UFABC

Universidade Federal do ABC

Trabalho de Graduação em
Engenharia de Informação

Classificação de Risco ESG com Dados Públicos: Uma Análise Comparativa de Modelos de Machine Learning

Lucas Sanchez Bitencourt
Michael Franklin Saito da Silva

Santo André, SP
Abril de 2026

Lucas Sanchez Bitencourt
Michael Franklin Saito da Silva

Classificação de Risco ESG com Dados Públicos: Uma Análise Comparativa de Modelos de Machine Learning

Trabalho de Graduação apresentado ao Curso de Graduação em Engenharia da Informação da Universidade Federal do ABC como requisito parcial para obtenção do título de Engenheiro da Informação.

Universidade Federal do ABC

Orientador:
Prof. Dr. Ricardo Suyama

Santo André, SP
2026

Resumo

Este trabalho investiga a viabilidade de classificação de risco ESG (Environmental, Social and Governance - critérios ambientais, sociais e de governança) a partir de dados públicos estruturados, com foco em sua aplicação como ferramenta de triagem e apoio à decisão no contexto de risco corporativo. A pesquisa adota abordagem quantitativa e exploratório-aplicada, estrutururando um pipeline que integra dados financeiros de empresas do S&P 500 (via Yahoo Finance) a indicadores públicos de risco ESG, após etapas de tratamento, padronização e consolidação da variável-alvo em três níveis (Low, Medium e High).

O problema foi modelado como classificação supervisionada multiclasse, utilizando Random Forest como algoritmo principal, e comparado aos modelos XGBoost (eXtreme Gradient Boosting) e KNN (K-Nearest Neighbors), com aplicação de SMOTE (Synthetic Minority Over-sampling Technique) e validação cruzada estratificada. A avaliação considerou métricas como acurácia, F1-score, AUC-ROC (Area Under the Receiver Operating Characteristic Curve) e AUC-PR (Area Under the Precision-Recall Curve).

Os resultados indicam desempenho consistente para fins de triagem, com maior capacidade de discriminação nas classes Low e Medium e limitações na identificação da classe High, e a análise de importância de atributos e explicações via SHAP (SHapley Additive exPlanations) demonstrou que o risco ESG, na base pública utilizada, é predominantemente capturado por quatro eixos: estrutura financeira e alavancagem, desempenho operacional, liquidez e sinais de mercado, e governança corporativa. Conclui-se que a modelagem é tecnicamente viável e interpretável, embora a detecção robusta de alto risco dependa do enriquecimento com informações qualitativas e contextuais.

Palavras-chave: ESG; risco ESG; classificação de risco; aprendizado de máquina; Random Forest; SHAP; dados públicos.

Sumário

Capítulo 1

Introdução

1.1 Introdução

1.2 Contextualização e Relevância do Tema ESG

Nos últimos anos, a incorporação de fatores ambientais, sociais e de governança (Environmental, Social and Governance - ESG) tem se consolidado como um elemento central na análise de desempenho e risco de empresas no mercado financeiro. Assim, investidores, instituições financeiras e órgãos reguladores têm ampliado a consideração desses fatores na tomada de decisão, buscando não apenas retorno financeiro, mas também sustentabilidade e responsabilidade corporativa [United Nations Principles for Responsible Investment 2019].

O conceito de ESG refere-se a um conjunto de critérios utilizados para avaliar o impacto das atividades empresariais sob três dimensões principais: ambiental, que está relacionada ao uso de recursos naturais, emissões e impactos ecológicos; social, que abrange relações de trabalho, diversidade e impacto na sociedade; e governança, que está associada à estrutura de gestão, transparência e práticas éticas. A integração desses critérios à análise financeira tradicional permite uma avaliação mais abrangente dos riscos e oportunidades associados às empresas [United Nations Principles for Responsible Investment 2019, Deloitte 2021].

Nesse contexto, a análise de risco de crédito, tradicionalmente baseada em indicadores financeiros, passa a incorporar fatores ESG como variáveis relevantes na avaliação da capacidade de pagamento e da sustentabilidade de longo prazo das organizações. Empresas com baixo desempenho em critérios ESG podem estar mais expostas a riscos regulatórios, reputacionais e operacionais, o que pode impactar diretamente sua estabilidade financeira [Deloitte 2021].

Paralelamente, o avanço das técnicas de ciência de dados e aprendizado de máquina tem possibilitado o desenvolvimento de modelos preditivos capazes de identificar padrões

complexos em grandes volumes de dados. Esses modelos podem ser aplicados na classificação de empresas quanto ao seu nível de risco ESG, auxiliando instituições financeiras na tomada de decisões mais informadas e alinhadas a práticas sustentáveis [LeoBreiman 2001].

Diante desse cenário, este trabalho propõe o desenvolvimento de modelos para classificação de risco ESG a partir de dados financeiros públicos, utilizando técnicas de aprendizado de máquina. A abordagem busca integrar informações provenientes de diferentes fontes, como indicadores financeiros extraídos do Yahoo Finance e ratings públicos de risco ESG, permitindo avaliar a relação entre características financeiras e níveis de risco ESG.

1.3 Definição do Problema de Pesquisa

Apesar do avanço na incorporação de critérios ESG em análises corporativas, ainda persiste um desafio significativo para as instituições financeiras: a disponibilidade de informações padronizadas e acessíveis. Muitas vezes, o acesso a bases de dados tratadas e consolidadas exige custos elevados, restringindo a aplicação prática de metodologias que integrem fatores ambientais, sociais e de governança em modelos de crédito.

Nesse contexto, surge a seguinte problemática: como desenvolver e analisar um sistema de classificação de risco ESG baseado exclusivamente em informações públicas, que possa ser integrado a modelos de crédito de instituições financeiras, contribuindo para decisões mais estratégicas e sustentáveis em concessão de empréstimos e investimentos?

Esse problema é relevante porque, ao possibilitar a utilização de dados públicos e acessíveis, reduz-se a dependência de soluções proprietárias, ampliando o alcance de práticas sustentáveis no setor financeiro e fortalecendo a gestão de riscos de crédito alinhada a critérios ESG.

1.4 Objetivos

O objetivo deste trabalho é analisar e desenvolver modelos de machine learning para classificação de risco ESG em empresas, utilizando dados públicos estruturados, de modo a investigar a viabilidade dessa abordagem para apoio à análise de risco corporativo.

Como objetivos específicos, destacam-se:

- Implementar e avaliar modelos de machine learning para classificação de risco ESG, com foco nos algoritmos Random Forest, KNN e XGBoost;

- Apresentar uma análise comparativa de desempenho entre os modelos Random Forest, KNN e XGBoost;
- Identificar os atributos mais relevantes para a classificação ESG com base em dados públicos;
- Discutir os resultados à luz da literatura acadêmica e das práticas de mercado financeiro.

1.5 Justificativa

A justificativa deste trabalho se apoia em dois pilares: prático e científico.

No âmbito prático, a adoção de práticas sustentáveis no mercado financeiro é uma demanda crescente, estimando-se que instituições financeiras que incorporam critérios ESG em suas análises conseguem não apenas reduzir riscos de inadimplência, mas também atrair investidores preocupados com responsabilidade socioambiental [PwCBrasil 2022].

No âmbito científico, a pesquisa contribui para a literatura ao propor uma abordagem metodológica de classificação de risco ESG baseada em técnicas de aprendizado de máquina, as quais têm se mostrado eficazes na identificação de padrões complexos em grandes volumes de dados [Géron 2019].

Além disso, a incorporação de fatores ESG na análise de risco de crédito tem sido amplamente discutida na literatura, destacando sua relevância para a avaliação mais abrangente das organizações [S&P Global Ratings, 2021].

Nesse contexto, a utilização de dados públicos representa uma contribuição relevante, pois pode democratizar o acesso a ferramentas de análise e reduzir a dependência de bases proprietárias, frequentemente custosas e restritivas, além de promover maior transparência na mensuração do risco ESG [PwCBrasil 2022, Deloitte 2021].

Capítulo 2

Referencial Teórico

2.1 ESG: Definições e Pilares

O conceito de ESG refere-se a um conjunto de critérios que avaliam o desempenho de uma organização para além de métricas financeiras tradicionais, contemplando sua atuação ambiental, responsabilidade social e práticas de governança corporativa. Esses fatores têm se consolidado como referenciais essenciais para investidores, reguladores e instituições financeiras, uma vez que permitem identificar riscos e oportunidades associados à sustentabilidade e à perenidade dos negócios [United Nations Principles for Responsible Investment 2019, Deloitte 2021].

De acordo com a FEBRABAN, o tema ESG vem ganhando centralidade no sistema financeiro brasileiro, alinhando-se a movimentos internacionais como os Princípios para o Investimento Responsável (PRI) da ONU. Ainda que não exista um padrão único de mensuração, há consenso sobre a relevância de seus três pilares fundamentais: ambiental, social e governança, como ilustrado na figura 2.1.

Figura 2.1: Pilares ESG: Ambiental, Social e Governança.

2.1.1 Fator Ambiental (Environmental)

O fator ambiental avalia como as organizações gerenciam seus impactos sobre o meio ambiente, englobando aspectos como emissões de gases de efeito estufa, eficiência energética, gestão de resíduos e uso responsável de recursos naturais. Empresas que negligenciam essas dimensões podem enfrentar riscos regulatórios, multas e danos reputacionais, além de impactos diretos em sua competitividade [S&P Global Ratings, 2021].

Para instituições financeiras, considerar variáveis ambientais na análise de crédito significa reduzir riscos de inadimplência decorrentes de passivos ambientais e alinhar portfólios de investimentos a políticas globais de mitigação das mudanças climáticas [FEBRABAN 2025].

2.1.2 Fator Social (Social)

O pilar social está relacionado às práticas empresariais no tratamento de colaboradores, clientes, fornecedores e comunidades; e inclui indicadores como condições de trabalho, diversidade e inclusão, respeito aos direitos humanos, engajamento comunitário e segurança dos produtos. Negligências nessas dimensões podem resultar em conflitos trabalhistas, boicotes de consumidores e dificuldades de acesso a mercados internacionais [SerasaExperian 2023].

No setor financeiro, incorporar aspectos sociais em modelos de crédito permite identificar riscos relacionados a práticas trabalhistas inadequadas ou violações de conformidade que possam comprometer a reputação da empresa analisada [PwCBrasil 2022].

2.1.3 Fator de Governança (Governance)

O fator de governança compreende estruturas de gestão, ética corporativa, transparência na divulgação de informações e mecanismos de controle interno. A ausência de boas práticas de governança pode levar a fraudes, má alocação de recursos e instabilidade organizacional [Deloitte 2021].

Segundo a Hackernoon (2024), a governança corporativa está diretamente relacionada à confiabilidade dos dados de ESG, pois empresas com maior transparência tendem a disponibilizar informações consistentes e auditáveis, fundamentais para análises de risco de crédito.

Assim, os três pilares ESG constituem uma estrutura integrada para avaliação do risco corporativo, que vai além do aspecto financeiro e possibilita maior robustez na análise de crédito realizada por instituições financeiras.

2.2 A Importância da Análise de Risco ESG no Mercado Financeiro

A integração dos fatores ESG na análise de crédito tem se tornado uma prática crescente no mercado financeiro, tanto em âmbito global quanto nacional, e tradicionalmente, os modelos de risco de crédito se baseavam em indicadores financeiros, como fluxo de caixa, endividamento e histórico de pagamento. Contudo, a inclusão de variáveis ambientais, sociais e de governança tem ampliado a capacidade preditiva desses modelos, permitindo identificar riscos não capturados por métricas convencionais [Deloitte 2021, United Nations Principles for Responsible Investment 2019].

Estudos apontam que empresas com baixa performance em critérios ESG estão mais suscetíveis a sofrer impactos financeiros negativos, seja por sanções regulatórias, ações judiciais, perda de reputação ou redução do acesso a capital [S&P Global Ratings, 2021]. Nesse sentido, investidores e instituições financeiras têm buscado incorporar sistematicamente essas dimensões na precificação de crédito e na gestão de portfólios.

No Brasil, a FEBRABAN destaca a relevância da sustentabilidade como elemento estratégico para o setor financeiro, estimulando a criação de políticas de crédito verde, taxonomias e guias de boas práticas. A ampliação do financiamento de projetos sustentáveis, como energias renováveis e agricultura de baixo carbono, é um reflexo da pressão crescente por alinhamento às agendas internacionais, como o Acordo de Paris [FEBRABAN 2019].

Além do viés regulatório, a análise de risco ESG também é percebida como diferencial competitivo, que de acordo com a empresa PwC Brasil, instituições financeiras que incorporam critérios ESG em seus modelos de crédito podem reduzir inadimplência, aumentar a atratividade para investidores e melhorar a imagem institucional perante clientes e a sociedade [PwCBrasil 2022].

Por outro lado, a ausência de mecanismos padronizados de mensuração ESG ainda representa um desafio relevante. A empresa Serasa Experian evidencia que, em segmentos como o crédito rural, a falta de análise ESG pode expor os concedentes a riscos financeiros significativos, inclusive multas bilionárias por descumprimento de conformidades ambientais, sociais e trabalhistas [SerasaExperian 2023]. Isso demonstra que a análise de risco ESG não deve ser vista apenas como uma prática opcional, mas como componente essencial da gestão de riscos financeiros.

Assim, a importância da análise de risco ESG no mercado financeiro reside em sua capacidade de tornar o processo de concessão de crédito mais robusto, transparente e alinhado às demandas de sustentabilidade, promovendo simultaneamente a estabilidade do sistema financeiro e a responsabilidade socioambiental.

2.3 ESG e Risco de Crédito: Abordagens e Desafios

A incorporação dos fatores ESG em modelos de crédito busca complementar a análise tradicional de risco, ampliando a visão sobre potenciais fontes de inadimplência e perda financeira. Enquanto a avaliação clássica considera variáveis como histórico de pagamentos, alavancagem financeira e indicadores contábeis, a perspectiva ESG introduz dimensões qualitativas e quantitativas relacionadas à sustentabilidade corporativa [United Nations Principles for Responsible Investment 2019, Deloitte 2021].

Entre as principais abordagens para integração ESG em risco de crédito, destacam-se:

- **Modelos de ratings internos**, nos quais instituições financeiras atribuem pontuações ESG às empresas com base em questionários, relatórios corporativos e indicadores públicos; embora essa metodologia seja flexível, depende fortemente da qualidade das informações fornecidas pelas empresas analisadas.

- **Integração de dados de terceiros**, utilizando bases de dados especializadas, como as disponibilizadas pela S&P Global ou MSCI. Essas soluções oferecem padronização e abrangência, mas implicam custos elevados e, por vezes, falta de transparência na metodologia de cálculo [S&P Global Ratings, 2021].

- **Modelos quantitativos baseados em dados públicos**, que buscam construir métricas ESG a partir de informações acessíveis em relatórios financeiros, balanços, notícias e bases regulatórias. Essa alternativa democratiza o acesso à análise ESG, mas enfrenta desafios relacionados à coleta, tratamento e padronização dos dados [Hackernoon 2024].

Apesar dos avanços, diversos desafios de padronização dos dados persistem, no qual podemos mencionar a **heterogeneidade das métricas ESG**: não existe uma regra global única para mensuração desses fatores, o que dificulta comparações entre setores e regiões [PwCBrasil 2022]. Outro obstáculo é a **escassez de dados públicos estruturados**, especialmente em mercados emergentes, onde muitas empresas não possuem obrigação regulatória de divulgar relatórios de sustentabilidade [FEBRABAN 2019].

Além disso, existe o risco de **greenwashing**, quando organizações divulgam informações incompletas ou distorcidas para aparentar maior comprometimento com práticas sustentáveis, comprometendo a confiabilidade das análises de risco [SerasaExperian 2023]. Por fim, do ponto de vista técnico, ainda há o desafio de integrar variáveis ESG em modelos estatísticos robustos, que consigam equilibrar aspectos qualitativos e quantitativos sem incorrer em vieses excessivos.

Assim, a integração ESG em risco de crédito representa tanto uma oportunidade de aprimorar a gestão financeira quanto um desafio metodológico e regulatório. O avanço

desse campo dependerá da padronização das métricas, da ampliação do acesso a dados confiáveis e do desenvolvimento de modelos analíticos capazes de traduzir fatores ESG em indicadores consistentes para a tomada de decisão.

2.4 Fontes de Dados Públicas para Análise Corporativa

A disponibilidade de informações padronizadas e acessíveis é um dos principais desafios na análise de risco ESG, e embora existam provedores especializados que oferecem bases consolidadas, o acesso a esses dados geralmente implica custos elevados, o que limita sua utilização por instituições de menor porte. Nesse sentido, o uso de **fontes de dados públicos** tem ganhado relevância ao democratizar a análise de sustentabilidade corporativa.

Entre as principais fontes de dados públicos utilizadas em estudos e práticas de análise ESG, destacam-se:

- **Relatórios financeiros e de sustentabilidade corporativa**: documentos publicados pelas próprias empresas, contendo informações sobre desempenho ambiental, social e de governança. Embora sejam fontes diretas, podem apresentar riscos de vieses ou práticas de *greenwashing*.
- **Órgãos reguladores e associações setoriais**: entidades como a Comissão de Valores Mobiliários (CVM) no Brasil e a Securities and Exchange Commission (SEC) nos Estados Unidos disponibilizam relatórios obrigatórios de companhias abertas, que podem ser usados para extrair indicadores ESG. No contexto nacional, a FEBRABAN tem promovido estudos e guias de boas práticas em sustentabilidade no setor financeiro [FEBRABAN 2019].
- **Agências internacionais de classificação**: a S&P Global Ratings disponibiliza indicadores de risco ESG que podem ser utilizados para avaliar empresas listadas em índices como o S&P 500 [S&P Global Ratings, 2021]. Esses dados permitem análises comparativas e integração a modelos de risco de crédito.
- **Bases públicas de mercado**: a biblioteca `yfinance`, em Python, permite a coleta automatizada de informações financeiras de empresas listadas, incluindo balanços, preços de ações e indicadores contábeis [Yahoo Finance, 2025]. Quando combinadas a ratings ESG públicos, essas informações podem compor bases robustas de análise.
- **Estudos institucionais e de consultorias**: relatórios produzidos por consultorias, como Deloitte e PwC, e entidades como a Serasa Experian, que divulgam periodi-

camente análises sobre risco ESG em segmentos específicos, como crédito rural e cooperativas de crédito [Deloitte 2021, PwCBrasil 2022, SerasaExperian 2023].

A utilização dessas fontes públicas representa um passo importante para ampliar o acesso a ferramentas de análise de risco ESG, entretanto, é necessário ressaltar que a heterogeneidade e a falta de padronização dos dados coletados impõem desafios adicionais ao seu processamento. Por isso, a integração de informações provenientes de diferentes origens deve ser acompanhada de técnicas de tratamento e pré-processamento, de forma a garantir consistência e confiabilidade dos indicadores utilizados na modelagem preditiva.

2.5 Fundamentos de Aprendizado de Máquina (Machine Learning)

O aprendizado de máquina (*machine learning*) corresponde a um ramo da inteligência artificial que desenvolve algoritmos capazes de aprender padrões a partir de dados e realizar previsões ou tomadas de decisão sem que sejam explicitamente programados para cada tarefa [Géron 2019]. Essa abordagem tem ganhado relevância em diferentes áreas, incluindo o setor financeiro, por permitir a construção de modelos mais flexíveis e robustos inclusive na análise de risco de crédito.

2.5.1 Aprendizado Supervisionado e Tarefas de Classificação

O aprendizado supervisionado é uma das abordagens mais utilizadas em *machine learning*, e para esse paradigma, o modelo é treinado a partir de um conjunto de dados rotulado, ou seja, composto por pares de entrada (variáveis explicativas) e saída (variável-alvo), onde o objetivo é o algoritmo aprender a relação entre essas variáveis de forma a generalizar para novos dados não vistos anteriormente [Géron 2019, McKinney 2017].

Entre as tarefas do aprendizado supervisionado, a classificação ocupa papel central quando o problema envolve a previsão de categorias discretas. Em contextos financeiros, por exemplo, a classificação é amplamente utilizada em modelos de *credit scoring*, na identificação de fraudes e na categorização de clientes segundo perfis de risco. Nesse tipo de problema, o modelo recebe como entrada atributos como indicadores financeiros, setoriais ou corporativos, e deve predizer a qual classe a observação pertence, como “baixo risco” ou “alto risco” [LeoBreiman 2001].

No caso específico deste trabalho, a tarefa de classificação é aplicada à análise de risco ESG, em que o objetivo é atribuir às empresas categorias relacionadas ao seu nível de risco socioambiental e de governança. A definição da variável-alvo ocorre a partir de ratings ESG públicos, que funcionam como rótulos supervisionados para o treinamento

dos modelos. A partir disso, diferentes algoritmos de classificação podem ser aplicados e comparados, como o Random Forest o modelo principal a ser investigado, e os modelos KNN e XGBoost, atuando como referências comparativas.

2.5.2 O Algoritmo Random Forest

O algoritmo Random Forest, proposto por Breiman (2001), é um método de aprendizado supervisionado baseado em *ensemble learning*, que combina múltiplas árvores de decisão com o objetivo de melhorar a acurácia e a capacidade de generalização do modelo. Sua principal característica consiste na construção de um conjunto (*forest*) de modelos base independentes, cujas previsões são agregadas por meio de votação majoritária (classificação) ou média (regressão).

Para compreender o funcionamento do Random Forest, é fundamental inicialmente entender o conceito de árvore de decisão, que é uma estrutura hierárquica composta por nós de decisão e folhas, onde cada nó interno representa um teste sobre uma variável de entrada, enquanto cada ramo corresponde ao resultado desse teste. As folhas representam a decisão final ou classe atribuída. O processo de construção da árvore envolve a seleção iterativa das variáveis que melhor separam os dados, geralmente utilizando métricas como índice de Gini ou entropia, com o objetivo de maximizar a pureza dos nós [Quinlan 1993].

