

PREVISÃO DE RATINGS ESG COM INDICADORES CONTÁBEIS E MACHINE LEARNING

RESUMO

Este estudo investigou a previsão de *ratings* ambientais, sociais e de governança (*environmental, social, and governance* - ESG) de empresas não-financeiras brasileiras de capital aberto, utilizando algoritmos de Machine Learning (ML). Foram empregados indicadores contábeis e o beta de mercado como variáveis preditoras e uma amostra de 104 empresas durante os anos de 2010 a 2023. Os algoritmos Random Forest, K-Nearest Neighbors, Support Vector Machine e One-vs-Rest foram utilizados para essa finalidade. Os resultados revelaram que as variáveis contábeis e o beta condicional exercem um papel importante na previsão de *ratings* ESG. Além disso, observou-se que a inclusão da métrica ESG e seus pilares individuais defasados elevaram substancialmente a capacidade preditiva dos modelos. Esta pesquisa contribui com novas evidências empíricas apontando para a utilidade de indicadores contábeis na previsão da classificação ESG com técnicas de ML.

Palavras-chave: ESG; Previsão; Indicadores Contábeis; Beta; Machine Learning

1 INTRODUÇÃO

A deterioração ambiental e a piora dos efeitos do aquecimento global têm trazido a demanda por mais políticas para redução das emissões de carbono e danos ambientais, como também, para minimizar seus reflexos negativos e isso ocorre não apenas no contexto governamental, mas também no nível das firmas (Taskin, Cagli, & Mandaci, 2021; Taskin, Sariyer, Acar, & Cagli, 2025). Conforme Taskin et al. (2021), no mundo de hoje um problema crítico são as mudanças climáticas.

Ainda, as dimensões de responsabilidade social, como o respeito aos direitos humanos e dos trabalhadores, e de governança, como a proteção dos direitos dos acionistas e transparência, também exigem esforços corporativos. O acompanhamento da prática ambiental, social e de governança (*environmental, social, and governance* - ESG) das firmas têm ganhado expressivo destaque, estando ligado ao interesse crescente dos *stakeholders* pela responsabilização e sustentabilidade (Saha; Khan, 2024).

Aliado a isso, a preocupação com o tema ESG tem sido impulsionada com a disseminação dos 17 Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU) e com o advento do Acordo de Paris, durante a 21ª Conferência das Partes (COP21) da Convenção-Quadro das Nações Unidas sobre Mudança do Clima, em 2015, da qual o Brasil é signatário e que objetiva manter o aumento da temperatura média global menor do que 2°C acima dos níveis pré-industriais e realizar esforços para limitar o aumento da temperatura a 1,5°C acima dos níveis pré-industriais. Neste último, cada nação define sua Contribuição Nacionalmente Determinada (*Nationally Determined Contribution* - NDC) e atualmente a brasileira é reduzir as emissões líquidas de gases de efeito estufa (GEE) entre 59% e 67% até 2035 (Brasil, 2024). Nessa perspectiva, por exemplo, empresas têm declarado suas intenções de neutralizar suas emissões de GEE.

Estudos acadêmicos como Aydoğuş, Gulay e Ergun (2022) têm positivamente relacionado às atividades ESG das firmas com o desempenho financeiro. Na meta-análise de Whelan et al. (2021), tem-se que dos mais de mil artigos analisados considerando o período de 2015-2020 e o foco em medidas como *return on equity* (ROE), *return on assets* (ROA) ou preço de ação, 58% encontraram uma relação positiva, 13% relação neutra, 21% resultados mistos e somente 8% o negativo. Há também relação da métrica ESG com outras variadas características das firmas, tais como melhorias no valor de mercado (Ahmad, Mobarek, & Roni, 2021), redução no custo de capital (Wong, Nguyen, & Kwansa, 2024) e qualidade de inovação (Li & Li, 2024). De maneira geral, tem ocorrido o reconhecimento da classificação ESG como indicador fundamental de desempenho não-financeiro (Saha & Khan, 2024) e surge a necessidade de explorar a sua previsibilidade.

Com isso, o estudo aqui realizado tem o foco na previsão de *ratings* ESG a partir da utilização de técnicas de *Machine Learning* (ML), que se mostram especialmente vantajosas frente aos métodos estatísticos tradicionais conforme aponta Bzdok et al. (2018). Segundo os autores, tais técnicas fazem poucas suposições de distribuições nos dados e podem ser eficazes mesmo quando os dados são coletados sem um delineamento experimental rigoroso e na presença de relações e interações não lineares complexas.

Para essa investigação foram utilizados indicadores das demonstrações contábeis e o beta de mercado, uma área de pesquisa que tem surgido para compreender e prever

as classificações de ESG. Estudos como D'Amato, D'Ecclesia e Levantesi (2021), D'Amato, D'Ecclesia, & Levantesi (2022) e, mais recentemente, Cini e Ferrari (2025) apontam para a aplicabilidade de dados advindos das demonstrações financeiras na previsão de *scores* ESG usando ML. Esses autores fornecem evidências em um cenário internacional com firmas constituintes do STOXX Europe 600, enquanto que evidências no contexto brasileiro ainda são escassas.

Baseando-se em Cini e Ferrari (2025), foi adotada a classificação ESG como variável de interesse na previsão, obtida a partir do *score* ESG numérico, e foram implementados os algoritmos de classificação Random Forest (RF), K-Nearest-Neighbors (KNN), Support Vector Machine (SVM) e One-vs-Rest (OVR). A amostra foi composta por firmas não-financeiras brasileiras de capital aberto listadas na Brasil, Bolsa, Balcão (B3) durante o período de 2010 a 2023. De forma geral, identificou-se que indicadores contábeis e o beta condicional se mostraram significativos na previsão de *ratings* ESG de firmas brasileiras.

Essa pesquisa traz contribuições em, pelo menos, três direções. Primeiro, avança na literatura que aponta para a utilidade de variáveis contábeis na previsão de *ratings* ESG usando ML. De forma geral, estudos empíricos nessa linha fazem uso de dados de empresas em países desenvolvidos enquanto que aqui é analisado o mercado de capitais de um país em desenvolvimento, como é o caso brasileiro. Segundo, fornece evidências a partir de informações que são facilmente acessadas como critério mais objetivo para a previsão de práticas ESG. E, em terceiro, contribui com investidores, fornecedores, clientes e reguladores no acompanhamento da evolução do ESG dessas empresas.

Além dessa introdução, o artigo está estruturado da seguinte forma: a Seção 2 apresenta uma revisão da literatura com os estudos relacionados; a Seção 3 detalha os procedimentos metodológicos adotados; a Seção 4 traz a análise e discussão dos resultados da pesquisa; finalmente, a Seção 5 conclui.

