil começa 2024 otimista, mas violência e desigualdade seguem preocupando muito. *Exame*, São Paulo, 4 jan. 2024. Disponível em: <https://exame.com/brasil/brasil-comeca-2024-otimista-mas-violencia-e-desigualdade-seguem-preocupando-muito/>. Acesso em: 20 out. 2024.

EXAME. ESG direciona decisão de 99% dos investidores no Brasil, diz pesquisa. *Exame*, São Paulo, 2024. Disponível em: <https://exame.com/bussola/esg-direciona-decisao-de-99-dos-investidores-no-brasil-diz-pesquisa/>. Acesso em: 12 nov. 2024.

FERNANDES, Fernando. *A study on deep learning and its applications in modern computing*. 2022. Tese (Doutorado em Ciência da Computação) – Universidade de São Paulo, São Paulo, 2022. Disponível em: https://www.teses.usp.br/teses/disponiveis/6/6143/tde-27012022-140548/publico/FernandesFT_DR_R.pdf. Acesso em: 13 nov. 2024.

GÉRON, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*. 2. ed. Sebastopol: O'Reilly Media, 2019. E-book.

GIESE, G.; LEE, L.-E.; MELAS, D.; NAGY, Z.; NISHIKAWA, L. Foundations of ESG investing: how ESG affects equity valuation, risk, and performance. *The Journal of Portfolio Management*, Wellesley, v. 45, n. 5, p. 69–83, jul. 2019. DOI: 10.3905/jpm.2019.45.5.069.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep learning*. Cambridge: MIT Press, 2016.

KHAN, Mozaffar; SERAFEIM, George; YOON, Aaron S. Corporate sustainability: first evidence on materiality. *NBER Working Paper*, n. 17950, 2016. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2575912. Acesso em: 20 abr. 2025.

LEE, Ook; JOO, Hanseon; CHOI, Hayoung; CHEON, Minjong. Proposing an integrated approach to analyzing ESG data via machine learning and deep learning algorithms. *Sustainability*, Basel, v. 14, n. 14, p. 8745, 2022. DOI: 10.3390/su14148745. Acesso em: 15 nov. 2024.

LOVATO, M. L. Greenwashing no Brasil: quando a sustentabilidade ambiental se resume a um rótulo. *Revista Eletrônica do Curso de Direito da UFSM*, Santa Maria, v. 8, n. 1, p. 162–171, 2013. Disponível em: <https://periodicos.ufsm.br/revistadireito/article/view/8257/4978>. Acesso em: 15 nov. 2024.

MCKINSEY & COMPANY. *2023 ESG report*. New York, 2023. Disponível em: <https://www.mckinsey.com/about-us/social-responsibility/2023-esg-report/2023-esg-report-mck.pdf>. Acesso em: 18 out. 2024.

NATURA & CO. *Relatório integrado Natura & Co 2023*. São Paulo, 2023. Disponível em: https://images.rede.natura.net/html/relatorio-anual/2023/Relatorio-Integrado-Natura-e-Co-2023_PT.pdf. Acesso em: 20 abr. 2025.

NG, Andrew. *Machine learning yearning*, 2018. Disponível em: https://nessie.ilab.sztaki.hu/~kornai/2020/AdvancedMachineLearning/Ng_MachineLearningYearning.pdf. Acesso em: 11 nov. 2024.

NIKE, Inc. *FY23 Nike, Inc. impact report*. Beaverton, 2023. Disponível em: <https://about.nike.com/en/newsroom/releases/fy23-nike-inc-impact-report>. Acesso em: 9 nov. 2024.

OLIVEIRA, J. F. da R.; VIANA JUNIOR, D. B. C.; PONTE, V. M. R.; DOMINGOS, S. R. M. Indicadores de desempenho e valor de mercado: uma análise nas empresas listadas na BM&FBOVESPA. *Revista Ambiente Contábil*, Natal, v. 9, n. 2, p. 240–258, 2017. DOI: 10.21680/2176-9036.2017v9n2ID10787. Disponível em: <https://periodicos.ufrn.br/ambiente/article/view/10787>. Acesso em: 10 maio 2025.