Entretanto, árvores de decisão individuais tendem a apresentar alta variância, sendo sensíveis a pequenas variações nos dados de treinamento, o que pode levar ao sobreajuste (*overfitting*), e nesse contexto, o Random Forest surge como uma solução baseada na técnica de *bagging* (*bootstrap aggregating*), que reduz a variância ao combinar múltiplos modelos treinados de forma independente [LeoBreiman 2001].

O processo de construção do Random Forest, conforme ilustrado na Figura 2.2, pode ser descrito da seguinte forma:

1. Geração de múltiplos subconjuntos de dados a partir da amostra original, utilizando reamostragem com reposição (*bootstrap*);

2. Treinamento de uma árvore de decisão para cada subconjunto gerado;

3. Em cada divisão da árvore, seleção aleatória de um subconjunto de atributos (*feature randomness*), o que aumenta a diversidade entre as árvores;

4. Agregação das previsões individuais por meio de votação majoritária.

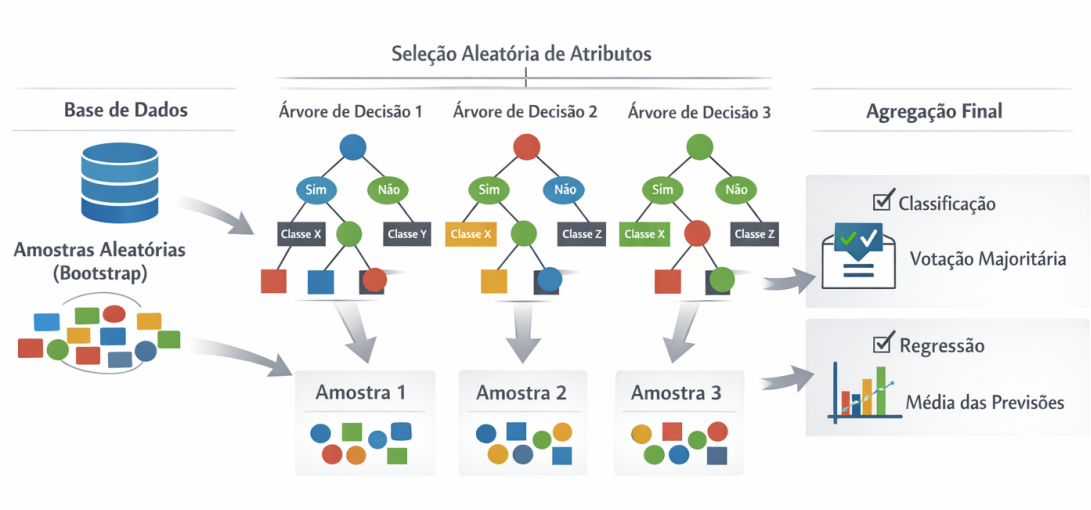


Figura 2.2: Ilustração do funcionamento do modelo Random Forest.

Essa combinação de amostragem aleatória e seleção aleatória de atributos permite que as árvores apresentem menor correlação entre si, o que contribui diretamente para a redução do erro do modelo.

O Random Forest é composto por diversas árvores de decisão que operam de forma paralela, sendo cada uma responsável por uma predição independente, cuja combinação resulta na decisão final do modelo.

Essa estratégia confere ao Random Forest vantagens significativas em relação a algoritmos baseados em árvore única, tais como:

- **Redução de sobreajuste** (*overfitting*), devido à diminuição da variância do modelo;
- **Capacidade de lidar com bases de dados complexas**, contendo variáveis numéricas e categóricas;
- **Estimativa da importância das variáveis**, permitindo identificar quais atributos mais influenciam a predição;
- **Robustez a ruídos e dados faltantes**, mantendo desempenho consistente mesmo com imperfeições nos dados.

No processo de desenvolvimento, a implementação do Random Forest pode ser potencializada pelo uso de técnicas de ajuste de hiperparâmetros. Neste trabalho, foi empregado um método de ajuste dos hiperparâmetros baseado na busca em grade, já implementada no pacote scikit-learn (método **GridSearchCV**), que realiza uma busca exaustiva sobre combinações de parâmetros pré-definidos, associada à validação cruzada. Essa abordagem permite selecionar automaticamente a configuração que otimiza métricas como F1-score, precisão e revocação [Géron 2019].

Entre os principais hiperparâmetros avaliados no Random Forest estão:

- número de árvores na floresta;
- profundidade máxima das árvores;
- número mínimo de amostras por nó folha;
- número de atributos considerados em cada divisão.

No contexto da análise de risco de crédito ESG, o Random Forest mostra-se particularmente adequado, pois:

1. Permite a integração de múltiplas dimensões (financeiras, ambientais, sociais e de governança) em um único modelo;
2. Suporta bases com grande número de atributos, característica comum em dados financeiros e indicadores ESG;
3. Fornece métricas de importância das variáveis, fundamentais para interpretação do impacto de fatores ESG no risco de crédito [Lundberg e Lee 2017].

Apesar de suas vantagens, o Random Forest apresenta algumas limitações, o modelo pode se tornar computacionalmente custoso em grandes volumes de dados e não fornece, de forma nativa, explicações detalhadas para cada decisão individual [LeoBreiman 2001, Géron 2019]. Dessa forma, técnicas complementares de interpretabilidade, como o método SHAP, tornam-se relevantes para análise mais aprofundada [Lundberg e Lee 2017]. Ainda assim, por meio do uso de validação cruzada e ajuste de hiperparâmetros, é possível garantir maior rigor metodológico e capacidade de generalização do modelo.

2.5.3 Modelos Comparativos de Classificação

Foram selecionados dois modelos de referência para efeito de comparação com Random Forest: o *K-Nearest Neighbors* (KNN) e o XGBoost (XGB). A escolha desses algoritmos se justifica por sua popularidade em problemas de classificação e por apresentarem características complementares ao Random Forest.

K-Nearest Neighbors (KNN)

O algoritmo K-Nearest Neighbors (KNN), proposto por Cover e Hart (1967), é um método supervisionado de classificação baseado na proximidade entre instâncias, em que sua lógica consiste em atribuir a uma nova observação a classe mais frequente entre seus k vizinhos mais próximos, de acordo com uma métrica de distância, sendo a distância Euclidiana uma das mais utilizadas.

Diferentemente de outros algoritmos de aprendizado de máquina, o KNN não realiza um processo explícito de treinamento, mas trata-se de um método do tipo *lazy learning* (aprendizado preguiçoso), no qual o modelo simplesmente armazena todos os dados de treinamento e adia a generalização para o momento da predição. Dessa forma, não há construção de uma função preditiva durante o treinamento, e toda a complexidade computacional é transferida para a fase de inferência [Cover e Hart 1967].

O funcionamento do algoritmo, conforme ilustrado na Figura 2.3, pode ser descrito pelas seguintes etapas:

1. Armazenamento do conjunto de dados de treinamento;
2. Cálculo da distância entre a nova observação e todas as instâncias do conjunto de dados;
3. Seleção dos k vizinhos mais próximos com base na menor distância;
4. Determinação da classe predominante entre os vizinhos selecionados;
5. Atribuição da classe resultante à nova observação.

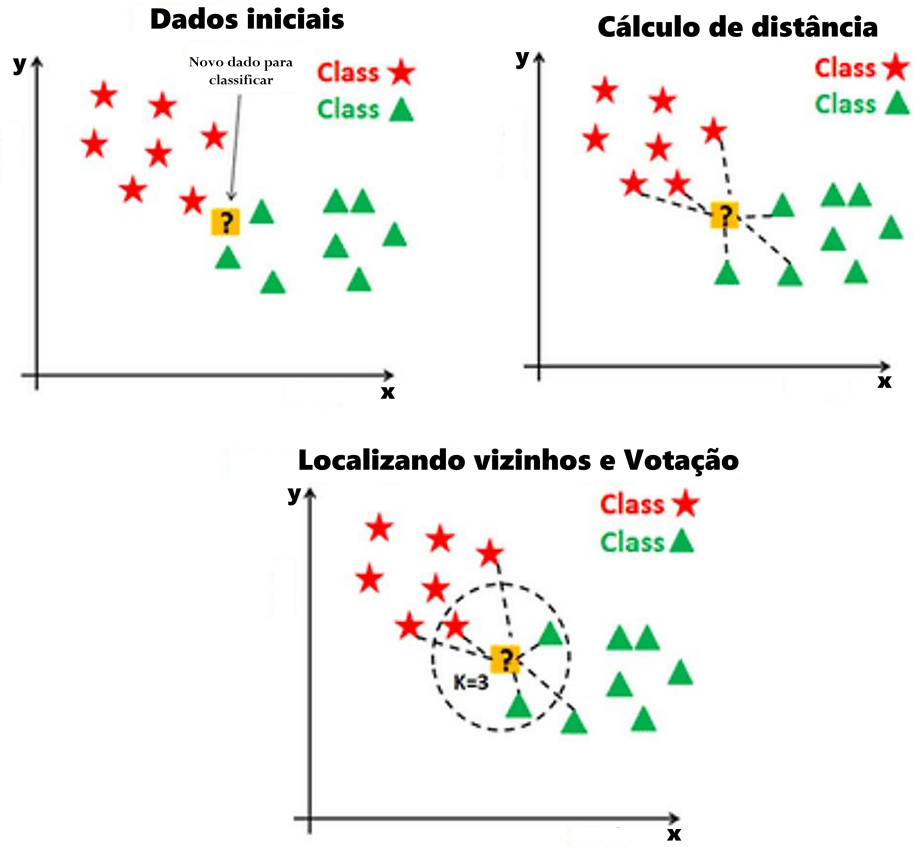


Figura 2.3: Representação ilustrativa do funcionamento do algoritmo K-Nearest Neighbors (KNN).

Fonte: Adaptado de Cover e Hart 1967.

A métrica de distância mais comum é a distância Euclidiana, definida como:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2.1}$$

onde x e y representam dois vetores no espaço de características.

O KNN classifica uma nova instância com base na proximidade espacial em relação aos dados existentes, formando regiões de decisão definidas pela densidade das classes no espaço de atributos.

A escolha do parâmetro k é um fator crítico para o desempenho do modelo. Valores baixos de k tornam o modelo mais sensível a ruídos, podendo levar ao sobreajuste, enquanto valores elevados tendem a suavizar as fronteiras de decisão, podendo resultar em subajuste. Dessa forma, a escolha de k deve ser realizada com base em validação empírica.

As principais características do KNN incluem:

- **Simplicidade**: é um algoritmo intuitivo e de fácil implementação, não requerendo treinamento explícito de parâmetros complexos;
- **Flexibilidade**: pode ser aplicado em diferentes contextos, desde reconhecimento de padrões até análise de risco de crédito;
- **Dependência da métrica de distância**: o desempenho está diretamente relacionado à escolha da métrica utilizada;
- **Sensibilidade à escala dos dados**: exige pré-processamento, como normalização, para evitar distorções causadas por variáveis em diferentes magnitudes.

Apesar de suas vantagens, o KNN apresenta limitações, como o alto custo computacional em bases de dados extensas, uma vez que requer o cálculo de distâncias para cada nova previsão. Além disso, sua performance pode ser prejudicada em cenários de alta dimensionalidade, fenômeno conhecido como *maldição da dimensionalidade* [Cover e Hart 1967].

XGBoost (XGB)

O XGBoost (Extreme Gradient Boosting), desenvolvido por Chen e Guestrin 2016, é um algoritmo baseado na técnica de *gradient boosting*, no qual múltiplos modelos fracos, geralmente árvores de decisão, são treinados de forma sequencial, de modo que cada novo modelo busca corrigir os erros cometidos pelos modelos anteriores. Essa abordagem permite a construção de modelos altamente preditivos, com elevada capacidade de generalização.

Diferentemente de métodos como o Random Forest, em que as árvores são treinadas de forma independente, o XGBoost constrói as árvores de forma sequencial. A cada iteração,

o modelo ajusta uma nova árvore aos resíduos (erros) do modelo anterior, utilizando o gradiente da função de perda para guiar o processo de aprendizado.

O processo de treinamento do XGBoost pode ser descrito pelas seguintes etapas:

1. Inicialização do modelo com uma predição base, geralmente a média das observações;
2. Cálculo do erro residual entre os valores reais e as predições atuais;
3. Treinamento de uma nova árvore de decisão para modelar esses resíduos;
4. Atualização da predição global, adicionando a contribuição da nova árvore ponderada por uma taxa de aprendizado (*learning rate*);
5. Repetição do processo até atingir um número definido de árvores ou critério de parada.

Matematicamente, o modelo pode ser representado como a soma de funções base:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \tag{2.2}$$

onde cada f_k representa uma árvore de decisão no espaço funcional $\mathcal{F}$.

Além disso, o XGBoost incorpora um termo de regularização na função objetivo, que penaliza a complexidade do modelo, contribuindo para evitar o sobreajuste:

$$\mathcal{L} = \sum_{i} l(y_i, \hat{y}_i) + \sum_{k} \Omega(f_k) \tag{2.3}$$

em que l representa a função de perda (por exemplo, log-loss) e Ω é o termo de regularização.

Conforme ilustrado na Figura 2.4, o XGBoost realiza um processo iterativo de refinamento das predições, no qual cada árvore corrige os erros da anterior, resultando em uma predição final obtida pela soma ponderada das árvores.

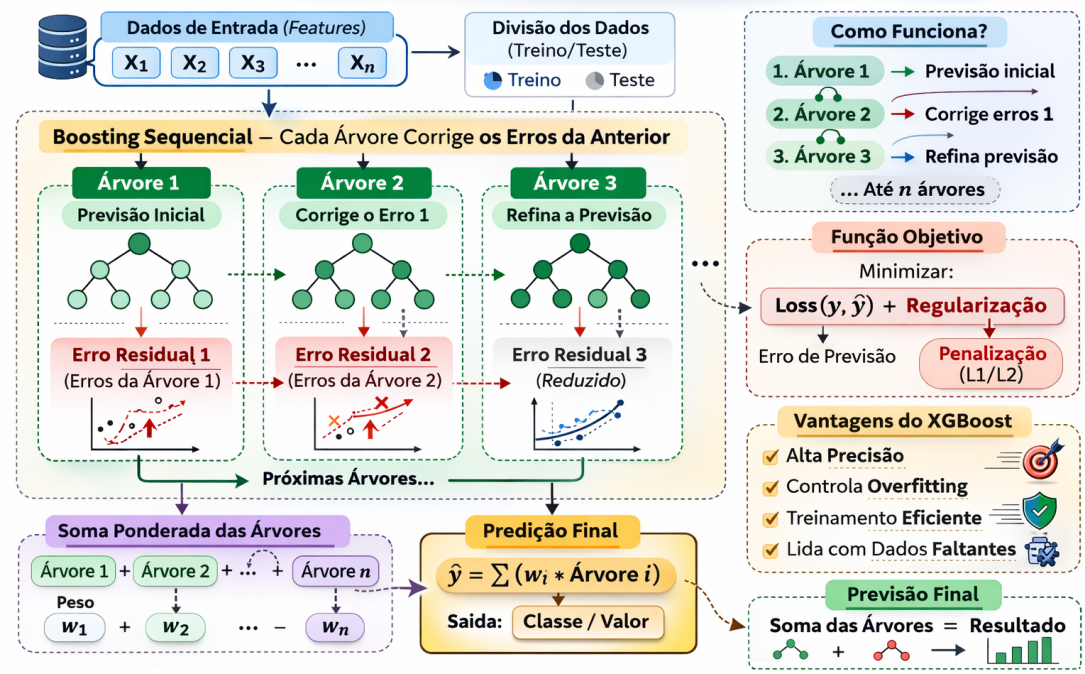


Figura 2.4: Infográfico do modelo XGBoost.

Fonte: Elaborado pelo autor, adaptado de Chen e Guestrin 2016.

Entre as principais vantagens do XGBoost, destacam-se:

- **Alta performance**: amplamente reconhecido por seu excelente desempenho em aplicações práticas e competições de ciência de dados;
- **Regularização integrada**: reduz o risco de sobreajuste por meio de penalizações na função objetivo;
- **Eficiência computacional**: utiliza técnicas de paralelização e otimização para acelerar o treinamento;
- **Flexibilidade**: suporta diferentes funções de perda e tarefas de classificação ou regressão.

Entre os principais hiperparâmetros do XGBoost, destacam-se:

- taxa de aprendizado;
- número de árvores;
- profundidade máxima das árvores;
- parâmetro de regularização.

Apesar de suas vantagens, o XGBoost demanda maior esforço de configuração e ajuste de hiperparâmetros em comparação ao Random Forest, o que pode torná-lo mais complexo

de aplicar em contextos práticos. No entanto, seu uso neste trabalho permite avaliar se os ganhos de desempenho justificam a maior complexidade metodológica, e no contexto da análise de risco de crédito ESG, o XGBoost apresenta potencial relevante por sua capacidade de capturar relações não lineares complexas entre variáveis financeiras e fatores ESG, contribuindo para modelos mais precisos e robustos.

2.5.4 SMOTE como Técnica de Balanceamento de Dados

Em muitos problemas de classificação, especialmente no contexto financeiro, ocorre o fenômeno do **desbalanceamento de classes**, no qual determinadas categorias apresentam muito mais observações do que outras, por exemplo, isso ocorre quando há um número elevado de empresas classificadas como "risco médio" e um número reduzido nas classes "alto risco" e "baixo risco". Nessas situações, modelos de aprendizado de máquina tendem a ser enviesados para a classe majoritária, comprometendo a capacidade de identificar corretamente casos menos frequentes, porém críticos.

Para contornar esse problema, diversas técnicas de balanceamento têm sido desenvolvidas. Entre elas, destaca-se o SMOTE, proposto por Nitesh Chawla, que é um método de **superamostragem sintética** que cria novas instâncias artificiais da classe minoritária a partir da interpolação entre exemplos reais existentes [Chawla et al. 2002].

Diferentemente da simples replicação de observações, essa técnica gera novos pontos ao longo do segmento de reta que conecta uma instância da classe minoritária a seus vizinhos mais próximos. Conforme ilustrado na Figura 2.5, o algoritmo seleciona um ponto da classe minoritária e gera novas amostras entre esse ponto e seus vizinhos, ampliando a representatividade da classe de forma mais distribuída no espaço de atributos.

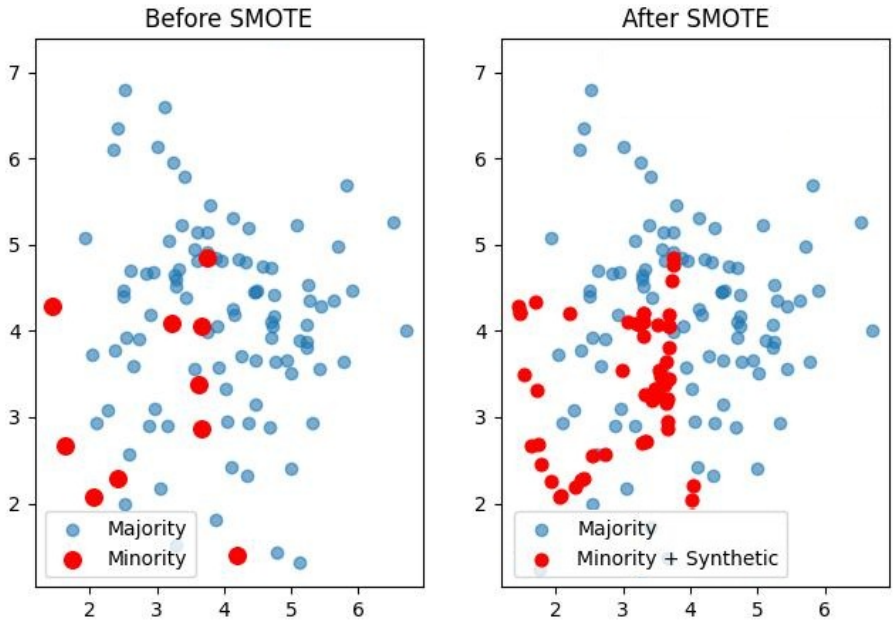


Figura 2.5: Representação ilustrativa do funcionamento do SMOTE, destacando a geração de amostras sintéticas.

Fonte: Adaptado de Chawla et al. 2002.

As principais vantagens do SMOTE incluem:

- **Melhora da capacidade de generalização** dos modelos em cenários de classes desbalanceadas;
- **Integração simples em pipelines** de aprendizado de máquina, podendo ser combinado com diferentes algoritmos de classificação;
- **Flexibilidade** para ajustar a proporção entre classes conforme necessário.

Contudo, o SMOTE também apresenta limitações. A criação de instâncias artificiais pode introduzir ruído em regiões onde as classes se sobrepõem, prejudicando a definição da fronteira de decisão. Além disso, sua eficácia pode ser reduzida em bases de alta dimensionalidade, nas quais a noção de distância entre pontos se torna menos representativa.

No contexto deste trabalho, a aplicação do SMOTE é particularmente relevante, pois a distribuição dos níveis de risco ESG tende a ser assimétrica, com maior concentração de empresas classificadas como risco intermediário. Dessa forma, o uso dessa técnica busca garantir que os modelos de classificação sejam capazes de identificar padrões relevantes também nas classes minoritárias, promovendo uma avaliação mais equilibrada e representativa do risco de crédito ESG.

2.5.5 Técnica para Tratamento de Dados Desbalanceados: StratifiedKFold

Além de técnicas de superamostragem, como o SMOTE, outra estratégia amplamente utilizada para lidar com bases de dados desbalanceadas é a validação cruzada estratificada, implementada pelo método `StratifiedKFold` da biblioteca `scikit-learn`. Essa técnica assegura que, em cada subdivisão do conjunto de dados, a proporção das classes seja preservada, evitando que determinadas partições apresentem distribuições distorcidas [Géron 2019].

O funcionamento da validação cruzada estratificada consiste em dividir os dados em k subconjuntos (ou *folds*), garantindo que cada um mantenha aproximadamente a mesma proporção de observações em cada classe presente no conjunto original. Conforme ilustrado na Figura 2.6, cada partição contém uma distribuição balanceada das classes, permitindo que o modelo seja treinado e avaliado de maneira mais consistente.