2 REVISÃO DA LITERATURA

Conforme salienta Cini e Ferrari (2025), o termo ESG busca corresponder três pilares fundamentais de finanças éticas: o ambiental (*environmental* - E), o qual se preocupa com mudanças climáticas e poluição; o social (S), que considera aspectos

como política de gênero e segurança no ambiente de trabalho; e a governança (*governance* - G), o qual está relacionado com independência do conselho e conformidade com as leis. Com essa relevância para variados *stakeholders*, diferentes pesquisas têm se debruçado sobre sua previsibilidade.

Especificamente, García et al. (2020) usaram variáveis contábeis e do mercado de ações para previsão do grupo ESG ao qual a firma pertence, utilizando a teoria aproximada dos conjuntos (RSA – *Rough Set Analysis*). Os autores utilizaram uma amostra de 1.688 observações de empresas europeias de capital aberto e os dados foram coletados da Refinitiv. Seus resultados apontaram para a utilidade das medidas nessa previsão quando as firmas foram segregadas em três ou quatro grupos igualmente balanceados. Krappel et al. (2021) identificaram que com base em dados das demonstrações contábeis e outras informações gerais (por exemplo, setor de atuação e localização geográfica) e usando um conjunto de técnicas com base em Catboost, XGBoost e rede neural tipo *feedforward* foi possível explicar 54% da variância das pontuações de ESG da Refinitiv. Assim, os autores destacaram que com base em informações publicamente disponíveis é possível obter classificações ESG automatizadas.

D’Amato, D’Ecclesia e Levantesi (2021) focaram no uso de indicadores baseados em dados do balanço patrimonial e da demonstração do resultado do exercício para previsão de *scores* ESG da Bloomberg e, com isso, analisaram o desempenho de indicadores contábeis para essa finalidade, tornando-se uma alternativa prática em relação às agências tradicionais de *rating*. As autoras investigaram uma amostra de 109 empresas que fizeram parte do índice STOXX Europe 600 durante o período de 2014 a 2018 e aplicaram o algoritmo RF, além de análises preditivas comparativas frente ao modelo linear generalizado (*generalized linear model* - GLM). Seus resultados indicaram que os indicadores contábeis apresentaram uma capacidade preditiva significativa e desempenho superior do RF.

De forma análoga, D’Amato, D’Ecclesia e Levantesi (2022) usaram dados contábeis e de mercado na previsão de *scores* ESG da Refinitiv com o algoritmo RF e comparações com o GLM. O banco de dados compreendeu uma amostra de 401 empresas que compuseram o índice STOXX Europe 600 (67% das integrantes do índice) no período de 2009 a 2019. As autoras encontraram resultados alinhados com

sua pesquisa anterior, na qual dados contábeis demonstraram poder preditivo significativo e o RF como abordagem mais indicada.

Raza et al. (2022) também se concentraram na previsão de *scores* ESG da Refinitiv a partir de métricas contábeis com *machine learning*, o que reduziria a necessidade de modelos subjetivos de avaliação ESG. Especificamente, os autores fizeram uso de uma amostra de 300 companhias não-financeiras dos Estados Unidos, Reino Unido e Alemanha de 2008 a 2020 e empregaram regressão polinomial, KNN, naïve Bayes, RF, SVM (linear e Rbf) e Redes Neurais Artificiais. Raza et al. (2022) identificaram as variáveis usadas a partir do balanço patrimonial e demonstração do resultado do exercício como relevantes para essa tarefa de previsão de *score* ESG e verificaram que a rede neural artificial obteve o melhor desempenho em termos de raiz do erro quadrático médio e erro absoluto médio.

Mais recentemente, Cini e Ferrari (2025) investigaram o uso de medidas contábeis e do beta condicional para prever a classificação de ESG no ano $t+1$ com ML. Os autores analisaram essa previsão com e sem a inclusão da classe ESG e seus pilares defasados como *inputs* e também consideraram o ESG de forma agregada em macroclasses e detalhada em subclasses. Para tanto, aplicaram o algoritmo RF como modelo principal para classificação e KNN e OVR de forma comparativa considerando empresas no índice STOXX Europe 600 no período de 2016 a 2021. Seus resultados apontaram que em nível macro, o RF obteve 0,80 e 0,62 de acurácia, considerando ou não a inclusão do ESG e seus pilares na modelagem, respectivamente. Tal desempenho superou os observados nos modelos comparativos. O RF também obteve melhor acurácia (0,57) na previsão das subclasses. Além disso, os autores apontaram que a inclusão do beta condicional trouxe ganhos para a análise, trazendo um conteúdo informacional relevante.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 Dados e Amostra

Para fins de realizar a análise proposta, foram coletados dados sobre o ESG score, os pilares E, S e G, separadamente, o ESG de controvérsia e das demonstrações contábeis anuais das empresas de capital aberto brasileiras listadas na Brasil, Bolsa,

Balcão (B3). O código de 2 dígitos do *North American Industry Classification System* (NAICS) foi adotado como critério para identificação das indústrias e foram excluídas empresas com códigos 52 e 53, as quais se referem a finanças, seguros e imóveis, e aquelas com valores ausentes. Essas informações foram obtidas da LSEG Data & Analytics (previamente denominada de Refinitiv).

Ademais, seguindo o artigo de Cini e Ferrari (2025), foi inserido o beta condicional dos ativos nas análises, o qual representa uma medida de risco sistêmico. Para essa estimação, utilizou-se o *Capital Asset Pricing Model* (CAPM) com correlações dinâmicas e foram adotados o índice Bovespa (Ibovespa) como *proxy* da carteira de mercado e o Certificado de Depósito Interbancário (CDI) como *proxy* do ativo livre de risco. Tais variáveis foram obtidas a partir da página eletrônica do Yahoo Finanças e do sistema de séries temporais do Banco Central do Brasil, respectivamente, enquanto que os retornos discretos dos ativos foram coletados da plataforma Comdinheiro.

O período de estudo abrangeu os anos de 2010 até 2023, o qual representa um compromisso em cobrir um maior lapso temporal, com a adoção obrigatória do IFRS no Brasil em 2010 (Mapurunga, Ponte, & Santos, 2014), e o último exercício fiscal com informação contábil disponível e está coerente com outras pesquisas nessa linha (por exemplo, Cini e Ferrari (2025) usam um intervalo de seis anos apenas). A amostra final do estudo abrangeu 104 empresas com 872 observações firma-ano.

3.2 Variáveis

Em termos de variáveis, o artigo adotou a proposta de Cini e Ferrari (2025), usando o valor de mercado da firma para normalização dos dados contábeis dada a heterogeneidade dos valores. O detalhamento é apresentado no Quadro 1.

Como pode ser observado no Quadro 1, foi adotado o ESG *score* disponível na plataforma da LSEG Data & Analytics, o qual define a categoria que a empresa se situa a partir de um dado intervalo tendo como base o ESG numérico. A Tabela 1 apresenta essas delimitações para o enquadramento e definição do *rating* ESG. Na Tabela 2 da seção de resultados se encontra a proporção de casos em cada intervalo definido.

