



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

# Ponderação do efeito das externalidades climáticas no risco de crédito

William Oliveira Camelo

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Orientador  
Prof. Dr. João Gabriel de Moraes Souza

Brasília  
2023

Ficha catalográfica elaborada automaticamente,  
com os dados fornecidos pelo(a) autor(a)

0048p Oliveira Camelo, William  
Ponderação do efeito das externalidades climáticas no  
risco de crédito / William Oliveira Camelo; orientador João  
Gabriel de Moraes Souza. -- Brasília, 2023.  
143 p.

Dissertação(Mestrado Profissional em Computação Aplicada)  
-- Universidade de Brasília, 2023.

1. Externalidades. 2. Risco Climático. 3. Gestão de  
Risco. 4. Análise de Crédito. 5. Aprendizagem de Máquina. I.  
de Moraes Souza, João Gabriel, orient. II. Título.



Universidade de Brasília

Instituto de Ciências Exatas  
Departamento de Ciência da Computação

## Ponderação do efeito das externalidades climáticas no risco de crédito

William Oliveira Camelo

Dissertação apresentada como requisito parcial para conclusão do  
Mestrado Profissional em Computação Aplicada

Prof. Dr. João Gabriel de Moraes Souza (Orientador)  
CIC/UnB

Prof. Dr. João Carlos Félix Souza  
Universidade de Brasília

Dr. Mathias Schneid Tessmann  
Inst. Brasileiro de Ensino, Desenv. e Pesquisa

Prof. Dr. Gladston Luiz da Silva  
Coordenador do Programa de Pós-graduação em Computação Aplicada

Brasília, 20 de março de 2023

# Dedicatória

Dedico este trabalho a todos aqueles que por mim tiveram de se adaptar às minhas falhas e abdicar de seu bem mais preciso, seu tempo.

“Não corrigir as próprias falhas é cometer a pior delas.” - Confúcio.

# Agradecimentos

É fascinante como o desempenho dos algoritmos de aprendizagem de máquina se assemelham ao comportamento humano. Também podemos observar que a medida que esses algoritmos evoluem, seu desenvolvimento torna-se mais próximo do funcionamento dos mecanismos sociais que agregam ou impulsionam ideias. O estudo de aprendizagem de máquina nos ajuda a compreender que, assim como nos modelos, as formas de pensar não estão erradas, mas devem ser constantemente avaliadas para não sofrerem de anomalias que possam invalidar seu objetivo ou restringir sua utilização a casos específicos. Faz-se importante notar que existem afirmações de "consciência" em inteligências artificiais, às quais nos trazem mais questões a serem resolvidas sobre nossa existência do que respostas.

*"O maior inimigo do conhecimento não é a ignorância, é a ilusão do conhecimento."*  
- *Stephen Hawking.*

Dos mistérios que envolvem nossa existência, não tenho certeza, mas do meu universo, posso dizer que o mundo só faz sentido quando você descobre ser filho, pai, mãe, irmão ou amigo de alguém.

*"O Universo não seria grande coisa, se não fosse o lar das pessoas que amamos".* - *Stephen Hawking.*

Por isso, faço meus agradecimentos a minha mãe, Lúcia Eliza Tavares de Oliveira, meu pai, Waldecy Camelo, a meu irmão, Wesley Oliveira Camelo e família, pela trajetória que me deram nesta vida e por me possibilitarem estar aqui hoje. Também agradeço aos meus amigos Márcio da Silva Gama, Jorge Gildi e amigos que diretamente auxiliaram na conquista desta etapa. Estendo meus agradecimentos a meu orientador, professor Dr. João Gabriel de Moraes Souza, pelo apoio em todos os momentos que necessitei de seu auxílio na elaboração da dissertação e aos demais professores do mestrado que acreditarem em meu potencial.

# Resumo

Investimentos em operações financeiras podem impactar setores sociais, econômicos e ambientais. Estes impactos podem não estar incorporados ao preço final de seus contratos. Tais impactos são chamados de externalidades e podem produzir efeitos positivos ou negativos. A análise do risco de crédito socioambiental reside na possibilidade de atividades econômicas gerarem impactos ou danos socioambientais, capazes de causar perdas financeiras que comprometem o pagamento de financiamentos, bem como responsabilizar o agente financeiro a recuperar o dano socioambiental. No caso do risco climático, o empreendimento pode estar excessivamente exposto, diretamente ou indiretamente, a eventos climáticos extremos capazes de comprometer seu fluxo de caixa e capacidade de pagamentos. Por isso, veio a necessidade de instrumentos de medição capazes de avaliar o comportamento das externalidades em relação ao risco de crédito das operações financeiras. O objetivo deste trabalho é fornecer uma avaliação da influência das variáveis que compõem o risco climático ao compor o risco de crédito das instituições financeiras. Para isso, foram inseridos indicadores de externalidades que representassem o risco climático a uma base de 152.936 operações e 24 observações, utilizando as metodologias Taxonomia Verde e Régua de Sensibilidade ao Risco Climático. O comportamento das operações foi analisado de forma detalhada quanto ao risco da operação, risco do cliente, volume de crédito e quantidade de operações em relação aos indicadores de risco climático propostos. Os dados também foram submetidos aos modelos de machine learning: regressão Logística (Logit), Árvore de Decisão, AdaBoost, Random Forast, Gradient Boosting e XGBoost. O modelo Gradient Boosting atingiu melhores resultados em suas métricas e o modelo XGBoost ficou em segundo, porém, com maior área sob a curva ROC. A explicabilidade foi obtida por meio do método Shapley Value. O modelo Gradient Boosting não apresentou explicabilidade satisfatória. O modelo XBoost foi utilizado para explicabilidade do projeto. Como resultado, conclui-se que externalidades climáticas podem impactar significativamente o risco de crédito das operações e devem ser foco de atenção.

**Palavras-chave:** externalidades; gestão de risco; investimentos sustentáveis, risco climático; análise de crédito, aprendizagem de máquina, shapley value

# Abstract

Investments in financial operations can impact social, economic and environmental sectors. These impacts may not be incorporated into the final price of your contracts. Such impacts are called externalities and can produce positive or negative effects. The socio-environmental credit risk analysis resides in the possibility of economic activities generating socio-environmental impacts or damages, capable of causing financial losses that compromise the payment of financing, as well as making the financial agent co-responsible for recovering the socio-environmental damage. In the case of climate risk, the enterprise may be excessively exposed, directly or indirectly, to extreme weather events capable of compromising its cash flow and payment capacity. Therefore, there was a need for measuring instruments capable of assessing the behavior of externalities in relation to the credit risk of financial operations. The objective of this work is to provide an evaluation of the influence of the variables that compose climate risk when composing the credit risk of financial institutions. For this purpose, indicators of externalities that represent climate risk were inserted into a base of 152,936 operations and 24 observations, using the Green Taxonomy and Climate Risk Sensitivity Ruler methodologies. The behavior of operations was analyzed in detail regarding the risk of the operation, customer risk, volume of credit and number of operations in relation to the proposed climate risk indicators. The data were also submitted to machine learning models: Logistic regression (Logit), Decision Tree, AdaBoost, Random Forst, Gradient Boosting and XGBoost. The Gradient Boosting model achieved better results in its metrics and the XGBoost model was second, however, with a larger area under the ROC curve. Explainability was obtained using the Shapley Value method. The Gradient Boosting model did not show satisfactory explanation. The XBoost model was used to explain the project. As a result, it is concluded that climatic externalities can significantly impact the credit risk of operations and should be the focus of attention.

**Keywords:** externalities, risk management, sustainable investments, climate risk, credit analysis, machine learning, shapley value

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Problema da Pesquisa . . . . .	4
1.2	Pergunta de Pesquisa . . . . .	4
1.3	Objetivos . . . . .	5
1.3.1	Objetivo Geral . . . . .	5
1.3.2	Objetivos Específicos . . . . .	5
1.4	Estrutura de Pesquisa . . . . .	5
<b>2</b>	<b>Embasamento Teórico</b>	<b>7</b>
2.1	Referencial Teórico . . . . .	7
2.1.1	Processo de Busca . . . . .	7
2.2	Estabelecimento do Contexto . . . . .	16
2.2.1	Contextualização dos agentes internacionais . . . . .	16
2.2.2	Contextualização dos agentes nacionais . . . . .	18
2.2.3	Risco de Crédito . . . . .	22
2.2.4	Política de Gerenciamento de Riscos e Controles Internos . . . . .	24
<b>3</b>	<b>Metodologia</b>	<b>28</b>
3.1	Materiais e Métodos . . . . .	28
3.1.1	Classificação do Risco Climático: . . . . .	28
3.1.2	Processo de avaliação de riscos . . . . .	32
3.1.3	Modelo de explicabilidade da influência das variáveis do risco climático ao compor o risco de crédito . . . . .	32
3.1.4	Processamento dos Dados . . . . .	33
3.1.5	Técnicas de Predição e Classificação . . . . .	36
3.1.6	Execução . . . . .	68
<b>4</b>	<b>Resultados e Discussão</b>	<b>75</b>
4.1	Classificação da Exposição ao Risco Climático . . . . .	75
4.2	Processo de avaliação e tratamento dos riscos . . . . .	85

4.3	Modelos . . . . .	93
4.4	Explicabilidade . . . . .	97
<b>5</b>	<b>Considerações Finais</b>	<b>119</b>
5.1	Limitações . . . . .	119
5.2	Conclusões . . . . .	120
	<b>Referências</b>	<b>123</b>

# Lista de Figuras

1.1	Estrutura da pesquisa . . . . .	6
2.1	Acordo de Paris e Objetivos do Desenvolvimento Sustentável . . . . .	17
2.2	Processo de gestão de riscos . . . . .	25
2.3	Modelo Referencial de Linhas de Defesa . . . . .	26
3.1	Natureza das atividades e Raiting de crédito . . . . .	30
3.2	Risco Locacional e Risco Não-Locacional . . . . .	30
3.3	Relação entre Prazo da Operação e Volume . . . . .	31
3.4	Relação entre a Relevância e a Proporcionalidade . . . . .	32
3.5	Funcionamento do KNN . . . . .	35
3.6	Dados de treino e teste . . . . .	38
3.7	Underfitting . . . . .	39
3.8	Overfitting . . . . .	39
3.9	Modelo ajustado . . . . .	40
3.10	Ridge e Lasso . . . . .	43
3.11	Lasso, Ridge e Elastic Net . . . . .	45
3.12	Regressão Logit . . . . .	46
3.13	Árvore de Decisão . . . . .	48
3.14	Funcionamento da Árvore de Decisão . . . . .	49
3.15	Curva de Lorenz . . . . .	51
3.16	Entropia . . . . .	52
3.17	Gini vs Entropia . . . . .	54
3.18	Boosting Trees . . . . .	55
3.19	Bagging: processo de separação . . . . .	56
3.20	Random Forast . . . . .	58
3.21	Gradient Boosting . . . . .	59
3.22	Acurácia . . . . .	61
3.23	Precisão . . . . .	62
3.24	Especificidade . . . . .	63

3.25	Revocação (Recall)	64
3.26	Ponto de ótimo Corte	65
3.27	Curva 1	65
3.28	Curva 2	65
3.29	Curva 3	66
4.1	Matriz de Confusão e Curva ROC: Regressão Logística	94
4.2	Matriz de Confusão e Curva ROC: Árvore de Decisão	94
4.3	Matriz de Confusão e Curva ROC: Random Forast	95
4.4	Matriz de Confusão e Curva ROC: Gradient Boosting	95
4.5	Matriz de Confusão e Curva ROC: AdaBoost	96
4.6	Matriz de Confusão e Curva ROC: XGBoost	96
4.7	SHAP value: Sumário	98
4.8	Shapley Value: Exposição ao Risco Climático e Risco de Crédito	102
4.9	Shapley Value: Exposição ao Risco Climático e Risco da Operação AA	105
4.10	Shapley Value: Exposição ao Risco Climático e Valor do Contrato	107
4.11	Shapley Value: Exposição ao Risco Climático e Economia Verde Moderada Contribuição (Social + Ambiental)	109
4.12	Shapley Value: Exposição ao Risco Climático e Economia Verde Moderada Contribuição (Ambiental)	111
4.13	Shapley Value: Exposição ao Risco Climático e Economia Verde Alta Con- tribuição (Social)	113
4.14	Shapley Value: Exposição ao Risco Climático e Economia Verde (Baixa ou Nenhuma Contribuição)	115
4.15	Shapley Value: Exposição ao Risco Climático e Exposição ao Risco Ambi- ental (Baixa ou Nenhuma Exposição)	117

# Lista de Tabelas

2.1	Combinação dos termos de pesquisa . . . . .	8
2.2	Externalidades Socioambientais . . . . .	8
2.3	Risco de Crédito . . . . .	12
3.1	Camada de Operações . . . . .	29
3.2	Lasso vs Ridge . . . . .	44
3.3	Ganho de Informação vs Impureza de Gini . . . . .	54
3.4	Principais características entre Bagging, Boosting . . . . .	57
3.5	Matriz de Confusão . . . . .	61
4.1	Taxonomia Verde: Exposição ao Risco climático . . . . .	75
4.2	Taxonomia Verde: Economia Verde . . . . .	77
4.3	Economia Verde: Exposição ao Risco Climático . . . . .	78
4.4	Exposição ao Risco Climático: Agronegócio . . . . .	80
4.5	Classificação do Rating das Operações . . . . .	82
4.6	Classificação do Rating dos Clientes . . . . .	82
4.7	Agupamento dos dados . . . . .	82
4.8	Exposição ao Risco Climático: Cultivos e Bovinos . . . . .	82
4.9	Exposição ao Risco Climático e Risco da Operação . . . . .	83
4.10	Exposição ao Risco Climático e Risco do Cliente . . . . .	84
4.11	Classificação do Risco . . . . .	86
4.12	Rating da operação . . . . .	87
4.13	Locacional . . . . .	87
4.14	Clientes sem CNAE . . . . .	88
4.15	CNAE sem subtipo de operação . . . . .	89
4.16	Volume de operações . . . . .	90
4.17	Relevância do Cliente . . . . .	91
4.18	Melhorias operacionais . . . . .	92
4.19	Avaliação de desempenho . . . . .	93

# Lista de Abreviaturas e Siglas

**ASG** Ambiental, Social e Governança.

**BACEN** Banco Central do Brasil.

**BID** Banco Interamericano de Desenvolvimento.

**BNDES** Banco Nacional de Desenvolvimento Econômico e Social.

**CMN** Conselho Monetário Nacional.

**CNAE** Códigos nacionais de atividade econômica.

**Febraban** Federação Brasileira de Bancos.

**FSB** Conselho de Estabilidade Financeira.

**GEE** Gases geradores do efeito estufa.

**IBGE** Instituto Brasileiro de Geografia e Estatística.

**IFC** International Finance Corporation.

**IVDNS** Índice de vulnerabilidade aos desastres naturais relacionados às secas no contexto da mudança do clima.

**LGPD** Lei Geral de Proteção dos Dados Pessoais (Lei no13.709/18).

**MAPA** Ministério da Agricultura, Pecuária e Abastecimento.

**MRLD** Modelo Referencial de Linhas de Defesa.

**ODS** Acordo de Paris e Objetivos do Desenvolvimento Sustentável.

**PIB** Produto Interno Bruto.

**PRI** Princípios para o Investimento Responsável.

**PRSA** Política de Responsabilidade Socioambiental.

**PSI** Princípios do Equador, os Princípios para Sustentabilidade em Seguro.

**RCP** cenários de altas emissões.

**RSRC** Régua de Sensibilidade ao Risco Climático.

**SCR** Sistema de Informações de Crédito.

**SFN** Sistema Financeiro Nacional.

**Sire** Secretaria de Inteligência e Relações Estratégicas da Embrapa.

**TCFD** Task Force on Climate-related Financial Disclosures.

**WWF** World Wide Fund for Nature Inc..

# Capítulo 1

## Introdução

O estudo de indicadores de externalidades surgiu para elucidar os impactos nos setores sociais, econômicos e ambientais que não estão diretamente envolvidos com os contratos de investimentos [Vasconcellos & Garcia, 2014]. Entende-se que externalidades são os efeitos positivos ou negativos não atrelados ao cálculo de viabilidade de um contrato de investimento [Salvatore, 2000].

Segundo [Mota, 2014], os fatores que influenciam os investimentos não ficam restritos a apenas uma área, eles podem sofrer interferências de diversos cenários e fontes.

Desse modo, na ausência de práticas de medidas preventivas, as atividades econômicas podem estar sujeitas a externalidades com proporções capazes de gerar perdas financeiras que comprometem sua capacidade financeira, incorrer em problemas judiciais e de imagem, com conseqüente responsabilização dos agentes econômicos sobre os efeitos das externalidades negativas sobre a sociedade [Paiva, 2003].

Em um contexto econômico, as instituições financeiras também estão sujeitas a uma variedade de impostos corretivos, conhecidos como *taxa Pigouviana*, com o objetivo de punir agentes causadores de lucros indevidos, alinhar incentivos a benefícios sociais, entre outros motivos. Porém, percebe-se que a maior parte desses impostos tem sido aplicada a externalidades ambientais [Shackelford et al., 2010]. Essa percepção corrobora o observado por [Painter, 2020], em que países com maior probabilidade de serem afetados por eventos ambientais extremos possuem custos mais elevados para emissão de títulos de longo prazo.

Também observa-se que o investidor aumentou sua atenção sobre os riscos envolvidos com as mudanças climáticas desde a publicação do Relatório Stern de 2006, que informa que alterações climáticas irão impactar profundamente os investimentos com efeitos em toda economia mundial. Esse relatório também informa em uma de suas principais conclusões que com apenas 1% de investimento do **Produto Interno Bruto (PIB)** investidos em ações preventivas, poderiam ser evitados 20% de perdas do mesmo **PIB** [Stern, 2006].

Essas percepções afetam diretamente o risco de crédito, que precisa ser minimizado por uma gestão de risco baseada em procedimentos que comportem análises subjetivas e objetivas com capacidade de fornecer dados para adequação, seleção, análise, precificação e monitoramento da inadimplência, em caso de ocorrência de fatores adversos [Santos, 2007].

O risco de crédito associado ao socioambiental também possui a premissa de que investidores esperam retornos mais altos, e os credores cobram taxas também mais elevadas em investimentos que não estão associados à preocupações ambientais, ao serem comparados com empresas que possuem tais preocupações [Chava, 2014].