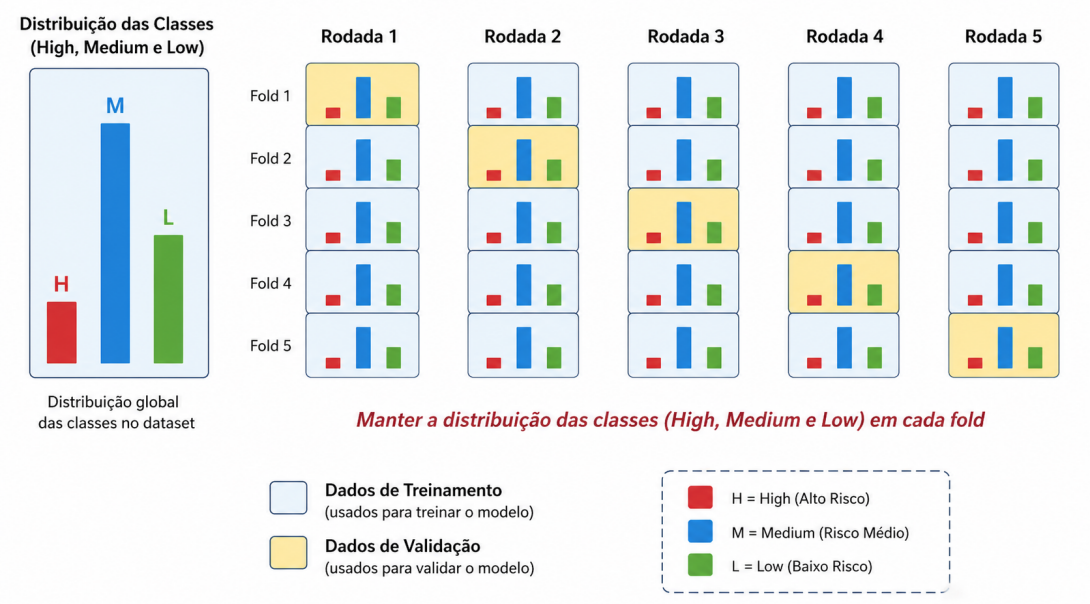


Figura 2.6: Esquema da validação cruzada estratificada (StratifiedKFold), evidenciando a preservação da proporção das classes em cada subconjunto de dados.

Fonte: Elaborado pelo autor, adaptado de Géron 2019.

Em seguida, o modelo é treinado k vezes, em cada iteração utilizando $k-1$ subconjuntos para treinamento e o subconjunto restante para teste. Ao final, as métricas de avaliação são calculadas como a média das iterações, fornecendo uma estimativa mais robusta do desempenho do modelo.

As principais vantagens do uso do `StratifiedKFold` incluem:

- **Preservação da proporção entre classes**, essencial em bases com forte desbalanceamento, como ocorre em ratings ESG;

- **Maior robustez estatística**, uma vez que o desempenho é avaliado em múltiplas divisões dos dados;
- **Integração simples** com outros métodos, como o `GridSearchCV`, permitindo calibração de hiperparâmetros de forma confiável.

No contexto da classificação de risco ESG, o uso do `StratifiedKFold` torna-se fundamental, pois garante que as classes de risco mais raras, como empresas com risco ESG elevado, estejam representadas em todas as fases de treinamento e validação. Dessa forma, reduz-se o risco de vieses na avaliação e assegura-se que os modelos sejam testados em cenários mais representativos da realidade.

2.5.6 Técnicas para Interpretação de Modelos: SHAP

A interpretabilidade é um dos aspectos mais relevantes em aplicações de aprendizado de máquina no setor financeiro, uma vez que decisões relacionadas à concessão de crédito e avaliação de risco devem ser transparentes e justificáveis. Modelos baseados em árvores, como Random Forest e XGBoost, embora apresentem elevada capacidade preditiva, são frequentemente classificados como *caixas-pretas*, dificultando a compreensão do impacto individual de cada variável na predição final.

Para contornar essa limitação, foi adotada neste trabalho a técnica SHAP (*SHapley Additive exPlanations*), proposta por Scott M. Lundberg e Su-In Lee, que é fundamentada na teoria dos jogos cooperativos, mais especificamente nos valores de Shapley, que representam uma forma de distribuir de maneira justa a contribuição de cada participante (neste caso, cada variável) para o resultado de uma função [Lundberg e Lee 2017].

No contexto de modelos de aprendizado de máquina, os valores de Shapley são utilizados para quantificar a contribuição de cada variável para a predição de um determinado exemplo. Essa contribuição é calculada considerando todas as possíveis combinações de variáveis, avaliando o impacto marginal da inclusão de cada atributo no modelo. Assim, o SHAP atribui a cada variável um valor que indica o quanto ela contribuiu positivamente ou negativamente para a predição.

Matematicamente, o valor de Shapley para uma variável i pode ser definido como:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[f(S \cup \{i\}) - f(S) \right] \tag{2.4}$$

onde S representa um subconjunto de variáveis, N é o conjunto total de atributos e $f(\cdot)$ é a função do modelo. Essa formulação garante que a contribuição de cada variável seja calculada de forma consistente e justa.

O funcionamento do SHAP pode ser descrito de forma simplificada pelas seguintes etapas:

1. Seleção de uma instância específica a ser interpretada;
2. Avaliação do modelo considerando diferentes subconjuntos de variáveis;
3. Cálculo da contribuição marginal de cada variável em relação à predição;
4. Agregação dessas contribuições para obter os valores de Shapley.

Entre as principais vantagens do SHAP, destaca-se, inicialmente, sua consistência teórica, uma vez que os valores de Shapley garantem uma distribuição justa da importância das variáveis, respeitando propriedades matemáticas relevantes, como eficiência e simetria. Além disso, o método oferece interpretabilidade tanto em nível global quanto local, permitindo analisar, de um lado, a contribuição média dos atributos em todo o conjunto de dados e, de outro, o impacto de cada variável em predições específicas. Outra vantagem importante é sua compatibilidade com modelos complexos, podendo ser aplicado em diferentes algoritmos de aprendizado supervisionado, incluindo modelos baseados em árvores, como Random Forest, bem como KNN e XGBoost, o que amplia sua utilidade em diferentes contextos de modelagem.

Além disso, para modelos baseados em árvores, existem implementações otimizadas do SHAP (*TreeSHAP*), que permitem o cálculo eficiente dos valores de Shapley sem a necessidade de avaliar todas as combinações possíveis de variáveis, reduzindo significativamente o custo computacional [Lundberg e Lee 2017].

No contexto da análise de risco ESG, o SHAP possibilita identificar quais fatores ambientais, sociais e de governança exercem maior influência sobre a classificação de uma empresa em determinado nível de risco. Dessa forma, torna-se possível compreender não apenas *o que* o modelo prediz, mas também *por que* determinada decisão foi tomada, promovendo maior transparência, interpretabilidade e confiabilidade no uso de modelos de aprendizado de máquina em instituições financeiras.

2.5.7 Métricas de Avaliação de Modelos de Classificação

A etapa de avaliação de desempenho é fundamental em projetos de aprendizado de máquina, pois permite verificar a qualidade preditiva dos modelos e compará-los de forma sistemática. No contexto do risco de crédito ESG, em que existem classes desbalanceadas e diferentes custos associados a erros de classificação, a escolha de métricas adequadas torna-se ainda mais relevante [Saito e Rehmsmeier 2015].

As principais métricas utilizadas neste trabalho são apresentadas a seguir:

- **Acurácia** (*accuracy*): representa a proporção de previsões corretas em relação ao total de instâncias avaliadas. É definida como:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.5}$$

onde TP (verdadeiros positivos), TN (verdadeiros negativos), FP (falsos positivos) e FN (falsos negativos) correspondem aos elementos da matriz de confusão. Embora intuitiva, essa métrica pode ser enganosa em bases desbalanceadas, pois um modelo pode apresentar alta acurácia ao privilegiar a classe majoritária.

- **Acurácia Balanceada** (Balanced Accuracy): corresponde à média do desempenho obtido em cada classe, atribuindo igual importância a todas elas, independentemente de seu tamanho.Pode ser definida como:

$$\text{Balanced Accuracy} = \frac{1}{C} \sum_{i=1}^{C} \text{Recall}_i \tag{2.6}$$

em que:

  - C representa o número de classes;
  - Recall_i representa a revocação da classe i.

No caso de problemas binários, a acurácia balanceada pode ser expressa como a média entre a sensibilidade e a especificidade:

$$\text{Acurácia Balanceada} = \frac{\text{Sensibilidade} + \text{Especificidade}}{2} \tag{2.7}$$

Essa métrica é especialmente útil em bases de dados desbalanceadas, pois evita que o desempenho do modelo seja superestimado devido à predominância de classes majoritárias, fornecendo uma avaliação mais representativa da capacidade de discriminação do classificador.

- **Precisão** (*precision*): mede a proporção de instâncias corretamente classificadas como positivas em relação ao total de instâncias previstas como positivas:

$$\text{Precisão} = \frac{TP}{TP + FP} \tag{2.8}$$

Essa métrica é particularmente relevante quando falsos positivos são indesejados, como em situações em que uma empresa é classificada incorretamente como de baixo risco.

- **Revocação** (*recall*): indica a proporção de instâncias positivas corretamente identificadas em relação ao total de instâncias positivas reais:

$$\text{Revocação} = \frac{TP}{TP + FN} \tag{2.9}$$

Essa métrica é essencial em cenários onde falsos negativos são críticos, como a não identificação de empresas com alto risco ESG.

- **Medida F1** (*F1-score*): corresponde à média harmônica entre precisão e revocação, equilibrando ambas as métricas:

$$F1 = 2 \cdot \frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \tag{2.10}$$

O F1-score é amplamente utilizado em bases desbalanceadas, pois considera simultaneamente os erros do tipo falso positivo e falso negativo.

- **Matriz de confusão**: é uma representação tabular que descreve o desempenho do modelo ao comparar os valores previstos com os valores reais. Seus elementos fundamentais (TP, TN, FP, FN) permitem analisar detalhadamente os tipos de erro cometidos pelo modelo, sendo base para o cálculo das demais métricas.

- **Área sob a curva ROC (AUC-ROC)**: mede a capacidade do modelo em distinguir entre classes, sendo definida como a área sob a curva que relaciona a taxa de verdadeiros positivos (*True Positive Rate*) e a taxa de falsos positivos (*False Positive Rate*):

$$\text{TPR} = \frac{TP}{TP + FN}, \quad \text{FPR} = \frac{FP}{FP + TN} \tag{2.11}$$

Embora amplamente utilizada, a AUC-ROC pode apresentar resultados otimistas em bases desbalanceadas.

- **Área sob a curva Precision-Recall (AUC-PR)**: avalia o desempenho do modelo considerando a relação entre precisão e revocação, sendo especialmente indicada em cenários com forte desbalanceamento de classes [Saito e Rehmsmeier 2015]. Diferentemente da AUC-ROC, essa métrica foca diretamente na performance da classe minoritária, tornando-a mais apropriada para problemas de risco.

- **Coeficiente de Correlação de Matthews (Matthews Correlation Coefficient - MCC)**: mede a qualidade da classificação ao considerar simultaneamente verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, sendo particularmente útil em bases desbalanceadas.

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{2.12}$$

  O MCC varia entre -1 e 1, em que 1 representa uma classificação perfeita, 0 indica desempenho equivalente ao acaso e valores negativos indicam discordância sistemática entre previsões e valores reais.

  Essa métrica é especialmente relevante em problemas com desbalanceamento entre classes, pois fornece uma avaliação mais robusta do desempenho global do modelo.

- **Coeficiente Kappa (Cohen's Kappa - κ)**: mede o grau de concordância entre as classificações previstas pelo modelo e os valores reais, descontando a concordância esperada ao acaso.

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{2.13}$$

  em que:

  - P_o representa a proporção de concordância observada;
  - P_e representa a proporção de concordância esperada ao acaso.

  O coeficiente Kappa varia entre -1 e 1, em que valores próximos de 1 indicam alta concordância, valores próximos de 0 indicam concordância equivalente ao acaso e valores negativos sugerem discordância entre as classificações.

  Essa métrica é especialmente útil para avaliar classificadores em cenários multiclasse e com desbalanceamento de classes, pois fornece uma medida mais robusta de concordância além da acurácia simples.

Segundo Saito e Rehmsmeier (2015), a curva Precision-Recall tende a ser mais informativa que a curva ROC em problemas de classes desbalanceadas, como é o caso do risco ESG. Por essa razão, este trabalho prioriza métricas como F1-score e AUC-PR na comparação entre Random Forest, KNN e XGBoost, assegurando uma avaliação mais realista e robusta do desempenho dos modelos.

Capítulo 3

Metodologia de Desenvolvimento

3.1 Abordagem Metodológica

A metodologia adotada neste trabalho possui caráter **quantitativo** e **exploratório-aplicado**, uma vez que busca não apenas compreender, mas também desenvolver um modelo computacional de classificação de risco ESG, onde o objetivo central é avaliar a relevância da incorporação de fatores ambientais, sociais e de governança em modelos de crédito por meio da aplicação de algoritmos de aprendizado de máquina sobre dados públicos.

O desenvolvimento da pesquisa foi estruturado em etapas sequenciais, configurando um **pipeline de ciência de dados** que abrange desde a coleta de informações até a análise de resultados. As principais fases podem ser sintetizadas da seguinte forma:

1. **Coleta de dados**: utilização de bases públicas para obtenção de informações financeiras e de sustentabilidade corporativa. Os dados financeiros foram extraídos por meio da biblioteca `yfinance`, com foco em empresas listadas no índice S&P 500, enquanto os indicadores de risco ESG foram obtidos a partir de relatórios públicos e bases complementares.

2. **Integração e pré-processamento**: unificação das bases coletadas em um único *dataset*, com padronização de atributos, tratamento de valores ausentes e preparação da variável-alvo (*ESG Risk Level*), utilizada como rótulo supervisionado.

3. **Modelagem preditiva**: aplicação do algoritmo Random Forest, em função de sua robustez, interpretabilidade relativa e capacidade de lidar com variáveis heterogêneas. Além disso, implementou-se uma análise comparativa com outros dois algoritmos, K-Nearest Neighbors (KNN) e XGBoost (XGB), visando validar a escolha do Random Forest.

4. **Tratamento de desbalanceamento**: utilização da técnica SMOTE (*Synthetic Minority Over-sampling Technique*) para mitigar a assimetria na distribuição das classes de risco ESG, garantindo melhor desempenho na classificação das classes minoritárias.

5. **Validação e avaliação**: adoção de validação cruzada estratificada e uso de métricas como acurácia, precisão, *recall*, F1-score, matriz de confusão, AUC-ROC e AUC-PR, de modo a garantir uma avaliação abrangente da performance dos modelos.

6. **Interpretação dos resultados**: análise da importância dos atributos (*feature importance*) com base no modelo Random Forest, permitindo identificar quais fatores financeiros e de sustentabilidade exercem maior impacto na classificação de risco ESG.

Essa abordagem metodológica possibilita avaliar de forma prática a viabilidade da utilização de dados públicos para análise de risco ESG, ao mesmo tempo em que garante rigor científico na construção e comparação dos modelos. O caráter exploratório se manifesta na busca por compreender o comportamento dos dados e a eficácia de diferentes algoritmos, enquanto o caráter aplicado está presente no desenvolvimento de um protótipo de sistema de classificação com potencial de uso em instituições financeiras.

3.2 Ambiente e Ferramentas Utilizadas

O desenvolvimento deste trabalho foi realizado em ambiente computacional baseado na linguagem de programação `Python` (versão 3.x), devido à sua ampla adoção em projetos de ciência de dados e aprendizado de máquina, bem como pela disponibilidade de bibliotecas especializadas para tratamento, modelagem e visualização de dados.

As principais ferramentas e bibliotecas utilizadas foram:

- **Pandas** [McKinney 2017]: utilizada para manipulação e tratamento de dados, possibilitando operações de leitura, limpeza, integração e transformação de tabelas.

- **NumPy**: empregada para operações numéricas e manipulação de matrizes, fornecendo suporte para os cálculos envolvidos na preparação e modelagem dos dados.

- **Matplotlib** e **Seaborn** [Hunter et al. 2025, Waskom 2025]: responsáveis pela construção de gráficos e visualizações, permitindo explorar a distribuição dos dados e interpretar resultados dos modelos.

- **Scikit-learn** [Géron 2019]: biblioteca central para a implementação de algoritmos de aprendizado supervisionado, incluindo Random Forest e KNN. Também foi utilizada

para a execução de técnicas de validação cruzada (`StratifiedKFold`), ajuste de hiperparâmetros (`GridSearchCV`) e cálculo de métricas de avaliação.

- **Imbalanced-learn** [Chawla et al. 2002]: utilizada para aplicar a técnica SMOTE (*Synthetic Minority Over-sampling Technique*), essencial para corrigir o desbalanceamento entre classes da variável-alvo.
- **XGBoost** [Chen e Guestrin 2016]: biblioteca específica para implementação do algoritmo XGBoost, explorado como modelo comparativo ao Random Forest.
- **SHAP** [Lundberg e Lee 2017]: pacote responsável pela interpretação dos modelos preditivos, permitindo compreender a contribuição de cada variável nas predições.
- **yfinance** [Aroussi 2025]: utilizada na coleta de dados financeiros de empresas listadas, automatizando o acesso a informações de mercado diretamente do Yahoo Finance.

Os experimentos foram executados em ambiente local, com sistema operacional baseado em `Windows 10`, processador Intel i7, 16 GB de memória RAM e GPU integrada. Para gerenciamento de código, foram utilizadas interfaces interativas de desenvolvimento em `Jupyter Notebook` e `VS Code`.

Esse conjunto de ferramentas possibilitou a implementação de um fluxo completo, desde a coleta e tratamento de dados até a modelagem, avaliação e interpretação dos resultados, garantindo reprodutibilidade e eficiência no desenvolvimento da pesquisa.

3.3 Processo de Desenvolvimento (Pipeline)

O processo de desenvolvimento deste trabalho foi estruturado como um pipeline de ciência de dados, organizado em etapas sequenciais e interdependentes, que abrangem desde a coleta de dados até a geração dos resultados finais, e essa abordagem permite maior reprodutibilidade, clareza metodológica e integração entre as diferentes fases do projeto, sendo amplamente adotada em aplicações de aprendizado de máquina e análise preditiva.

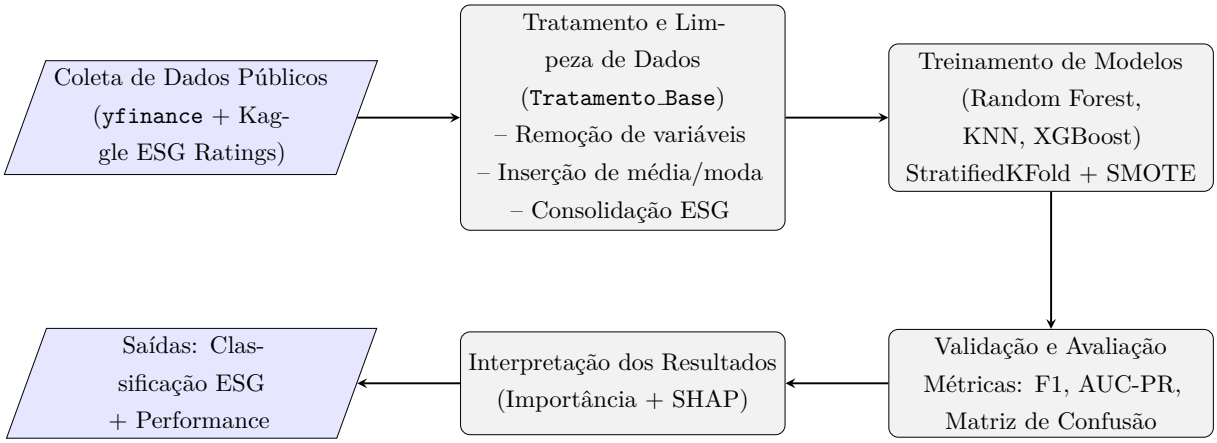


Figura 3.1: Fluxograma do processo de desenvolvimento do modelo de classificação de risco ESG.

Conforme ilustrado na Figura 3.1, o pipeline inicia-se com a coleta de dados públicos provenientes das fontes Yahoo Finance e bases de ratings ESG, seguido pelas etapas de tratamento e preparação dos dados, nas quais são realizadas operações como remoção de variáveis irrelevantes, inserção de valores ausentes e consolidação das classes de risco ESG. Essas etapas são fundamentais para garantir a qualidade dos dados de entrada, fator crítico para o desempenho de modelos de classificação.

Na sequência, ocorre a etapa de treinamento dos modelos, na qual são aplicados algoritmos supervisionados, como Random Forest, K-Nearest Neighbors e XGBoost. Durante essa fase, são incorporadas técnicas como validação cruzada estratificada e estratégias de balanceamento de classes, com o objetivo de reduzir vieses e melhorar a capacidade de generalização dos modelos.

Posteriormente, os modelos são submetidos à etapa de validação e avaliação, utilizando métricas apropriadas para problemas de classificação, como F1-score, AUC-PR e matriz de confusão, permitindo uma análise mais robusta do desempenho preditivo, especialmente em cenários com desbalanceamento de classes.

Por fim, os resultados são interpretados por meio da análise de importância de atributos e técnicas de explicabilidade, possibilitando compreender o impacto das variáveis no processo decisório do modelo. Essa etapa é especialmente relevante no contexto de risco ESG, em que a transparência dos modelos contribui para maior confiabilidade na tomada de decisão.