Quadro 1
Variáveis Usadas no Estudo

Variável	Descrição
ESG	<i>Score</i> ESG
ENV	<i>Score</i> para o pilar ambiente
SOC	<i>Score</i> para o pilar social
GOV	<i>Score</i> para o pilar governança
CONTR	Controvérsias sobre os pilares ESG
VENDAS	Vendas líquidas
LAJIDA	Lucro antes de juros, impostos, depreciação e amortização
LL	Lucro após impostos
LAIR	Lucro antes dos impostos
CAIXA	Caixa das atividades operacionais - acumulado
PL	Patrimônio líquido
DPLN	Dummy com valor igual a 1 para patrimônio líquido negativo
PE	Passivo exigível
DIVT	Dívida total
DIVL	Dívida líquida
ROE	Retorno sobre o patrimônio líquido obtido como patrimônio líquido sobre o ativo total (return on equity - ROE)
ROA	Retorno sobre o ativo obtido como lucro após impostos sobre o ativo total (return on assets - ROA)
INT	Intangível - líquido
CAPEX	Despesas de capital (capital expenditure - CAPEX) - discreto
BETA	Beta condicional

Nota: Todas as variáveis contábeis foram normalizadas pelo valor de mercado da firma, exceto ROE, ROA e BETA.

Fonte: Autoria própria (2025)

Tabela 1
Intervalo para a categoria de ESG score

Intervalo numérico	ESG rating
$0.000000 \leq score \leq 0.083333$	<i>D-</i>
$0.083333 < score \leq 0.166666$	<i>D</i>
$0.166666 < score \leq 0.250000$	<i>D+</i>
$0.250000 < score \leq 0.333333$	<i>C-</i>
$0.333333 < score \leq 0.416666$	<i>C</i>
$0.416666 < score \leq 0.500000$	<i>C+</i>
$0.500000 < score \leq 0.583333$	<i>B-</i>
$0.583333 < score \leq 0.666666$	<i>B</i>
$0.666666 < score \leq 0.750000$	<i>B+</i>
$0.750000 < score \leq 0.833333$	<i>A-</i>
$0.833333 < score \leq 0.916666$	<i>A</i>
$0.916666 < score \leq 1.000000$	<i>A+</i>

Fonte: Autoria própria (2025)

Assim, ao invés de buscar prever um resultado numérico, o estudo se concentrou em usar uma abordagem de previsão de variáveis textuais no ano $t+1$, as quais compreendem os *ratings* ESG.

Para a estimação do beta foi utilizado o modelo *Dynamic Conditional Correlation - Generalized Autoregressive Conditional Heteroskedasticity* (DCC-GARCH) de Engle (2002), considerando as correlações condicionais entre os preços diários de cada título e o índice Ibovespa, considerando retornos em excesso a taxa do CDI.

3.3 Modelos de *Machine Learning*

No âmbito da classificação do *rating* ESG, procedeu-se à avaliação de quatro modelos de ML. O intuito desta análise comparativa foi discernir o modelo com o

desempenho superior para a referida tarefa. As subseções seguintes descrevem os modelos empregados.

3.3.1 Random Forest (RF)

Proposto Breiman (2001), o RF é um algoritmo baseado em um conjunto de árvores de decisão, onde múltiplas árvores são construídas a partir de amostras aleatórias dos dados. De acordo com Bengio et al. (2016), o RF melhora a robustez das árvores de decisão combinando várias árvores de forma aleatória. Para cada árvore, um subconjunto aleatório de tamanho n é extraído com reposição do conjunto de dados de treinamento - é o processo chamado de Bootstrap - que é feito para garantir a diversidade entre as árvores individuais e reduzir a variância do modelo. Após isso ocorre a seleção aleatória de variáveis, em que cada nó, um subconjunto aleatório de m variáveis preditoras ($m < p$, onde p é o total de variáveis preditoras) é escolhido para determinar os melhores pontos de divisão. Ao final, o resultado é obtido por votação, dado por:

$$\hat{y} = \arg \max_j \sum_{b=1}^B I(f_b(x) = j)$$

onde:

- $f_b(x)$ é a predição da b -ésima árvore
- $I(\cdot)$ é a função indicadora
- \hat{y} é a previsão de classe

3.3.2 K-Nearest Neighbors (k-NN)

O KNN (Cover & Hart, 1967) é um algoritmo clássico de aprendizado supervisionado usado principalmente para problemas de classificação. Como define Kramer (2013), a ideia central é simples: classificar um ponto de dados com base na classe ou valor de seus k vizinhos mais próximos no espaço de atributos. Dado um ponto de dados desconhecido x , o KNN calcula a distância entre esse ponto e todos os outros pontos do conjunto de treinamento. A distância é calculada com diferentes métricas, como a distância Euclidiana ou a de Manhattan.

Uma vez que as distâncias são calculadas, o algoritmo seleciona os k pontos mais próximos de x . A classe do ponto x é determinada por uma votação majoritária entre os k vizinhos mais próximos. Por fim, o ponto x será classificado na classe que aparecer com mais frequência entre os k vizinhos.

3.3.3 Support Vector Machine (SVM)

O SVM (Cortes & Vapnik, 1995) é um modelo de aprendizado de máquina particularmente eficaz quando há alta dimensionalidade dos dados. Como definido por Noble (2006) o algoritmo trabalha construindo hiperplanos separando duas ou mais classes - maximizando a margem entre as classes, ou seja, a distância entre o hiperplano e os pontos mais próximos de cada classe, chamados de vetores de suporte. O objetivo do SVM é encontrar um hiperplano $W^t X + b = 0$ que maximize a margem e que seja capaz de classificar corretamente todos os pontos. A função de otimização do SVM, para encontrar o hiperplano, é dada por:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

sujeito à condição de separabilidade:

$$y_i(w^t x_i + b) \geq 1, \forall i$$

3.3.4 One-vs-Rest (OVR)

O modelo One-Versus-Rest (OVR) (Klautau & Rifkin, 2004), também conhecido como One-Vs-All (OVA), é uma abordagem para classificação multiclasse que treina um classificador binário para cada classe, distinguindo-a das demais. Em um problema com K classes, o método consiste em treinar K classificadores binários, onde cada um aprende a diferenciar sua classe específica das outras.

Para cada classe i , é treinado um classificador binário $f_i(x)$, onde:

$$f_i(x) = w_i^t x + b_i$$

a predição final para um novo exemplo x é obtida selecionando o classificador com a maior saída:

$$\hat{y} = \arg \max_i f_i(x)$$

ou seja, é atribuído x à classe cujo classificador binário retorna o maior valor de decisão.