Segundo [Ali et al., 2020], as instituições financeiras que mantêm práticas internas e externas de gestão de cadeia de suprimentos verdes têm impactos positivos no desempenho ambiental e experimentam melhor desempenho financeiro.

Também é observado que a divulgação de informações ambientais tem o poder de influenciar as decisões de alocação de investimentos em diferentes horizontes de investimentos e investidores [Holm & Rikhardsson, 2008].

A importância da conscientização e divulgação dessas informações cresce a medida que passamos por mudanças climáticas. A 16ª edição do Índice de Risco Climático, publicado em 2021, alerta que as mudanças climáticas atingiram um patamar que não pode ser mais ignorado por nenhuma região, fornece indicadores quanto ao nível de exposição e vulnerabilidade a eventos climáticos extremos e informa que a mitigação e adaptação a esses eventos é do interesse de todos os países do mundo [Eckstein et al., 2021].

Segundo [Huang et al., 2018], o empreendimento pode estar excessivamente exposto diretamente ou indiretamente a eventos climáticos extremos capazes de comprometer o seu fluxo de caixa e capacidade de pagamento de financiamentos, mas poucos estudos examinaram diretamente os impactos do risco relacionado ao clima nas instituições financeiras de capital aberto ou fechado.

As externalidades climáticas têm o potencial de impactar o risco de crédito de forma significativa. Pois em matéria de riscos, a exposição às mudanças do clima é considerada como uma grande ameaça e tem a capacidade de potencializar os riscos já existentes [Krepsky & Scherer, 2019].

No Brasil, a preocupação com as questões ambientais é refletida em uma legislação robusta, estando presente em sua carta magna, a constituição federal de 1988, trazendo em seu artigo 225 a preocupação com a legislação ambiental, e na Lei de Crimes Ambientais (Lei 9.605/98) que estabeleceu a punibilidade das contravenções dentre diversos outros dispositivos legais [Borges et al., 2009].

Mas, embora tenhamos uma legislação complexa, ela possui brechas apontadas pelo relatório [ONU Meio Ambiente](#) que constatou uma incapacidade em implementar e fazer

cumprir a legislação, fazendo com que a fraca aplicação seja uma tendência e, com isso, auxiliando no agravamento dos problemas ambientais [BPBES, 2019].

Isso induz o mercado nacional a uma preferência por setores que são de consumo intensivo de recursos naturais, pois a sobreposição da legislação ambiental pelos interesses econômicos de empresas ou influências políticas posiciona o direito ambiental em segundo plano de importância, o que acaba beneficiando os criminosos com aplicações não severas da lei, nos casos em que não há a ausência de aplicação [MARION, 2013].

Como exemplo dos efeitos de externalidades negativas que podem ser geradas por investimentos sem instrumentos de avaliação de externalidades ambientais, podemos citar Cubatão, localizada no estado de São Paulo, que ficou conhecida mundialmente como "Vale da morte", por ter sido classificada por vários anos como a cidade mais poluída do mundo, segundo a ONU. Cubatão também foi exemplo ao aplicar a legislação ambiental associada a instrumentos de monitoração de investimentos, desse modo, foi reconhecida como símbolo de recuperação na conferência Eco-92, perdendo seu posto de cidade mais poluída do mundo e em 2014 deixou de ser considerada a cidade mais poluída do Brasil [BBC, 2017].

Segundo [Verdiano et al., 2017] a legislação ambiental ainda não atingiu seu objetivo e precisa de um esforço para suprir as diversas brechas existentes.

Essa situação se agrava ao ser observado que o Brasil é afetado pelas mudanças climáticas atualmente e será ainda mais vulnerável em projeções futuras. Tendo como áreas mais vulneráveis a Amazônia e o Nordeste [Marengo, 2008].

Ter instrumentos de projeções de riscos considerando as mudanças climáticas se torna cada vez mais urgente para a economia do país que foi considerado em 2020 como sendo o terceiro maior produtor de alimentos e fibras do mundo e é o segundo maior exportador do agronegócio mundial. O **Ministério da Agricultura, Pecuária e Abastecimento (MAPA)** informou que o agronegócio foi responsável por uma participação recorde de 48% do superavit nacional em 2020. Portanto, conhecer os cenários hidrológicos futuros e suas incertezas são cruciais para economia brasileira [FieldView, 2022].

Tendo o exposto, o presente estudo oferece aprimoramento dos instrumentos de medição e exposição às externalidades climáticas sobre as operações de crédito das instituições financeiras ao avaliar de forma detalhada o impacto da incorporação da medição do risco climático às operações financeiras no processo estratégico. Para isso, uma base de 152.936 operações e 24 observações foi utilizada para analisar o comportamento das operações quanto ao risco da operação, o risco do cliente, o volume de crédito e quantidade de operações em relação aos indicadores de risco climático propostos. Os indicadores incorporados utilizaram a metodologia Taxonomia Verde e a **Régua de Sensibilidade ao Risco Climático (RSRC)** para sua concepção, ambos as metodologias foram fornecidas

pela **Federação Brasileira de Bancos (Febraban)**. Após as análises quantitativas, os dados foram submetidos aos modelos de machine learning que tratam problemas de classificação: Regressão Logística (Logit), Árvore de Decisão, AdaBoost, Random Forest, Gradient Boosting e XGBoost. Os resultados do modelo com melhor desempenho foram avaliados para aferir a importância das variáveis que compõem o risco climático e assim inferir seu impacto ao compor os indicadores de risco de crédito das operações financeiras por meio do método Shapley Value. Os riscos e desafios enfrentados durante a utilização da metodologia proposta também foram identificados e avaliados.

## 1.1 Problema da Pesquisa

O risco climático vem sendo considerado por diversas entidades internacionais como um dos mais relevantes em termos de impacto e probabilidade. Com potencial de ruptura de cadeias globais de produção, este risco é considerado um dos riscos relevantes para o conglomerado de instituições financeiras, que enfrentam desafios para emissão de relatórios que elucidem ao mercado e aos investidores os riscos que suas operações estão sujeitas. A urgência em fornecer dados sobre os impactos causados pelas mudanças climáticas aumenta a medida que estudos mostram que a parcela das carteiras de crédito expostas a esse risco é comparável ao capital das instituições [Battiston et al., 2017].

Segundo [Roncoroni et al., 2021], uma política de transição desonerada com atenção aos riscos ambientais e condições de mercado mais fortes permitem alcançar políticas climáticas mais ambiciosas mantendo o mesmo nível de risco financeiro. A forma como as mudanças climáticas impactarão direta e indiretamente os sistemas agrícolas e alimentares e as vulnerabilidades relacionadas é permeada de incertezas. Por isso, Construir resiliência agora é fundamental para estar preparado para mudanças futuras [Young et al., 2012].

Este estudo visa a elaboração de ferramenta capaz de classificar as operações quanto ao grau de exposição ao risco climático e elucidar o impacto de suas variáveis sobre o risco de crédito das operações. Dessa forma, com os dados disponibilizados, será possível elaborar relatórios descritivos do impacto do risco climático sobre as operações, detalhar a influência que cada variável que o compõe possui no risco de crédito e fornecer projeções para serem usadas como insumo na configuração da carteira de crédito das instituições.

## 1.2 Pergunta de Pesquisa

- A variável de risco climático impacta negativamente o risco de crédito das operações de investimento?

## **1.3 Objetivos**

### **1.3.1 Objetivo Geral**

O objetivo geral deste trabalho é fornecer uma avaliação da influência das variáveis que compõem o risco climático ao compor o risco de crédito das instituições financeiras.

### **1.3.2 Objetivos Específicos**

Para atingir o objetivo geral, será necessário realizar os seguintes objetivos específicos:

1. Identificar na literatura as metodologias utilizadas para classificar os riscos envolvendo os temas ambientais e seu impacto no risco de crédito.
2. Investigar os principais desafios que envolvem a classificação do risco climático.
3. Gerar indicadores ambientais que representem o risco climático.
4. Analisar o comportamento das operações quanto ao risco da operação, risco do cliente, volume de crédito e quantidade de operações em relação aos indicadores de risco climático propostos.
5. Realizar processo de avaliação e tratamento dos riscos envolvidos na classificação do risco climático.
6. Desenvolver modelo capaz de explicar a influência das variáveis utilizadas na classificação do risco climático ao compor o risco de crédito.
7. Fornecer análise detalhada do modelo.
8. Operacionalizar o processo de medição e avaliação do risco climático.

## **1.4 Estrutura de Pesquisa**

A pesquisa foi estruturada em cinco capítulos assim como é referenciado na Figura 1.1.



Figura 1.1: Estrutura da pesquisa

Fonte: Produzido pelo autor

**Capítulo 1** : *Introdução*. Introduce o tema do estudo, apresentando os motivos que levaram ao seu desenvolvimento e o esperado ao final dos trabalhos.

**Capítulo 2** : *Embasamento Teórico*. Apresentação das técnicas utilizadas para seleção dos estudos considerados de relevância e capazes de elucidar os conceitos necessários para entendimento deste trabalho.

**Capítulo 3** : *Metodologia*. Descreve o que será feito durante os trabalhos, os materiais e métodos utilizados para auferir e inferir os resultados.

**Capítulo 4** : *Resultados e Discussão*. Aplicação da metodologia proposta, comparando os resultados aos estudos relacionados e resultados esperados.

**Capítulo 5** : *Considerações Finais*. Informa as conclusões finais das análises realizadas, apresenta os riscos e desafios encontrados para concepção do trabalho.

# Capítulo 2

## Embasamento Teórico

### 2.1 Referencial Teórico

Para revisão dos trabalhos, deve-se identificar os trabalhos relacionados ao tema estudado, com a finalidade de gerar conhecimentos que acrescentem valor ao estudo. Os trabalhos foram revisados de acordo com sua capacidade de elucidar e interpretar com clareza as questões levantadas para o estudo proposto [Kitchenham & Charters, 2007].

#### 2.1.1 Processo de Busca

Foram relacionados trabalhos que tivessem a capacidade de explicar a relevância do estudo das externalidades ambientais e elucidar sua relação com os ganhos e perdas financeiras. Desse modo, o processo de busca foi dividido em dois grupos: *Externalidade Socioambiental*, que elucidada a importância do estudo das externalidades ambientais e como elas podem afetar os compromissos das instituições, gerando riscos ou ampliando vulnerabilidades existentes; e *Risco de Crédito*, que elucidada as formas como as instituições buscam identificar, analisar e medir suas vulnerabilidades para compor o risco de crédito, inclusive com técnicas de aprendizagem de máquina.

O filtro foi realizado em julho de 2021, utilizando as palavras "*Externalities*", "*Risk Management*", "*Sustainable Investments*", "*Climate Risk*", "*Credit Analysis*", "*Machine Learning*" e "*Shapley Value*" como alternativas, dando preferência para artigos com período máximo de dez anos de publicação nas plataformas Web of Science, Scopus e Google Scholar, separados dentro das categorias "Externalities", "Risk Management", "Climate Risk" e "Sustainable Investments".

O processo de separação pode ser visto na Tabela 2.1, que detalha de forma exemplificativa como foram utilizados os termos descritos e seus respectivos resultados. Os resultados foram contabilizados e estudados de acordo com sua relevância.

Tabela 2.1: Combinação dos termos de pesquisa

<b>Palavras-Chave</b>	<b>Base de Dados</b>
Categorias: "Externalities"AND "Risk Management"OR "Climate Risk"AND "Sustainable Investments"	
"Externalities"	10.326
"Externalities"AND "Risk Management"	373
"Externalities"AND "Risk Management"AND "Sustainable Investments"	11
"Credit Risk"	7.820
"Credit Risk"AND "Credit Analysis"	2,338
"Credit Risk"AND "Credit Analysis"AND "Machine Learning"	34
"Credit Risk"AND "Credit Analysis"AND "Shapley Value"	1

Fonte: Elaborado pelo autor

Do total, após a separação e retirada dos artigos repetidos ou considerados fora do escopo, foram escolhidos dez artigos para cada categoria, assim como é ilustrado nas Tabelas 2.2 e 2.3.

Tabela 2.2: Externalidades Socioambientais

<b>Autores</b>	<b>Síntese</b>
----------------	----------------

<p>Environmental Externalities and Cost of Capital. [Chava, 2014]</p>	<p>Faz uma análise do impacto do perfil ambiental de uma empresa em seu custo de capital próprio e capital de terceiros. Relata que investidores exigem retornos esperados significativamente mais altos sobre ações excluídas por telas ambientais e mudanças climáticas em comparação com empresas sem tais preocupações ambientais e que credores também cobram uma taxa de juros significativamente mais alta nos empréstimos bancários concedidos a empresas com essas preocupações ambientais. Evidencia que o perfil ambiental de uma empresa não é simplesmente um componente omitido de risco de inadimplência.</p>
<p>Internalization of environmental externalities: Development of a method for elaborating the statement of economic and environmental results. [Eidelwein et al., 2018]</p>	<p>A internalização das externalidades ambientais pelos resultados econômicos das organizações torna-se relevante à medida que os gestores percebem os impactos e riscos que a degradação da natureza exerce sobre seus negócios. Avanços nos métodos de internalização das externalidades ainda são escassos. O estudo aplicou uma metodologia com foco em sistemas produtivos pra identificar que as externalidades ambientais representaram um custo equivalente a 12,5% da receita líquida consolidada.</p>
<p>Valuing the environmental externalities of rice and wheat cultivation in the lower reaches of the Yangtze River. [Lv et al., 2010]</p>	<p>Informa que externalidades ambientais geradas pela agricultura atraem atenção considerável, mas a maioria das pesquisas está focadas somente nos efeitos negativos ao meio ambiente. Por isso, há uma grande necessidade de reavaliar os papéis ambientais gerais da agricultura para otimizar seu desempenho ambiental, ampliar e estimular formulação de políticas. Analisou as externalidades do cultivo de arroz-trigo em Zhenjiang para fornecer um procedimento analítico com a medição de dimensões físicas e avaliação monetária. Como resultado apresentou que a colheita de arroz estudada possui enormes externalidades ambientais positivas, mesmo contendo impactos ambientais negativos.</p>

<p>Valuing the environmental externalities of oasis farming in Left Banner. [Wei et al., 2010]</p>	<p>Este estudo investigou uma abordagem para estimar as dimensões físicas das externalidades ambientais de um sistema de cultivo de milho em oásis no noroeste da China e a valoração monetária das externalidades ambientais. Os resultados mostram um custo-benefício social pequeno frente aos custos agrícolas.</p>
<p>Foreign direct investments, environmental externalities and capital segmentation in a rural economy. [Antoci et al., 2015]</p>	<p>Este artigo examina os possíveis efeitos dos fluxos de investimento externo no desenvolvimento das economias rurais locais, levando em consideração duas características recorrentes de muitos países em desenvolvimento: segmentação do mercado de capitais e externalidades ambientais. Compara a diferença da atividade produtiva do setor externo e local e também informa as condições sob as quais o capital pode promover diversidade e melhorar o bem-estar das populações locais.</p>
<p>Integrating ecology and economics in understanding responses in securing land use externalities internalization in water catchments. [Sanga &amp; Mungatana, 2016]</p>	<p>Informa ser um desafio internalizar externalidades em bacias hidrográficas dos países em desenvolvimento. Desenvolveu um modelo com base em estruturas dinâmicas de sistemas para avaliar os impactos a longo prazo da captação de água na bacia de água Uluguru. Comparou o modelo com outros instrumentos econômicos. Como resultado informa que a internalização das externalidades sem comprometer a distribuição de benefícios e bem-estar dos proprietários é mais eficiente em com o subsídio downstream-upstream.</p>
<p>Wind power externalities: A meta- analysis. [Mattmann et al., 2016]</p>	<p>Apresenta a primeira meta-análise quantitativa da literatura de avaliação não mercantil sobre os efeitos externos associados à produção de energia eólica. Analisa valores econômicos relativos de diferentes tipos de externalidades. Os resultados indicam significativos efeitos positivas e negativas das externalidades visuais e estimativas de bem-estar. Mostrou que a política de promoção de turbinas eólicas com outras políticas verdes facilitam a expansão da energia eólica.</p>

<p>Fair trade, Agrochemical Input Use, and Effects on Human Health and the Environment.</p> <p>[<a href="#">Sellare et al., 2020</a>]</p>	<p>O estudo informa que padrões voluntários de sustentabilidade podem melhorar o bem-estar socioeconômico de pequenos agricultores em países em desenvolvimento e ajudar a reduzir impactos negativos da produção agrícola na saúde e no meio ambiente. Também relata que os padrões de sustentabilidade focavam apenas em indicadores econômicos e argumenta que o Comércio Justo (FairTrade) e outros padrões podem afetar o uso e insumos agroquímicos por meio de efeitos positivos e negativos à saúde e meio ambiente. Avaliou os dados de agricultores e trabalhadores rurais na Costa do Marfim para analisar os efeitos do Comércio Justo quanto ao uso de fertilizantes e pesticidas e seu impacto na saúde e meio ambiente. Identificou que o Comércio Justo aumentou a quantidade de insumos químicos e níveis agregados de toxicidade, porém, reduziu a incidência de sintomas agudos relacionados a pesticidas entre os avaliados. Os resultados sugerem que suposições simplistas sobre os efeitos dos padrões de sustentabilidade na saúde e no meio ambiente podem ser inadequadas.</p>
<p>Integrating life-cycle assessment and choice analysis for alternative fuel valuation.</p> <p>[<a href="#">Winden et al., 2014</a>]</p>	<p>Monetiza os danos de externalidades ambientais e seu risco à saúde associado ao uso do biocombustível etanol a base de milho. Os resultados revelam que a produção e consumo de etanol de amido de milho produzem reduções nos resultados ambientais e na saúde.</p>
<p>Evaluation of social externalities in regional communities affected by coal seam gas projects: A case study from Southeast Queensland.</p> <p>[<a href="#">Phelan et al., 2017</a>]</p>	<p>Examina a avaliação das externalidades sociais em comunidades afetadas por quatro grandes projetos de gás de camada de carvão na região da Bacia de Surat, no sudeste de Queensland, na Austrália. Utiliza abordagem de método misto transversal e seus resultados mostram que as percepções da comunidade sobre a desigualdade econômica, senso coletivo de incerteza e futuros impactos no padrão de vida indica um aumento das preocupações e perda de confiança no governo local, inibindo a capacidade da comunidade se planejar para o futuro.</p>

Fonte: Elaborado pelo autor

Tabela 2.3: Risco de Crédito

Autores	Síntese
Análise do modelo creditrisk+ em uma amostra de portfólio de crédito. [Mileo et al., 2013]	Analisa os fundamentos teóricos e o desempenho do Modelo CreditRisk+, metodologia de gestão de risco de crédito criada por bancos, em uma amostra de portfólio de crédito. Compara as medidas de avaliação de risco de perdas de determinado ano aos dados de perda simulados para o ano posterior, e realiza testes de estresse, para verificar a sensibilidade do Modelo a mudanças no cenário econômico. Os resultados indicam que o modelo CreditRisk+ subestimou os resultados analisados.
Modelo de classificação de risco de crédito de empresas. [Brito & Assaf Neto, 2008]	Diversas novas técnicas de mensuração de riscos e tomadas de decisão têm sido desenvolvidas. Nesse contexto, a pesquisa desenvolveu um modelo de classificação com base na amostra do mercado brasileiro entre 1994 e 2004. Utilizou a regressão logística e sua validação fez uso do modelo Jackknife e Curva Roc. Seus resultados indicam que o modelo de classificação mostrou melhores resultados de previsão.
An empirical comparison of machine-learning methods on bank client credit assessments. [Munkhdalai et al., 2019a]	Estabelece comparações entre os modelos baseados em especialistas humanos e aprendizagem de máquina utilizando dados reais de consumidores e fornecedores fazendo o uso de vários algoritmos de aprendizagem de máquina, desde regressão logística ao redes neurais profundas e XGBoost. Os resultados mostram que se as instituições tivessem usado modelos de aprendizagem de máquina, suas perdas teriam sido menores.
Regulatory learning: How to supervise machine learning models? an application to credit scoring. [Guégan & Hassani, 2018]	Informa que a estrutura atual de abordagens escolhidas pelas instituições é incompatível com a chegada do Big Data e que a seleção de modelos pode evoluir dinamicamente. Desas forma, faz um análise dos problemas atuais relacionados ao ambiente de Big Data e modelos de aprendizagem de máquina e destaca os presentes problemas da estrutura atual.