3.3.1 Coleta de Dados via API (Yahoo Finance)

A coleta de dados financeiros foi realizada utilizando a biblioteca yfinance em Python, que permite a extração estruturada de informações públicas de empresas listadas em mercados financeiros. O universo amostral foi composto pelas empresas pertencentes ao índice S&P 500, cuja escolha se justifica pela ampla disponibilidade de dados financeiros

padronizados e pela presença consolidada dessas organizações em avaliações de risco ESG.

Inicialmente, foi definida uma lista contendo os tickers das empresas do S&P 500, os quais foram previamente padronizados a fim de evitar inconsistências durante as etapas subsequentes de integração com a base de dados ESG. A partir dessa lista, foi realizada a extração automatizada das informações por meio da API do yfinance, utilizando, para cada empresa, o objeto .info, que reúne um conjunto abrangente de atributos financeiros, operacionais e de mercado.

Durante esse processo, foi adotado um critério de seleção das variáveis baseado em três aspectos principais: (i) disponibilidade consistente entre as empresas analisadas, (ii) baixa taxa de valores ausentes e (iii) relevância econômico-financeira para a caracterização das organizações. Com base nesses critérios, foram priorizados indicadores relacionados à estrutura de capital, desempenho operacional e percepção de mercado.

Dentre as variáveis coletadas, destacam-se indicadores de mercado, como marketCap, beta e enterpriseValue, métricas financeiras como totalRevenue, ebitda e grossMargins, bem como indicadores operacionais, incluindo profitMargins, returnOnAssets e returnOnEquity. Esses atributos permitem representar diferentes dimensões do desempenho corporativo, servindo como base para a modelagem do risco ESG.

Os dados coletados foram organizados em formato tabular (DataFrame), no qual cada linha representa uma empresa e cada coluna corresponde a um atributo. Essa estrutura facilita a manipulação, o tratamento e a posterior aplicação de algoritmos de aprendizado supervisionado. Adicionalmente, foi realizado um tratamento inicial de inconsistências, com a remoção de registros que apresentavam ausência completa de informações relevantes, assegurando maior qualidade e confiabilidade à base de dados final.

Essa etapa constitui o fundamento informacional do estudo, permitindo representar quantitativamente o perfil econômico-financeiro das empresas e viabilizando a integração com os indicadores de risco ESG nas etapas subsequentes.

3.3.2 Integração com Ratings ESG

Os dados financeiros coletados foram integrados a uma base complementar contendo informações de risco ESG (Environmental, Social and Governance), disponibilizada publicamente na plataforma Kaggle sob o título “S&P 500 ESG Risk Ratings”. Essa base reúne indicadores associados ao risco socioambiental e de governança das empresas, incluindo variáveis como a pontuação total de risco ESG (Total ESG Risk Score), os componentes individuais de risco ambiental, social e de governança (Environment, Social, Governance Risk Scores) e o nível agregado de risco (ESG Risk Level).

A integração entre as bases foi realizada utilizando o identificador comum ticker, que representa o código de negociação das empresas no mercado financeiro. Inicialmente, foi

realizada a padronização dos identificadores, convertendo todos os tickers para um formato uniforme (string em letras maiúsculas), além da remoção de inconsistências como sufixos ou variações de nomenclatura, garantindo a correta correspondência entre os registros das duas bases.

Em seguida, foi realizada a junção dos conjuntos de dados por meio de uma operação do tipo inner join, utilizando o campo ticker como chave primária. Essa estratégia assegura que apenas empresas com informações disponíveis tanto na base financeira quanto na base ESG sejam mantidas no conjunto final, evitando a introdução de registros incompletos ou inconsistentes.

Após a integração, foram conduzidas etapas de validação dos dados, incluindo a verificação de duplicidade de registros e a remoção de entradas inconsistentes, com o objetivo de garantir a integridade e a confiabilidade do dataset consolidado.

A variável-alvo utilizada nos modelos de classificação, denominada *esg_risk_level*, foi definida a partir da coluna ESG Risk Level presente na base ESG. Essa variável categoriza as empresas em diferentes níveis de risco, como baixo, médio e alto, permitindo estruturar o problema como uma tarefa de aprendizado supervisionado.

Dessa forma, a integração das bases viabiliza a construção de um *dataset* unificado que relaciona características econômico-financeiras às dimensões de risco ESG, permitindo que os modelos aprendam padrões associados à classificação do risco a partir de dados públicos estruturados.

3.3.3 Pré-processamento e Engenharia de Atributos

O tratamento inicial dos dados teve como objetivo preparar as variáveis para os modelos de aprendizado supervisionado, garantindo consistência, qualidade e adequação estatística da base. Essa etapa incluiu a exclusão de atributos redundantes, o tratamento de valores ausentes, a padronização das variáveis e a consolidação da variável-alvo ESG.

Durante o processo de limpeza, as variáveis foram removidas de forma sistemática com base em critérios metodológicos voltados à redução de viés e à melhoria da capacidade de generalização do modelo. Inicialmente, foram excluídos identificadores únicos, como symbol e uuid, por apresentarem risco de memorização e potencial vazamento de informação, uma vez que poderiam permitir ao modelo diferenciar empresas individualmente em vez de aprender padrões estruturais.

Adicionalmente, foram removidos metadados administrativos e de contato, como zip, address, phone e website, por não possuírem relação causal com o risco ESG e por poderem introduzir vieses geográficos irrelevantes. Variáveis textuais e de alta cardinalidade, como descrições de negócios e estruturas organizacionais, também foram descartadas, pois exigiriam técnicas específicas de processamento de linguagem natural e poderiam aumentar

a complexidade do modelo sem ganhos proporcionais de desempenho.

Outro grupo relevante corresponde às variáveis de mercado intradiário e de tempo real, como currentPrice, open e previousClose, que foram excluídas devido ao risco de vazamento temporal em relação ao período de referência do rating ESG. Da mesma forma, indicadores derivados de opiniões de analistas, como recommendationMean e targetMeanPrice, foram removidos por refletirem avaliações exógenas potencialmente correlacionadas com o próprio risco ESG, o que poderia introduzir circularidade no modelo.

Também foram descartadas variáveis associadas a eventos corporativos com marca temporal, como datas de dividendos e chamadas de resultados, devido ao possível desalinhamento temporal com os indicadores ESG. Variáveis relacionadas à configuração de mercado, como exchange e region, bem como informações de setor e indústria, foram removidas com o objetivo de evitar que o modelo aprendesse atalhos estruturais (por exemplo, associação direta entre setor e nível de risco), comprometendo sua capacidade de generalização.

Além disso, variáveis com baixa variabilidade ou alta taxa de valores ausentes (superior a 70%) foram eliminadas, assim como campos considerados artefatos do processo de coleta de dados. Esse conjunto de decisões permitiu reduzir significativamente a dimensionalidade da base, mantendo apenas variáveis com potencial preditivo relevante.

No que se refere ao tratamento de valores ausentes, foi adotada uma estratégia de inserção simples (single imputation), na qual variáveis numéricas foram preenchidas pela média amostral e variáveis categóricas pela moda. Essa abordagem foi escolhida por sua simplicidade, reprodutibilidade e compatibilidade com diferentes algoritmos, permitindo manter o tamanho amostral e evitar a introdução de ruídos adicionais nos dados.

Para o modelo K-Nearest Neighbors (KNN), foi implementado com o scikit learn pelo método StandardScaler a transformação dos dados para média zero e desvio padrão unitário. Essa etapa é essencial para algoritmos baseados em distância, garantindo que variáveis em diferentes escalas não distorçam o cálculo de proximidade entre observações.

Por fim, foi realizada a engenharia da variável-alvo. A variável original ESG Risk Level, composta por cinco categorias (Negligible, Low, Medium, High e Severe), foi consolidada em três níveis: Low, Medium e High. As classes Negligible e Low foram agrupadas em Low, enquanto Severe foi incorporada à classe High. Essa transformação teve como objetivo reduzir o desbalanceamento, aumentar o suporte amostral das classes extremas e melhorar a estabilidade estatística das métricas de avaliação.

Em conjunto, essas etapas resultaram em uma base de dados mais limpa, consistente e adequada à modelagem supervisionada, permitindo que os algoritmos capturem padrões estruturais relevantes entre características financeiras e o risco ESG, minimizando a influência de ruídos, redundâncias e vieses.

3.3.4 Modelagem e Treinamento dos Classificadores

A etapa de modelagem teve como objetivo treinar e avaliar diferentes algoritmos de classificação supervisionada capazes de predizer o nível de risco ESG a partir das variáveis financeiras e de mercado previamente tratadas. Para isso, foram selecionados três modelos com características complementares: *Random Forest*, *K-Nearest Neighbors* (KNN) e *XGBoost*, implementados, respectivamente, nos scripts `ML_RF`, `ML_KNN` e `ML_XGB`.

O modelo *Random Forest* foi escolhido como principal abordagem devido à sua robustez, capacidade de lidar com variáveis heterogêneas e resistência a *outliers*, além de fornecer medidas internas de importância dos atributos, o que favorece a interpretabilidade dos resultados. O modelo foi configurado inicialmente com `número de estimadores = 100`, `profundidade máxima não limitada`, `mínimo de amostras para divisão = 2` e `mínimo de amostras por folha = 1`, sendo posteriormente submetido a ajuste de hiperparâmetros por meio do *GridSearchCV*, com variações principalmente no número de árvores e na profundidade máxima. Essa estratégia permitiu encontrar uma configuração mais adequada ao problema, equilibrando capacidade preditiva e risco de sobreajuste.

O modelo *K-Nearest Neighbors* (KNN) foi utilizado como *baseline*, representando uma abordagem baseada em proximidade no espaço de atributos. O algoritmo foi configurado com `número de vizinhos = 5` e métrica de distância euclidiana (`metric = euclidean`). Sua inclusão no estudo permite avaliar o desempenho de um classificador mais simples e diretamente dependente da estrutura geométrica dos dados, sendo particularmente sensível à escala das variáveis, o que justifica a etapa prévia de normalização descrita na Seção 3.3.3.

Por sua vez, o modelo *XGBoost* foi utilizado com uma abordagem baseada em *gradient boosting*, amplamente reconhecida por sua elevada performance em problemas de classificação tabular. Após o processo de otimização por *GridSearchCV*, o modelo apresentou como melhores hiperparâmetros: número de estimadores igual a 300, taxa de aprendizagem igual a 0,05, profundidade máxima igual a 6, proporção de amostragem igual a 0,8, proporção de amostragem de colunas igual a 1,0, regularização L1 igual a 0,0 e regularização L2 igual a 5,0.

Essa configuração indica uma estratégia de aprendizado mais gradual, com maior número de árvores e menor taxa de aprendizagem, favorecendo a generalização do modelo. Além disso, a profundidade máxima igual a 6 permite capturar relações não lineares entre as variáveis, enquanto a amostragem parcial das observações e a regularização L2 contribuem para reduzir o risco de sobreajuste.

O processo de treinamento foi conduzido utilizando validação cruzada estratificada (*StratifiedKFold*), garantindo que a proporção das classes fosse preservada em cada subdivisão do conjunto de dados. Foi adotado um esquema com $k = 5$ *folds*, no qual, a cada

iteração, o modelo é treinado em uma partição dos dados e validado em outra, permitindo uma avaliação mais robusta e menos sensível a variações específicas da amostra.

Para lidar com o desbalanceamento entre as classes de risco ESG, foi aplicada a técnica SMOTE (*Synthetic Minority Over-sampling Technique*) exclusivamente nos dados de treinamento de cada *fold*. Essa abordagem evita vazamento de informação para o conjunto de validação e permite gerar amostras sintéticas da classe minoritária, equilibrando a distribuição das classes durante o processo de aprendizado.

O treinamento foi realizado de forma iterativa ao longo dos *folds*, sendo que, em cada iteração, o modelo era ajustado nos dados de treino (após aplicação do SMOTE) e avaliado nos dados de validação. Ao final do processo, os resultados foram agregados, permitindo estimar o desempenho médio dos modelos e sua variabilidade.

Essa estratégia de modelagem e treinamento assegura não apenas a comparabilidade entre os algoritmos, mas também a robustez estatística dos resultados, garantindo que as conclusões obtidas não sejam dependentes de uma única divisão dos dados, mas sim representativas do comportamento geral dos modelos frente ao problema de classificação de risco ESG.

3.3.5 Validação e Avaliação de Performance

A etapa de validação e avaliação de performance teve como objetivo mensurar de forma abrangente o desempenho dos modelos de classificação, considerando não apenas a capacidade global de acerto, mas também o comportamento específico em cada classe de risco ESG. Dada a natureza desbalanceada do problema, essa etapa foi estruturada de modo a evitar vieses de avaliação e garantir maior robustez estatística dos resultados.

Para mitigar o desbalanceamento entre as classes, foi utilizada a técnica SMOTE (*Synthetic Minority Over-sampling Technique*), aplicada exclusivamente aos dados de treinamento em cada iteração da validação cruzada. Essa abordagem permite gerar instâncias sintéticas da classe minoritária, aumentando seu suporte amostral e favorecendo o aprendizado de padrões associados a essa classe, sem introduzir vazamento de informação para os dados de validação.

A avaliação dos modelos foi conduzida com base em um conjunto de métricas complementares, selecionadas de forma a capturar diferentes dimensões do desempenho preditivo. A acurácia foi utilizada como medida global de desempenho, representando a proporção total de classificações corretas. No entanto, considerando o desbalanceamento das classes, métricas adicionais foram incorporadas para uma análise mais detalhada.

A precisão foi utilizada para avaliar a confiabilidade das previsões positivas, ou seja, a proporção de classificações corretas dentre aquelas atribuídas a uma determinada classe, sendo particularmente relevante para o controle de falsos positivos. O *recall*, por sua vez,

mede a capacidade do modelo de identificar corretamente as instâncias reais de cada classe, sendo essencial para avaliar a detecção de casos relevantes, especialmente na classe de alto risco.

O F1-score foi empregado como uma medida de equilíbrio entre precisão e *recall*, sintetizando o desempenho do modelo em cenários nos quais há necessidade de conciliar cobertura e confiabilidade. Complementarmente, a matriz de confusão foi utilizada para analisar de forma detalhada os padrões de erro, permitindo identificar quais classes são mais frequentemente confundidas e em que direção ocorrem essas confusões.

Além dessas métricas, foram consideradas medidas baseadas em probabilidades. A AUC-ROC (*Area Under the Receiver Operating Characteristic Curve*) foi utilizada para avaliar a capacidade discriminativa global dos modelos ao longo de diferentes limiares de decisão, enquanto a AUC-PR (*Area Under the Precision-Recall Curve*) foi empregada como métrica complementar, especialmente relevante em cenários desbalanceados, nos quais a curva ROC pode apresentar interpretações excessivamente otimistas.

A utilização conjunta dessas métricas permite uma avaliação multidimensional do desempenho dos modelos, capturando tanto a capacidade global de classificação quanto o comportamento específico em classes minoritárias. Essa abordagem mostra-se particularmente adequada ao contexto de risco ESG, no qual a correta identificação de casos de alto risco possui relevância crítica, mesmo quando essa classe representa uma fração reduzida da base de dados.

A estratégia adotada assegura uma análise robusta, comparável e alinhada às características do problema, fornecendo subsídios consistentes para a interpretação dos resultados apresentados no Capítulo 4.

Capítulo 4

Resultados e Discussão

Neste capítulo são apresentados e discutidos os resultados obtidos a partir da aplicação do pipeline de modelagem descrito no Capítulo 3, com o objetivo de avaliar a capacidade preditiva dos modelos na classificação do risco ESG de empresas a partir de dados públicos.

A análise concentra-se no desempenho do modelo *Random Forest*, definido como abordagem principal, sendo comparado aos modelos de referência *K-Nearest Neighbors* (KNN) e *XGBoost* (XGB). Essa comparação permite avaliar a robustez das diferentes abordagens de aprendizado supervisionado frente às características do problema, especialmente considerando o desbalanceamento das classes e a natureza tabular dos dados.

Os resultados são apresentados de forma estruturada, contemplando inicialmente a análise exploratória dos dados e a distribuição das classes ESG, seguida pela avaliação quantitativa dos modelos por meio de métricas como acurácia, precisão, *recall*, F1-score, AUC-ROC e AUC-PR, obtidas a partir de validação cruzada estratificada. Adicionalmente, são analisadas as matrizes de confusão, com o objetivo de compreender os padrões de erro e o comportamento dos modelos em cada classe de risco.

Por fim, é realizada a análise interpretativa do modelo *Random Forest*, com foco na importância dos atributos e na explicabilidade das previsões por meio da técnica SHAP, permitindo identificar quais variáveis exercem maior influência na determinação do risco ESG. Essa etapa complementa a avaliação quantitativa, contribuindo para a compreensão dos fatores que direcionam as decisões do modelo e reforçando a aplicabilidade prática dos resultados.

4.1 Análise Exploratória e Tratamento dos Dados

A etapa inicial da pesquisa teve como objetivo consolidar, validar e qualificar a base de dados a ser utilizada na modelagem do risco ESG, garantindo consistência estatística e adequação ao aprendizado supervisionado [Géron 2019]. A base foi construída a partir da

integração entre informações financeiras públicas do *Yahoo Finance* e os indicadores de risco ESG da *Sustainalytics*, resultando inicialmente em um conjunto com 503 empresas do índice S&P 500 e 189 variáveis [Aroussi 2025].

Durante a análise exploratória, observou-se a presença significativa de valores ausentes, redundância de atributos e variáveis com baixa relevância preditiva. Esse diagnóstico motivou a aplicação das etapas de tratamento descritas no Capítulo 3, incluindo filtragem de variáveis, inserção de valores ausentes e padronização dos dados.

Como resultado, a base final foi consolidada com 499 empresas e 97 variáveis numéricas, representando dimensões relevantes como rentabilidade, liquidez, alavancagem, volatilidade e aspectos associados à governança corporativa.

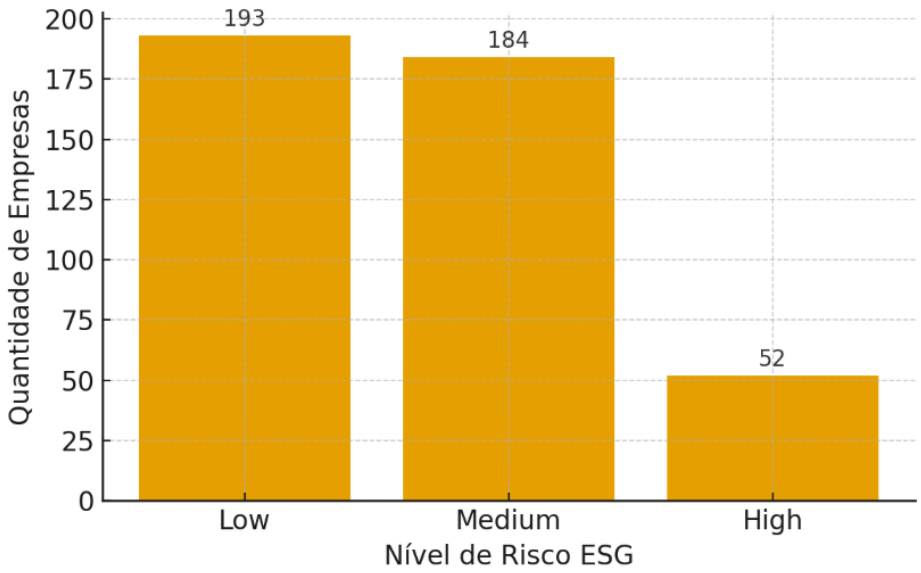


Figura 4.1: Distribuição de Classes de Risco ESG.
Fonte: Elaborado pelo autor (2025).

A Figura 4.1 apresenta a distribuição final das classes de risco ESG após o processo de tratamento e engenharia da variável-alvo. Observa-se que a classe de baixo risco (*Low*) concentra 193 empresas, enquanto a classe de risco médio (*Medium*) possui 184 empresas e a classe de alto risco (*High*) reúne 52 empresas. Essa distribuição evidencia um desbalanceamento moderado, com predominância de empresas classificadas nos níveis mais baixos de risco.

Esse comportamento é coerente com o perfil das empresas do S&P 500, que, em geral, apresentam maior maturidade em práticas de governança, transparência e gestão de riscos socioambientais. No entanto, a menor representatividade da classe de alto risco reforça a necessidade de adoção de estratégias específicas durante a modelagem, como técnicas de balanceamento de classes, a fim de evitar vieses nos modelos preditivos.

Do ponto de vista metodológico, a análise da Figura 4.1 é fundamental para compreender o comportamento da variável-alvo, uma vez que a distribuição das classes impacta diretamente a escolha das métricas de avaliação e a estratégia de treinamento dos modelos, conforme discutido no Capítulo 3.

O processo de tratamento resultou em ganhos significativos na qualidade da base de dados, refletidos nos seguintes aspectos:

- **Redução de dimensionalidade**: remoção de mais de 80 variáveis redundantes, com base na análise exploratória inicial, que diminuindo a dimensionalidade do conjunto de dados e mitigando o risco de sobreajuste;
- **Padronização dos dados**: conversão de todas as variáveis para formato numérico compatível com algoritmos de aprendizado supervisionado;
- **Tratamento de valores ausentes**: eliminação de valores ausentes nas variáveis preditoras por meio de inserção controlada;
- **Correção de inconsistências**: tratamento de variáveis originalmente representadas como intervalos (*ranges*), com separação entre valores mínimos e máximos;
- **Padronização da variável-alvo**: ajuste semântico das classes ESG, preservando a coerência conceitual da variável de classificação.

Essas etapas asseguraram a construção de uma base de dados limpa, consistente e reprodutível, permitindo que os modelos treinados capturem padrões estruturais entre o desempenho financeiro das empresas e seu respectivo nível de risco ESG, minimizando a influência de ruídos ou artefatos nos dados.