3.3 Métricas de desempenho

Para avaliar e comparar a eficácia dos modelos, foram empregadas as seguintes métricas de desempenho: acurácia, precisão, revocação e F1-score. Essas métricas são amplamente utilizadas em tarefas de classificação e fornecem uma análise abrangente do desempenho preditivo dos modelos. De acordo com Rainio, Teuho e Klén (2024), a acurácia mede a proporção de previsões corretas sobre o total de observações. E é definida como $\alpha = \frac{TP + TN}{TP + TN + FP + FN}$, onde:

- TP (True Positives): verdadeiros positivos (exemplos corretamente classificados como positivos).
- TN (True Negatives): verdadeiros negativos (exemplos corretamente classificados como negativos).
- FP (False Positives): falsos positivos (exemplos negativos classificados incorretamente como positivos).
- FN (False Negatives): falsos negativos (exemplos positivos classificados incorretamente como negativos).

A precisão indica a proporção de exemplos classificados como positivos que realmente pertencem à classe positiva. Sua definição é: $\pi = \frac{TP}{TP + FP}$. A revocação ou sensibilidade se preocupa em medir a capacidade do modelo de identificar corretamente todas as ocorrências positivas usando a fórmula: $\rho = \frac{TP}{TP + FN}$. Por último, o F1-Score é a média harmônica entre precisão e revocação, proporcionando um equilíbrio entre essas duas métricas - definida como: $\phi = 2 \times \frac{\pi \times \rho}{\pi + \rho}$ (Rainio et al., 2024).

Com o objetivo de avaliar a capacidade dos modelos de prever o *score* ESG, a modelagem foi realizada de duas formas: Classificação detalhada e agrupada. Na primeira, foram consideradas todas as classes e subclasses do *score* ESG (A, A+, A-, B, B+, B-, C, C+, C-, D, D+, D-). Essa abordagem permite capturar variações mais sutis na classificação, mas também torna a tarefa mais complexa e sensível a problemas de

desbalanceamento. Na segunda, as subclasses foram agrupadas em suas categorias principais, reduzindo o número de classes para A, B, C e D. Esse agrupamento simplifica a modelagem e pode melhorar a estabilidade da previsão, reduzindo os impactos do desbalanceamento

Para a previsão do *rating* ESG em $t+1$, foram consideradas as seguintes estratégias de estimação:

- I: inclusão das variáveis relacionadas às questões ESG, índices contábeis e beta condicional;
- II: inclusão dos índices contábeis e beta condicional.

A implementação dos modelos foi realizada na linguagem Python 3.13. Diversas bibliotecas foram empregadas para manipulação de dados, modelagem e avaliação dos resultados. Para o tratamento e análise exploratória dos dados, foi utilizado pandas 2.2.3 (Reback et al., 2020) e numpy 2.2.0 (Harris et al., 2020), enquanto as funções estatísticas, como cálculo de assimetria e curtose, foram extraídas do scipy.stats 0.14.4 (Perktold, & Seabold, 2010). A visualização dos dados foi facilitada pelo uso do matplotlib 3.9 (Hunter, 2007) e seaborn 0.13.2 (Waskom, 2021). A divisão dos dados, modelagens, e avaliações foi feita com a biblioteca Scikit-learn 1.6.1 (Pedregosa et al., 2011).

Os dados foram divididos em 80% (698 observações) para treino e 20% (174 observações) para teste. O tempo de estimação dos modelos foi de, no máximo, 2 minutos, garantindo uma execução rápida sem comprometer a qualidade das previsões. Esse desempenho foi possível devido à otimização dos hiperparâmetros e ao uso eficiente dos recursos computacionais, permitindo uma análise ágil e reprodutível.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Esta seção começa com uma análise exploratória dos dados, destacando a distribuição das empresas em cada categoria do *score* ESG, apresentando tanto a estrutura detalhada (A, A+, A-, B, B+, B-, C, C+, C-, D, D+, D-) quanto a versão agrupada (A, B, C, D), além da evolução da média do *score* ESG ao longo dos anos, permitindo compreender a distribuição das classes e desafios como o desbalanceamento. Em seguida, são avaliados os resultados da modelagem preditiva, comparando o

desempenho de diferentes algoritmos por meio de métricas como acurácia, precisão, revocação e F1-score, além de discutir a influência das abordagens de classificação e o impacto de variáveis explicativas. Por fim, exploram-se as implicações dos resultados, destacando desafios na previsão do *score* ESG, limitações do estudo e potenciais melhorias para pesquisas futuras, bem como aplicações práticas dos modelos desenvolvidos.

4.1 Análise exploratória

A Tabela 2 apresenta a distribuição percentual das classificações ESG em subclasses. Observa-se que a maior parte das empresas se encontra na categoria B e suas variações (B, B-, B+), totalizando 42% (14,75%, 14,13% e 13,11%, respectivamente). A categoria C também possui representatividade significativa, somando 33% (C+: 13,01%, C: 11,78% e C-: 8,19%). Já a categoria A, que representa as empresas com melhor desempenho em critérios ESG, possui participação mais restrita, com 12% (A-: 9,01%, A: 2,45% e A+: 0,10%). Por fim, a categoria D, que indica menor aderência às boas práticas ESG, representa 10,89% do total (D+: 7,37%, D: 4,30% e D-: 1,74%).

A análise dessas distribuições revela que a maioria das empresas se concentra nas categorias intermediárias (B e C), enquanto as classificações extremas (A e D) apresentam menor frequência. Essa estrutura sugere que, embora haja um avanço na incorporação de práticas ESG, a quantidade de empresas com desempenho diferenciado ainda é reduzida, ao passo que a parcela de empresas com mais baixa pontuação permanece significativa.

A Figura 1 ilustra, por meio de um gráfico de linha temporal, a evolução anual da pontuação média de ESG para diversos setores econômicos presentes no banco de dados. O setor de Informação demonstra as maiores pontuações ao longo do período analisado, com uma variação entre 55 e 65 pontos. Setores como Mineração e Extração de Petróleo e Gás também exibem valores médios elevados, situando-se entre 50 e 60 pontos. Em contraste, os setores de Saúde e Educação apresentam maior volatilidade, com pontuações oscilando entre 35 e 45 pontos. O setor de Acomodação e Alimentação registra as menores médias de ESG, mantendo-se na faixa de 15 a 20 pontos ao longo dos anos. Similarmente, os setores de Construção e Administração e Gestão de Empresas apresentam valores médios próximos a 20 e 30 pontos. Em suma, o gráfico

destaca uma persistente disparidade entre os setores, com alguns apresentando consistentemente um desempenho superior em ESG, enquanto outros permanecem com pontuações inferiores em toda a série temporal.