<p>A novel gsci-based ensemble approach for credit scoring. [Chen et al., 2020]</p>	<p>Apresenta um novo modelo heterogêneo baseado em Shapley generalizado e na integral de Choquet. Com base no operador de agregação definido e no modelo de programação linear, uma abordagem GSCI é proposta para o ensemble credit scoring. Ilustra a eficiência e viabilidade da abordagem GSCI com treze modelos de aprendizado de máquina em quatro conjuntos de dados públicos de pontuação de crédito e três dados líderes P2P do mundo real. Além disso, testes robustos e análises comparativas são feitos para demonstrar a adaptabilidade e desempenho do modelo ensemble baseado em GSCI.</p>
<p>Dynamic ensemble classification for credit scoring using soft probability. [Feng et al., 2018]</p>	<p>Informa que classificadores múltiplos têm alcançado desempenhos melhores que individuais. Mas que ensembles seletivos não consideram os custos relativos aos erros Tipo I e II para pontuações de crédito ao selecionarem classificadores que trazem maiores riscos. Além disso, as seleções dos conjunto de dados são baseadas no desempenho do classificador, independente de seu comportamento no conjunto de teste. Este estudo visa preencher essa lacuna e superar limitações ao propor um método de classificação de conjuntos dinâmicos para pontuação de crédito com base em probabilidade suave. Os classificadores são combinados por amostras no conjunto de teste com base em seus resultados e comparados individualmente com classificadores conhecidos. Os resultados experimentais demonstram a capacidade e eficiência do método proposto para melhorar o desempenho de previsão em relação aos modelos de referência.</p>

<p>Explainable machine learning in credit risk management.</p> <p>[Bussmann et al., 2021]</p>	<p>Propôs um modelo explicável de inteligência artificial que pode ser usado na gestão do risco de crédito. O modelo aplica redes de correlação aos valores de Shapley para que as previsões sejam agrupadas de acordo com as similaridades nas explicações subjacentes. Informa que tomadores de risco podem ser agrupados por conjuntos e características financeiras semelhantes que podem ser utilizadas para explicar sua pontuação de crédito e prever seu comportamento.</p>
<p>Statistical and machine learning models in credit scoring: A systematic literature survey.</p> <p>[Dastile et al., 2020]</p>	<p>Informa que o método de regressão logística é tradicional na previsão de crédito. No entanto, na literatura já podem ser encontrados modelos de aprendizagem de máquina capazes de substituir suas previsões. Apesar dos avanços, existem problemas relacionados a explicabilidade dos resultados, havendo a necessidade de uma pesquisa completa na literatura para revisar os modelos estatísticos e de aprendizagem de máquina em pontuações de crédito. Este estudo se baseou em 74 estudos publicados entre 2010 e 2018 e descobriu que um conjunto de classificadores tem melhor desempenho do que um único.</p>

<p>Bank green lending and credit risk: an empirical analysis of china's green credit policy. [Zhou et al., 2020]</p>	<p>Este estudo investiga empiricamente a relação entre os empréstimos verdes dos bancos e seu risco de crédito e como os regulamentos chineses de finanças verdes contribuem para a solvência de bancos individuais e a resiliência do sistema financeiro, ao analisar 41 bancos chineses de 2007 a 2018. Investigou os riscos financeiros associados aos empréstimos verdes e se um aumento nos empréstimos verdes diminuiria o índice de empréstimos inadimplentes. Utilizou análises de regressão para verificar o efeito do aumento dos créditos verdes na inadimplência. Os resultados revelam que o índice de empréstimos verdes tem um impacto negativo no índice de inadimplência e introduz um novo valor para a literatura atual sobre o impacto das políticas de empréstimo verde e fornece uma nova perspectiva que suporta a sustentabilidade financeira nos Emirados Árabes Unidos.</p>
<p>The drivers of systemic risk in financial networks: a data-driven machine learning analysis. [Alexandre et al., 2021]</p>	<p>O objetivo deste artigo foi avaliar o papel das variáveis financeiras e da topologia de rede como determinantes do risco sistêmico (SR), que foi calculado aplicando a metodologia diferencial DebtRank. Os determinantes de SR específicos foram avaliados por XGBoost e floresta aleatória (random forest) e sua interpretabilidade utilizou o método Shapley. O estudo relata que a importância de uma determinada característica na condução do SR pode variar e com isso variando a dimensão do choque que terá o risco que está sendo avaliado, quer seja por perdas de impacto sistêmico ou vulnerabilidades. A pesquisa também concluiu que, enquanto a importância das características topológicas para a previsão do impacto sistêmico dos bancos aumenta com o nível do choque inicial, ela diminui para as cooperativas de crédito.</p>

Fonte: Elaborado pelo autor

## 2.2 Estabelecimento do Contexto

O contexto em que a pesquisa está inserida é estabelecido pelo relacionamento das diversas entidades internacionais e nacionais que lidam com os impactos gerados pelas mudanças climáticas e também pelas instituições responsáveis por monitorar o risco de crédito dos meios de pagamento. Por esse motivo, entender os conceitos ligados ao risco climático e risco de crédito torna-se necessário para esta pesquisa.

Tendo o exposto, este capítulo aborda os conceitos fundamentais para entendimento desta pesquisa, tais como a contextualização das principais entidades responsáveis pelo monitoramento do risco climático, os órgãos nacionais reguladores do risco de crédito das instituições financeiras e demais conceitos relevantes.

Essas informações foram separadas em um contexto macro de risco de crédito e inseridas ao contexto da instituição financeira foco deste estudo. Portanto, serão apresentados os principais acordos internacionais que tratam do assunto, como o sistema financeiro nacional está estruturado para atender às iniciativas internacionais e a visão do mercado. Também serão explicados os conceitos relacionados ao risco climático, risco socioambiental e risco agroclimático. Por fim, serão detalhadas as etapas que compõe a formação do risco de crédito e as linhas de defesa que tratam de eventuais oscilações financeiras.

Entender esses conceitos é importante para compreender como os mecanismos do risco de crédito das instituições financeiras se relacionam com o tema do estudo.

### 2.2.1 Contextualização dos agentes internacionais

A Organização das Nações Unidas estabelece os três pilares (econômico, social e ambiental) para o desenvolvimento sustentável dos países e formaliza o **Acordo de Paris e Objetivos do Desenvolvimento Sustentável (ODS)** com a Agenda 30, composta por 17 objetivos com 169 metas e 232 indicadores que apontam a urgência de direcionar a sociedade em um caminho mais sustentável. Esses objetivos são um apelo para acabar com a pobreza, proteger o meio ambiente e o clima, e garantir que as pessoas possam desfrutar de paz e prosperidade. O Pacto Global tem a missão de engajar as empresas para esta nova agenda de desenvolvimento, pois entende que o setor privado tem um papel essencial nesse processo por ser grande detentor do poder econômico, propulsor de inovações e tecnologias que influenciam e engajam os mais diversos públicos, desde governos, fornecedores, colaboradores e consumidores. A Figura 2.1 ilustra os **ODS** que regem sobre as iniciativas internacionais de melhores práticas sociais, ambientais, econômicas e institucionais, incluindo sugestões de instrumentos de medição e avaliação de impactos.



Figura 2.1: Acordo de Paris e Objetivos do Desenvolvimento Sustentável  
Fonte: [Evolverde](#)

Segundo a [ONU, 2015], os objetivos sustentáveis representam uma abordagem que contempla as necessidades sociais e proteção de direitos. Para isso, incluem medidas de proteção dos ecossistemas, criação de cidades sustentáveis, integração de políticas de mudanças climáticas em estratégias nacionais, enfrentamento da violência, abuso, exploração e tráfico, bem como medidas que estabelecem o fim da desnutrição, o fortalecimento das redes de saúde, acesso universal à água, saneamento e educação. Esses objetivos visam o futuro e a pleno desenvolvimento das novas gerações. Seu conteúdo pode ser encontrado em sua íntegra no site [Nações Unidas Brasil](#).

Dentre esses objetivos, a ação número 13 prevê a adoção de medidas urgentes para combater as alterações climáticas e seus impactos, ao apontar a necessidade de reforçar as capacidades de adaptação aos riscos relacionados ao clima e às catástrofes naturais, bem como visar a necessidade de criação de mecanismos que auxiliem uma gestão eficaz considerando as mudanças climáticas.

Para auxiliar nessa questão, temos a [Task Force on Climate-related Financial Disclosures \(TCFD\)](#), criada pelo [Conselho de Estabilidade Financeira \(FSB\)](#), que é encarregada da função de engajar as empresas a informarem seus investidores sobre os riscos relacionados às mudanças climáticas e o modo em que são gerenciadas, visando melhorar a divulgação de informações financeiras relacionadas ao clima.

Em 2017, a [TCFD](#) emitiu recomendações financeiras e não financeiras a serem utilizadas de forma voluntária a nível global, capazes de promover informações úteis às avaliações e precipitações apropriadas dos riscos e oportunidades associadas à evolução das mudanças do clima. O objetivo dessas diretrizes é melhorar a transparência em matéria de riscos relacionados às mudanças climáticas e sua utilidade para o setor financeiro, ao conside-

rar os riscos relacionados ao clima e fornecer parâmetros de como cada organização está fazendo seu gerenciamento. A **TCFD** possui adesão de mais de 1200 membros em todo o mundo e vem ganhando cada vez mais apoiadores [TCFD, 2020].

## 2.2.2 Contextualização dos agentes nacionais

O **Sistema Financeiro Nacional (SFN)** é composto por instituições e instrumentos financeiros, que têm o objetivo de atuar na transferência de recursos de pessoas, empresas e governos que estejam superavitários para agentes deficitários. Este sistema está regulamentado pela Lei 4595/1964<sup>1</sup>, sendo estruturado da seguinte forma: **Conselho Monetário Nacional (CMN)**, **Banco Central do Brasil (BACEN)**, Banco do Brasil, **Banco Nacional de Desenvolvimento Econômico e Social (BNDES)** e Instituições Financeiras Públicas e Privadas [BACEN, 2021].

Segundo o Decreto nº 1.307<sup>2</sup> de 09 de novembro de 1994, o **CMN** é o órgão normativo superior do **SFN**, responsável por formular as políticas da moeda e do crédito, com o objetivo de manter a estabilidade financeira e promover o desenvolvimento econômico e social do país. O **BACEN** possui o papel de entidade fiscalizadora com as funções de gerenciar a política econômica, fiscalizar o **SFN** e garantir a estabilidade e o poder de compra da moeda. Dessa forma, boas práticas nas implementações de suas medidas devem ser observadas, incluindo as que tratam de sustentabilidade.

Tendo o exposto, visando garantir o aprimoramento dos regulamentos sobre a política de responsabilidade e estrutura de gerenciamento de riscos sociais, ambientais e climáticos aplicáveis às instituições financeiras, foram publicadas atualizações que ampliam os normativos relacionados ao tema nas Resoluções do **CMN** nº 4.943<sup>3</sup>, nº 4.944<sup>4</sup> e nº 4.945<sup>5</sup>, além de disposições sobre divulgação, relatórios e oportunidades nas Resoluções do **BACEN** 139<sup>6</sup>, 140<sup>7</sup> e 153<sup>8</sup>.

---

<sup>1</sup>[http://www.planalto.gov.br/ccivil\\_03/leis/14595.htm](http://www.planalto.gov.br/ccivil_03/leis/14595.htm)

<sup>2</sup>[http://www.planalto.gov.br/ccivil\\_03/decreto/1990-1994/D1307.htm](http://www.planalto.gov.br/ccivil_03/decreto/1990-1994/D1307.htm)

<sup>3</sup><https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20CMN&numero=4943>

<sup>4</sup><https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20CMN&numero=4944>

<sup>5</sup><https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20CMN&numero=4945>

<sup>6</sup><https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20BCB&numero=139>

<sup>7</sup><https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20BCB&numero=140>

<sup>8</sup><https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Instru%C3%A7%C3%A3o%20Normativa%20BCB&numero=153>

Conforme estabelecido em Resolução N<sup>o</sup> 4.327<sup>9</sup> que trata sobre a **Política de Responsabilidade Socioambiental (PRSA)** das instituições financeiras e demais instituições autorizadas a funcionar, o **BACEN** também dispõe sobre o risco socioambiental que é representado pela possibilidade de ocorrência de perdas em decorrência de danos socioambientais.

A **PRSA** orienta as diretrizes e comportamentos, considerando a atuação das empresas, governos e sociedades sobre iniciativas dedicadas à redução de riscos e aproveitamento de oportunidades associadas às questões socioambientais, o que inclui as mudanças climáticas; promoção de atividades relacionadas aos Direitos Humanos e princípios fundamentais do trabalho; universalização dos direitos sociais e respeito à diversidade nas relações negociais. As instituições devem se ater a essa resolução, bem como aos **Princípios do Equador**, os **Princípios para Sustentabilidade em Seguro (PSI)** e os **Princípios para o Investimento Responsável (PRI)**, e se ater aos padrões de desempenho do **International Finance Corporation (IFC)** e as Diretrizes de Meio Ambiente, Saúde e Segurança do Grupo Banco Mundial para produtos financeiros que englobem projetos industriais e de infraestrutura de grande porte.

Também é importante ressaltar que o **BACEN** é uma instituição que pertencente aos apoiadores da **TCFD**, desde 2020, e se mantém alinhado à agenda global e às iniciativas lideradas por organismos definidores de padrões financeiros internacionais, em prol da estabilidade do **SFN** [**BACEN**, 2021].

Diante disso, para guiar as instituições financeiras no cumprimento dos normativos impostos pelos agentes regulamentadores, a **Federação Brasileira de Bancos (Febraban)** atua como principal entidade representativa dos bancos brasileiros, fornecendo guias e informativos que visem o aperfeiçoamento de normativos, a melhoria contínua dos serviços e redução de riscos.

Seguindo as recomendações da iniciativa **FSB**<sup>10</sup> que visam apoiar a divulgação de informações sobre mudanças climáticas e permitir que as instituições e empresas melhorem sua gestão ao risco e oportunidades, a **Febraban** propôs um conjunto de ações visando a implementação das recomendações da **TCFD**<sup>11</sup>, que identificam os setores com maior exposição aos riscos climáticos e a priorização no reporte de suas informações.

Entende-se como sendo risco climático as mudanças no clima que influenciam na temperatura, radiação solar e ecologia das plantas, que possam afetar a disponibilidade de insumos para o plantio, gerar aparecimento de pragas de insetos, bem como de insetos benéficos que se alimentam de ervas daninhas [**Allara et al.**, 2012]. Essas mudanças potencializam os riscos já existentes, bem como possibilitam a ocorrência de novos eventos

---

<sup>9</sup>[https://www.bcb.gov.br/pre/normativos/res/2014/pdf/res\\_4327\\_v1\\_0.pdf](https://www.bcb.gov.br/pre/normativos/res/2014/pdf/res_4327_v1_0.pdf)

<sup>10</sup><https://www.fsb.org>

<sup>11</sup><https://www.fsb-tcfd.org>

não previstos. Em matéria de riscos, a exposição às mudanças do clima é considerada uma grande ameaça. [Krepsky & Scherer, 2019].

O setor de agricultura é um exemplo que deve ser priorizado. Para esse setor, a **Febraban** em parceria com a WayCarbon<sup>12</sup> fornece o conceito de risco agroclimático baseado no zoneamento de áreas aptas à produção de culturas, considerando atributos climáticos e de solo, além de características ecofisiológicas das culturas. As informações disponibilizadas orientam sobre as adversidades climáticas que podem afetar os investimentos de forma positiva ou negativa. Segundo [Selvaraju, 2012], o risco climático na agricultura é a probabilidade de um perigo hidrometeorológico que afete a subsistência de agricultores, criadores de gado, pescadores e habitantes da floresta.

Esse risco vem sendo evidenciado em relatórios e reportes que informam a urgência da necessidade de adoção de medidas para prever seus impactos. Um desses relatórios é o **Índice de vulnerabilidade aos desastres naturais relacionados às secas no contexto da mudança do clima (IVDNS)**<sup>13</sup> fornecido pela WWF que informa que as mudanças que ocorrem nos padrões climáticos podem afetar diretamente os sistemas de precipitação e causar impactos negativos e que, no Brasil, a ocorrência de períodos de secas é um problema secular, principalmente no Nordeste.

Esse relatório também ressalta que há uma frequência nas tendências de reduções de precipitações pelo mundo, acompanhadas de simulações climáticas com previsões de climas mais intensos para o futuro.