4.2 Comparação da Performance dos Modelos de Classificação

4.2.1 Acurácia, Precisão, Recall, F1-Score, MCC e Cohen's kappa

Após a aplicação do balanceamento das classes por meio do SMOTE, foram avaliados três algoritmos supervisionados de classificação: *Random Forest* (RF), *XGBoost* (XGB) e *K-Nearest Neighbors* (KNN). O procedimento experimental foi conduzido com validação cruzada estratificada (*StratifiedKFold*), utilizando três partições ($k = 3$) e 1.000 execuções com diferentes sementes aleatórias, com o objetivo de reduzir a variabilidade associada à divisão dos dados e obter estimativas mais robustas e estáveis do desempenho dos modelos.

A comparação entre os classificadores foi realizada a partir de um conjunto de métricas complementares, composto por acurácia, *balanced accuracy*, precisão, *recall*, F1-score, MCC (*Matthews Correlation Coefficient*) e Cohen's kappa. A adoção conjunta dessas métricas se justifica pela natureza multiclasse e desbalanceada do problema, uma vez que medidas isoladas, como a acurácia, podem não ser suficientes para representar adequadamente o desempenho dos modelos em classes com menor frequência, como a de alto risco ESG [Saito e Rehmsmeier 2015, Géron 2019].

Nesse contexto, as métricas selecionadas permitem uma análise mais abrangente do comportamento dos classificadores. Enquanto a acurácia e a *balanced accuracy* fornecem uma visão global do desempenho, precisão, *recall* e F1-score permitem avaliar de forma mais detalhada o equilíbrio entre acertos e erros em cada classe. Já o MCC e o Cohen's kappa complementam essa análise ao oferecer medidas de concordância e correlação mais robustas, especialmente úteis em cenários com distribuição desigual entre classes.

As subseções e tabelas apresentadas a seguir discutem esses resultados à luz do comportamento empírico dos modelos e das particularidades do problema de classificação de risco ESG, buscando identificar não apenas qual algoritmo obteve melhor desempenho global, mas também quais limitações e vantagens cada abordagem apresenta em termos de sensibilidade às classes minoritárias, estabilidade preditiva e potencial de aplicação prática.

Acurácia e Balanced Accuracy

A acurácia mede a proporção global de classificações corretas realizadas pelo modelo, considerando todas as classes de forma agregada. Por outro lado, a *balanced accuracy* calcula a média do desempenho por classe, atribuindo o mesmo peso a cada uma delas, sendo, portanto, mais adequada em cenários com distribuição desbalanceada, como o observado neste estudo.

Os resultados obtidos para essas métricas são apresentados na Tabela 4.1, que sintetiza o desempenho médio dos modelos ao longo das execuções experimentais.

Tabela 4.1: Desempenho médio dos modelos de classificação ESG

Modelo	**Acurácia**	**Acurácia Balanceada**
Random Forest	$0{,}59 \pm 0{,}02$	$0{,}50 \pm 0{,}02$
XGBoost	$0{,}60 \pm 0{,}02$	$0{,}51 \pm 0{,}02$
KNN	$0{,}44 \pm 0{,}02$	$0{,}44 \pm 0{,}02$

Fonte: Elaborado pelo autor (2025).

Conforme observado na Tabela 4.1, os modelos baseados em árvores, *Random Forest*

($0{,}59 \pm 0{,}02$) e *XGBoost* ($0{,}60 \pm 0{,}02$), apresentaram desempenho superior em termos de acurácia global quando comparados ao KNN ($0{,}44 \pm 0{,}02$). Esse resultado indica maior capacidade desses modelos em capturar padrões gerais nos dados, especialmente considerando a natureza tabular e heterogênea das variáveis utilizadas.

No entanto, ao analisar a *balanced accuracy*, observa-se uma redução consistente nos valores para todos os modelos, com resultados próximos de 0,50 para *Random Forest* e *XGBoost*, e 0,44 para o KNN. Essa diferença entre acurácia e *balanced accuracy* evidencia que o desempenho não está uniformemente distribuído entre as classes, indicando que os modelos apresentam maior capacidade de acerto nas classes majoritárias (*Low* e *Medium*) em detrimento da classe minoritária (*High*).

Esse comportamento está diretamente relacionado à distribuição das classes apresentada na Figura 4.1, na qual a classe de alto risco (*High*) representa uma parcela significativamente menor da amostra. Como consequência, mesmo após a aplicação de técnicas de balanceamento durante o treinamento, os modelos ainda enfrentam maior dificuldade em aprender padrões representativos dessa classe.

Além disso, a proximidade dos valores de *balanced accuracy* em torno de 0,50 sugere um desempenho próximo ao aleatório quando considerada a média entre classes, o que reforça a complexidade do problema de classificação de risco ESG. Esse resultado pode ser interpretado como reflexo da elevada heterogeneidade das empresas classificadas como alto risco, bem como da possível sobreposição de características financeiras entre diferentes níveis de risco ESG.

Portanto, embora a acurácia indique um desempenho global razoável, a análise conjunta com a *balanced accuracy* revela limitações importantes na capacidade dos modelos de generalizar adequadamente para todas as classes, especialmente aquelas menos representadas.

Precisão (Precision)

A precisão (*precision*) mede a proporção de previsões corretas dentre aquelas classificadas como pertencentes a uma determinada classe. Em outras palavras, indica o quão confiáveis são as previsões positivas realizadas pelo modelo para cada categoria. Essa métrica é especialmente relevante em cenários nos quais falsos positivos possuem impacto significativo, como no contexto de risco ESG.

No presente estudo, uma alta precisão na classe *High* implica que, quando o modelo classifica uma empresa como de alto risco, essa previsão tende a ser correta, reduzindo o risco de penalizações indevidas ou decisões equivocadas em processos de análise, triagem ou investimento.

Os resultados de precisão por classe e por modelo são apresentados na Tabela 4.2.

Tabela 4.2: Precisão dos modelos de classificação por classe ESG

Classe	**Métrica**	**Random Forest**	**XGBoost**	**KNN**
Low	Precision	$0{,}64 \pm 0{,}02$	$0{,}64 \pm 0{,}02$	$0{,}63 \pm 0{,}03$
Medium	Precision	$0{,}59 \pm 0{,}02$	$0{,}61 \pm 0{,}02$	$0{,}52 \pm 0{,}03$
High	Precision	$0{,}29 \pm 0{,}06$	$0{,}30 \pm 0{,}06$	$0{,}17 \pm 0{,}02$

Fonte: Elaborado pelo autor (2025).

A partir da Tabela 4.2, observa-se que os modelos *Random Forest* e *XGBoost* apresentam desempenho consistente nas classes *Low* e *Medium*, com valores de precisão superiores a 0,59, chegando a aproximadamente 0,64 na classe de baixo risco. Esse resultado indica que, para essas classes, os modelos apresentam boa confiabilidade nas previsões, com baixa incidência de falsos positivos.

No entanto, ao analisar a classe *High*, verifica-se uma queda significativa na precisão para todos os modelos, com valores próximos de 0,28 (*Random Forest*), 0,31 (*XGBoost*) e 0,17 (KNN). Esse comportamento indica que uma parcela relevante das empresas classificadas como alto risco pelos modelos não pertence, de fato, a essa categoria, evidenciando maior ocorrência de falsos positivos.

Essa limitação está diretamente relacionada à distribuição desbalanceada das classes, conforme apresentado na Figura 4.1, em que a classe de alto risco representa uma fração reduzida da amostra. Como consequência, os modelos tendem a apresentar maior incerteza ao identificar corretamente essa classe, mesmo após a aplicação de técnicas de balanceamento durante o treinamento [Chawla et al. 2002, Saito e Rehmsmeier 2015].

Além disso, sob a perspectiva do domínio ESG, esse resultado pode ser interpretado como reflexo da natureza dos dados utilizados. As variáveis financeiras capturam com maior precisão aspectos relacionados à estrutura de capital, rentabilidade e governança, que tendem a estar associados a empresas de menor risco. Por outro lado, a identificação de empresas de alto risco ESG frequentemente depende de fatores menos estruturados e mais qualitativos, como eventos controversos, práticas socioambientais específicas ou riscos regulatórios, que não são plenamente refletidos em indicadores financeiros tradicionais.

Comparativamente, o modelo *XGBoost* apresenta ligeira vantagem na classe *High* em relação ao *Random Forest*, sugerindo maior capacidade de capturar padrões mais complexos associados ao risco elevado. Já o KNN apresenta desempenho inferior em todas as classes, reforçando sua limitação em cenários com alta dimensionalidade e variáveis heterogêneas.

Portanto, a análise da precisão evidencia que, embora os modelos apresentem boa confiabilidade na identificação de empresas de baixo e médio risco, ainda há limitações relevantes na classificação da classe de alto risco.

Recall (Sensibilidade)

O *recall* (sensibilidade) mede a capacidade do modelo de identificar corretamente os casos reais pertencentes a uma determinada classe. Em outras palavras, indica a proporção de instâncias positivas que foram corretamente classificadas pelo modelo. Essa métrica é particularmente relevante em cenários nos quais o custo de falsos negativos é elevado.

No contexto de risco ESG, o *recall* assume papel central na identificação de empresas classificadas como alto risco (*High*), uma vez que a não detecção dessas empresas pode resultar em decisões inadequadas de investimento, exposição a riscos reputacionais ou falhas em processos de monitoramento preventivo.

Os resultados de *recall* por classe e por modelo são apresentados na Tabela 4.3.

Tabela 4.3: Recall dos modelos de classificação por classe ESG

Classe	**Métrica**	**Random Forest**	**XGBoost**	**KNN**
Low	Recall	$0{,}69 \pm 0{,}02$	$0{,}70 \pm 0{,}32$	$0{,}44 \pm 0{,}03$
Medium	Recall	$0{,}59 \pm 0{,}03$	$0{,}60 \pm 0{,}03$	$0{,}42 \pm 0{,}03$
High	Recall	$0{,}22 \pm 0{,}05$	$0{,}23 \pm 0{,}04$	$0{,}47 \pm 0{,}05$

Fonte: Elaborado pelo autor (2025).

A análise da Tabela 4.3 evidencia diferenças importantes no comportamento dos modelos. Para as classes *Low* e *Medium*, os modelos *Random Forest* e *XGBoost* apresentam valores de *recall* relativamente elevados (acima de 0,58), indicando boa capacidade de recuperação dos casos dessas classes majoritárias. O modelo KNN, por sua vez, apresenta desempenho inferior nessas categorias, com valores significativamente mais baixos.

No entanto, o comportamento se altera ao analisar a classe *High*. Nesse caso, o modelo KNN apresenta o maior *recall* (0,467), indicando maior capacidade de identificar empresas efetivamente classificadas como alto risco. Entretanto, conforme observado na Seção anterior (Tabela 4.2), esse ganho em *recall* ocorre em detrimento de uma precisão bastante reduzida, o que implica elevada incidência de falsos positivos.

Esse padrão reflete um comportamento típico de modelos baseados em distância em espaços de alta dimensionalidade, nos quais a separação entre classes tende a se tornar menos discriminativa, levando o modelo a classificar um maior número de instâncias como pertencentes à classe minoritária.

O modelo *XGBoost*, por sua vez, apresenta um *recall* intermediário na classe *High* (0,292), mantendo desempenho mais equilibrado entre as classes. Já o *Random Forest* apresenta o menor *recall* nessa classe (0,216), indicando maior dificuldade em recuperar instâncias de alto risco.

A análise conjunta das Tabelas 4.2 e 4.3 evidencia claramente o *trade-off* entre precisão e *recall*. Enquanto o KNN maximiza a detecção da classe *High* (maior *recall*), ele o faz com baixa confiabilidade (baixa precisão). Por outro lado, modelos como *Random Forest* e *XGBoost* apresentam maior controle sobre falsos positivos, porém com menor capacidade de identificar todos os casos de alto risco.

Sob a perspectiva aplicada ao ESG, esse *trade-off* possui implicações diretas. Em cenários de triagem e monitoramento, pode ser preferível um modelo com maior *recall* para a classe *High*, ainda que com algum aumento de falsos positivos, uma vez que a não identificação de empresas de alto risco pode gerar consequências mais severas do que classificações excessivamente conservadoras.

Nesse sentido, o modelo *XGBoost* apresenta um compromisso mais equilibrado entre detecção e confiabilidade, enquanto o KNN representa uma abordagem mais sensível, porém menos precisa. Essa análise evidencia a importância de considerar múltiplas métricas na avaliação dos modelos, evitando conclusões baseadas em apenas um único critério de desempenho.

F1-Score

O F1-score representa a média harmônica entre precisão e *recall*, sendo uma métrica amplamente utilizada para avaliar o equilíbrio entre a capacidade de detecção de uma classe e a confiabilidade das previsões realizadas. Diferentemente da acurácia, o F1-score é particularmente útil em cenários com desbalanceamento de classes, pois penaliza simultaneamente falsos positivos e falsos negativos.

Os resultados de F1-score por classe e por modelo são apresentados na Tabela 4.4.

Tabela 4.4: F1-score dos modelos de classificação por classe ESG

Classe	**Métrica**	**Random Forest**	**XGBoost**	**KNN**
Low	F1-score	$0{,}66 \pm 0{,}02$	$0{,}67 \pm 0{,}02$	$0{,}52 \pm 0{,}02$
Medium	F1-score	$0{,}59 \pm 0{,}02$	$0{,}61 \pm 0{,}02$	$0{,}47 \pm 0{,}03$
High	F1-score	$0{,}24 \pm 0{,}05$	$0{,}26 \pm 0{,}05$	$0{,}25 \pm 0{,}03$

Fonte: Elaborado pelo autor (2025).

A análise da Tabela 4.4 evidencia que o modelo *XGBoost* apresenta o melhor desempenho geral, especialmente nas classes *Low* (0,689) e *Medium* (0,606), indicando maior equilíbrio entre precisão e *recall* nessas categorias. O modelo *Random Forest* apresenta desempenho próximo na classe *Low* (0,662), porém com queda acentuada na classe *Medium* (0,245), sugerindo menor capacidade de manter simultaneamente boa detecção e baixa taxa de erros nessa classe.

O modelo KNN, por sua vez, apresenta desempenho inferior nas classes majoritárias, com F1-score de 0,520 (*Low*) e 0,467 (*Medium*), refletindo as limitações já observadas nas métricas anteriores, especialmente sua sensibilidade à estrutura dos dados.

Ao analisar a classe *High*, observa-se que todos os modelos apresentam F1-scores relativamente baixos, com valores de 0,216 (*Random Forest*), 0,262 (*XGBoost*) e 0,246 (KNN). Esses resultados indicam que, apesar das diferenças observadas entre precisão e *recall* nas seções anteriores (Tabelas 4.2 e 4.3), nenhum dos modelos consegue alcançar um equilíbrio satisfatório entre essas métricas para a classe minoritária.

Esse comportamento reforça o diagnóstico previamente identificado: a classe de alto risco apresenta maior complexidade e menor separabilidade no espaço de atributos, o que dificulta sua correta identificação pelos modelos. Além disso, a menor representatividade dessa classe na base de dados contribui para a redução do desempenho agregado, mesmo com a aplicação de técnicas de balanceamento.

A análise conjunta das Tabelas 4.2, 4.3 e 4.4 evidencia de forma clara o *trade-off* entre precisão e *recall*. O modelo KNN, por exemplo, apresentou maior *recall* na classe *High*, porém com baixa precisão, resultando em um F1-score apenas intermediário. Já o *XGBoost*, ao apresentar valores mais equilibrados entre essas duas métricas, obtém o melhor desempenho agregado nessa classe, ainda que em níveis moderados.

Do ponto de vista aplicado ao contexto ESG, o F1-score confirma que modelos baseados em árvores, como *Random Forest* e *XGBoost*, apresentam maior capacidade de capturar relações complexas entre variáveis financeiras e níveis de risco, mantendo um equilíbrio mais consistente entre detecção e confiabilidade das previsões.

Portanto, os resultados apresentados na Tabela 4.4 indicam que o modelo *XGBoost* oferece o melhor compromisso entre precisão e *recall*, especialmente nas classes majoritárias, enquanto a identificação da classe de alto risco permanece como um desafio relevante, exigindo possíveis avanços metodológicos ou a incorporação de novas fontes de dados para melhoria do desempenho.

MCC e Cohen's Kappa

O coeficiente de correlação de Matthews (*Matthews Correlation Coefficient* — MCC) mede a correlação entre as predições do modelo e os rótulos reais, considerando todas as classes simultaneamente. Já o coeficiente kappa de Cohen (*Cohen's* κ) avalia o grau de concordância entre as classificações previstas e reais, ajustando o resultado pela concordância esperada ao acaso. Ambas as métricas são particularmente adequadas para problemas multiclasse e desbalanceados, pois fornecem uma visão mais robusta do desempenho global em comparação à acurácia isolada.

Os resultados dessas métricas são apresentados na Tabela 4.5.

Tabela 4.5: Métricas globais de concordância dos modelos de classificação ESG

Modelo	**MCC**	**Cohen's κ**
Random Forest	$0{,}30 \pm 0{,}03$	$0{,}30 \pm 0{,}03$
XGBoost	$0{,}32 \pm 0{,}03$	$0{,}32 \pm 0{,}03$
KNN	$0{,}17 \pm 0{,}03$	$0{,}16 \pm 0{,}03$

Fonte: Elaborado pelo autor (2025).

A partir da Tabela 4.5, observa-se que os modelos baseados em árvores apresentam os melhores resultados em termos de concordância global. O modelo *XGBoost* obteve MCC de $0{,}31 \pm 0{,}028$ e Cohen's κ de $0{,}32 \pm 0{,}028$, enquanto o *Random Forest* apresentou valores muito próximos, com MCC de $0{,}30 \pm 0{,}029$ e κ de $0{,}31 \pm 0{,}028$. Esses resultados indicam uma correlação moderada entre as previsões e os valores reais, sugerindo que os modelos capturam padrões relevantes além do que seria esperado por uma classificação aleatória.

Em contraste, o modelo KNN apresentou desempenho inferior, com MCC de $0{,}19 \pm 0{,}027$ e κ de $0{,}16 \pm 0{,}026$, indicando baixa concordância e menor capacidade de generalização. Esse comportamento é consistente com os resultados observados nas métricas anteriores, nas quais o KNN demonstrou maior sensibilidade à distribuição dos dados e menor estabilidade preditiva.

A análise dessas métricas complementa as interpretações anteriores ao evidenciar que, mesmo com acurácia em torno de 0,60 (Tabela 4.1), o nível de concordância real entre predições e rótulos permanece moderado. Esse resultado reforça a importância de utilizar métricas mais robustas, como MCC e Cohen's κ, especialmente em problemas com desbalanceamento de classes, como o de risco ESG.

Do ponto de vista aplicado, esses valores indicam que os modelos *Random Forest* e *XGBoost* apresentam desempenho consistente e significativamente superior ao aleatório, ainda que com espaço para melhorias, principalmente na distinção da classe de alto risco, conforme discutido nas análises de precisão, *recall* e F1-score (Tabelas 4.2, 4.3 e 4.4).

Adicionalmente, os resultados evidenciam que ganhos moderados em métricas globais podem ser relevantes em aplicações práticas de ESG, especialmente em cenários de triagem e monitoramento automatizado. Nesses contextos, a combinação entre desempenho preditivo, interpretabilidade e estabilidade do modelo torna-se fundamental para garantir confiança nas decisões apoiadas por sistemas de aprendizado de máquina.

4.2.2 Curva ROC

A curva ROC (*Receiver Operating Characteristic*) é um instrumento amplamente utilizado para avaliar a capacidade discriminativa de modelos de classificação probabilística,

representando a relação entre a taxa de verdadeiros positivos (*True Positive Rate* — TPR) e a taxa de falsos positivos (*False Positive Rate* — FPR) ao longo de diferentes limiares de decisão.

No contexto deste estudo, a curva ROC permite analisar se os modelos são capazes de ordenar corretamente as empresas de acordo com a probabilidade de pertencimento a cada classe de risco ESG, sendo particularmente útil para avaliar o desempenho em termos de separabilidade entre classes. No entanto, conforme discutido anteriormente, sua interpretação deve ser realizada com cautela em cenários desbalanceados, sendo complementada por outras métricas, como a curva Precision–Recall [Saito e Rehmsmeier 2015].

A Figura 4.2 apresenta a curva ROC do modelo *Random Forest*, considerando a abordagem *One-vs-Rest* para o problema multiclasse.

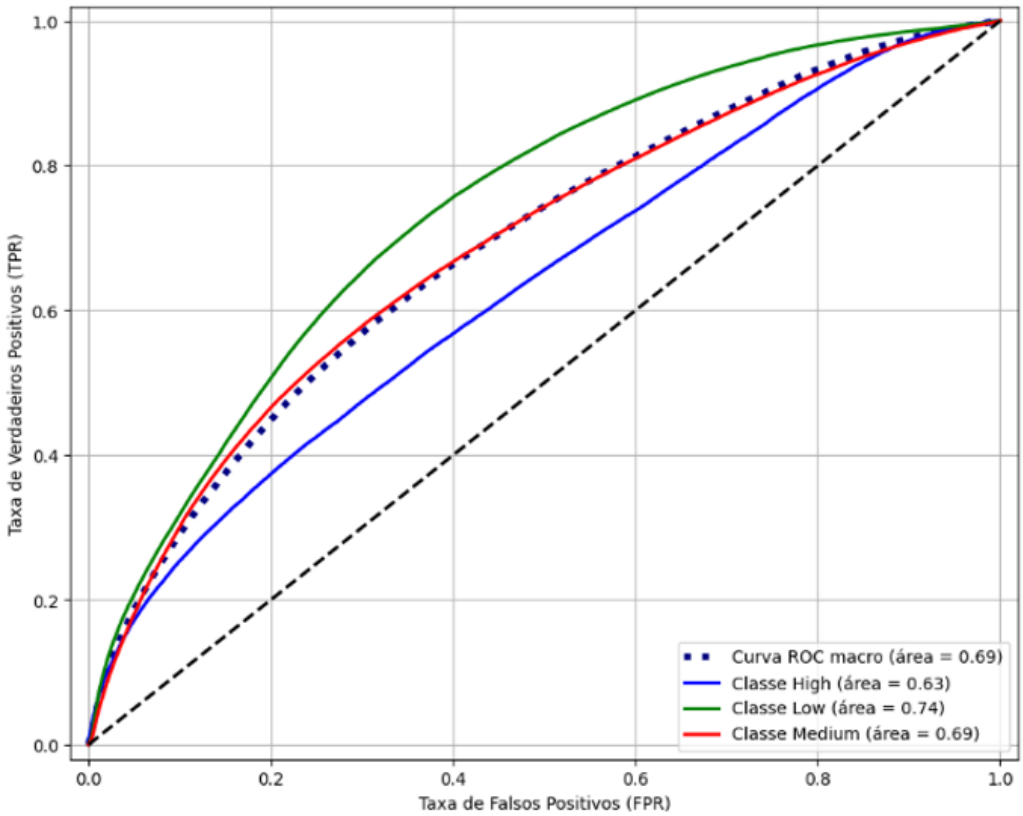


Figura 4.2: Curva ROC do modelo Random Forest

Fonte: Elaborado pelo autor (2025).