Tabela 2

Distribuição de empresas em cada categoria ESG detalhado

ESG rating	Porcentagem
A+	0,10%
A	2,45%
A-	9,01%
B+	13,11%
B	14,75%
B-	14,13%
C+	13,01%
C	11,78%
C-	8,19%
D+	7,37%
D	4,30%
D-	1,74%

Fonte: Autoria própria (2025)

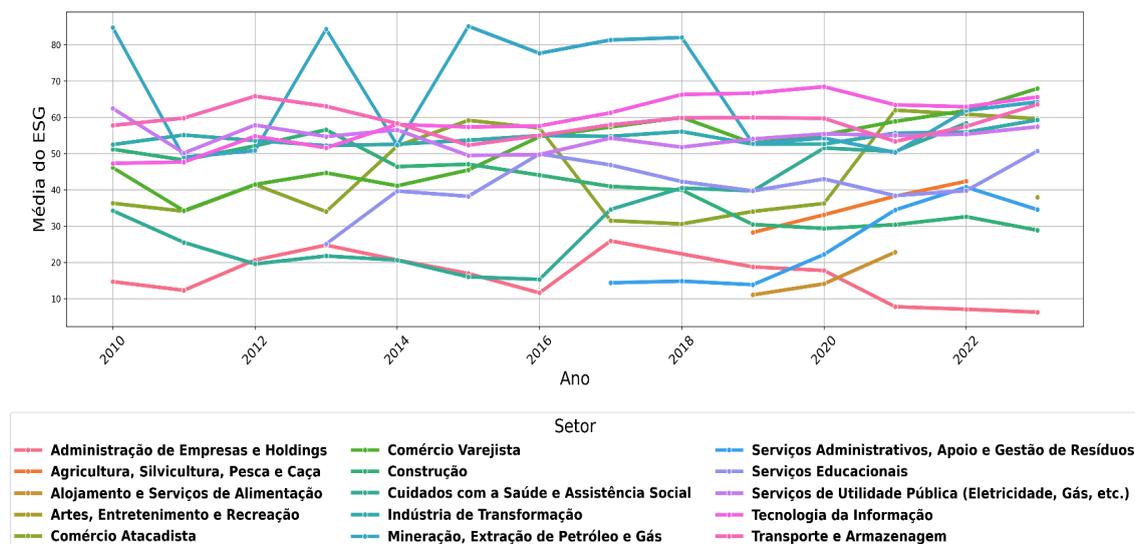


Figura 1. Evolução do ESG por setor (2010 - 2023)

Fonte: Autoria própria (2025)

A Figura 1 ilustra, por meio de um gráfico de linha temporal, a evolução anual da pontuação média de ESG para diversos setores econômicos presentes no banco de dados. O setor de Informação demonstra as maiores pontuações ao longo do período analisado, com uma variação entre 55 e 65 pontos. Setores como Mineração e Extração de Petróleo e Gás também exibem valores médios elevados, situando-se entre 50 e 60 pontos. Em contraste, os setores de Saúde e Educação apresentam maior volatilidade, com pontuações oscilando entre 35 e 45 pontos. O setor de Acomodação e Alimentação registra as menores médias de ESG, mantendo-se na faixa de 15 a 20 pontos ao longo dos anos. Similarmente, os setores de Construção e Administração e Gestão de Empresas apresentam valores médios próximos a 20 e 30 pontos. Em suma, o gráfico destaca uma persistente disparidade entre os setores, com alguns apresentando consistentemente um desempenho superior em ESG, enquanto outros permanecem com pontuações inferiores em toda a série temporal.

A Tabela 3 apresenta a matriz de transição de Markov, que detalha as probabilidades de migração entre os diferentes estados (classificações ESG) ao longo de um ano. Essa matriz revela uma dinâmica significativa nas categorias intermediárias de ESG, indicando uma considerável movimentação de empresas nesses rankings. Por outro lado, a dificuldade em alcançar os níveis mais elevados de ESG é reforçada pela matriz, que demonstra uma única transição envolvendo o rating mais alto (A+), e essa ocorre em sentido decrescente.

Em comparação com os resultados de Cini e Ferrari (2025), a matriz de transição construída com dados de empresas brasileiras revela uma concentração notável em sua diagonal principal. Esse padrão sugere uma maior estabilidade das empresas em suas classificações ESG ao longo do tempo, indicando uma baixa mobilidade entre os diferentes ratings. Essa característica de menor dinamismo pode representar um desafio para a modelagem preditiva via técnicas de Machine Learning, pois a limitada variação da variável dependente pode dificultar a capacidade dos modelos em identificar padrões significativos de transição entre os níveis de classificação ESG.

4.2 Resultados da Estimação

A apresentação dos resultados será dividida em duas estratégias distintas: I e II. A Tabela 4 detalha o desempenho dos algoritmos na tarefa de previsão do rating ESG, com a Estratégia I, que emprega o *score* ESG e seus pilares individuais como variáveis preditoras. Observa-se que os modelos aprendem essa relação intrínseca, beneficiando-se de um ganho substancial de informação desses componentes. Contudo, essa forte correlação entre as variáveis pode levar ao overfitting se não controlada adequadamente. Para mitigar esse risco, o número total de árvores e a profundidade máxima do modelo foram reduzidos. Os resultados alcançados com essa estratégia demonstram uma acurácia de 96% para o RF e de 93% para o SVM.

Tabela 3

Matriz de transição para o ESG score

t	A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-
A+	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
A	0,00	0,81	0,19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
A-	0,00	0,07	0,66	0,24	0,03	0,00	0,00	0,00	0,00	0,00	0,00	0,00
B+	0,00	0,01	0,21	0,59	0,17	0,03	0,00	0,00	0,00	0,00	0,00	0,00
B	0,00	0,00	0,04	0,20	0,57	0,18	0,01	0,00	0,00	0,00	0,00	0,00
B-	0,00	0,00	0,01	0,07	0,24	0,50	0,16	0,03	0,00	0,00	0,00	0,00
C+	0,00	0,00	0,00	0,01	0,06	0,26	0,51	0,14	0,02	0,00	0,00	0,00
C	0,00	0,00	0,00	0,00	0,01	0,09	0,25	0,46	0,18	0,01	0,01	0,00
C-	0,00	0,00	0,00	0,00	0,00	0,01	0,12	0,32	0,44	0,10	0,00	0,00
D+	0,00	0,00	0,00	0,00	0,01	0,03	0,01	0,07	0,13	0,58	0,15	0,00
D	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,05	0,13	0,36	0,41	0,05
D-	0,00	0,00	0,00	0,00	0,00	0,00	0,07	0,00	0,00	0,07	0,20	0,67
	$t + 1$											

Fonte: Autoria própria (2025)

Por outro lado, a Tabela 5 apresenta os resultados da Estratégia II, que utiliza apenas indicadores contábeis e o beta condicional como preditores. Nesta abordagem, a capacidade preditiva do modelo é mais realista, com a acurácia do Random Forest

declinando para 35% e a do SVM para 28%. Esses resultados indicam uma menor robustez do modelo na previsão do rating ESG quando se utilizam apenas essas variáveis. Para alcançar uma previsão mais precisa nesse nível de granularidade das classes, seria necessário treinar o modelo com um volume consideravelmente maior de dados do que o atualmente disponível.