O **IVDNS** foi construído para indicar a vulnerabilidades dos municípios brasileiros com base em mudanças climáticas entre 2011 a 2040 e avalia os componentes que causam a seca meteorológica, como alterações em precipitação, dificuldades de manter o equilíbrio hidrológico e ocorrência de secas extremas, o contexto socioeconômico e ambiental que contribui na ampliação ou redução dos efeitos das mudanças climáticas, a capacidade de responder e reorganizar os sistemas sociais quando exposto às secas. Este estudo apresenta resultados utilizando as metodologias HadGEM e Miroc, que classificam o aumento da vulnerabilidade da região em alto, médio e baixo em **cenários de altas emissões (RCP)** em suas pontuações de 4.5 e 8.5.

Com as projeções utilizadas para realizar os cálculos do **IVDNS**, foi possível observar que a vulnerabilidade aos desastres naturais de secas tentem a aumentar em todo território brasileiro, para todos os períodos futuros, caso nada seja feito sobre os padrões socioeconômicos e de governança atuais. Esse estudo traz alerta para urgência na implementação de medidas de adaptação e políticas públicas de gestão de risco. Ressalta-se que esse índice avalia resultados que apontam para um aumento da vulnerabilidade a

---

<sup>12</sup><https://waycarbon.com>

<sup>13</sup>[https://www.wwf.org.br/natureza\\_brasileira/reducao\\_de\\_impactos2/clima/copy\\_of\\_mudancas\\_climaticas2\\_20062017\\_1938/](https://www.wwf.org.br/natureza_brasileira/reducao_de_impactos2/clima/copy_of_mudancas_climaticas2_20062017_1938/)

ocorrências de desastres em cada região. Isso não significa que um local indicado como crítico terá o maior número de ocorrências de desastres, e sim terá um maior aumento na ocorrência.

Tendo o exposto, a **Febraban** desenvolveu às metodologias Taxonomia Verde e **Régua de Sensibilidade ao Risco Climático (RSRC)**. A **RSRC** foi desenvolvida com o intuito de servir de guia para as instituições financeiras desenvolverem suas metodologias de classificação de risco climático. Seu funcionamento é separado por três camadas (setorial, clientes e operação) que permitem analisar a sensibilidade da carteira de crédito aos riscos climáticos. Essa ferramenta aumenta seu grau de análise progressivamente a cada camada, enquanto exemplifica as ações que devem ser realizadas para medição do risco climático, objetivando facilitar a priorização de ações relativas ao gerenciamento dos riscos e facilitar a análise de materialidade das divulgações recomendadas pela **TCFD** para bancos, seu detalhamento encontra-se no capítulo **3.1.1**.

A Taxonomia verde foi gerada a partir de estudo, que contou com a parceria do Centro de Estudos em Sustentabilidade, da Fundação Getúlio Vargas (FGVces) e o Departamento de Economia da PUC Rio, para consolidar a metodologia capaz de mensurar os recursos financeiros na Economia Verde, considerando as operações de crédito que abrangem todo o território nacional. Também contou com a participação dos bancos de sua Comissão Setorial de Responsabilidade Social e Sustentabilidade (CRSS), do Grupo de Trabalho Clima e Economia Verde e o apoio técnico da SITAWI, ao ser revista em 2020. Essa metodologia considerou as recomendações do Banco Mundial, aproveitou os trabalhos e orientações emitidas na implementação das recomendações da **TCFD**, no setor bancário, e teve como base a Taxonomia da União Europeia, os Climate Bond Standards e os Social Bond Principles. Seu principal objetivo é caracterizar o crédito sob a ótica socioambiental e climática em três modalidades:

- Economia Verde (EV): classifica potenciais impactos positivos a sociedade e ao meio ambiente, considerando as contribuições ambientais e sociais. Pode receber a classificação de alta contribuição, moderada contribuição ou baixa/nenhuma contribuição.
- Exposição às mudanças climáticas ou risco climático (ERC): classifica a exposição das carteiras quanto ao risco climático. Pode receber a classificação de alta exposição, moderada exposição ou baixa/nenhuma exposição.
- Exposição ao risco ambiental: classifica a exposição das carteiras quanto ao risco ambiental. Pode receber a classificação de alta exposição, moderada exposição ou baixa/nenhuma exposição.

A Taxonomia permite aos bancos mensurarem, de maneira simples, a exposição de suas carteiras de acordo com critérios socioambientais. Isso pode ser feito para finalida-

des prudenciais ou para avaliações de potenciais impactos positivos. Os detalhes dessa metodologia foram acessados em julho de 2021 e encontram-se disponíveis em [Taxonomia Verde](#). As medições observadas estão em [Guia Explicativo da Taxonomia Verde](#).

Em face às observações realizadas, produzir estudos sobre as vulnerabilidades dos investimentos para entender os impactos que podem ameaçar os sistemas de produção, bem como identificar os possíveis cenários futuros se torna indispensável, haja vista que a economia brasileira está concentrada no agronegócio.

### 2.2.3 Risco de Crédito

Os riscos que envolvem as atividades bancárias são todos aqueles inerentes às suas atividades, podendo estar relacionados a diversos fatores que influenciam o mercado, como procedimentos operacionais, socioambiental, novas tecnologias, dentre outros tipos de riscos que direta ou indiretamente afetem a composição de risco da instituição.

A respeito de investimentos sustentáveis, avaliar a visão do mercado é um importante mecanismo gerador de insumos, para entender os desafios a serem enfrentados. Como parte desses desafios, podemos citar as informações divulgadas pela plataforma do [Banco Interamericano de Desenvolvimento \(BID\)](#), em que informa que apenas 2% dos títulos são considerados verdes, devido a falta de classificações confiáveis. Desse modo, favorecendo a prática de "lavagem-verde" (greenwashing), quando os produtos são alegados sustentáveis somente para fins de marketing. A plataforma também relata que há uma falta de liquidez das operações pela baixa quantidade de títulos [[Érika Motoda, 2021](#)].

Segundo [[Coutinho, 1994](#)], a necessidade de instrumentos capazes de mitigar ou adaptar os gastos com o processo, impactos das mudanças climáticas e estimular a extinção de emissões induzem a competitividade.

O estudo terá seu foco na composição do Risco de Crédito e como ele pode ser afetado pelas variáveis de Risco Climático mensuradas. Mas para isso, além do supracitado, é preciso entender as etapas que compõem o Risco de Crédito, que serão detalhadas a seguir.

#### Definição do Risco de Crédito

Segundo [[Blatt, 1999](#)], a palavra crédito origina-se do substantivo *creditum*, que significa confiança. Desse modo, o Risco de Crédito está associado à confiança gerada ao se emprestar o crédito a terceiros, e é medido pela probabilidade do poupador de crédito receber o que se é devido pelo tomador de crédito.

Para [[Gitman et al., 2010](#)], Risco de crédito reside na possibilidade de exposição a perdas ou danos de um ativo, em que o fluxo de caixa pode não atingir o retorno esperado.

Por isso, para elucidar as diversas situações que influenciam esse tipo de risco, é necessário entender o processo decisório que permeia as informações obtidas pra liberação de crédito.

Essas informações são obtidas por documentação empresarial, visitas a clientes e demais ações que forneçam insumos para os indicadores que forem definidos como importantes para tomadas de decisão da instituição [Capelletto & Corrar, 2008].

## **Análise de Crédito**

Tendo o exposto, é realizada a análise de crédito, que consiste na identificação e medição por métodos estatísticos da exposição ao risco de não se obter os retornos esperados dos tomadores e transações realizadas. Essa mensuração é realizada com o auxílio de ferramentas e aplicações capazes de realizar a mensuração individual ou grupo de carteiras [Neto & Lima, 2009].

Para nortear a análise de crédito, [Weston & Brigham, 2000] define a análise dos cinco fatores que tem sido um enfoque tradicionalmente utilizado. Esses fatores são:

**Caráter** : investiga-se o respeito ao compromisso com as dívidas.

**Capacidade** : identifica-se o poder de compra de um indivíduo.

**Capital** : verifica-se o endividamento do cliente, sua rentabilidade e solidez financeira.

**Garantias** : avalia-se a garantia deixada pelo cliente.

**Condições** : analisa-se o contexto econômico em que o cliente se encontra.

Para [Brito, 2019], estudar o Risco de Crédito é fundamental, por ser afetado por diferentes fatores e ser um dos principais riscos inerentes às atividades de instituições financeiras. Realizar seu gerenciamento é crucial para mitigar efeitos causados pelo risco de inadimplência.

Segundo [Mileo et al., 2013], o risco de inadimplência em massa em instituições bancárias é um dos principais fatores responsáveis por sua insolvência. E, por isso, buscar alternativas de aprimoramento de seus instrumentos de crédito que sejam capazes de refletir diversos cenários é crucial.

O analista deve utilizar as informações prestadas e, junto com sua experiência, deve escolher os modelos adequados para realizar a análise de crédito.

## **Modelos de Predição**

Segundo [Brito & Assaf Neto, 2008], os modelos de análise podem ser definidos em três categorias:

**Classificação** : expressada em forma de rating ou score, avalia o risco de um tomador ou operação por meio de medida que represente um risco de *default*. Normalmente utilizados em concessão de crédito por instituições financeiras.

**Estocásticos** : objetiva avaliar o comportamento estocástico do risco de crédito. Apresenta uma variável que se comporta de forma aleatória ao longo do período, devido a sua capacidade evolutiva. Normalmente utilizados em mercado de títulos e derivativos.

**Portfólio** : estimam as perdas de valor por distribuições estatísticas. Quantificam o risco a partir dessa distribuição. Permitem uma avaliação de forma agregada do risco de crédito devido a sua capacidade de considerar os efeitos produzidos pela correlação entre os ativos da carteira.

Este estudo visa identificar a relevância do risco climático na probabilidade de solvência das operações financeiras e de seus tomadores. Por esse motivo, o modelo definido para este estudo é o de Classificação, tendo em vista sua larga utilização nos processos decisórios de concessão de crédito e seu grande propósito ser o de medir a probabilidade de um evento considerado *default* ocorrer ao longo de determinado tempo. São modelos baseados em dois grupos distintos, os *default*, ou seja, os que incorreram em insolvência, e os que não incorreram em *default*, ou seja, os considerados solventes.

Segundo [Gonçalves et al., 2013], esses modelos normalmente utilizam previsões estatísticas de análises multivariadas, regressões lineares e não lineares que auxiliam no processo de gestão do crédito, permitindo a criação de equações que classifiquem indicadores de bons ou maus pagadores.

Mas para avaliar o risco de inadimplência não basta somente usar métodos convencionais, é necessário fazer uso das novas tecnologias e métodos estatísticos mais sofisticados de machine learning, para incorporar respostas eficientes aos ambientes apresentados que estão cada vez mais dinâmicos [Jordan & Mitchell, 2015]. Segundo [Munkhdalai et al., 2019b], a técnica de machine learning associada à inteligência artificial produziu um salto no campo de análise e previsão de dados, e [Guegan & Hassani, 2018] afirma que esses algoritmos são capazes de aprender com os dados e fazer previsões simulando os processos de tomada de decisão do cérebro humano. Dessa forma, esses algoritmos são capazes de adaptar-se aos diversos cenários. Essa adaptação é um fator considerado essencial para otimização das análises sobre concessão de crédito.

## 2.2.4 Política de Gerenciamento de Riscos e Controles Internos

A Gestão de Riscos deve ser tratada como componente essencial da gestão das instituições, ela auxilia no alcance de objetivos, previsão de riscos e orienta implementação de respostas.

A instituição escolhida segue norteada pela ISO 31000, norma que oferece uma abordagem de aplicação ampla e visa harmonizar o processo de gestão de riscos, acrescidas de normativos que regem as instituições financeiras. Dessa forma, são realizadas avaliações periódicas nos processos de gerenciamento dos riscos com a finalidade de verificar se estão de acordo com as orientações estratégicas, as políticas específicas e as normas internas e regulatórias. Esse processo também é revisado anualmente por parte da instituição, conforme as normas de supervisão e de regulação bancárias.

As informações da instituição seguem a **Lei Geral de Proteção dos Dados Pessoais (Lei no13.709/18) (LGPD)**<sup>14</sup>, lei que entrou em vigor em setembro de 2020, e implementou uma série de procedimentos para garantir o devido tratamento dos dados vulneráveis. A **LGPD** também define dez princípios para o tratamento de dados, que contemplam todo o ciclo de utilização da informação, englobando desde a formalização da finalidade do dado, sua privacidade, garantias de integridade, transparência, segurança e uso correto das informações.

Em termos gerais, o processo de gerenciamento e controle de riscos consiste nas ações de definição do escopo, contextualização, processo de gestão de riscos (identificação, análise e avaliação), implementação de respostas, monitoramento e comunicação às partes interessadas, durante todo o processo. Suas etapas estão indicadas na Figura 2.2.

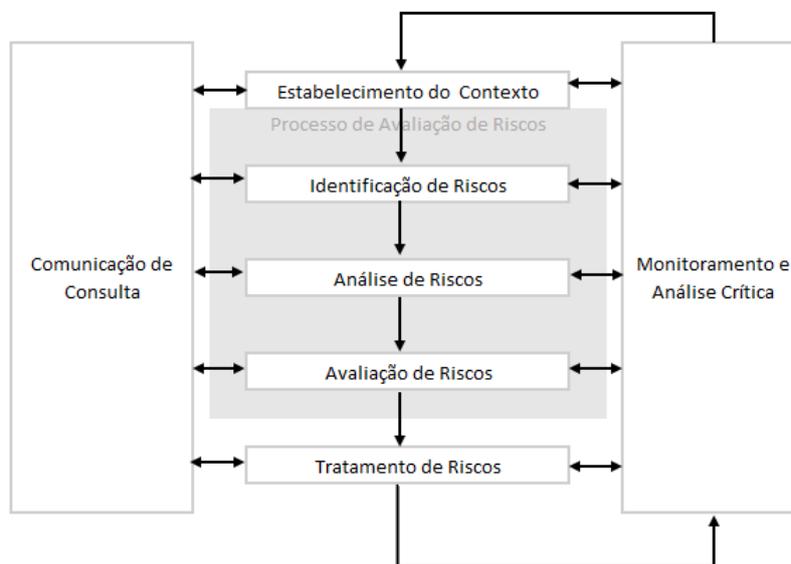


Figura 2.2: Processo de gestão de riscos  
Fonte: MAPA

<sup>14</sup>[http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/113709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm)

A Figura 2.2 ilustra o relacionamento das etapas do gerenciamento e controle de riscos e como elas se comunicam durante o processo.

O estabelecimento da contextualização compreende o contexto externo: o meio externo com potencial de afetar a instituição, como cultura, social, política, legislação dentre outros; e o contexto interno: ambiente que a instituição é regida, como sua estrutura organizacional, seu regimento, capacidades e recursos, dentre outros. A Identificação de Riscos é responsável pelo processo de reconhecimento de riscos. A análise do risco realiza a investigação das causas e fontes dos riscos, quais suas consequências, positivas ou negativas, identificam sua probabilidade e busca entender o impacto que pode ser gerado, caso a situação aconteça. Também temos a avaliação de riscos que auxilia na tomada de decisão, com base no que foi identificado no processo de análise. Nesta etapa, os riscos são priorizados e direcionados. Em seguida, os riscos passam pelo tratamento de riscos que implementam os direcionamentos realizados.

Durante a realização dessas etapas, também se destacam as etapas de comunicação e consulta, processo contínuo de fornecimento de informações às partes interessadas, e o monitoramento e análise crítica que realiza verificação contínua, observando as mudanças do cenário, apontando melhorias e apurando resultados.

A instituição objeto desta pesquisa também segue o Modelo Referencial de Linhas de Defesa, como uma forma eficaz de melhorar a comunicação e conscientização dos papéis e responsabilidades definidas pela gestão. Nesta abordagem, há três linhas de defesa, assim como mostra a Figura 2.3.



Figura 2.3: Modelo Referencial de Linhas de Defesa

Fonte: DNIT

A Figura 2.3 ilustra a interação das linhas de defesa e seus agentes responsáveis. Seus papéis também são delimitados por funções e competências, sendo que a 1ª Linha de Defesa compreende as funções que gerenciam e têm propriedade sobre os riscos. É consti-

tuída por Tomadores de Riscos – áreas que devem observar orientações corporativas sobre o gerenciamento de riscos das áreas de gestão de riscos e controles internos de 2<sup>a</sup> Linha de Defesa, abrangendo a identificação, mensuração, avaliação, controle, monitoramento, mitigação e reporte dos riscos em seus processos; e Gestor Específico de Riscos: áreas que definem e orientam de forma corporativa quanto aos processos chave (incluindo visão de negócios, riscos e controles).

A 2<sup>a</sup> Linha de Defesa, corresponde às funções corporativas típicas de gestão de riscos e controles internos, sendo responsáveis pela Regulação e Supervisão Corporativa da Gestão de Riscos e pela Regulação e Supervisão do Sistema de Controles Internos e Compliance. São áreas responsáveis pelas orientações corporativas, pela supervisão e pelo assessoramento do processo de gestão de riscos e controles internos às áreas de primeira linha, especializadas na tomada de decisão.

A 3<sup>a</sup> Linha de Defesa, abrange a função auditoria interna, responsável por avaliar a correlação de todos os ciclos de gerenciamento de riscos, de controles internos e de governança da Organização, incluindo a forma de atuação da primeira e segunda linhas de defesa.

# Capítulo 3

## Metodologia

### 3.1 Materiais e Métodos

As análises serão divididas em três etapas: classificação do risco climático, processo de avaliação de riscos e modelo de explicabilidade da influência das variáveis do risco climático ao compor o risco de crédito.

Os instrumentos necessários para realização das etapas descritas estão detalhados a seguir nos 3.1.1, 3.1.2, 3.1.3, 3.1.4 e 3.1.5. Sua forma de execução está detalhada no tópico 3.1.6.

#### 3.1.1 Classificação do Risco Climático:

Os indicadores de risco climático foram incorporados utilizando o material fornecido pela Taxonomia Verde detalhada no item 2.2.2 e pela medição utilizando a camada de operações da RSRC informada no item 2.2.2 que será detalhada a seguir.

A análise exploratória dos dados classificados foi realizada para identificar como os indicadores propostos se relacionavam com o risco da operação, risco do cliente, volume financeiro, quantidade de operações e inadimplência. Seus resultados estão detalhados no item 4.1.

#### Régua de Sensibilidade ao Risco Climático

Nessa proposta, as informações foram separadas em dois grupos: a relevância que é formada pela associação do rating da operação, o risco climático locacional da operação, o código da natureza da operação; e pela proporcionalidade que é formada pela associação do prazo e volume da operação. Em seguida, o risco climático é obtido pela associação entre a variável de relevância e a variável de proporcionalidade. As informações citadas foram estruturadas na Tabela 3.1 a seguir.