A partir da Figura 4.2, observa-se que o modelo *Random Forest* apresenta uma AUC macro de aproximadamente 0,69, indicando capacidade discriminativa moderada. Ao analisar as curvas por classe, verifica-se que o desempenho varia de forma significativa:

- Classe *Low*: AUC $\approx 0{,}74 \rightarrow$ maior capacidade de separação;

- Classe *Medium*: AUC $\approx 0{,}69 \rightarrow$ desempenho intermediário;
- Classe *High*: AUC $\approx 0{,}63 \rightarrow$ menor capacidade discriminativa.

Essa diferença evidencia que o modelo é mais eficiente na distinção de empresas de baixo risco em relação às demais classes, enquanto apresenta maior dificuldade na separação da classe de alto risco. Esse comportamento é consistente com os resultados apresentados nas seções anteriores, especialmente nas análises de *recall* e F1-score, nas quais a classe *High* apresentou desempenho inferior.

Em termos geométricos, observa-se que as curvas se posicionam acima da linha de referência aleatória (linha diagonal), indicando que o modelo apresenta desempenho superior ao acaso em todas as classes. No entanto, a menor distância da curva da classe *High* em relação a essa linha reforça sua menor separabilidade no espaço de atributos.

Comparativamente, o modelo *XGBoost* apresentou AUC macro ligeiramente superior ($\approx 0{,}71$), com melhoria discreta na classe *High* ($\approx 0{,}65$), sugerindo maior capacidade de capturar relações não lineares mais complexas. Já o modelo KNN apresentou AUC macro inferior ($\approx 0{,}63$), evidenciando menor capacidade discriminativa, especialmente na identificação da classe minoritária.

A análise da curva ROC, em conjunto com as métricas anteriormente discutidas, reforça que os modelos avaliados são adequados para tarefas de triagem e priorização de empresas, nas quais o ordenamento por risco é mais relevante do que a classificação pontual em um único limiar.

Entretanto, a identificação mais robusta de empresas de alto risco ESG permanece um desafio, possivelmente associado à limitação dos dados utilizados, predominantemente financeiros, que não capturam integralmente fatores qualitativos relevantes, como eventos controversos, riscos reputacionais e aspectos socioambientais não estruturados.

4.2.3 Curva Precision–Recall

A curva Precision–Recall (*PR*) é especialmente adequada para a avaliação de modelos em cenários desbalanceados, pois enfatiza o desempenho na classe minoritária e fornece uma visão mais informativa do que a curva ROC nesses contextos [Saito e Rehmsmeier 2015]. Essa curva explicita o compromisso entre precisão e *recall* ao longo dos diferentes limiares de decisão, permitindo analisar simultaneamente a capacidade de detecção e a confiabilidade das previsões.

A Figura 4.3 apresenta a curva Precision–Recall do modelo *Random Forest*, adotado como modelo principal devido à sua robustez e interpretabilidade.

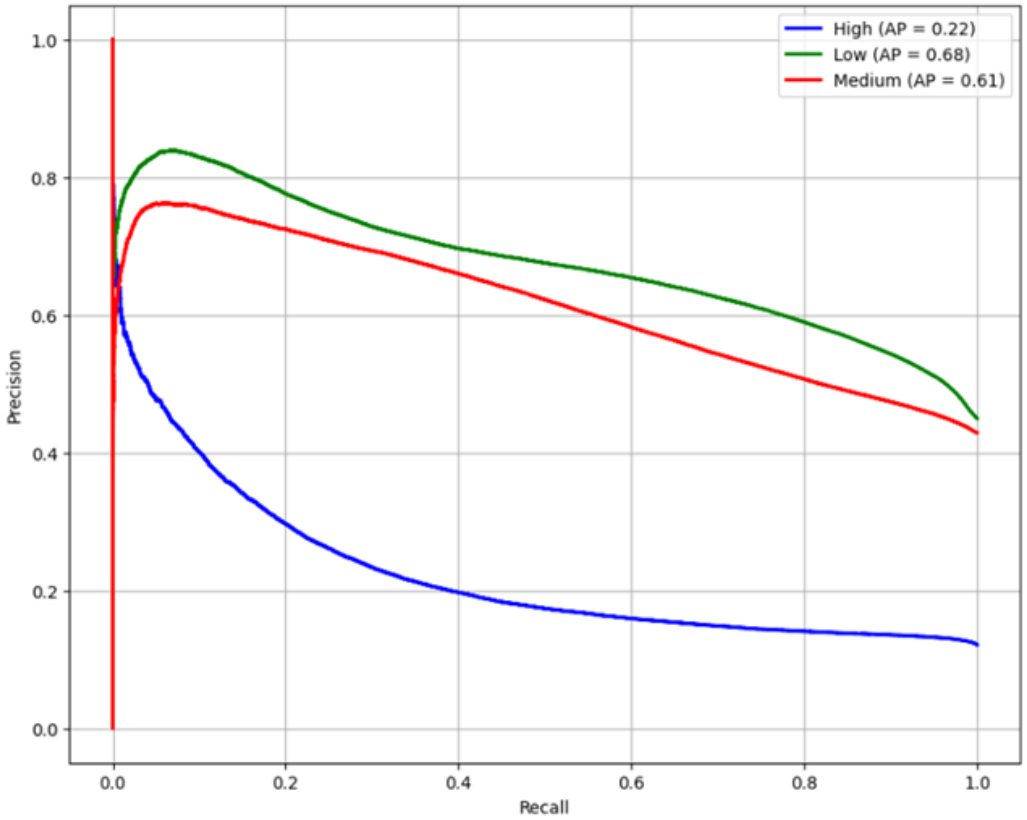


Figura 4.3: Curva Precision–Recall por classe ESG

Fonte: Elaborado pelo autor (2025).

A partir da Figura 4.3, observa-se que o modelo apresenta comportamentos distintos entre as classes. A curva da classe *Low* mantém níveis elevados de precisão mesmo com o aumento do *recall*, evidenciando alta estabilidade e confiabilidade na identificação de empresas de baixo risco. Esse comportamento indica que o modelo consegue distinguir com clareza esse grupo, o que está alinhado com a maior representatividade dessa classe na base de dados.

Para a classe *Medium*, observa-se uma queda gradual da precisão à medida que o *recall* aumenta. Esse padrão é esperado, uma vez que essa classe representa um grupo intermediário, com características mais heterogêneas e fronteiras menos bem definidas em relação às classes *Low* e *High*.

Por outro lado, a curva da classe *High* apresenta comportamento significativamente distinto, com rápida queda de precisão conforme o *recall* aumenta. Esse padrão indica que, para capturar um maior número de empresas de alto risco, o modelo passa a gerar uma quantidade elevada de falsos positivos, reduzindo a confiabilidade das previsões nessa classe.

Essa interpretação é reforçada pelos valores de *Average Precision* (AP), que correspondem à área sob a curva PR. A síntese desses resultados é apresentada na Tabela 4.6.

Tabela 4.6: Average Precision (AP) médio dos modelos por classe ESG

Classe	**Random Forest**	**XGBoost**	**KNN**
Low	0,68	0,70	0,64
Medium	0,61	0,65	0,52
High	0,22	0,26	0,17
AP Macro	0,50	0,54	0,44

Fonte: Elaborado pelo autor (2025).

Conforme observado na Tabela 4.6, o modelo *Random Forest* apresenta AP macro de aproximadamente 0,50, com desempenho significativamente superior nas classes *Low* (0,68) e *Medium* (0,61), e desempenho reduzido na classe *High* (0,22). Esses resultados são consistentes com as análises anteriores de precisão, *recall* e F1-score, reforçando a dificuldade dos modelos em capturar adequadamente a classe minoritária.

Comparativamente, o modelo *XGBoost* apresenta o melhor desempenho global, com AP macro de 0,54 e ganhos mais expressivos na classe *High* (0,26), indicando maior sensibilidade a padrões mais complexos. No entanto, essa melhora ocorre com aumento da complexidade do modelo, o que pode impactar aspectos de interpretabilidade e governança, relevantes no contexto ESG.

O modelo KNN, por sua vez, apresenta o menor desempenho, com AP macro de 0,44 e resultados inferiores em todas as classes, evidenciando limitações já observadas nas demais métricas, especialmente em cenários de alta dimensionalidade.

Sob a perspectiva aplicada ao problema de risco ESG, as curvas Precision–Recall evidenciam uma assimetria estrutural: os modelos apresentam alto desempenho na identificação de empresas com baixo risco e desempenho intermediário para risco médio, porém enfrentam dificuldades significativas na detecção da classe de alto risco.

Essa limitação pode ser atribuída não apenas ao desbalanceamento da base de dados, mas também à natureza dos atributos utilizados. Como as variáveis são predominantemente financeiras, elas capturam com maior precisão aspectos relacionados à estabilidade e governança, enquanto riscos ESG severos frequentemente estão associados a fatores qualitativos, como controvérsias, eventos reputacionais e questões regulatórias, que não são plenamente refletidos nos dados estruturados utilizados.

Dessa forma, a análise das curvas PR reforça que, embora os modelos sejam adequados para tarefas de triagem e priorização, a identificação mais precisa de empresas de alto risco ESG pode demandar o enriquecimento da base de dados com informações qualitativas ou fontes externas adicionais.

4.2.4 Interpretação da Matriz de Confusão

As matrizes de confusão permitem analisar de forma direta o comportamento dos modelos, evidenciando não apenas os acertos, mas também os padrões de erro entre as classes *Low*, *Medium* e *High*. No presente estudo, foi utilizada a versão normalizada por linha, o que permite interpretar cada linha como a distribuição das previsões condicionadas à classe real, sendo equivalente à análise de *recall* por classe.

A Figura 4.4 apresenta a matriz de confusão do modelo *Random Forest*, destacando os padrões de classificação e os principais fluxos de erro entre as categorias de risco ESG.

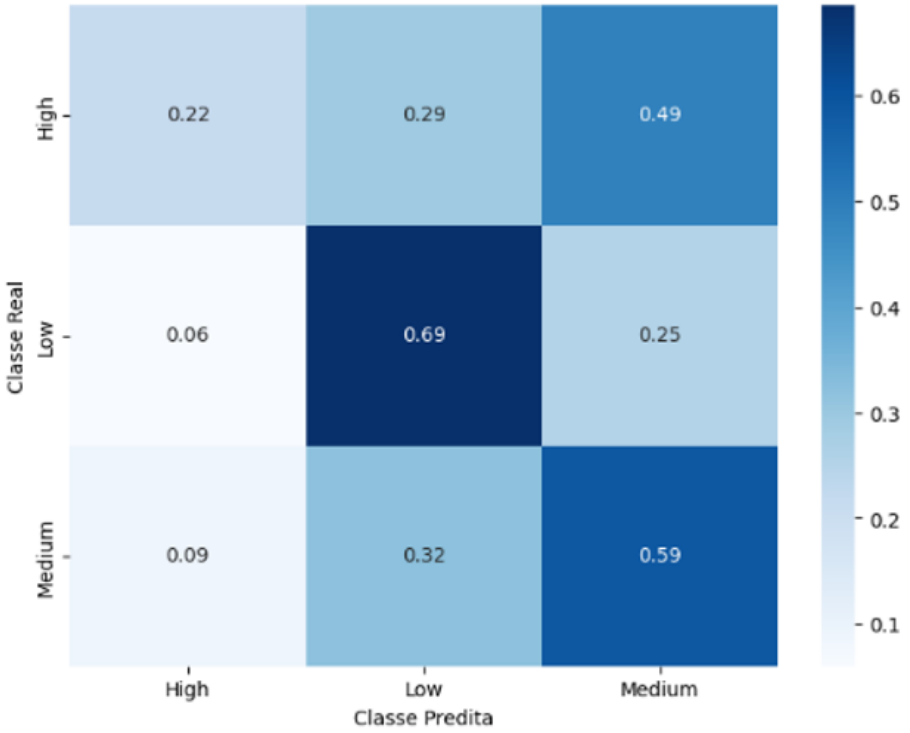


Figura 4.4: Matriz de confusão do modelo Random Forest

Fonte: Elaborado pelo autor (2025).

A partir da Figura 4.4, observa-se que os valores da diagonal principal representam os acertos do modelo, correspondendo aos *recalls* por classe. Nesse sentido, o modelo apresenta desempenho mais elevado na classe *Low* (0,69), seguido da classe *Medium* (0,59) e, por fim, da classe *High* (0,22). Esses valores são consistentes com os resultados apresentados anteriormente na Tabela 4.3, reforçando a maior dificuldade do modelo em identificar corretamente a classe de alto risco.

A análise dos erros revela padrões estruturais relevantes. A principal confusão ocorre na classe *High*, na qual aproximadamente 49% das empresas são classificadas como *Medium* (*High* $\rightarrow$ *Medium*). Esse comportamento indica que o modelo tende a suavizar o risco elevado, deslocando essas empresas para a classe intermediária, o que evidencia dificuldade

em distinguir claramente os níveis mais altos de risco ESG.

De forma complementar, observa-se que a classe *Medium* apresenta uma taxa significativa de confusão com a classe *Low* (*Medium* $\rightarrow$ *Low* = 0,32), sugerindo que parte das empresas classificadas como risco intermediário possui características mais próximas de empresas de baixo risco, ao menos sob a ótica das variáveis utilizadas.

Por outro lado, nota-se que há baixo fluxo de erro entre extremos, especialmente no caso de *Low* $\rightarrow$ *High* (0,06). Esse comportamento é desejável, pois indica que o modelo raramente classifica empresas de baixo risco como alto risco, reduzindo a ocorrência de falsos alarmes críticos.

Do ponto de vista conceitual, esse padrão revela uma hierarquia de dificuldade na classificação, na qual a classe *Low* é mais facilmente identificável, enquanto a distinção entre as classes *Medium* e *High* apresenta maior sobreposição no espaço de atributos. Nesse contexto, a chamada ambiguidade refere-se justamente à presença de características semelhantes entre essas duas classes, dificultando a definição de fronteiras discriminativas claras para o modelo. Essa estrutura é coerente com os resultados das métricas anteriores e com a distribuição das classes observada na Figura 4.1.

A dificuldade na identificação da classe *High* pode ser atribuída não apenas ao desbalanceamento original da base, mas também à limitada separabilidade dos dados utilizados para representar empresas de maior risco. Mesmo com o uso do SMOTE para balanceamento, as amostras sintéticas geradas são construídas por interpolação a partir das observações existentes e, portanto, preservam os padrões do espaço original. Assim, se os exemplos da classe *High* já apresentam sobreposição com a classe *Medium*, as amostras sintéticas tendem a reproduzir esse mesmo comportamento, sem necessariamente aumentar o poder discriminativo da classe minoritária. Esse aspecto ajuda a explicar por que a principal confusão observada permanece concentrada em *High* $\rightarrow$ *Medium*, mesmo após o balanceamento.

Além disso, como os atributos utilizados são predominantemente financeiros, eles tendem a capturar com maior precisão aspectos relacionados à estabilidade, liquidez e governança, enquanto níveis elevados de risco ESG frequentemente dependem de fatores qualitativos ou eventos específicos menos refletidos nessas variáveis. Comparativamente, o modelo *XGBoost* apresenta padrão semelhante ao *Random Forest*, com leve melhora na classe *Medium* e desempenho ainda limitado na classe *High*. Já o modelo KNN apresenta maior *recall* para a classe *High*, porém com maior dispersão de erros entre classes, indicando um ganho em sensibilidade acompanhado de menor confiabilidade global.

De forma geral, a análise da matriz de confusão confirma que os modelos apresentam bom desempenho na identificação de empresas de baixo risco, desempenho intermediário para risco médio e limitações relevantes na detecção de alto risco.

4.3 Análise da Importância dos Atributos (Feature Importance)

A análise de importância dos atributos tem como objetivo identificar quais variáveis exercem maior influência na classificação do risco ESG, bem como compreender de que forma contribuem para diferenciar as classes *Low*, *Medium* e *High*. Essa etapa complementa a avaliação quantitativa apresentada na Seção 4.2, permitindo interpretar os mecanismos internos dos modelos.

Para essa análise, foram utilizadas medidas de importância global específicas de cada modelo, baseadas na redução de impureza no *Random Forest*, *gain* no *XGBoost* e *Permutation Importance* no KNN, além de técnicas de explicabilidade baseadas em SHAP (*Shapley Additive exPlanations*) para os modelos baseados em árvores. O uso combinado dessas abordagens permite avaliar não apenas a relevância das variáveis, mas também a direção e a magnitude de seus efeitos sobre as predições.

A Figura 4.5 apresenta a importância global das 30 variáveis mais relevantes no modelo *Random Forest*, enquanto as Figuras 4.6 e 4.7 mostram a decomposição dos efeitos via SHAP, permitindo analisar como cada variável contribui para a classificação das empresas.

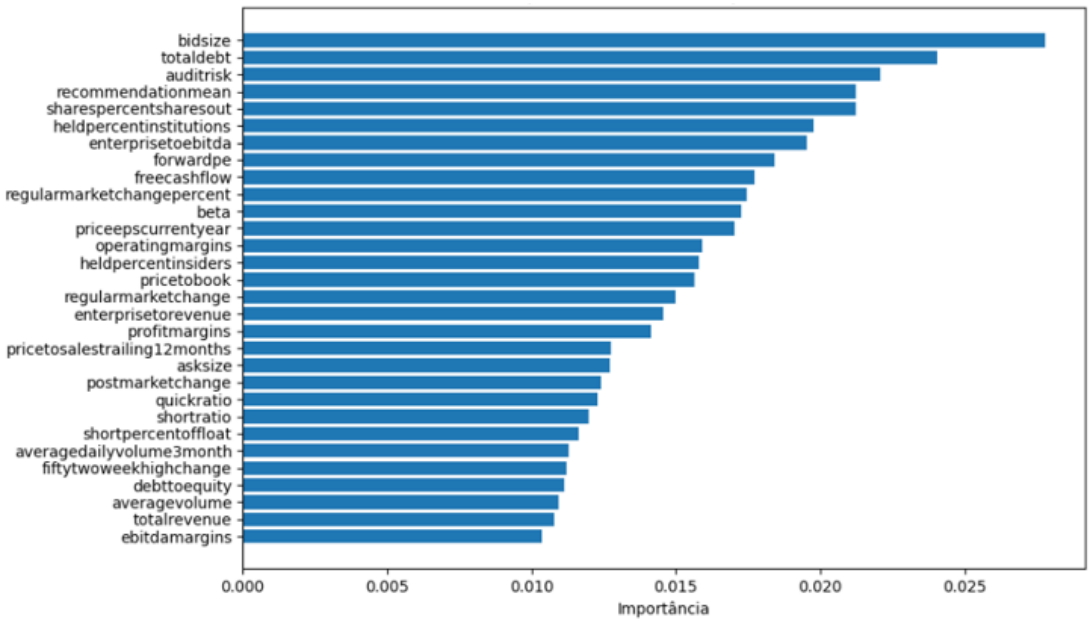


Figura 4.5: Importância global das 30 variáveis mais relevantes do modelo Random Forest

Fonte: Elaborado pelo autor (2025).