Por meio de um processo iterativo, identificou-se que o valor de $k = 4$ proporcionou o melhor desempenho do algoritmo KNN em termos de acurácia. Conforme esperado, seus resultados foram inferiores aos obtidos pelos modelos RF e SVM, que demonstraram maior robustez preditiva. O KNN atingiu uma acurácia de 72% na Estratégia I e 22% na Estratégia II. Por fim, o modelo *baseline* OVR apresentou o pior desempenho entre os avaliados, com acurácias de 42% e 21% nas Estratégias I e II, respectivamente.

A melhor estimativa, o RF da estratégia I (Tabela 4), obteve precisão e *recall* de 100% em 6 das 11 subclasses, evidenciando a capacidade do modelo em identificar corretamente essas classes e minimizar falsos positivos/negativos. Essas métricas indicam uma dificuldade em prever a classe C- (66% de precisão, 70% de recall, F1-score de 68%), possivelmente por semelhança com outras. Por último, nenhum modelo conseguiu prever corretamente a classe A+, tendo em vista sua escassez de casos, conforme demonstrado na Tabela 1.

Dessa forma, os modelos de ML utilizados demonstram um desempenho superior na previsão de empresas que já possuem ratings ESG estabelecidos, em virtude da inerente estabilidade dessas classificações no curto prazo, dado que foi avaliada a transição de um ano para outro. Contudo, mesmo diante de cenários envolvendo empresas sem ratings preexistentes ou quando essa informação não está disponível para os modelos de ML, a capacidade preditiva diminui, embora ainda preserve sua relevância, indicando que as informações contábeis e de mercado podem ser utilizadas para previsão de ratings, mas nesse caso precisariam ser complementadas com outras informações.

Complementando os achados, foi analisado o *feature importance* do RF para as duas estratégias adotadas. Ao extrair a informação do quanto a variável preditora é importante para a previsão do modelo, utilizando os componentes ESG, observou-se,

conforme a Figura 2, que essas variáveis tiveram uma importância alta em relação aos índices contábeis e somados, os componentes exercem 60% de importância na previsão do ESG. Enquanto isso, os índices contábeis e o beta dividem 40% da importância, com a receita e o intangível tendo os maiores níveis (3%). Ao retirar os componentes ESG, conforme a Figura 3, evidencia-se que o intangível alcançou 10% de importância, seguido pelo Capex, vendas e o beta, com 8%, 7% e 6% de importância, respectivamente.

4.3 Análise adicional

Cini e Ferrari (2025) adotam dois conjuntos distintos de variáveis dependentes: o ESG em sua forma detalhada (A+, A, A-, B+, B, B-, C+, C, C-, D+, D e D-) e o ESG em classes macro (A, B, C e D). A utilização do ESG agrupado tende a melhorar o desempenho geral dos modelos preditivos, pois reduz o número de classes a serem previstas, diminuindo tanto a complexidade do problema quanto o desbalanceamento entre categorias.

Sendo assim, buscou-se analisar os resultados considerando as macroclasses. A matriz de transição para essa categorização revelou, similarmente ao constatado na Tabela 3, uma elevada permanência das empresas em seus respectivos ratings: 100% das empresas classificadas como A mantiveram-se nessa categoria, seguidas por 96% das empresas com score B que permaneceram em B, 89% das empresas em C que não migraram, e 79% das empresas em D que também mantiveram sua classificação. Essa notável estabilidade contribuiu para que, na Estratégia I, o modelo RF alcançasse uma acurácia de 99% no conjunto de teste, com precisão e recall de 100% em quase todas as classes, seguido pelo SVM, com 97% de acurácia. Contudo, o modelo apresentou indícios de overfitting, atribuídos principalmente a dois fatores interligados: a baixa variabilidade na variável dependente e a utilização dos componentes ESG do período $t-1$ como preditores.

Tabela 4
Resultado com a Estratégia I para prever o rating ESG

RF (Acurácia: 96%)			KNN (Acurácia: 72%)			SVM (Acurácia: 93%)			OVR (Acurácia: 42%)			
Classe	Precisão	Recall	F1-Score	Precisão	Recall	F1-Score	Precisão	Recall	F1-Score	Precisão	Recall	F1-Score
A+	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
A	76%/	100%	85%	66%	66%	66%	100%	77%	85%	100%	35%	50%
A-	100%	100%	100%	71%	100%	82%	82%	100%	92%	61%	80%	69%
B+	95%	100%	96%	69%	69%	69%	94%	75%	82%	45%	57%	50%
B	100%	90%	95%	80%	78%	79%	89%	87%	88%	40%	25%	30%
B-	91%	100%	95%	75%	73%	74%	82%	84%	83%	39%	62%	45%
C+	100%	100%	100%	78%	70%	74%	75%	89%	81%	41%	70%	55%
C	92%	100%	96%	82%	71%	76%	85%	85%	85%	43%	41%	42%
C-	66%	70%	68%	59%	59%	59%	100%	82%	90%	15%	15%	15%
D+	69%	64%	65%	67%	70%	69%	85%	82%	83%	57%	38%	42%
D	81%	69%	75%	78%	78%	78%	84%	84%	84%	77%	22%	35%
D-	100%	100%	100%	100%	23%	40%	80%	100%	88%	100%	78%	83%

Fonte: Autoria própria (2025)

Tabela 5:
Resultado com a Estratégia I para prever o rating ESG

	RF (Acurácia: 35%)			KNN (Acurácia: 22%)			SVM (Acurácia: 28%)			OVR (Acurácia: 21%)		
Classe	Precisão	Recall	F1-Score	Precisão	Recall	F1-Score	Precisão	Recall	F1-Score	Precisão	Recall	F1-Score
A+	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
A	34%	34%	34%	0%	0%	0%	20%	35%	25%	0%	0%	0%
A-	40%	28%	32%	21%	35%	30%	21%	25%	24%	40%	15%	20%
B+	37%	34%	35%	21%	36%	25%	26%	25%	26%	16%	17%	17%
B	49%	51%	50%	40%	24%	31%	35%	43%	39%	24%	26%	25%
B-	23%	53%	36%	75%	73%	74%	30%	36%	35%	16%	33%	22%
C+	50%	40%	43%	22%	24%	23%	21%	22%	21%	29%	35%	33%
C	29%	29%	29%	5%	4%	4%	35%	33%	34%	30%	8%	14%
C-	28%	11%	18%	0%	0%	0%	43%	22%	31%	0%	0%	0%
D+	35%	35%	35%	24%	9%	11%	15%	9%	10%	64%	18%	25%
D	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
D-	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Fonte: Autoria própria (2025)