Tabela 3.1: Camada de Operações

<b>Relevância</b>	<b>Proporcionalidade</b>
• Rating da operação	• Prazo da operação
• Risco climático locacional da operação	• Volume da operação
• Natureza das atividades (Descrição contida na lista de CNAE fornecida pela FEBRABAN)	
<i>Objetivos da medição da camada</i>	
<ul style="list-style-type: none"> <li>• Operações a serem priorizadas no gerenciamento de riscos climáticos</li> <li>• Motivos que contribuem para a maior sensibilidade do cliente</li> <li>• Necessidade de ajuste no processo de concessão de crédito e monitoramento das operações</li> <li>• Qualidade da carteira do setor econômico</li> </ul>	

Fonte: Elaborado pelo autor

### Variáveis de Relevância

*Rating da operação.* Representa a classificação do risco da operação entre AA – A, B – C ou D – H, quanto pior o rating, maior o impacto.

*Risco climático locacional da operação.* É a classificação de alto, médio ou baixo risco das unidades produtivas do cliente. Existem algumas ferramentas de classificação sugeridas como: Think Hazard, WRI Aqueduct Atlas e WWF Water Risk.

*Natureza das atividades financiadas.* Lista de **CNAE** correspondente ao informado pela **TCFD** de setores classificados como alta, média ou baixa exposição aos riscos climáticos.

Nesta primeira etapa, a Figura 3.1 mostra a relação entre a variável Natureza das atividades e o Rating da operação tendo como resultado a variável de *Risco não-locacional*.

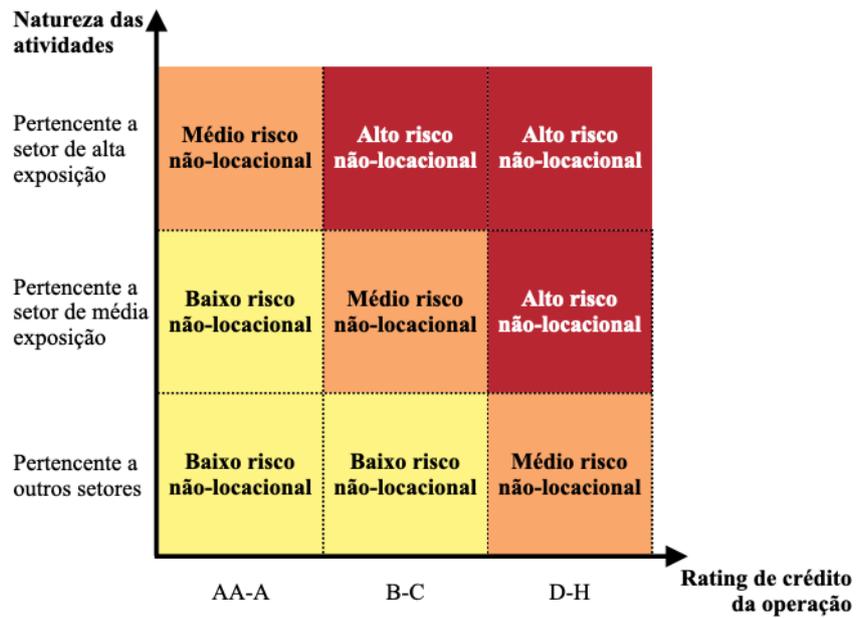


Figura 3.1: Natureza das atividades e Raiting de crédito  
 Fonte: Febraban

A Figura 3.2 relaciona a variável de risco não-locacional obtida no passo anterior com a variável de risco locacional tendo como resultado a variável *Relevância*.

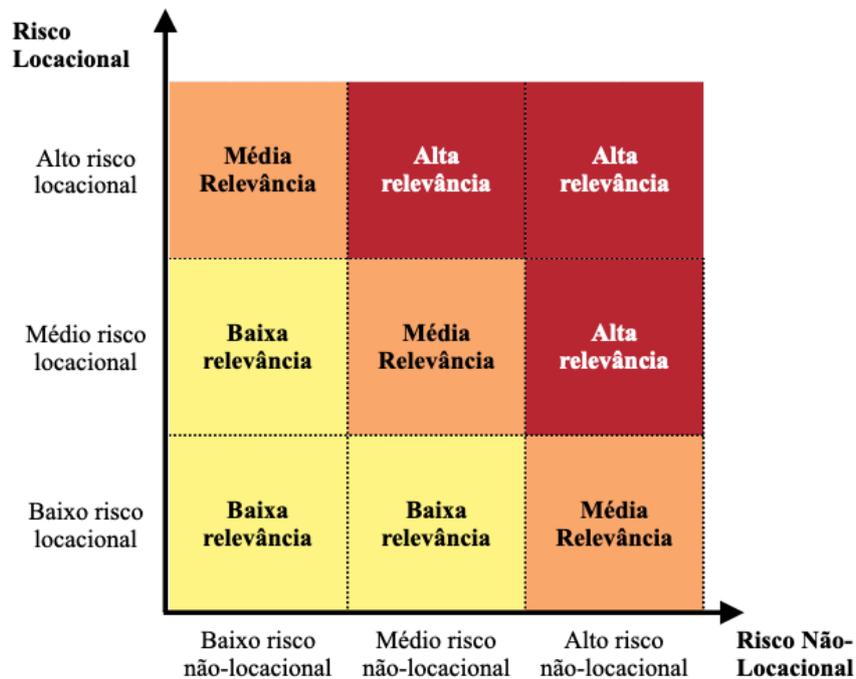


Figura 3.2: Risco Locacional e Risco Não-Locacional  
 Fonte: Febraban

## Variáveis de Proporcionalidade

Nesta segunda etapa, avalia-se a proporcionalidade no nível das operações, que é obtida por meio da avaliação da relação entre as duas variáveis: o prazo e o volume da operação.

*Prazo da operação.* Fornecerá um risco alto para operações maiores que 60 meses; médio para maiores ou iguais a 24 meses e menores ou iguais a 60 meses; e baixo para operações menores que 24 meses.

*Volume da operação.* Para esta variável existe uma relativização dos valores classificados entre S1 a S5 considerando o descrito na Resolução CMN 4553/2017 <sup>1</sup> que trata a classificação dos bancos.

A Figura 3.3 ilustra a relação das variáveis de proporcionalidade.

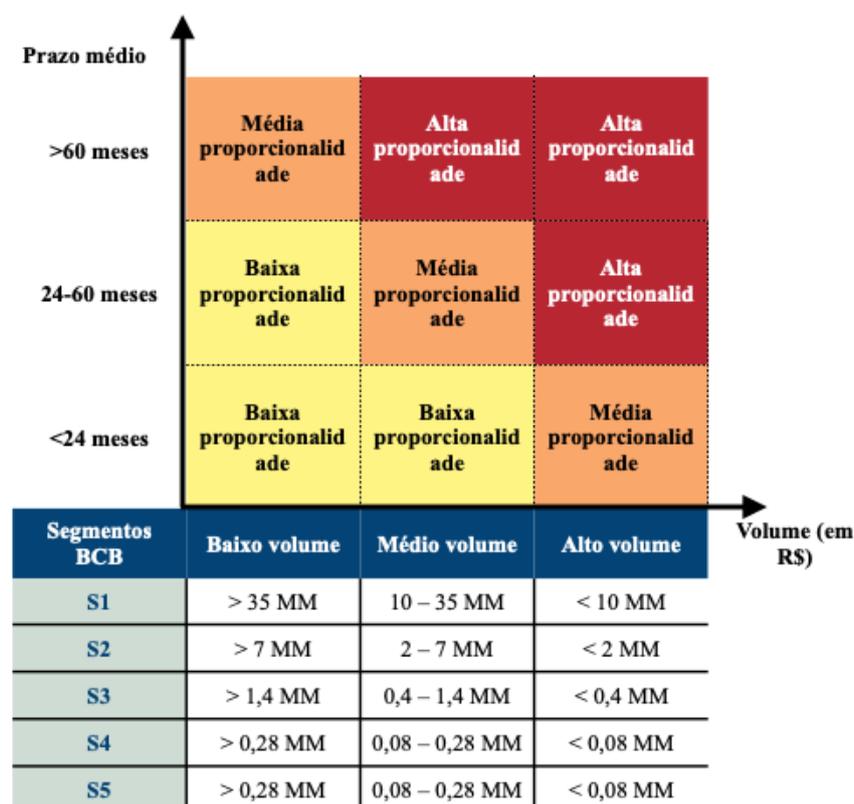


Figura 3.3: Relação entre Prazo da Operação e Volume

Fonte: Febraban

<sup>1</sup>[https://www.in.gov.br/materia/-/asset\\_publisher/Kujrw0TZC2Mb/content/id/20361114/do-1-2017-01-31-resolucao-n-4-553-de-30-de-janeiro-de-2017-20361101](https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/20361114/do-1-2017-01-31-resolucao-n-4-553-de-30-de-janeiro-de-2017-20361101)

Por fim, a Figura 3.4 ilustra a classificação do risco climático.

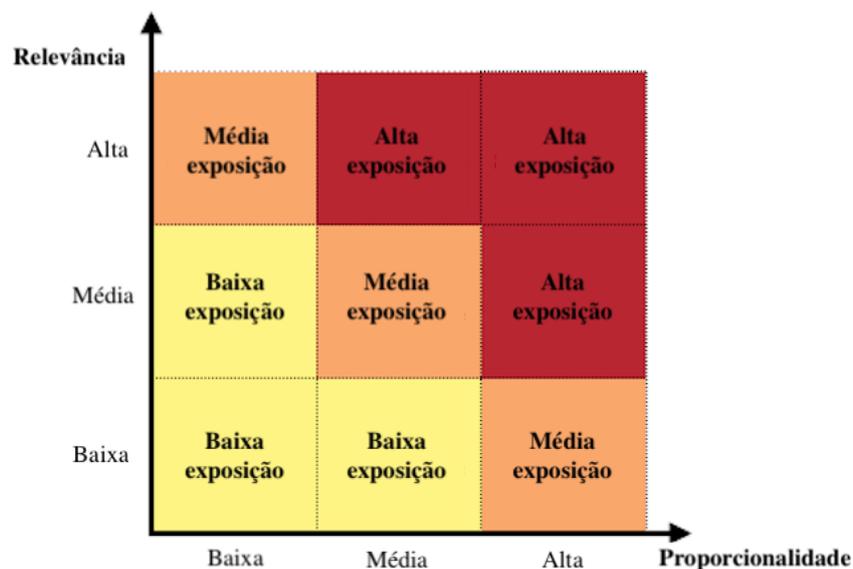


Figura 3.4: Relação entre a Relevância e a Proporcionalidade  
Fonte: Febraban

### 3.1.2 Processo de avaliação de riscos

Processo que envolve a realização do mapeamento dos riscos envolvidos na criação do indicador do risco climático, bem como a identificação de seus impactos e proposição de soluções de contorno.

Os resultados foram desenvolvidos por meio de levantamento documental, análise de base de dados e reuniões com a gerência de riscos da instituição incluindo a participação de entes parceiros.

As reuniões foram separadas por hierarquia contabilizando dez reuniões. Sendo cinco com a área de riscos da instituição e cinco com a área de sustentabilidade da instituição entre 2019 a 2021. Seus detalhes estão presentes no item 3.1.6.

### 3.1.3 Modelo de explicabilidade da influência das variáveis do risco climático ao compor o risco de crédito

Foi proposto um modelo de explicabilidade das variáveis que compõe o risco climático para fornecer indicadores de previsibilidade do risco de crédito e informar qual a influência que a associação de suas variáveis possui sobre o resultado final.

Para isso, foram utilizadas técnicas de processamento de dados, para retirar inconsistências, verificar ausência de dados e ajustar a formatação da base.

Também foram utilizadas técnicas de seleção e classificação de variáveis, para estabelecer o grau de importância dos recursos utilizados e reduzir sua dimensionalidade.

Segundo [Chandrashekar & Sahin, 2014], utilizar critérios de seleção capazes de medir a relevância dos recursos e possibilitar a retirada de variáveis irrelevantes é necessário para possibilitar uma generalização eficiente dos modelos. Pois, quando os recursos são reduzidos, não se utilizam dados novos, e sim os próprios dados de entrada para reduzir seu número. Desse modo, ao encontrar subconjuntos de dados considerados relevantes, reduz-se a necessidade computacional, auxilia no entendimento e melhora a generalização do modelo.

Os dados resultantes foram submetidos a um conjunto de algoritmos que tratam de problemas de classificação em modelos de aprendizagem de máquinas supervisionadas. Seus resultados foram submetidos a métricas de avaliação para identificar o modelo com melhor ajuste.

Segundo [James et al., 2013], devido a diversidade de algoritmos e problemas que vem sendo solucionados por sua utilização, realizar comparações entre os resultados de seus desempenhos preditivos é essencial. Tendo em vista que interação desses algoritmos a conjunto de dados pode destoar, ou seja, determinados algoritmos podem apresentar um bom desempenho com certos conjuntos de dados e esse desempenho pode não se refletir com outros conjuntos de dados, e também, assim como é ressaltado por [Kuhn et al., 2013], alguns modelos podem apresentar desempenho semelhante, o que gera a necessidade de avaliar os benefícios que cumpram com mais eficiência os objetivos propostos, quer seja pela facilidade de implementação, diversidade de funções ou interpretabilidade.

O modelo identificado com o melhor ajuste foi submetido a ferramentas de explicabilidade capazes de mensurar o comportamento das variáveis selecionadas ao compor o risco de crédito.

Os mecanismos utilizados estão detalhados nos itens 3.1.4 e 3.1.5, e a forma como foram utilizados no item 3.1.6.

### 3.1.4 Processamento dos Dados

Antes de realizar análises de modelos, é necessário realizar a estruturação dos dados em dados úteis. Ou seja, realizar análises descritivas, ajustar inconsistências de formato, lidar com dados ausentes e realizar a retirada de variáveis que não contribuem com a análise proposta. Para auxiliar nessa etapa, serão apresentados os conceitos de K-Nearest Neighbors, SMOOTE e alguma de suas variações.

- **K-Nearest Neighbors**

Segundo [Dong & Peng, 2013], Dados ausentes ou nulos são problemas muito comuns que podem ocorrer por falhas em sua obtenção. Para construir um bom modelo de aprendizagem de máquina, é necessário lidar com essa situação, pois essa situação pode limitar o poder estatístico do modelo, gerar falha ao rejeitar uma hipótese. Essa falha pode ocasionar o erro tipo I, quando a hipótese é rejeitada, mas de fato é verdadeira, ou o erro tipo II, quando a hipótese não é rejeitada, mas de fato é falsa, aumentar o erro padrão (média amostral em relação à média da população), dificultar a generalização, desbalancear amostras e distorcer a validade das conclusões.

Segundo [Rubin, 1976], os tipos de valores ausentes geralmente podem ser classificados como:

- Faltando completamente ao acaso (MCAR): Quando o dado ausente não possui nenhuma relação ou probabilidade com algum evento, nesse caso, sua ausência é totalmente aleatória. Pode ocorrer por fatores que levaram a perdas dos dados.
- Faltando ao acaso (MAR): Neste caso, a ausência depende das características dos dados que foram observados. Por exemplo, uma pesquisa realizada por ser enviada por recusa de determinados grupos em responder.
- Faltando não ao acaso (MNAR): Ocorre quando a ausência depende tanto das características dos dados quanto dos valores ausentes. Um dado pode depender parcialmente de outro que não tem o costume de ser medido em determinadas situações. Neste caso, há uma dificuldade em determinar o que gerou a ausência.

Existem diversas abordagens na literatura para preenchimento de dados ausentes, mas o método k nearest neighbor (kNN) é o mais popular devido a sua simples implementação e elevada performance [Zhang et al., 2017]. Este método realiza a substituição dos valores ausentes pelo valor médio da distância Euclidiana dos seus vizinhos próximos na base de treino.

A figura Figura 3.5 ilustra em sua área oval os pontos vizinhos mais próximos identificados para o dado em cinza.

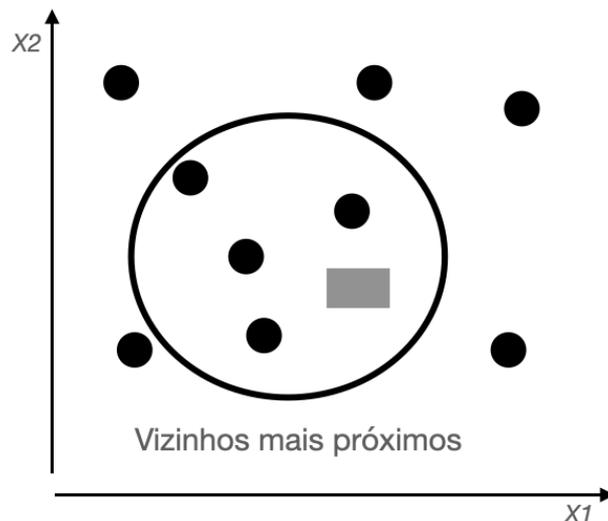


Figura 3.5: Funcionamento do KNN

Fonte: Elaborado pelo autor

A ideia que serve de base para o KNN é identificação do valor 'k' de amostras nos conjuntos de dados que são semelhantes, em que 'k' é o número de vizinhos fornecido para comparação. Dessa forma, os dados ausentes que serão imputados são os valores médios dos dados vizinhos encontrados no conjunto de dados.

A ferramenta `KNNImputer` fornecido pela bibliotéca `scikit-learn`<sup>2</sup> é amplamente utilizada para realização deste método.

#### • SMOOTE

Segundo [Chawla et al., 2002], uma base de dados é considerada desbalanceada quando ocorre uma grande diferença de ocorrências de casos de uma classe em relação as outras.

Em algoritmos de aprendizagem de máquina, essa situação pode gerar distorções que levam os modelos a não agregarem valor para as classes de menor representação, ocasionando falha no desempenho [Phua et al., 2004].

Há uma grande variedade de métodos para lidar com classes desbalanceadas, porém, lidar com essas classes por meio da redistribuição de classes por amostragem são abordagens bastante comuns.

Essas abordagens utilizam métodos que consistem em:

- Undersampling: eliminação de casos da classe majoritária.
- Oversampling: replicação de casos da classe minoritária.