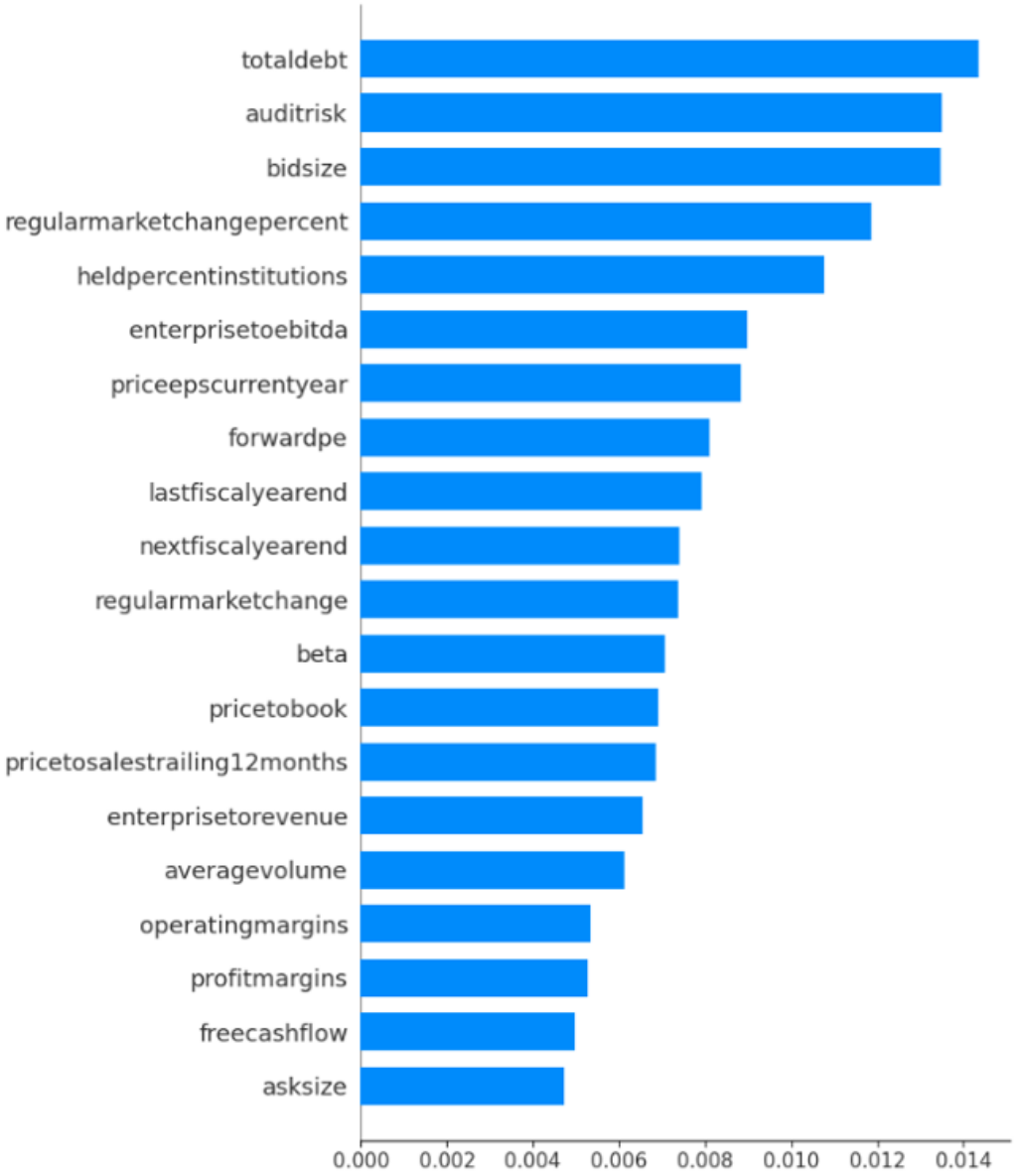


Figura 4.6: SHAP por classe (bar/beeswarm) – direção e magnitude dos efeitos para a classe Low do modelo Random Forest

Fonte: Elaborado pelo autor (2025).

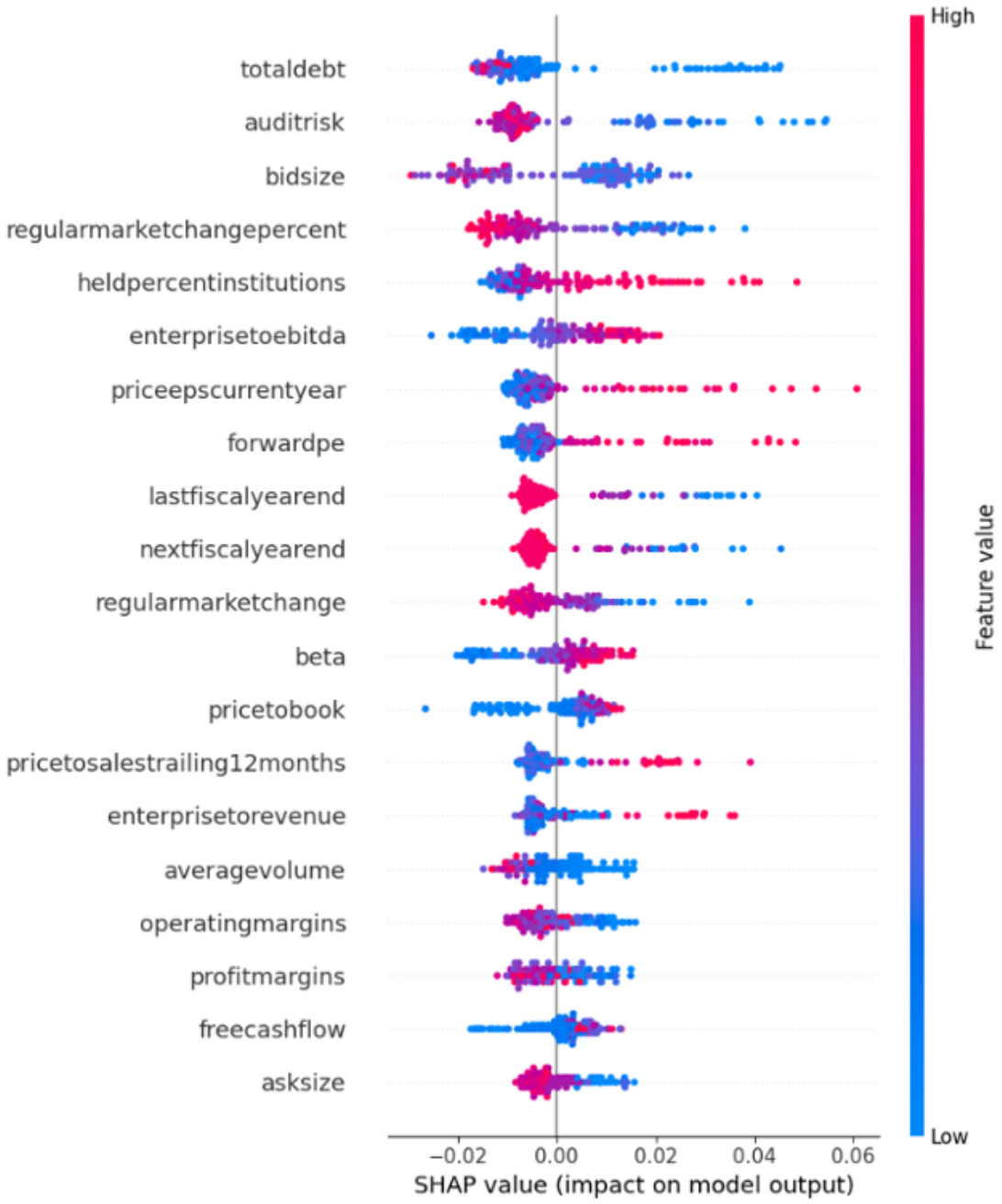


Figura 4.7: SHAP por classe (beeswarm) – direção e magnitude dos efeitos para a classe Low do modelo Random Forest

Fonte: Elaborado pelo autor (2025).

Random Forest

A partir da Figura 4.5, observa-se que o *Random Forest* destaca um conjunto consistente de variáveis, que podem ser organizadas em quatro eixos principais: estrutura financeira, desempenho operacional, liquidez e percepção de mercado e governança corporativa.

A análise visual evidencia que variáveis como `totalDebt`, `bidSize`, `auditRisk` e `enterpriseToEbitda` ocupam posições de destaque, indicando forte influência na deter-

minação do risco ESG.

- Estrutura financeira e alavancagem

Variáveis como `totalDebt`, `debtToEquity`, `enterpriseToEbitda` e `currentRatio` refletem a capacidade da empresa de sustentar suas obrigações financeiras. A leitura combinada com os gráficos SHAP (Figuras 4.6 e 4.7) indica que níveis elevados de endividamento tendem a deslocar as previsões em direção a classes de maior risco, enquanto estruturas financeiras mais equilibradas favorecem classificações de menor risco.

- Desempenho operacional e geração de caixa

Indicadores como `profitMargins`, `operatingMargins`, `ebitdaMargins` e `freeCashFlow` aparecem como fatores relevantes. Conforme observado nos gráficos SHAP, valores mais elevados dessas variáveis estão associados a contribuições positivas para a classificação como baixo risco, sugerindo que a eficiência operacional atua como fator de proteção.

- Liquidez e microestrutura de mercado

Variáveis como `bidSize`, `askSize`, `averageVolume` e `shortRatio` refletem o comportamento do mercado em relação aos ativos. A análise indica que maior liquidez e menor pressão vendedora tendem a contribuir para classificações de menor risco, enquanto sinais de instabilidade ou baixa liquidez deslocam as previsões em direção a maior risco.

- Governança corporativa

Variáveis como `auditRisk`, `recommendationMean` e `heldPercentInstitutions` reforçam o papel central da governança. Nos gráficos SHAP, observa-se que níveis mais elevados de risco de auditoria estão associados a impactos positivos no risco ESG, enquanto maior participação institucional tende a contribuir para classificações mais seguras.

A análise via SHAP (Figuras 4.6 e 4.7) permite aprofundar essa interpretação ao evidenciar não apenas a importância das variáveis, mas também a direção dos seus efeitos. Observa-se um padrão consistente: empresas classificadas como *High* tendem a apresentar maior alavancagem, menor liquidez e sinais mais frágeis de governança, enquanto empresas *Low* apresentam fundamentos mais sólidos e maior estabilidade operacional.

XGBoost

O modelo *XGBoost* apresenta padrões semelhantes ao *Random Forest*, como mostrado nos resultados obtidos no anexo, porém com maior ênfase em variáveis relacionadas à dinâmica de mercado. Esse comportamento sugere maior sensibilidade a interações não lineares e à combinação de fatores como volatilidade, liquidez e estrutura de capital.

A análise indica que o *XGBoost* tende a capturar o risco ESG não apenas como uma

característica estrutural da empresa, mas também como um fenômeno dinâmico, refletido em sinais de mercado e reprecificação dos ativos. Essa característica contribui para seu melhor desempenho em métricas como AUC e F1-score, conforme observado na Seção 4.2.

KNN

O modelo KNN foi utilizado como *baseline*, com a importância das variáveis estimada por meio de *Permutation Importance*. Os resultados apresentados no anexo, indicam maior relevância de variáveis relacionadas à escala e variação de preços, como indicadores de volatilidade e amplitude de mercado.

No entanto, a ausência de mecanismos internos para modelar interações complexas limita a capacidade do KNN de capturar relações mais sofisticadas entre variáveis financeiras e risco ESG. Esse comportamento é consistente com os resultados de desempenho apresentados anteriormente, nos quais o modelo apresentou menor capacidade de generalização.

Análise comparativa dos modelos

A análise conjunta dos modelos revela uma convergência importante: mesmo utilizando diferentes abordagens de aprendizado, os modelos identificam padrões semelhantes para a classificação do risco ESG.

De forma geral, o risco ESG, quando estimado a partir de dados públicos estruturados, está fortemente associado a quatro dimensões principais:

- Estrutura financeira e alavancagem;
- Rentabilidade e geração de caixa;
- Governança corporativa;
- Liquidez e comportamento de mercado.

O modelo *Random Forest* se destaca pelo equilíbrio entre desempenho e interpretabilidade, enquanto o *XGBoost* apresenta maior sensibilidade a padrões complexos. O KNN, por sua vez, atua como referência metodológica, evidenciando as limitações de abordagens baseadas em distância nesse contexto.

4.4 Análise do Modelo por Setores de Serviços

A análise setorial tem como objetivo investigar como os modelos de classificação de risco ESG se comportam diante das especificidades econômicas, operacionais e regulatórias

de diferentes segmentos da economia. Diferentemente da análise agregada apresentada na Seção 4.2, essa abordagem permite avaliar a capacidade dos modelos de capturar a materialidade do risco ESG em contextos distintos, nos quais a natureza do risco pode variar significativamente.

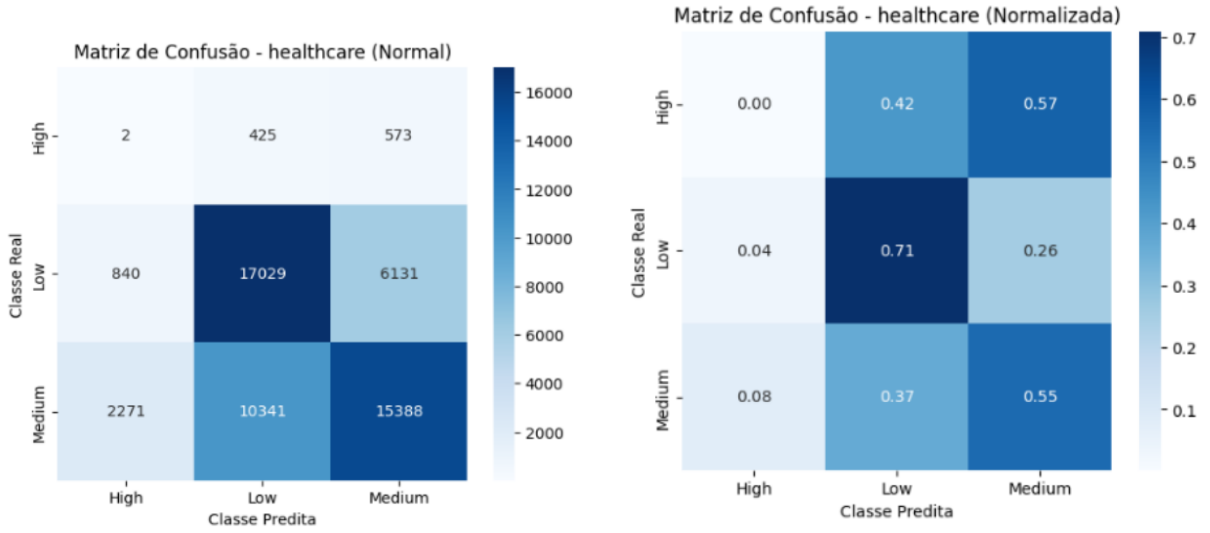
Essa perspectiva é particularmente relevante no contexto ESG, uma vez que a materialidade dos fatores ambientais, sociais e de governança não é uniforme entre setores. Em segmentos intensivos em capital e fortemente regulados, o risco tende a se manifestar de forma mais estruturada e observável, enquanto em setores baseados em ativos intangíveis, esse risco frequentemente assume caráter mais qualitativo, dependente de eventos e percepções externas.

A análise foi conduzida a partir de duas dimensões complementares:

1. Matrizes de confusão normalizadas por setor, que permitem avaliar a separabilidade entre as classes *Low*, *Medium* e *High*;

2. Valores médios de SHAP por classe, que permitem interpretar os principais vetores explicativos das previsões.

A Figura 4.8 apresenta as matrizes de confusão (valores absolutos e normalizados) para o setor *Healthcare*, enquanto a Figura 4.9 apresenta a decomposição das contribuições das variáveis via SHAP.



(a) Matriz de confusão – valores absolutos

(b) Matriz de confusão – valores normalizados

Figura 4.8: Matrizes de confusão do modelo Random Forest para o setor Healthcare

Fonte: Elaborado pelo autor (2025).

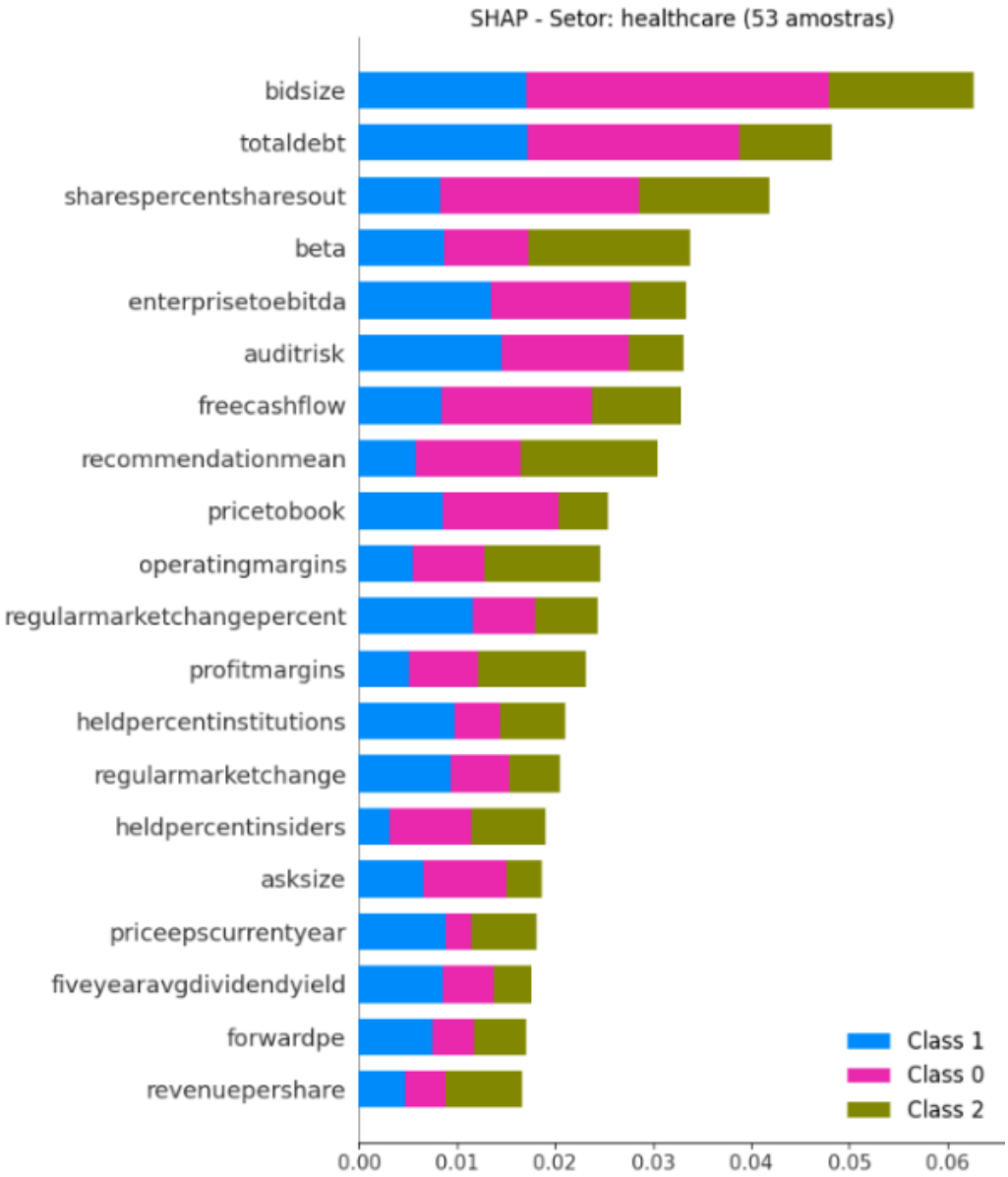


Figura 4.9: Importância das variáveis por classe (SHAP) para o setor Healthcare

Fonte: Elaborado pelo autor (2025).

4.4.1 Random Forest — Interpretação Setorial

O modelo *Random Forest*, adotado como referência interpretável, permite observar de forma clara como os padrões de classificação variam entre setores.

A partir da Figura 4.8 (matriz normalizada), observa-se que, no setor *Healthcare*, o modelo apresenta desempenho assimétrico entre as classes. A classe *Low* apresenta alto *recall* ($\approx 0{,}71$), indicando forte capacidade de identificação de empresas com perfil de risco reduzido. A classe *Medium* apresenta desempenho intermediário ($\approx 0{,}55$), enquanto a classe *High* apresenta *recall* nulo ($\approx 0{,}00$), sendo majoritariamente classificada como *Medium* ($\approx 0{,}57$) ou *Low* ($\approx 0{,}42$).

Esse padrão revela um comportamento crítico: o modelo não apenas apresenta dificuldade em identificar empresas de alto risco, mas tende sistematicamente a reclassificá-las como risco intermediário, indicando uma compressão das classes superiores em direção ao centro da distribuição.

Do ponto de vista estatístico, esse comportamento está alinhado com os resultados globais apresentados na Seção 4.2, nos quais a classe *High* apresentou menor *recall* e F1-score. No entanto, a análise setorial evidencia que essa limitação não é uniforme, sendo mais pronunciada em setores como *Healthcare*.

Sob a perspectiva do domínio ESG, esse resultado pode ser interpretado à luz da natureza do risco no setor. Em *Healthcare*, riscos ESG frequentemente estão associados a fatores como:

- Questões éticas (ensaios clínicos, acesso a medicamentos);
- Regulação sanitária;
- Controvérsias relacionadas a práticas comerciais;
- Impactos sociais indiretos.

Esses elementos possuem forte componente qualitativo e dificilmente são capturados por variáveis financeiras tradicionais, o que explica a limitação do modelo.

A análise da Figura 4.9 (SHAP) reforça essa interpretação ao evidenciar que as variáveis mais relevantes para o modelo permanecem predominantemente financeiras e de mercado, como:

- `totalDebt` (alavancagem);
- `bidSize` (liquidez e interesse de mercado);
- `enterpriseToEbitda` (valuation);
- `auditRisk` (governança).

Observa-se que:

- Valores elevados de endividamento deslocam as previsões para maior risco;
- Sinais positivos de liquidez e confiança de mercado favorecem classificações de menor risco;
- Variáveis de governança atuam como moderadores importantes.

Entretanto, a ausência de variáveis diretamente relacionadas a fatores ambientais e sociais limita a capacidade do modelo de capturar riscos ESG mais severos, especialmente aqueles não refletidos em indicadores financeiros.

Esse padrão indica que o *Random Forest* constrói uma leitura predominantemente estrutural do risco ESG, baseada em fundamentos econômico-financeiros e sinais de mercado, sendo altamente eficaz em identificar estabilidade (*Low*), moderadamente eficaz em capturar transições (*Medium*), mas limitado na identificação de extremos (*High*).

4.4.2 XGBoost — Interpretação Setorial

O modelo *XGBoost* apresenta comportamento semelhante ao *Random Forest*, porém com maior sensibilidade a interações não lineares entre variáveis.

Essa característica permite ao modelo capturar padrões mais complexos, especialmente em setores nos quais o risco ESG emerge da combinação de múltiplos fatores. Em termos setoriais, observa-se que o *XGBoost* apresenta ganhos mais evidentes em setores como *Energy* e *Utilities*, nos quais o risco está mais diretamente associado a variáveis estruturais.

No setor *Healthcare*, embora a dificuldade na identificação da classe *High* persista, o modelo apresenta melhor desempenho na distinção da classe *Medium*, sugerindo maior capacidade de modelar regiões intermediárias do espaço de decisão.

A análise via SHAP indica que o *XGBoost* atribui maior peso a variáveis relacionadas à dinâmica de mercado, como liquidez, volume e variações de preço, sugerindo que o modelo incorpora, ainda que indiretamente, a percepção do mercado sobre o risco ESG.