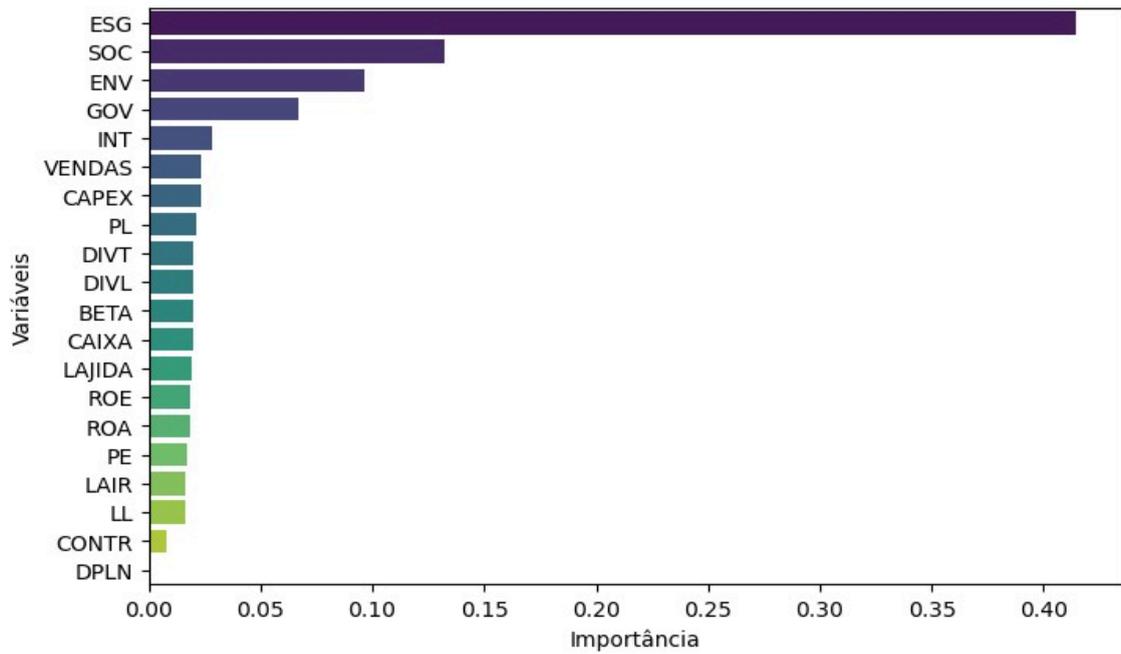


Figura 2. Feature Importance do RF para a estratégia I

Fonte: Autoria própria (2025)

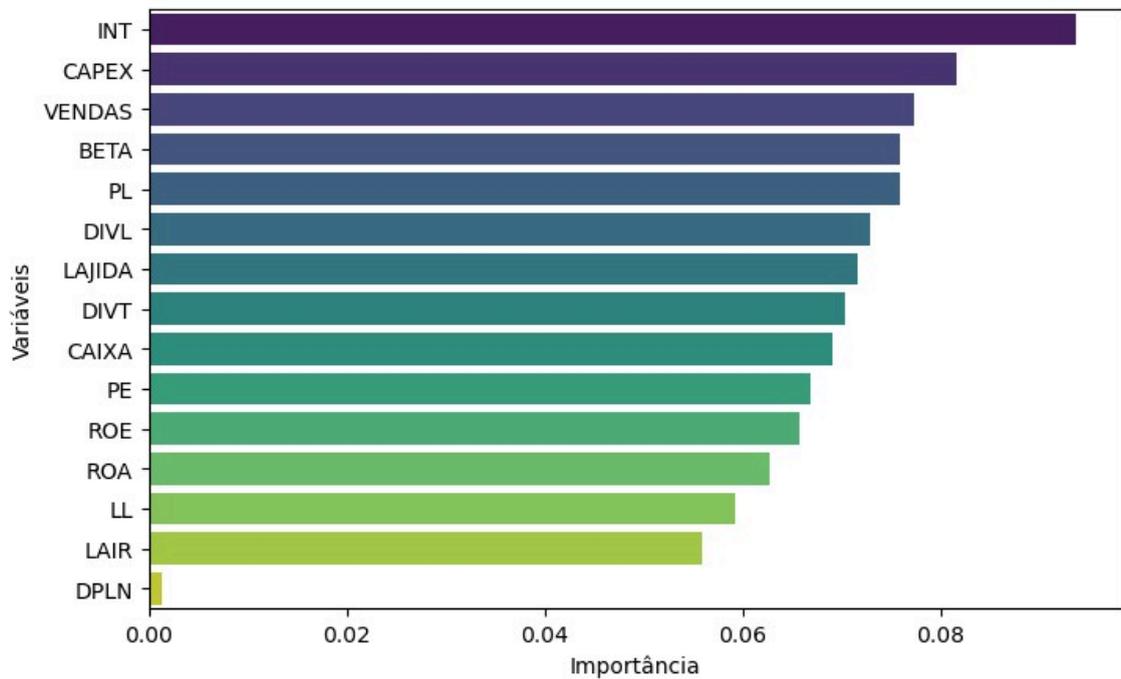


Figura 3. Feature Importance do RF para a estratégia II

Fonte: Autoria própria (2025)

Finalmente, para avaliar a robustez dos modelos, restringiu-se a análise aos dois setores com o maior número de observações empresa-ano: indústria de transformação e

serviços de utilidade pública. Ao utilizar o ESG defasado e seus pilares como preditores, em conjunto com as demais variáveis, obteve-se uma acurácia de 93,54% com o RF, 51,61% com o KNN, 97,84% com o SVM e 35,48% com o OVR. Ao remover o ESG e seus componentes das variáveis preditoras, o desempenho geral dos algoritmos sofreu uma redução significativa, para 20,43% no RF, 19,35% no KNN, 17,20% no SVM e 17% no OVR. Em suma, os padrões observados anteriormente se mantiveram, reforçando a tendência das empresas em permanecer em suas respectivas classificações ESG.

5 CONSIDERAÇÕES FINAIS

A crescente urgência trazida pelas mudanças climáticas, pela degradação ambiental, necessidade de respeitar direitos humanos e ter transparência tem intensificado o foco nas práticas de ESG, impulsionado por marcos globais como os ODS e o Acordo de Paris, nos quais o Brasil assumiu compromissos ambiciosos de redução de emissões. Nesse contexto, as questões ESG emergem como um indicador crucial do engajamento corporativo com a sustentabilidade, influenciando diversas partes interessadas. Contudo, os desafios persistem em relação à metodologia, padronização e confiabilidade das avaliações de ESG.