<sup>2</sup><https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html>

Na literatura, o método SMOOTE é frequentemente citado para fins de balanceamento de dados. Seu funcionamento utiliza o algoritmo KNN para gerar amostras sintéticas para classes minoritárias por oversampling aleatório.

Segundo [Chawla et al., 2002], a utilização da técnica SMOOTE e suas variações forneceram um desempenho superior em 47 de 48 experimentos.

[Batista et al., 2004] relata que uma de suas variações conhecidas é o método SMOOTE + ENN, sua execução baseia-se em gerar bons resultados ao combinar o método de minorias sintéticas SMOOTE, com o método ENN.

O método ENN consiste em remover a classe majoritária. Este método verifica se a classe majoritária do vizinho mais próximo é a mesma que a classe da observação. Se não for, tanto a observação, como o seu vizinho mais próximo serão removidos. É um método de subamostragem que remove a classe majoritária para corresponder à classe minoritária.

O SMOOTE + ENN funciona em duas etapas. Primeiro é utilizado a técnica SMOOTE sobre as classes minoritárias, em seguida utiliza-se o ENN sobre as classes majoritárias.

### 3.1.5 Técnicas de Predição e Classificação

A classificação do risco climático proporciona um novo indicador que influenciará o risco de crédito das operações. Portanto, o entendimento de suas variáveis e seu potencial de influenciar o processo de tomada de decisão é fundamental.

Modelos dessa categoria podem gerar dados complexos com elevado volume de variáveis. Dessa forma, inflacionando a sofisticação das análises que, por sua vez, irão produzir respostas complexas e de difícil interpretação.

Propor modelos de interpretação intuitiva e combinar modelos mantendo a fácil compreensão é essencial para evitar erros e consolidar fundamentos.

Segundo [Gouveia Jr, 1999], um modelo pode ser definido como sendo a representação da imagem de uma teoria em forma de sistema simplificado constituído para mapear entradas e saídas buscando investigar se existe alguma conexão entre as variáveis e seus resultados do problema proposto. Eles são úteis para previsões, por serem capazes de aprender a relação entre as entradas e as saídas, durante seu treinamento.

Segundo [Müller & Guido, 2016], modelos treinados são avaliados ao serem expostos a conjunto de dados de teste para realizarem previsões que serão comparadas com dados já conhecidos e seu desempenho analisado por métricas de validação.

Modelos de algoritmos de machine learning são comumente divididos em: aprendizado supervisionado, aprendizado não supervisionado, aprendizado semi-supervisionado e aprendizado de reforço.

- Supervisionado: o computador recebe um conjunto de dados conhecidos que permite que seu aprendizado realize a explicabilidade da variável Y por meio da variável X.
- Não Supervisionado: o computador recebe dados não conhecidos e necessita extrair padrões realizando agrupamentos (clusterizações) de X por similaridade.
- Semi-supervisionado: recebe um conjunto de dados parcialmente conhecidos em que visa rotular e interpretar os dados não conhecidos.
- Reforço: observa o ambiente e usa esses dados para identificar o comportamento ideal. Faz uso de abordagem iterativa que utiliza de ação e recompensa, para auxiliar o computador em suas ações.

Tendo o exposto, os fatores ligados a esse entendimento serão mapeados e categorizados a fim de possibilitar previsões que auxiliem nas tomadas de decisão. E, para isso, segundo [Breiman et al., 1984], os modelos de classificação e regressão por Árvores de Decisão (CART) aparecem como alternativa para exploração e predição por sua simplicidade, robustez e versatilidade de técnicas.

A construção desses modelos possibilita a explicação da variável resposta, pela interação do conjunto de suas covariáveis; e também possibilitam a utilização de técnicas de melhoria ou complemento visando modelos mais consolidados [Breiman et al., 2017].

Sua versatilidade é notoriamente comprovada e aplicada em todos os tipos de cenários, por ter um funcionamento intuitivo, não impor restrições quanto à natureza de distribuição de variáveis e sua simplicidade na interpretação de modelos e resultados.

Em modelos de Machine Learning, a escolha do tipo de aprendizagem é um dos fatores principais. Para o estudo, utilizaremos técnicas para avaliação de problemas de classificação, em que os fatores relacionados à influencia do indicador do risco climático no processo de crédito serão mapeados e categorizados visando a explicabilidade de suas previsões.

Para melhor entendimento, as métricas e termos mais relevantes utilizados em problemas de classificação serão apresentados a seguir, por serem considerados necessários para a avaliação do modelo proposto de acordo com o que já foi exposto nos capítulos anteriores.

**Regularização e Seleção de Variáveis :** Ridge, Lasso, Elastic Net.

**Abordagem de Classificação :** Logit, Árvore de decisão, Boosting, Bagging, AdaBoost, Random Forast, Gradient Boosting e XGBoost.

**Medidas de ajuste:** Acurácia, Precisão, Especificidade, Revocação, F1-Score e curva ROC/AUC.

**Medidas de explicabilidade :** Shapley Value.

### Regularização e Seleção de Variáveis:

No aprendizado de máquina podem ocorrer situações em que o modelo funcionará como o esperado no conjunto de dados testado, porém, quando apresentado a novos dados, anomalias podem ocorrer em seu funcionamento, levando a um desempenho inferior que o esperado. Esse fenômeno pode ocorrer por diversos motivos. Dentre eles, temos o sobreajuste (overfitting), quando o modelo se ajusta muito bem aos dados de treinamento apresentados, porém não realiza as previsões com o mesmo desempenho ao ser apresentado a dados novos. Neste caso, o modelo somente memorizou os padrões de um determinado dado de treinamento e não se tornou flexível o suficiente para ser generalizado e aplicado em dados reais; também temos o subajuste (underfitting), quando o modelo não se adapta nem mesmo aos dados de treinamento, não executando um bom desempenho [Müller et al., 2001].

Esses conceitos podem ser visualmente entendidos nas Figuras 3.6 a 3.9 a seguir, em que foram separados dados de treino e teste a fim de verificar o desempenho do modelo proposto:

A Figura 3.6 representa graficamente um conjunto de dados separados em treino e teste.

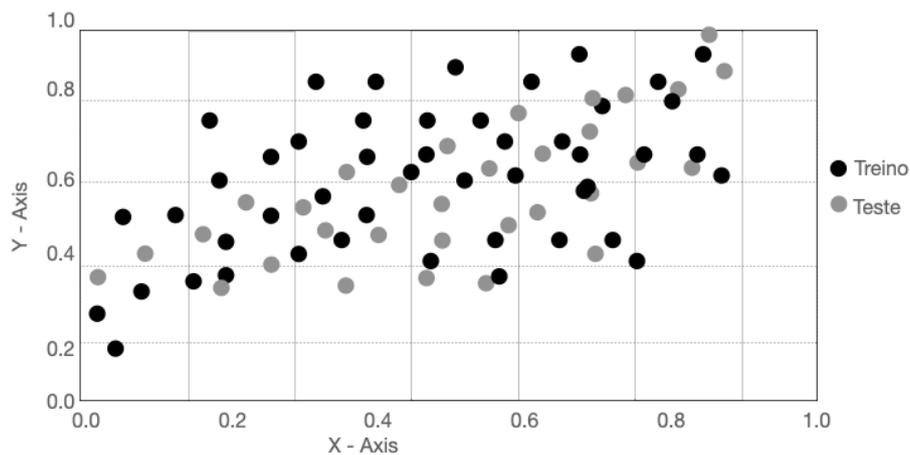


Figura 3.6: Dados de treino e teste

Fonte: Elaborado pelo autor

Na Figura 3.7, uma reta de previsão é traçada distante dos dados de treino, indicando que o modelo não encontrou uma boa relação entre os dados apresentados, mostrando um desempenho tão inferior ao esperado que não há a necessidade de ser aplicado sobre os dados de teste.

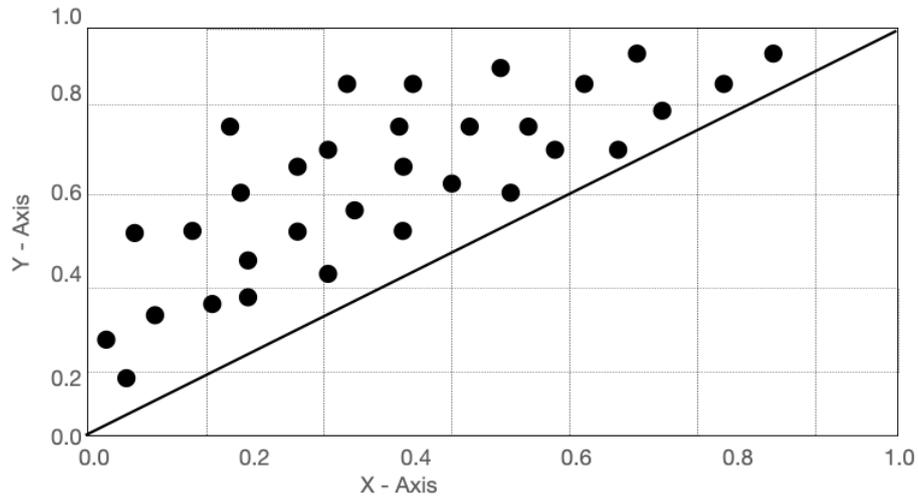


Figura 3.7: Underfitting  
Fonte: Elaborado pelo autor

No próximo exemplo, a Figura 3.8 apresenta o Overfitting, que ocorre quando o desempenho do modelo se ajusta tão bem aos dados de teste que não pode ser aplicado a outros dados.

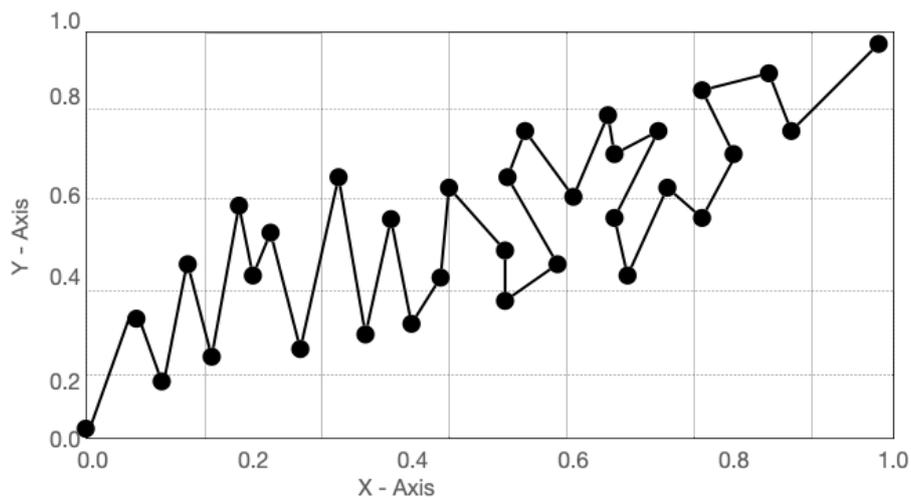


Figura 3.8: Overfitting  
Fonte: Elaborado pelo autor

A Figura 3.9 exemplifica um modelo com o ajuste desejado. Este modelo visa alinhar as expectativas de acertos às previsões, mantendo uma taxa de erros e acertos considerada ótima, a depender das expectativas do modelo desenhado.

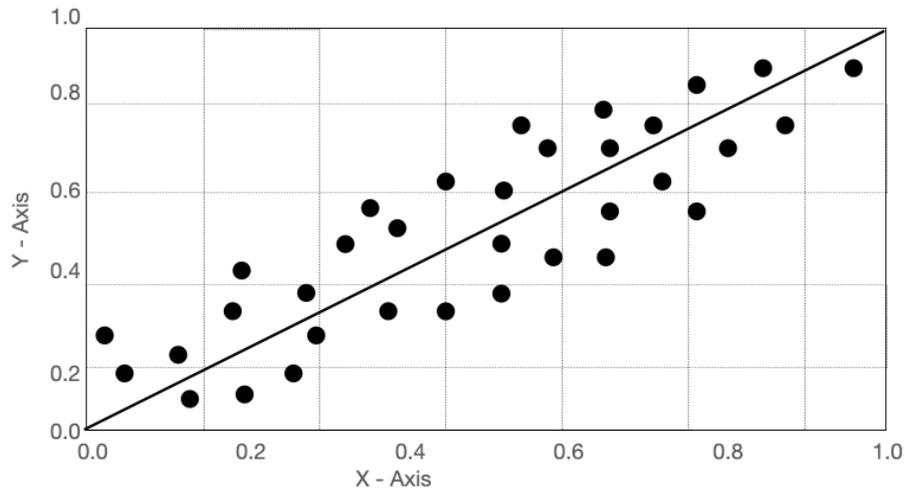


Figura 3.9: Modelo ajustado  
 Fonte: Elaborado pelo autor

Além do exposto, também precisamos entender a multicolinearidade, fenômeno que descreve a alta correlação entre as variáveis. Para isso, observamos que em uma regressão linear que calcula os parâmetros ótimos, temos que  $\hat{\beta}$  seja a melhor estimativa não viesada dos parâmetros a serem estimados. Como mostrado a fórmula a seguir:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (3.1)$$

Segundo [Paul, 2006], a multicolinearidade pode afetar principalmente a estabilidade dos coeficientes, ampliando o intervalos de confiança nos coeficientes de regressão. Desse modo, excluir ou adicionar novos dados pode alterar os coeficientes de forma significativa, gerando incertezas na correlação entre as variáveis. [Paul, 2006] também cita que a completa eliminação da multicolinearidade não é possível, mas ela pode ser reduzida pela aplicação de diferentes métodos, como aumentar a quantidade de novos dados, reespecificar o modelo, utilizar técnicas de redução de dados e fatores de regularização.

Segundo [Van Dusen, 2016], a multicolinearidade pode tornar modelos excessivamente sensíveis a mudanças de valores e, até mesmo, inverter o sinal de seus coeficientes.

A multicolinearidade é um efeito capaz de afetar seriamente as estimativas obtidas e a generalização do modelo. Portanto, identificar a exposição do modelo à multicolinearidade é essencial. [Shrestha, 2020], propõe que a multicolinearidade pode ser medida pelo Fator de Inflação da Variância - VIF representado pela fórmula:

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \quad (3.2)$$

Quanto maior for o valor do Fator de Inflação da Variância, maior será o efeito da multicolinearidade.

- Se  $\text{VIF} = 1$ , as variáveis independentes não estão correlacionadas entre si.
- Se  $1 < \text{VIF} < 5$ , correlação moderada.
- Se  $5 < \text{VIF} < 10$  alta correlação das variáveis.
- Se  $10 < \text{VIF}$ , haverá multicolinearidade entre os preditores no modelo, indicam que os coeficientes de regressão são pouco estimados com a presença de multicolinearidade.

Tendo o exposto, visando a regularização de conjuntos de dados desbalanceados, redução de dimensionalidade dos modelos e conseqüentemente a mitigação dos fenômenos citados, serão apresentadas as técnicas de regularização Ridge, Lasso e Elastic Net.

Essas técnicas podem ser definidas como uma forma de modelar conjunto de dados ao aproximar a relação de suas variáveis dependentes com uma ou mais variáveis independentes, por meio da introdução de termos regularizadores, visando a obtenção de melhores resultados.

#### • Regularização Ridge ou L2

Introduzida por [Hoerl & Kennard, 1970], sua concepção parte do princípio que a regressão de mínimos quadrados falha quando o número de variáveis usadas para predição é superior ao número de observações, isso ocorre por não haver uma diferenciação na importância das variáveis preditoras, dessa forma, são incluídas todas as variáveis durante o processo. O modelo ridge foi desenvolvido para acrescentar fatores que podem priorizar parcelas da equação, visando mitigar a imprecisão dos estimadores em modelos que possuam variáveis independentes altamente correlacionadas, ou seja, com alto valor de multicolinearidade. A regularização L2 acrescenta penalidades para soma dos quadrados dos coeficientes diferentes de zero, com a finalidade de adicionar viés suficiente para aproximar as estimativas dos resultados reais, evitar o overfitting e mitigar grandes oscilações.

Segundo [McDonald, 2009], a regressão de ridge adiciona penalidade igual ao quadrado da magnitude dos coeficientes, reduzindo todos os coeficientes ao mesmo fator sem que sejam eliminados. Dessa forma, estimula os valores a se aproximarem de zero ao introduzir uma nova informação  $\lambda I$  à diagonal produzida por  $(X^T X)$  com a finalidade de aumentar

a variância de forma artificial e conseqüentemente aprimorar sua generalização. Onde  $\lambda$  representa a quantidade de penalidade.

A regularização de Ridge pode ser apresentada da seguinte forma:

$$L_{\text{ridge}}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i^t \hat{\beta})^2 + \lambda \sum_{j=1}^p b_j^2 \quad (3.3)$$

Também podendo ser reescrita em notação matricial:

$$(Y - X\beta)^T(Y - X\beta) + \lambda\beta^T\beta \quad (3.4)$$

A notação matricial apresentada pode ser minimizada ao ser derivada, em que o termo  $\hat{\beta}$  é isolado ao igualar a relação a zero, sendo apresentada da seguinte forma:

$$\hat{\beta}_{\text{ridge}} = (X^T X + \lambda I)^{-1} X^T Y \quad (3.5)$$

Tendo o exposto, a regressão de Ridge visa extrair melhores resultados dos modelos com a introdução do fator de minimização que, ao ser introduzido no primeiro termo da equação, reduz o erro no set de treino e sua introdução no segundo penaliza a complexidade do modelo. No entanto, como informado, esta técnica não zera os coeficientes e, com isso, não elimina variáveis, porém auxilia no entendimento dos dados do modelo.

- **Regularização Lasso ou L1**

Também conhecida como norma de penalidade L1, difere-se da regressão Ridge por adicionar penalidade à soma dos valores absolutos dos coeficientes diferentes de zero. Dessa forma, irá penalizar coeficientes altos, podendo reduzir seus valores a 0. Desse modo, este método possui a capacidade de simplificar o modelo ao retirar variáveis. [Tibshirani, 1996].

Sua fórmula se assemelha a regressão de Ridge, porém, utiliza valores absolutos, assim como é mostrado na fórmula a seguir:

$$L_{\text{lasso}}(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i^j \hat{\beta})^2 + \lambda \sum_{j=1}^n |\hat{\beta}|_j \quad (3.6)$$

Também podendo ser representada em sua forma matricial:

$$(Y - X\beta)^T(Y - X\beta) + \lambda|\beta|_1 \quad (3.7)$$

## Ridge vs Lasso

Ambos os métodos citados permitem regularizar os dados e tratar variáveis correlacionadas, por meio do fator de penalização, em que:

- $\lambda = 0$ , todas as variáveis são consideradas, caso irrestrito, em que somente a soma dos mínimos quadrados será considerada.
- $\lambda \approx \infty > 0$ , a restrição passa a afetar as variáveis do sistema e, a medida que ocorre o aumento de  $\lambda$ , mais variáveis são afetadas.
- O viés (bias) aumenta a medida que  $\lambda$  aumenta.
- A acurácia (variance) aumenta a medida que  $\lambda$  reduz.