Essa característica amplia a capacidade preditiva do modelo, mas também introduz maior complexidade interpretativa, o que pode ser um fator relevante em aplicações práticas.

4.4.3 KNN — Interpretação Setorial

O modelo KNN, utilizado como *baseline*, apresenta comportamento fortemente condicionado à estrutura geométrica dos dados.

Em setores mais homogêneos, como *Utilities* e *Energy*, o modelo consegue identificar padrões de similaridade entre empresas, apresentando maior *recall* para a classe *High*. No entanto, esse ganho ocorre com aumento significativo de erros, refletindo baixa precisão e menor confiabilidade.

No setor *Healthcare*, observa-se maior dispersão de erros entre as classes, evidenciando a dificuldade do modelo em lidar com alta dimensionalidade e variáveis heterogêneas. Esse comportamento está associado ao fenômeno conhecido como “maldição da dimensionalidade”, no qual a noção de proximidade entre observações perde significado.

Dessa forma, o KNN atua como um referencial metodológico importante, evidenciando os limites de abordagens baseadas exclusivamente em similaridade estrutural.

4.4.4 Síntese Comparativa Setorial e ESG

A análise comparativa evidencia padrões robustos e consistentes entre os modelos. De forma geral, observa-se que:

- Setores com materialidade ESG tangível (*Energy*, *Utilities*, *Industrials*) apresentam melhor separabilidade entre classes;
- Setores regulados e financeiramente estáveis concentram previsões na classe *Medium*;
- Setores com riscos intangíveis (*Technology*, *Communication Services*, *Healthcare*) apresentam maior dificuldade na identificação da classe *High*.

No caso específico do setor *Healthcare*, os resultados indicam que o modelo é capaz de identificar com alta confiabilidade empresas de baixo risco, apresenta desempenho moderado na classe intermediária, mas falha na detecção de empresas de alto risco, refletindo limitações informacionais da base utilizada.

Em síntese, os resultados indicam que a modelagem quantitativa do risco ESG é viável, porém fortemente dependente da natureza do setor e da disponibilidade de variáveis representativas. Esse achado reforça a necessidade de:

- Abordagens calibradas por setor;
- Incorporação de variáveis qualitativas;
- Integração com fontes externas de informação ESG.

4.5 Considerações ESG e Técnicas

A análise integrada de métricas globais, curvas ROC e Precision–Recall, matrizes de confusão, importância de atributos, explicações via SHAP e avaliação setorial permite uma leitura consolidada: o risco ESG é parcialmente mensurável com dados financeiros e de mercado, mas permanece multifacetado e dependente de contexto. O que os modelos capturam com maior consistência são sinais ligados a disciplina financeira, governança e confiança institucional refletida no mercado.

O Random Forest se destacou como um modelo de leitura estrutural: suas previsões tendem a se apoiar em fundamentos como alavancagem, liquidez, rentabilidade e proxies

de governança. Isso é coerente com a literatura que relaciona governança e processos internos a melhor desempenho sustentável e menor exposição a riscos persistentes.

O XGBoost adicionou uma camada mais reativa e sensível a sinais dinâmicos do mercado, atribuindo mais peso a liquidez, estabilidade de preço e volatilidade. Esse comportamento favorece a detecção de riscos extremos em setores nos quais a materialidade ambiental e regulatória é mais "visível" quantitativamente, mas não elimina a limitação estrutural: riscos qualitativos e reputacionais ainda tendem a escapar quando não há variáveis que os representem adequadamente.

O KNN forneceu uma leitura complementar baseada em similaridade, útil para revelar homogeneidade setorial e padrões locais, mas insuficiente para generalização e para a captura de nuances de governança e materialidade social. Isso reforça a necessidade de modelos capazes de aprender interações complexas quando o problema envolve múltiplas dimensões e fronteiras difusas.

Do ponto de vista técnico-conceitual, os resultados indicam que diferentes famílias de variáveis atuam de forma complementar: fundamentos financeiros fornecem base estrutural; proxies de governança sustentam credibilidade e mitigação; e métricas de mercado funcionam como termômetro dinâmico de confiança institucional. Essa recorrência reforça a coerência do aprendizado mesmo em uma base predominantemente financeira.

Em síntese, os modelos analisados operacionalizam o risco ESG como um fenômeno econômico-financeiro e institucionalmente mediado. Eles evidenciam potencial para triagem e monitoramento, mas também expõem a fronteira atual da abordagem: a classe High permanece dependente de sinais qualitativos e de maior riqueza informacional.

4.6 Discussão dos Resultados à Luz do Mercado e da Literatura

Os resultados obtidos convergem com o diagnóstico contemporâneo das finanças sustentáveis: há evidências de que aspectos ESG se relacionam com desempenho financeiro e risco, mas o campo é marcado por heterogeneidade setorial, diferenças de disclosure e divergências de mensuração entre ratings. Assim, mesmo quando a modelagem encontra padrões consistentes, eles tendem a ser mais fortes em dimensões capturáveis por dados estruturados.

O Random Forest apresentou desempenho estável e interpretação coerente ao associar menor risco a fundamentos mais sólidos e governança mais confiável. Esse resultado dialoga com a literatura que destaca governança como pilar estruturante do desempenho sustentável e da redução de riscos de longo prazo, sobretudo quando há mecanismos de monitoramento institucional e transparência.

O XGBoost ampliou o poder discriminativo, especialmente em setores ambientalmente intensivos, sugerindo que a combinação entre fundamentos e sinais de mercado melhora a sensibilidade do modelo a riscos emergentes. A presença de variáveis de liquidez e estabilidade de preço entre os principais fatores explicativos sugere que o mercado incorpora parcialmente sinais ESG na precificação, ainda que de forma incompleta e sujeita a ruído informacional.

O KNN, apesar de limitado, ajudou a confirmar a existência de padrões locais e homogeneidade setorial, mas também mostrou que similaridade numérica não substitui informação qualitativa quando o risco depende de reputação, controvérsias e fatores institucionais. Isso é consistente com a ideia de que o ESG é multidimensional e não se reduz a um único eixo observável em variáveis financeiras.

Um achado transversal é a centralidade da governança: variáveis relacionadas a auditoria, presença institucional e estrutura de controle aparecem como determinantes recorrentes, reforçando empiricamente a visão de que accountability e monitoramento reduzem exposição a riscos persistentes. Esse ponto é particularmente relevante porque a governança funciona como elo entre fundamentos financeiros e capacidade de sustentar práticas ambientais e sociais no longo prazo.

4.7 Implicações Práticas e Limitações do Modelo Desenvolvido

Do ponto de vista aplicado, os achados indicam que modelos supervisionados interpretáveis podem apoiar processos de screening, due diligence e monitoramento de risco ESG, especialmente quando o objetivo é triagem e priorização de empresas para análises mais profundas. A utilidade é maior quando o risco é refletido por fundamentos, governança e sinais de mercado relativamente estáveis.

A identificação recorrente de variáveis ligadas à alavancagem, liquidez, auditoria e presença institucional sugere aplicações diretas em alertas preventivos e auditoria orientada a risco. Em contextos corporativos e regulatórios, a combinação de desempenho moderado com transparência é valiosa: o modelo não precisa "substituir" avaliação humana, mas reduzir o espaço de busca e melhorar a alocação de atenção.

A análise setorial reforça a necessidade de calibragem por contexto: setores com materialidade mais tangível respondem melhor ao framework quantitativo, enquanto segmentos dominados por intangíveis exigem enriquecimento informacional. Nesse sentido, o modelo pode ser adaptado como instrumento de diagnóstico setorial e benchmarking, desde que suas limitações sejam explicitadas para evitar interpretações indevidas.

Entre as limitações, destaca-se a dependência de dados financeiros e de mercado,

que reduz a capacidade de capturar dimensões ambientais e sociais diretas, além da própria divergência entre mensurações ESG e lacunas de disclosure. Soma-se a isso o desbalanceamento amostral e o caráter transversal da análise, que não modela dinâmicas temporais e eventos reputacionais ao longo do tempo.

Em síntese, o arcabouço proposto demonstra potencial para aplicações reais de triagem e apoio à decisão, mas deve ser entendido como uma camada quantitativa interpretável que se beneficia de dados adicionais, especialmente qualitativos, quando o objetivo é elevar a sensibilidade para alto risco e reduzir ambiguidade setorial.

Capítulo 5

Conclusão

O presente estudo teve como objetivo central o desenvolvimento, a avaliação e a interpretação de modelos de aprendizado de máquina aplicados à classificação do risco ESG (Environmental, Social and Governance) de empresas de capital aberto, a partir de dados financeiros e de mercado oriundos de fontes públicas. Ao longo da pesquisa, investigou-se em que medida a estrutura econômico-financeira, a governança corporativa e a percepção institucional refletida no mercado podem atuar como proxies quantitativas da sustentabilidade corporativa, contribuindo para a literatura de finanças sustentáveis e para aplicações práticas de inteligência artificial na gestão de risco empresarial.

Os resultados obtidos demonstram a viabilidade técnica e a consistência conceitual da abordagem proposta, e observou-se que o risco ESG, embora multifacetado e setorialmente condicionado, pode ser inferido de forma útil a partir de variáveis estruturadas quando analisado por modelos supervisionados interpretáveis, como Random Forest e XGBoost, combinados a técnicas de explicabilidade. A análise por SHAP foi essencial para tornar transparente o papel das variáveis e sustentar a coerência econômica das predições, reforçando a adequação metodológica do estudo.

O conjunto de análises de métricas globais de desempenho, curvas ROC e Precision–Recall, interpretações por classe, importância de atributos, explicações locais e globais e leituras setoriais, sustentam a hipótese central do trabalho: dimensões da sustentabilidade corporativa se articulam com saúde financeira, qualidade da governança e confiança institucional, e parte dessa interdependência pode ser capturada com razoável precisão por dados públicos estruturados. Ao mesmo tempo, os resultados evidenciam que a classe High permanece menos observável nesse tipo de base, o que preserva a necessidade de cautela interpretativa e de complementação informacional em aplicações de maior criticidade.

5.1 Síntese das Contribuições do Trabalho

Este trabalho apresenta contribuições relevantes em três dimensões complementares: científica, metodológica e aplicada.

Do ponto de vista científico, os achados reforçam empiricamente a literatura de finanças sustentáveis ao evidenciar que disciplina financeira e mecanismos de governança aparecem como pilares mensuráveis associados ao risco ESG em bases públicas. Em particular, os modelos baseados em árvores aprenderam padrões consistentes com a noção de que governança atua como eixo estruturante que conecta desempenho econômico e exposição a riscos ambientais e sociais, funcionando como componente-chave na sustentação de práticas corporativas ao longo do tempo.

No âmbito metodológico, a pesquisa propôs e validou uma estrutura analítica completa e replicável, integrando pré-processamento, tratamento de desbalanceamento, validação multisemente e comparação sistemática de algoritmos, com interpretabilidade via SHAP. Essa combinação supera abordagens estritamente preditivas ao incorporar leitura econômica e transparência sobre as decisões do modelo, requisito especialmente relevante para problemas de risco e decisões institucionais.

Adicionalmente, a análise de materialidade por setor evidenciou a heterogeneidade com que o risco ESG se manifesta nos dados quantitativos: setores ambientalmente intensivos e regulados apresentaram sinais mais claros de separação, enquanto setores baseados em intangíveis apresentaram maior limitação informacional. Essa evidência reforça o caráter contextual do ESG e a necessidade de análises sensíveis à materialidade setorial, evitando generalizações indevidas a partir de um único padrão agregado.

Por fim, a contribuição aplicada reside em demonstrar utilidade concreta dos modelos para triagem e suporte à decisão em gestão de risco corporativo, compliance e investimento responsável. A recorrência de variáveis associadas à alavancagem, liquidez, auditoria e presença institucional fornece subsídios operacionais para priorização de análise, monitoramento e alerta, aproximando ciência de dados e finanças sustentáveis em um mesmo arcabouço interpretável.

5.2 Atendimento aos Objetivos Propostos

Os objetivos estabelecidos no início do trabalho foram plenamente alcançados. O objetivo geral: desenvolver e avaliar modelos de aprendizado de máquina para classificação do risco ESG, foi atendido por meio da implementação, calibração e comparação de Random Forest, XGBoost e KNN, avaliados com métricas robustas, curvas ROC e Precision–Recall, e análises interpretativas baseadas em SHAP.

Quanto aos objetivos específicos, a pesquisa identificou de forma consistente variáveis relevantes para a predição do risco ESG, com destaque para indicadores financeiros, de governança e de mercado, e interpretou os resultados sob uma ótica econômica alinhada ao debate contemporâneo em sustentabilidade corporativa. A análise setorial permitiu compreender diferenças de materialidade e limitação informacional entre segmentos, enquanto a estrutura metodológica interpretável consolidou um arcabouço replicável e alinhado a boas práticas de modelagem responsável.

Dessa forma, o estudo não apenas atingiu seus objetivos formais, como também contribuiu para ampliar a compreensão sobre o uso de modelos interpretáveis em problemas ESG, evidenciando onde a inferência quantitativa é mais confiável e onde depende de complementação por dados qualitativos.

5.3 Limitações da Pesquisa

Apesar dos avanços alcançados, esta pesquisa apresenta limitações importantes. A principal decorre do uso de uma base predominantemente financeira e quantitativa, que não contempla indicadores ambientais e sociais diretos (como emissões de carbono ou métricas de diversidade), restringindo a captura de dimensões qualitativas do ESG especialmente em setores intensivos em ativos intangíveis e riscos reputacionais.

Outra limitação relevante refere-se ao desequilíbrio amostral, com sub-representação estrutural da classe High ESG Risk. Embora técnicas de balanceamento tenham sido aplicadas, persistem dificuldades na detecção de eventos raros e na separação do alto risco, o que é coerente com o caráter escasso e difuso desse sinal em bases públicas.

Adicionalmente, a análise possui caráter transversal e não incorpora dinâmica temporal do risco ESG. Mudanças graduais, deteriorações progressivas e choques reputacionais ao longo do tempo não são capturados por esse desenho, o que limita a utilidade do modelo para monitoramento longitudinal. Soma-se a isso a presença de multicolinearidade em proxies financeiras e a restrição de generalização entre setores, reforçando que o risco ESG é fortemente dependente do contexto econômico e regulatório.

Essas limitações não invalidam os resultados, mas delimitam seu alcance e reforçam a necessidade de cautela na interpretação, sobretudo quando o objetivo é detectar riscos severos com baixa observabilidade em dados estruturados.

5.4 Propostas para Trabalhos Futuros

Os resultados desta pesquisa abrem possibilidades relevantes de aprofundamento metodológico e ampliação empírica. Uma direção promissora consiste na integração de

dados qualitativos e não estruturados como notícias, relatórios corporativos e bases de controvérsias por meio de técnicas de NLP, ampliando a capacidade de capturar dimensões sociais e reputacionais pouco refletidas em variáveis financeiras.

Outra linha de avanço envolve a modelagem temporal do risco ESG. Arquiteturas sensíveis ao tempo, como LSTM (Long Short-Term Memory) e GRU (Gated Recurrent Unit, ou Unidade Recorrente com Portão), podem permitir monitorar evolução longitudinal, identificar transições de governança e detectar padrões de deterioração gradual, aproximando a modelagem da dinâmica real do risco e da precificação de incertezas ao longo do tempo.

Adicionalmente, a expansão da base e o desenvolvimento de modelos calibrados por setor ou região podem aumentar aderência operacional e reduzir vieses de generalização. Em paralelo, integrar as predições a frameworks consolidados de descoberta (como Sustainability Accounting Standards Board (SASB), Global Reporting Initiative (GRI) e Task Force on Climate-related Financial Disclosures (TCFD)) que permitem avaliar coerência entre práticas reportadas e sinais econômico-financeiros, contribuindo também para a detecção de potenciais inconsistências de divulgação.

Por fim, a operacionalização prática em painéis interativos e o uso de modelos híbridos combinando técnicas supervisionadas e não supervisionadas podem fortalecer a utilidade do arcabouço em contextos corporativos e regulatórios, aproximando ainda mais a pesquisa acadêmica das necessidades do mercado e da supervisão institucional.

Bibliografia

AROUSSI, R. *yfinance: Yahoo! Finance market data downloader*. 2025. Acessado em: 06 set. 2025. Disponível em: ⟨https://pypi.org/project/yfinance/⟩. Citado 2 vezes nas páginas page.3131 e page.4040.

CHAWLA, N. V. et al. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 2002. Disponível em: ⟨https://www.jair.org/index.php/jair/article/view/10302⟩. Citado 4 vezes nas páginas page.2121, page.2222, page.3131 e page.4444.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [s.n.], 2016. p. 785–794. Disponível em: ⟨https://arxiv.org/abs/1603.02754⟩. Citado 3 vezes nas páginas page.1818, page.2020 e page.3131.

COVER, T.; HART, P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, v. 13, n. 1, p. 21–27, 1967. Disponível em: ⟨https://isl.stanford.edu/~cover/papers/transIT/0021cove.pdf⟩. Citado 2 vezes nas páginas page.1717 e page.1818.

DELOITTE. *ESG and Credit Risk*. 2021. Acessado em: 06 set. 2025. Disponível em: ⟨https://www2.deloitte.com/content/dam/Deloitte/in/Documents/risk/in-ra-ESG-and-Credit-Risk-noexp.pdf⟩. Citado 7 vezes nas páginas page.55, page.77, page.88, page.99, page.1010, page.1111 e page.1313.

FEBRABAN. *Estudos em Sustentabilidade 2019*. 2019. Acessado em: 06 set. 2025. Disponível em: ⟨https://portal.febraban.org.br/pagina/3085/43/pt-br/estudos-sustentabilidade-2019⟩. Citado 3 vezes nas páginas page.1010, page.1111 e page.1212.

FEBRABAN. *Portal FEBRABAN*. 2025. Acessado em: 06 set. 2025. Disponível em: ⟨https://portal.febraban.org.br/⟩. Citado na página page.99.

GéRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2nd. ed. [S.l.]: O'Reilly Media, 2019. ISBN 978-1492032649. Citado 8 vezes nas páginas page.77, page.1313, page.1515, page.1616, page.2323, page.3030, page.3939 e page.4242.

HACKERNOON. *Creating a Systematic ESG Scoring System: Abstract and Introduction*. 2024. Acessado em: 06 set. 2025. Disponível em: ⟨https://hackernoon.com/creating-a-systematic-esg-scoring-system-abstract-and-introduction⟩. Citado na página page.1111.

HUNTER, J. D. et al. *Matplotlib: Visualization with Python*. 2025. Acessado em: 06 set. 2025. Disponível em: ⟨https://matplotlib.org/stable/contents.html⟩. Citado na página page.3030.

LEOBREIMAN. Random forests. *Machine Learning*, v. 45, p. 5–32, 2001. Disponível em: ⟨https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf⟩. Citado 4 vezes nas páginas page.66, page.1313, page.1414 e page.1616.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems (NeurIPS 2017)*. [s.n.], 2017. Disponível em: ⟨https://proceedings.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf⟩. Citado 4 vezes nas páginas page.1616, page.2424, page.2525 e page.3131.

MCKINNEY, W. *Python for Data Analysis*. 2nd. ed. [S.l.]: O'Reilly Media, 2017. ISBN 978-1491957660. Citado 2 vezes nas páginas page.1313 e page.3030.

PWCBRASIL. *Panorama do ESG nas cooperativas de crédito*. 2022. Acessado em: 06 set. 2025. Disponível em: ⟨https://www.pwc.com.br/pt/estudos/servicos/auditoria/2022/PwC_ESG_Cooperativas.pdf⟩. Citado 5 vezes nas páginas page.77, page.99, page.1010, page.1111 e page.1313.

QUINLAN, J. R. *C4.5: Programs for Machine Learning*. [S.l.]: Morgan Kaufmann, 1993. Citado na página page.1414.

SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, v. 10, n. 3, p. e0118432, 2015. Disponível em: ⟨https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118432⟩. Citado 6 vezes nas páginas page.2525, page.2727, page.4242, page.4444, page.4949 e page.5050.

SERASAEXPERIAN. *Analisar risco ESG dos produtores rurais pode proteger concedentes de crédito*. 2023. Acessado em: 06 set. 2025. Disponível em: ⟨https://www.serasaexperian.com.br/sala-de-imprensa/agronegocios/analisar-risco-esg-dos-produtores-rurais-pode-proteger-concedentes-de-credito-de-multas-de-ate-r90-bilhoes/⟩. Citado 4 vezes nas páginas page.99, page.1010, page.1111 e page.1313.

S&P Global Ratings,. *Definições e aplicação dos indicadores de crédito ESG*. 2021. Acessado em: 06 set. 2025. Disponível em: ⟨https://www.spglobal.com/_assets/documents/ratings/pt/pdf/2021-10-13-definicoes-e-aplicacao-dos-indicadores-de-credito-esg.pdf⟩. Citado 5 vezes nas páginas page.77, page.99, page.1010, page.1111 e page.1212.

United Nations Principles for Responsible Investment. *ESG in credit risk and ratings*. [S.l.], 2019. Acessado em: 06 set. 2025. Disponível em: ⟨https://www.unpri.org/download?ac=17119⟩. Citado 4 vezes nas páginas page.55, page.88, page.1010 e page.1111.

WASKOM, M. *Seaborn: Statistical Data Visualization*. 2025. Acessado em: 06 set. 2025. Disponível em: ⟨https://seaborn.pydata.org/⟩. Citado na página page.3030.

Yahoo Finance,. *Yahoo Finance Data API*. 2025. Dados coletados via biblioteca yfinance. Disponível em: ⟨https://finance.yahoo.com/⟩. Citado na página page.1212.

ANEXO – Repositório do Código-Fonte do Projeto

O código-fonte completo utilizado para o desenvolvimento do modelo de classificação de risco ESG está disponível no repositório público do GitHub no link abaixo:

Disponível em: ⟨https://github.com/lucasbitencourt/esg-risk-classification/tree/main/⟩

Acesso em: 04 mar. 2026.