Este estudo contribui para a literatura ao explorar a capacidade preditiva de técnicas de Machine Learning na previsão de ratings ESG utilizando dados contábeis e o beta de mercado de empresas não-financeiras brasileiras de capital aberto. Ao fazer isso, avança na investigação sobre a utilidade de variáveis contábeis nesse contexto, complementando estudos internacionais e oferecendo evidências para o mercado brasileiro, onde tais análises ainda são incipientes. Os achados da pesquisa, por sua vez, apontam para a relevância dos indicadores contábeis como preditores de ratings ESG, alinhando-se com a perspectiva de que esses dados oferecem clareza, consistência, robustez, e menor suscetibilidade a ruídos de mercado, além de maior transparência e confiabilidade por serem padronizados e auditados. Os modelos têm um melhor desempenho na previsão de empresas que já possuem ratings, pela característica desse tipo de classificação não sofrer grandes alterações no curto prazo. Quando confrontados com situações de empresas que não possuem ratings, ou essa informação não é disponibilizada aos modelos, a capacidade preditiva diminui, porém não deixa de ser importante.

A pesquisa, porém, não está isenta de limitações. A amostra depende da existência de *score* ESG associado a firma e isso reflete na redução da quantidade de observações disponíveis para análise e no acompanhamento dessas transições ao longo do tempo. Como sugestão para trabalhos futuros, pode-se considerar a comparação de outros algoritmos de Machine Learning e abordagens diferenciadas para o próprio acesso ao desempenho ESG das firmas, como modelos baseados em Redes Neurais e/ou pré-processamento com data augmentation. Também pode se avaliar a capacidade preditiva em intervalos maiores, como a cada dois anos ou mais.

REFERÊNCIAS

- Ahmad, N., Mobarek, A., & Roni, N. N. (2021). Revisiting the impact of ESG on financial performance of FTSE350 UK firms: Static and dynamic panel data analysis. *Cogent Business & Management*, 8(1), 1900500.
- Aydoğmuş, M., Gülay, G., & Ergun, K. (2022). Impact of ESG performance on firm value and profitability. *Borsa Istanbul Review*, 22, S119-S127.
- Brasil, Presidência da República. (2024, 13 de novembro). *Em Baku, Brasil destaca meta de redução de 67% nas emissões de gases de efeito estufa até 2035*. Recuperado de <https://www.gov.br/planalto/pt-br/acompanhe-o-planalto/noticias/2024/11/em-baku-brasil-destaca-meta-de-reducao-de-67-nas-emissoes-de-gases-de-efeito-estufa-at-e-2035>
- Bengio, Y. Goodfellow, I. Courville A. (2016). *Deep Learning*. MIT Press Book. Recuperado de http://imlab.postech.ac.kr/dkim/class/cs514_2019s/DeepLearningBook.pdf
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bzdok, D., Altman, N. & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, 15, 233–234. <https://doi.org/10.1038/nmeth.4642>
- Cini, F., & Ferrari, A. (2025). Towards the estimation of ESG ratings: A machine learning approach using balance sheet ratios. *Research in International Business and Finance*, 73, 102653.

- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297. <https://doi.org/10.1007/BF00994018>
- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1), 21-27. <https://doi.org/10.1109/TIT.1967.1053964>
- D'Amato, V., D'Ecclesia, R., & Levantesi, S. (2021). Fundamental ratios as predictors of ESG scores: A machine learning approach. *Decisions in Economics and Finance*, 44(2), 1087-1110.
- D'Amato, V., D'Ecclesia, R., & Levantesi, S. (2022). ESG score prediction through random forest algorithm. *Computational Management Science*, 19(2), 347-373.
- Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3), 339-350.
- García, F., González-Bueno, J., Guijarro, F., & Oliver, J. (2020). Forecasting the environmental, social, and governance rating of firms by using corporate financial performance variables: A rough set approach. *Sustainability*, 12(8), 3324.
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(03), 90-95. doi: 10.1109/MCSE.2007.55
- Kramer, O. (2013). K-Nearest Neighbors. In: *Dimensionality Reduction with Unsupervised Nearest Neighbors*. Intelligent Systems Reference Library, 51. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38652-7_2
- Krappel, T., Bogun, A., & Borth, D. (2021). Heterogeneous ensemble for ESG ratings prediction. arXiv preprint arXiv:2109.10085. Recuperado de <https://arxiv.org/abs/2109.10085>
- Li, Y., & Li, S. (2024). ESG performance and innovation quality. *International Review of Economics & Finance*, 92, 1361-1373. <https://doi.org/10.1016/j.iref.2024.02.063>

- Noble, W. S. (2006). What is a support vector machine?. *Nature Biotechnology*, 24(12), 1565-1567. <https://doi.org/10.1038/nbt1206-1565>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
<http://www.jmlr.org/papers/v12/pedregosa11a.html>
- Rainio, O., Teuvo, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1), 6086.
<https://doi.org/10.1038/s41598-024-56706-x>
- Raza, H., Khan, M. A., Mazliham, M. S., Alam, M. M., Aman, N., & Abbas, K. (2022). Applying artificial intelligence techniques for predicting the environment, social, and governance (ESG) pillar score based on balance sheet and income statement data: A case of non-financial companies of USA, UK, and Germany. *Frontiers in Environmental Science*, 10, 975487.
- Reback, J., McKinney, W., Van Den Bossche, J., Augspurger, T., Cloud, P., Klein, A., ... & Seabold, S. (2020). pandas-dev/pandas: Pandas 1.0. 5. *Zenodo*.
<https://doi.org/10.5281/zenodo.3898987>
- Rifkin, R., & Klautau, A. (2004). In defense of one-vs-all classification. *Journal of Machine Learning Research*, 5(Jan), 101-141.
- Saha, A. K., & Khan, I. (2024). Sustainable prosperity: unravelling the Nordic nexus of ESG, financial performance, and corporate governance. *European Business Review*, 36(6), 793-815.
- Santos, E. S., Ponte, V. M. R., & Mapurunga, P. V. R. (2014). Adoção obrigatória do IFRS no Brasil (2010): índice de conformidade das empresas com a divulgação requerida e alguns fatores explicativos. *Revista Contabilidade & Finanças*, 25(65), 161-176.
- Seabold, S., & Perktold, J. (2010). Statsmodels: econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference - *SciPy*, 7(1), 92-96. <https://doi.org/10.25080/ajora-92bf1922-011>

- Taskin, D., Cagli, E. C., & Mandaci, P. E. (2021). The impact of temperature anomalies on commodity futures. *Energy Sources, Part B: Economics, Planning, and Policy*, 16(4), 357-370. <https://doi.org/10.1080/15567249.2021.1922546>
- Taskin, D., Sariyer, G., Acar, E., & Cagli, E. C. (2025). Do past ESG scores efficiently predict future ESG performance?. *Research in International Business and Finance*, 74, 102706.
- Waskom, M. L., (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021, <https://doi.org/10.21105/joss.03021>
- Whelan, T., Atz, U., Van Holt, T., & Clark, C. (2021). ESG and financial performance. Uncovering the relationship by aggregating evidence from 1,000 plus studies published between 2015-2020, 1-19.
- Wong, R., Nguyen, H. T. M., & Kwansa, N. A. (2024). ESG performance and cost of capital: What do we know?. Evidence from the US. *International Journal of Monetary Economics and Finance*, 17(1), 74-96.