Apesar de terem fórmula semelhantes, possuem resultados diferentes. Assim como é possível ver na Figura 3.10,

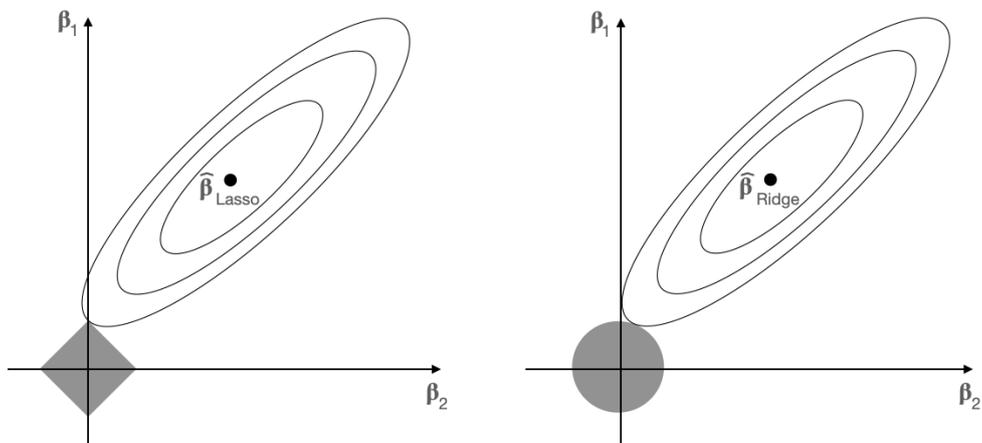


Figura 3.10: Ridge e Lasso  
Fonte: Elaborado pelo autor

As principais diferenças dos métodos citados estão detalhados a seguir na tabela Tabela 3.2.

Tabela 3.2: Lasso vs Ridge

<b>Lasso</b>	<b>Ridge</b>
Penalização L1.	Penalização L2.
Pode definir coeficientes com valores zero.	Aproxima os valores dos coeficientes a zero.
Potencial de eliminar variáveis.	Não elimina variáveis.
Se sai melhor com menor volume de parâmetros preditores que influenciam a predição da variável resposta.	Melhor desempenho com maior volume de parâmetros preditores que influenciem a predição da variável resposta grandes.
Utiliza a soma dos módulos dos coeficientes.	Utiliza o quadrado das somas dos coeficientes.
Pode resultar em modelos esparsos com poucos coeficiente e perda de informação, porém simplifica o modelo e o tornando de fácil interpretação.	A penalidade é desproporcionalmente maior para coeficientes altos, os recursos correlacionados tendem a receber coeficientes parecidos, porém não resulta em nenhuma eliminação de variáveis ou produção de coeficientes esparsos.

Fonte: Elaborado pelo autor

- **Elastic Net**

Esta metodologia surgiu para solucionar os problemas enfrentados pelas técnicas de regularização L1 e L2 e foi apresentada como uma alternativa capaz de buscar o melhor das duas técnicas [Wang et al., 2006].

Sua fórmula acrescenta tanto a regularização ridge  $\lambda_2 \|\beta\|^2$ , como a regularização lasso  $\lambda_1 \|\beta\|_1$  a função, visando mitigar as limitações citadas. Sua fórmula pode ser representada em notação simplificada da seguinte forma:

$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}}(\|y - X\beta\|^2 + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1) \quad (3.8)$$

A Figura 3.11 ilustra graficamente as diferenças entre as regressões Ridge, Lasso e Elastic Net.

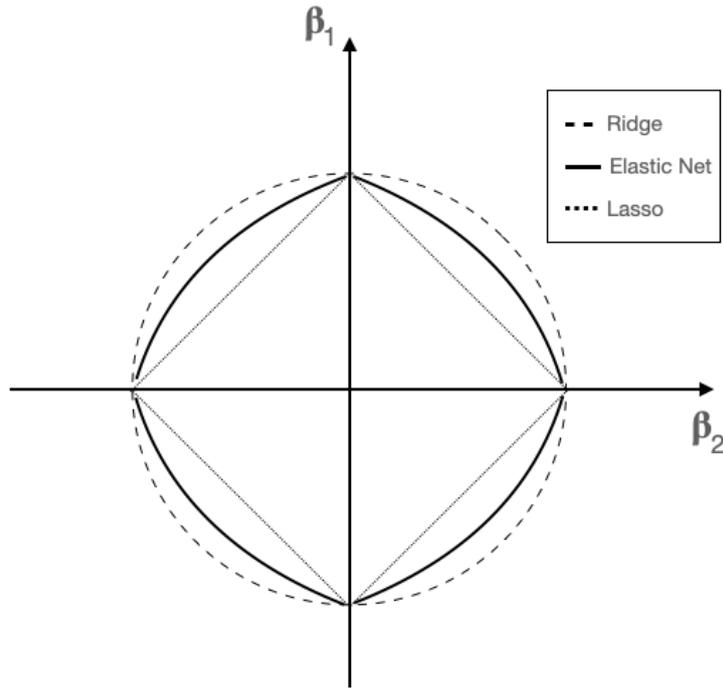


Figura 3.11: Lasso, Ridge e Elastic Net  
 Fonte: Elaborado pelo autor

## Abordagem de Classificação

- **Logit**

Conhecida como regressão logística ou classificador de máxima entropia, pode ser classificada como logit binomial, quando for composta por duas opções de classificação; e logit multinomial, quando existirem mais opções de classificação. Sua função é baseada no ajuste adequado de um modelo de variáveis categóricas, por meio da transformação logarítmica das probabilidades sobre a ocorrência de um evento acontecer ou não, dessa forma, produz uma curva em  $S$  capaz de contemplar a maioria dos pontos de cada opção do modelo [Hosmer Jr et al., 2013].

Segundo [Trueck & Rachev, 2009], é considerada uma poderosa técnica para problemas de classificação (dados categóricos) e tem sido usada para explicar a probabilidade de respostas binárias, ou seja, eventos que entraram ou não em *default* e sua relação entre as variáveis explicativas, tendo como uma de suas primeiras aplicação contextos de dificuldades financeiras. Essa técnica pondera as variáveis independentes ao atribuir uma pontuação a  $Y$  na forma de probabilidade de falha em uma amostra, de modo que a distribuição logística transforma o intervalo de regressão entre  $(0, 1)$  para os valores de  $Y$ . [Trueck & Rachev, 2009] também ressalta que a vantagem dessa abordagem é que não assume normalidade multivariada e matrizes de covariância iguais, assim como não

há necessidade relação linear entre as variáveis de entrada e saída.

A Figura 3.12 ilustra graficamente a curva em S realizada pela regressão logística.

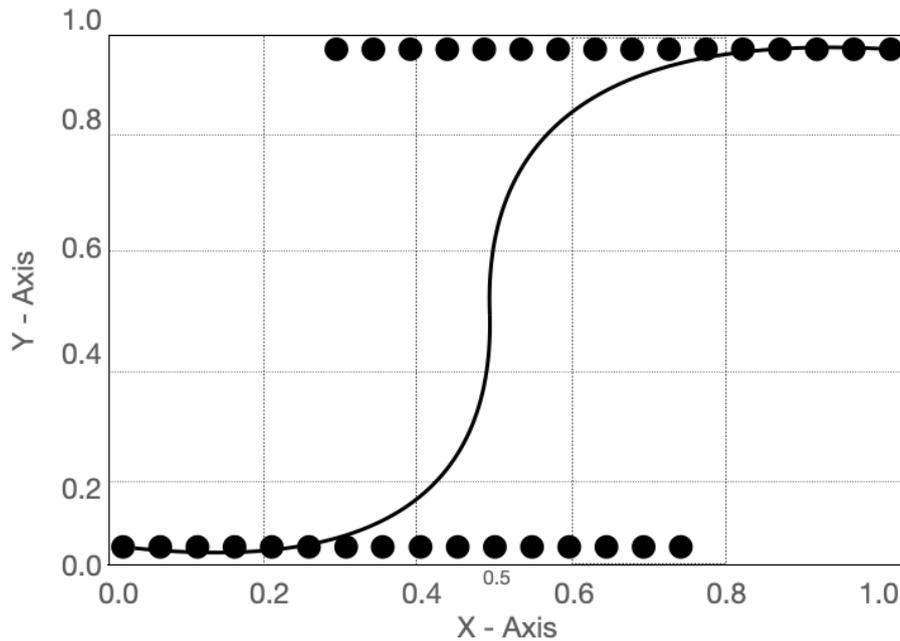


Figura 3.12: Regressão Logit  
Fonte: Elaborado pelo autor

Sua função pode ser descrita como a probabilidade de um determinado evento ( $p(x)$ ) ocorrer como variável resposta. Sua função pode ser representada a seguir pela probabilidade do evento  $Y = 1$  ocorrer condicionada ao valor de  $X$ :

$$P(Y = 1) = \frac{1}{1 + e^{-g(x)}} \quad (3.9)$$

Em que  $g(x)$  equivale a um conjunto de  $p$  variáveis independentes  $X_1, X_2, X_3 \dots X_p$ .

Desse modo,  $g(x) = (\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})$  e a equação pode ser reescrita da seguinte forma.

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}} \quad (3.10)$$

Quando:

- $g(x) \in -\infty$ ,  $P(Y=1) \geq 0,5$  então classifica-se  $Y=1$ .
- $g(x) \in +\infty$ ,  $P(Y=1) < 0,5$  então classifica-se  $Y=0$ .

Em regressões lineares, os parâmetros dos modelos são estimados por mínimos quadrados. No entanto, em regressões logísticas, os parâmetros  $\beta_1, \beta_2, \beta_3 \dots \beta_p$  são estimados

pelo método máxima entropia (verossimilhança) e sua interpretação é realizada por meio da razão da probabilidade de sucesso sobre a probabilidade de fracasso (ODDS), Em que  $P$  é a probabilidade do evento ocorrer.

$$odds = \frac{p}{(1 - p)} \quad (3.11)$$

Por sua vez, a quantificação da força de associação entre os dois eventos é obtida pela comparação entre as probabilidades de sucesso e a probabilidade de fracasso usando a função Razão de Chance (ODDS Ratio), representado pela função:

$$ratio = \frac{odds_1}{odds_2} \quad (3.12)$$

Parâmetros para uma boa implementação:

- Assumir que os pontos entre as variáveis de entrada e saída é linear;
- Variáveis independentes não podem ter multicolinearidade, tendo pouco ou nenhuma relação com a variável preditora;
- Grande volume de dados amostrais para gerar resultados mais confiáveis de análise;
- Não possuir heterocedaticidade;
- Valores de resíduo iguais a zero;
- Não pressupõe normalidade dos resíduos nem homogeneidade de variâncias.

- **Árvore de Decisão**

As Árvores de Decisão são algoritmos supervisionados conhecidos por sua flexibilidade e facilidade de interpretação, assim como é mostrado na figura Figura 3.13. Elas podem ser chamadas de Árvores de Classificação, quando tratarem problemas de classificação com variáveis categóricas; ou Árvores de Regressão, quando os problemas representam variáveis numéricas. Também podem ser otimizadas por diversos algoritmos que exploram paralelização e particionamento de dados [Su & Zhang, 2006].

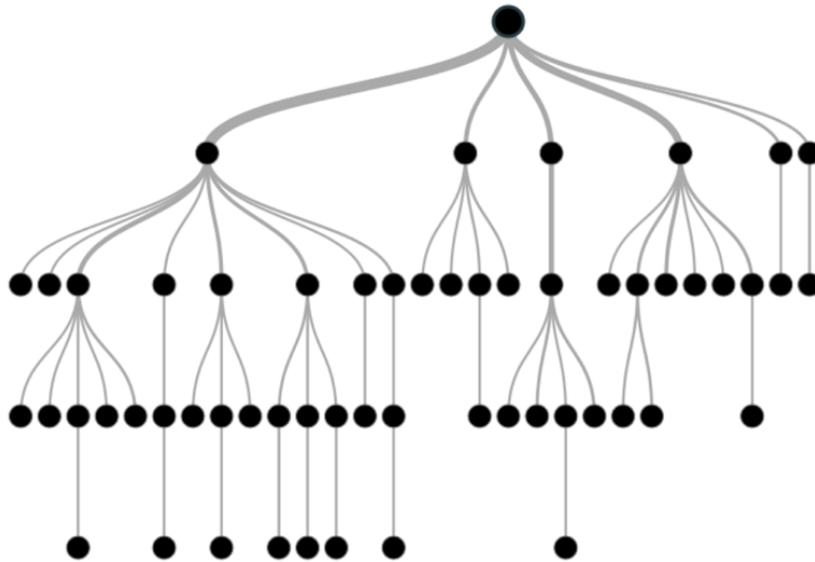


Figura 3.13: Árvore de Decisão  
 Fonte: [Data Science](#)

Segundo [Charbuty & Abdulazeez, 2021]. O funcionamento da Árvore de Decisão é caracterizado pelos termos:

- *Nó inicial*: representado pela amostra original de dados, que se dividirá em possíveis resultados.
- *Nó Intermediário*: são subamostras originadas a partir da aplicação da decisão tomada. Desses Nós podem originar novas subamostras.
- *Nó Final*: uma subamostra que não sofre processo de partição. Mostra o resultado final de um caminho de decisão.
- *Nó de probabilidade (Intermediários ou Folha)*: representado por um círculo que mostra a probabilidade de determinado resultado ocorrer.
- *Nó de decisão*: é representado por um quadrado, mostra uma decisão tomada.

Seu funcionamento é intuitivo:

- Um conjunto de dados de informações é atribuído ao Nó inicial.
- Os dados originais são submetidos a processos de decisão e, para cada processo, os dados são particionados de forma binária antes de seguir para o Nó subsequente, em que o processo se repete.
- O passo anterior ocorrerá até a chegada ao Nó Final.

A árvore é o resultado final gerado a partir das partições binárias sucessivas de uma amostra original. A Figura 3.14 ilustra a representação do funcionamento de uma Árvore de Decisão.

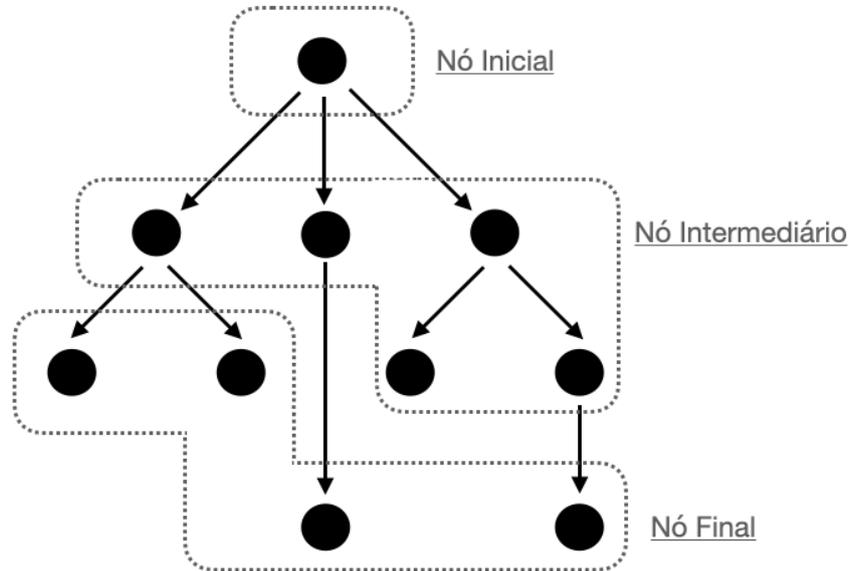


Figura 3.14: Funcionamento da Árvore de Decisão  
Fonte: Elaborado pelo autor

A fórmula da árvore de decisão pode ser representada da seguinte forma:

$$f : \{0, 1\}^n \rightarrow \{0, 1\} \text{ Ou } (\mathbf{x}, Y) = (x_1, x_2, x_3, \dots, x_k, Y) \quad (3.13)$$

Árvores de decisão são ótimos classificadores por sua performance aliada a sua simplicidade. Em estudo feito por [Caruana & Niculescu-Mizil, 2006], uma comparação em larga escala foi realizada com onze conjuntos de dados diferentes, seus resultados foram avaliados por oito métricas de desempenho diferentes e observou-se que os classificadores de aprendizagem binária supervisionada que apresentaram melhores performances foram: árvores de decisão impulsionadas (Boosted Tree), Florestas Aleatórias (Random Forast), Árvores de Decisão (Bagged Tree) e depois o Máquina de vetores de suporte (Support Vector Machine- SVM), respectivamente, dentre o total de classificadores avaliados (SVMs, neural nets, logistic regression, naive bayes, memory-based learning, random forests, decision trees, bagged trees, boosted trees, and boosted stumps).

[Guégan & Hassani, 2018] realiza uma extensa comparação entre as abordagens de aprendizado de máquina, discutindo o credit scoring implementado por instituições financeiras e os desafios encontrados para implementação dos modelos. Em seus resultados, foi sugerido que em média Random Forast pode parecer melhor, mas ressalta a atenção em caso de evolução do conjunto de dados.

[Munkhdalai et al., 2019b] faz uma comparação abordagens de predição de especialistas humanos e aprendizagem de máquina no domínio financeiro concluindo que se as instituições tivessem usado modelos próprios baseados em aprendizagem de máquina, suas perdas de crédito esperadas teriam sido menores e seriam mais sustentáveis, também aponta que a maior área sob a curva (AUC) e precisão foram obtidas pelo algoritmos de redes neurais e Extreme Gradient Boosting da extensa comparação realizada.

Em termos coloquiais, a Árvore de Decisão é um agrupamento de if-else, em que, para cada condição, um caminho será tomado. Seu propósito é realizar diversas divisões de dados em subconjuntos de modo que cada subconjunto fique mais puro a medida que o número de variáveis alvo é reduzido. Métodos como SVMs, regressão logística e redes neurais profundas resolvem os mesmos problemas, no entanto, eles são mais difíceis de interpretar e podem exigir muitos ajustes de hiper-parâmetros antes que um bom resultado seja obtido.