OXFAM BRASIL. *Desigualdade: S.A.* São Paulo, 2024. Disponível em: <https://www.oxfam.org.br/forum-economico-de-davos/desigualdade-s-a/>. Acesso em: 18 out. 2024.

OU, Ling; ZHANG, Zhibin; LI, Renyu; CHEN, Zhuo. Economic policy uncertainty and business performance: the moderating role of service transformation. *International Review of Economics & Finance*, v. 88, p. 531–546, nov. 2023. Elsevier BV. <http://dx.doi.org/10.1016/j.iref.2023.06.038>.

PAIXÃO, Gabriela Miana de Mattos; SANTOS, Bruno Campos; ARAUJO, Rodrigo Martins de; RIBEIRO, Manoel Horta; MORAES, Jermana Lopes de; RIBEIRO, Antonio L. Machine Learning na Medicina: Revisão e Aplicabilidade. *Arquivos Brasileiros de Cardiologia*, São Paulo, v. 118, n. 1, p. 95–102, jan. 2022.

PIAGET, Jean. *O nascimento da inteligência na criança*. Tradução: Maria Luísa Lima. 10. ed. Lisboa: Publicações Dom Quixote, 1986. 387 p. (Coleção Plural, n. 10).

PIRES, Bianca Mendes; SILVA, Marcos Fernandes Gonçalves da. *Uma avaliação crítica de ESG: conceito, evolução e prática*. São Paulo: Fundação Getúlio Vargas, 2023. Disponível em: <https://periodicos.fgv.br/ric/article/view/86133/84908>. Acesso em: 9 nov. 2024.

ST-AUBIN, Alexandre. *Machine learning: a primer*. Montreal: McGill University, 2024. Disponível em: <https://www.math.mcgill.ca/gsams/drp/papers/papers2024/Alexandre-St-Aubin.pdf>. Acesso em: 11 nov. 2024.

UNILEVER. *Climate action*. London, 2025. Disponível em: <https://www.unilever.com/sustainability/climate/>. Acesso em: 10 maio 2025.

VELTE, P. Environmental, social and governance (ESG) performance and earnings management: Empirical evidence from Germany. *Journal of Global Responsibility*, v. 8, n. 2, p. 169–178, 2017.

ZHOU, Zhi-Hua. *Machine Learning*. Cham: Springer, 2021. DOI: 10.1007/978-1-4471-2766-2.

ZHU, Jerry. *Semi-supervised learning*. Madison: University of Wisconsin-Madison, 2020.

SOBRE OS AUTORES

Maria Clara de Oliveira Ramos

Graduada em Análise e Desenvolvimento de Sistemas (UEPB) e graduanda em Letras – Inglês (UEPB). Atua na interseção entre tecnologia e comunicação, com experiência em atendimento ao cliente, suporte técnico e desenvolvimento web. Busca aplicar a base em programação e fluência linguística em projetos interdisciplinares, unindo tecnologia, linguagem e resolução de problemas.

Lattes: <http://lattes.cnpq.br/9242302425679869>

Contato: m.claraoliveiraramos@gmail.com

Allisson Silva dos Santos

Doutorando e Mestre em Administração pelo Programa de Pós-Graduação em Administração - PPGA da Universidade Federal da Paraíba - UFPB. Especialista em Metodologias Ativas pela Universidade Federal do Vale do São Francisco - UNIVASF. Bacharel em Administração pelo Instituto Federal de Educação, Ciência e Tecnologia da Paraíba - IFPB. Graduado em Sistemas para Internet pela Universidade Cesumar.

Lattes: <http://lattes.cnpq.br/8553703836893305>

Contato: allissonst@hotmail.com

Como citar este artigo

RAMOS, M. C. D. O; DOS SANTOS, A.S; Desempenho Financeiro e Esg: Uma Abordagem Comparativa Por Meio De Aprendizado De Máquina. **Revista de Administração da Universidade Estadual de Goiás (RAUEG)**. Anápolis, v.17 , n. Fluxo Cont., p. 1-16, agos 2026. Disponível em: link do artigo. Acesso em: dd mês ano (Ex.: 10 set. 2025).

Recebido em: 15/11/2025

Aprovado em: 20/08/2026

Sistema de Avaliação: Double Blind Review

Editor Científico: Dr. Marco Aurélio Pedrosa de Melo