Para entender como ocorre o processo que gera a melhor classificação ou divisão de cada Nó da Árvore de Decisão, serão apresentados os conceitos de seleção índice de Gini e Ganho de Informação, critérios de partição mais conhecidos para determinar o grau de incerteza quando a Árvore de Decisão é usada para classificação [Onoda, 2001].

### Impureza de Gini

Segundo [Provost & Fawcett, 2013], é utilizado para medir o nível de desordem dos elementos, medindo a frequência da classificação errônea de cada variável ao ser escolhida de forma aleatória.

O Índice de Gini é compreendido como sendo a área acumulada observada pela curva de Lorenz sendo representado matematicamente como a razão das áreas expressas como:

$$G = \alpha / (\alpha + B) \quad (3.14)$$

$\alpha$  - área de desigualdade observada.

$(\alpha + B)$  - área triangular de total igualdade de renda.

Sendo que  $\alpha$  representa a área localizada entre a linha de total igualdade das variáveis binomiais e a curva de Lorenz.  $B$  representa a área abaixo da curva de Lorenz, assim como mostra a Figura 3.15.

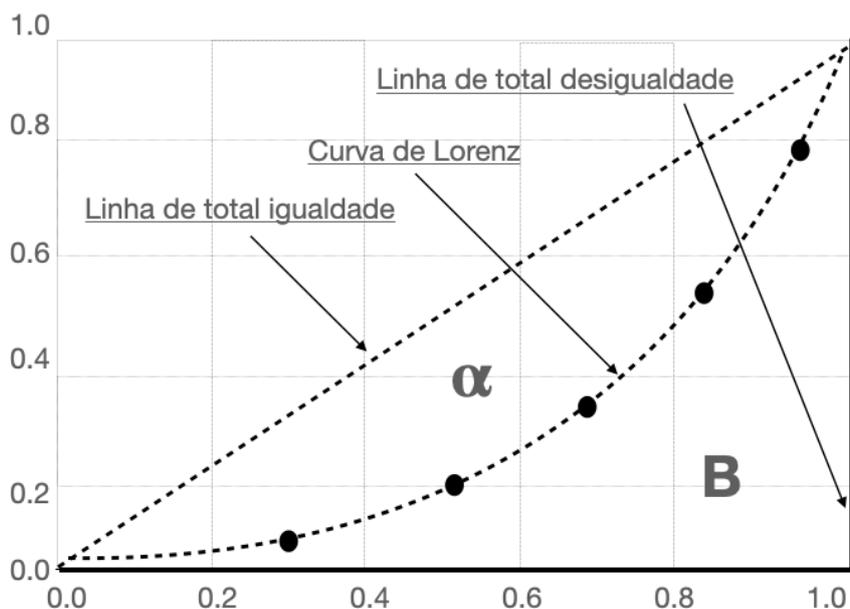


Figura 3.15: Curva de Lorenz  
Elaborado pelo autor

Quanto maior a área observada em  $\alpha$ , maior o índice de Gini e maior a desigualdade. O coeficiente de Gini pode ser calculado pela expressão:

$$Gini = 1 - \sum_{i=1}^n (p_i)^2 \quad (3.15)$$

$p_i$  = probabilidade de classificação de um elemento em uma classe.

$n$  = o número de classes.

- Seu valor mínimo é 0, quando o conjunto de dados é classificado como sendo de uma única classe e não podem ser mais divididos.
- Seu valor máximo é 1, quando há distribuição aleatória em várias classes.
- O valor 0,5 demonstra que há distribuição igualitária em classes binárias.

A classificação com o menor Índice de Gini será escolhida para compor o Nó da Árvore de Decisão.

## Entropia

Segundo [Provost & Fawcett, 2013], é usada para compor o cálculo do Ganho de Informação, representa a medição da impureza de cada subconjunto da árvore. É a probabilidade de obter um evento a partir de uma amostra de dados. Seu resultado está entre 0 e 1, e pode ser representado matematicamente pela fórmula:

$$entropia = - \sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (3.16)$$

$p(x_i)$  = proporção de exemplos em relação a todo.

Sua medição é dada pela subtração das probabilidades dos eventos binários ocorrerem, como pode ser visto a seguir:

$$entropia = -(p(x_{i+}) \log_2 p(x_{i+})) - (p(x_{i-}) \log_2 p(x_{i-})) \quad (3.17)$$

$p(x_{i+})$  = proporção dos eventos positivos e o total de eventos do conjunto.

$p(x_{i-})$  = proporção dos eventos negativos e o total de eventos do conjunto.

O Gráfico de distribuição da Entropia pode ser representado como mostra a Figura 3.16, em que o eixo vertical representa a entropia e o eixo horizontal a distribuição dos dados.

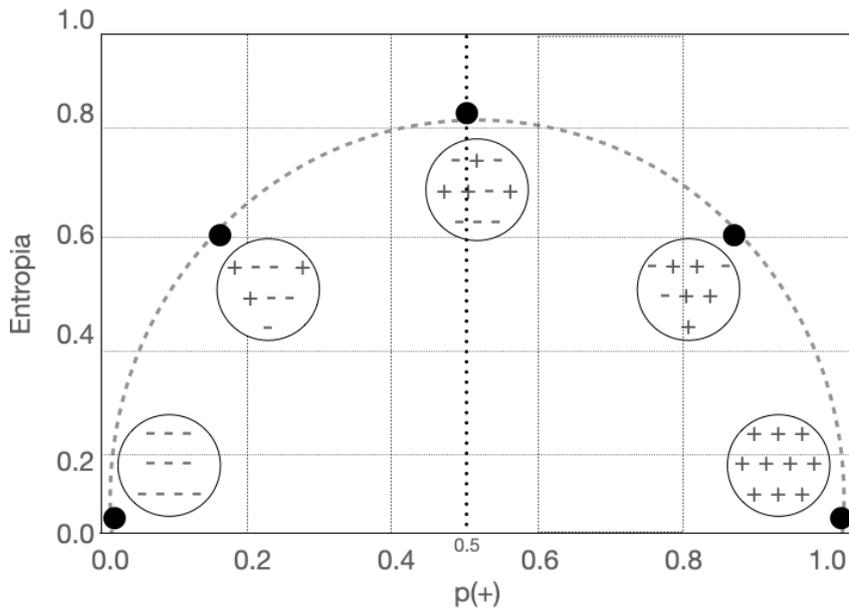


Figura 3.16: Entropia  
Elaborado pelo autor

É possível observar pela Figura 3.16 que, quanto mais próximo de 1 for a Entropia, maior a desordem dos dados, quanto menor for a entropia, menor seria a desordem. Isso significa que quanto mais próxima de 1 for a entropia, mais perto de se obter 50% de probabilidade nas decisões binárias do conjunto de dados apresentado.

## Ganho de Informação

Segundo [Provost & Fawcett, 2013], é usada para calcular o ganho de informação obtido pela seleção do conjunto de dados responsável para compor os ramos da Árvore, utiliza da entropia como parte integrante de seus cálculos e é aplicada em cada etapa de seleção para escolher o conjunto de dados na escolha da divisão inicial e em divisões subsequentes de forma mais consistente. É a medida que representa a efetividade de separação de registros de uma seleção do conjunto. Seu calculo é realizado quando se compara a entropia atual com a entropia obtida pela nova ramificação.

O ganho de informação pode ser expressa pela fórmula:

$$Gain(S, A) = Entropia(S) - \sum_{v \in \text{valores}(A)} p(A_v) * Entropia(A_v) \quad (3.18)$$

S = todo o conjunto.

A = atributo em questão.

$\sum_{v \in \text{valores}(A)}$  = soma dos pesos dos filhos (A) dado pelo  $N^\circ$  de Amostras do Filho pelo  $N^\circ$  de Amostras do Pai.

$Entropia(A_v)$  = representa a Entropia.

O ganho de informação é a subtração das entropias ponderadas dos ramos da entropia original, quanto maior o ganho de informação, melhor.

O peso atribuído ao nó filho é o número de amostras no nó dividido pelas amostras totais de todos os nós filhos.

$$peso(filho) = \frac{N^\circ \text{AmostrasFilho}}{N^\circ \text{AmostrasPai}} \quad (3.19)$$

Da mesma forma, o ganho de informação também pode ser calculado com a pontuação obtida pelo calculo de Gini.

## Ganho de Informação vs Impureza de Gini

Ambas as estratégias trabalham de formas similares medindo a incerteza e a quantidade esperada de informações que seriam necessárias para seleção de novas instâncias. Medidas base para qualquer algoritmo baseado em Árvores de Decisão.

A Figura 3.17, em que o eixo vertical representa o índice de impureza calculado pela Entropia e Índice de Gini e o eixo horizontal a distribuição dos dados:

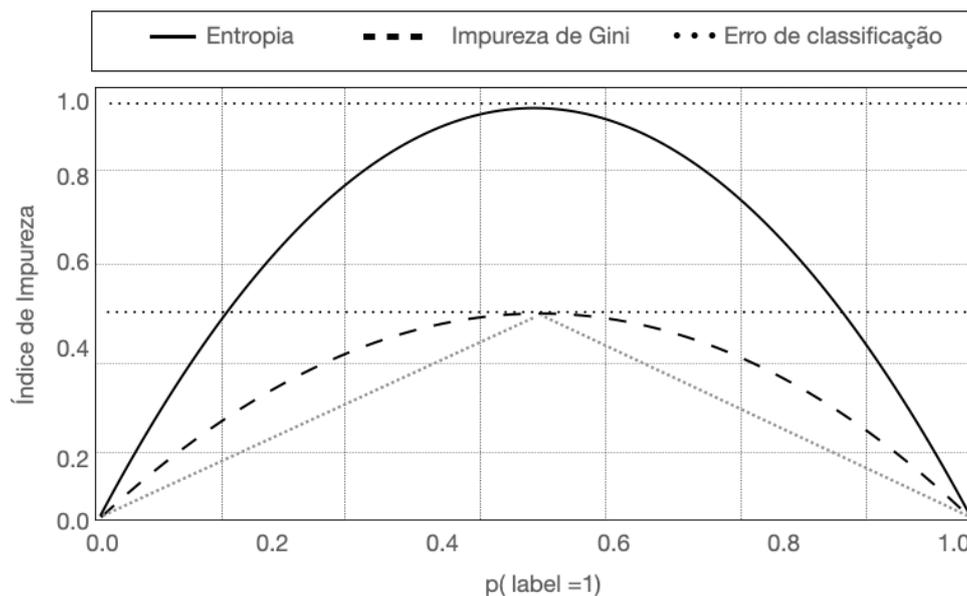


Figura 3.17: Gini vs Entropia  
 Fonte: Elaborado pelo autor

Considerando o expoto, a Tabela 3.3 apresenta algumas das principais diferenças entre as duas técnicas.

Tabela 3.3: Ganho de Informação vs Impureza de Gini

<b>Ganho de Informação</b>	<b>Impureza de Gini</b>
Balanceamento de dados por uso de logarítmica, é computacionalmente mais complexa que o Índice de Gini.	Computacionalmente menos intensivo, tende a ser preferível por limitações em volumes elevados de conjuntos de dados.
Por se utilizar de Entropia como balanceador de registros em cada Nó, possui maior gama de resultados. Capaz de calcular a entropia antes e depois de divisões de dados de dados e indica a impureza de seus elementos.	Nas Árvores de Classificação com partições binárias, tende a isolar num ramo os registros que representam as classes mais frequentes, ao limitar os resultados a apenas um.
Intervalo com valor mínimo 0 indo até 1, antes de começar sua redução.	Intervalo entre 0 até 0,5, antes de começar sua redução.
Utiliza os algoritmos ID3, C4.5.	Utiliza os algoritmos CART.

Fonte: Elaborado pelo autor

Segundo [Charbuty & Abdulazeez, 2021], as Árvores de Decisão podem ser divididas em métodos de Bagging (Ensacamento), Boosted Trees (Impulsionamento). Esses modelos são considerados métodos de aprendizado em conjunto (Ensemble Methods), ou seja, métodos que agregam um conjunto de algoritmos de aprendizado para obter um melhor desempenho preditivo. Pois, ao realizar a predição combinando modelos de classificação, o erro fornecido tende a ser menor, podendo reduzir o viés e a variância [Opitz & Maclin, 1999].

- **Impulsionamento (Boosted Trees)**

Consiste em treinar diversos modelos de forma individual e em sequencia, ou seja, cada modelo treinado irá aprender com os erros modelo anterior, utilizando de pesos para observações que foram piores estimadas no modelo anterior. Os resultados são combinados para chegar ao resultado final [Zhou, 2012]. Assim como mostra a Figura 3.18.

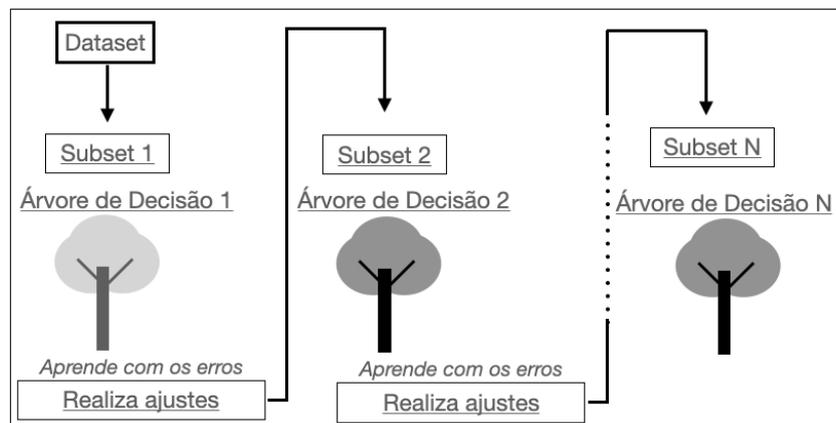


Figura 3.18: Boosting Trees  
 Fonte: Elaborado pelo autor

Este algoritmo visa construir um classificador de desempenho alto ao discriminar classificadores de baixo desempenho.

A fórmula do método de impulsionamento visa reduzir o erro quadrático médio  $\frac{1}{n} \sum_i (\hat{y}_i - y_i)^2$ , em uma função  $\hat{y} = F(x)$ , e sua fórmula pode ser expressa a seguir:

$$F_{m+1}(x_i) = F_m(x_i) + h_m(x_i) = y_i \tag{3.20}$$

$\hat{y}_i$  = valor a ser previsto.

$y_i$  = o valor observado  $F_{m_x}$ .

$n$  = número de amostras em  $y$ .

$m$  = estágios.

- **Bagging (Bootstrap Aggregating)**

Introduzido por Leo Breiman em 1994, seu funcionamento parte da combinação de previsões de diversas árvores de decisão criadas a partir de um conjunto de dados. As árvores tiveram sua amostra de dados remodelada de forma uniforme e em paralelo, a partir da fonte. Cada árvore é independente e treinada por um subconjunto aleatório dos dados. Em seguida, as árvores votam para chegar a um consenso [Breiman, 1996].

Este método é composto das etapas: aggregating e bootstrapping. Seu funcionamento consiste em gerar conjunto de amostras a partir da reamostragem bootstrap de dados, em que cada amostra será ajustada com um modelo preditivo de forma individual, em seguida, ocorre a agregação das previsões realizadas.

Tendo o exposto, ressalta-se a citação de [Opitz & Maclin, 1999], ao sugerir que utilizar combinações de preditores tende a fornecer um erro menor que a utilização de modelos individuais. Esta afirmação vai ao encontro do que proposto por [Breiman, 1996], que também informa que a utilização do método bagging reduz a variância e o erro quadrático médio das previsões.

A Figura 3.19 ilustra a separação do conjunto de dados com substituição de forma paralela e aleatória pela técnica bootstrapping. É importante ressaltar que quando as amostras são selecionadas sem substituição, ocorre dependência entre seleções, e este não é o objetivo.

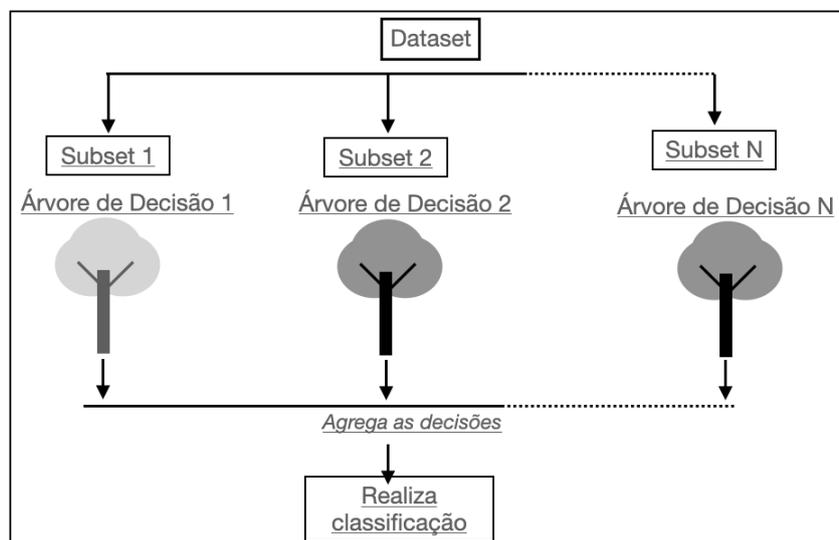


Figura 3.19: Bagging: processo de separação

Fonte: Elaborado pelo autor

O método Bagging pode ser dividido em três etapas:

1. Realizar a técnica bootstrap para gerar B conjunto de observações ajustadas por um modelo preditivo.
2. Treinar o modelo para obter as predições em X.
3. Realizar a agregação dos resultados e calcular a média das predições.

A fórmula do método Bagging pode ser observada a seguir, em que  $\hat{h}_{bag}(x)$  representa o modelo de árvore gerado a partir da b-ésima amostra de bootstrap  $\hat{h}^b(x)$ .

$$\hat{h}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{h}^b(x) \quad (3.21)$$

Tendo o expoto, na Tabela 3.4 a seguir foram separadas as principais características dos modelos Bagging, Boosting.

Tabela 3.4: Principais características entre Bagging, Boosting

<b>Boosting</b>	<b>Bagging</b>
Não utiliza Bootstrap.	Utilizam Bootstrap.
Amostras dependentes.	Amostras independentes.
Peso atribuído aos casos mais difíceis.	Não há atribuição de peso.
Visa redução de viés e pode aumentar overfitting.	Visa redução de overfitting e pode aumentar o viés.
Indicado para modelos com baixo desempenho por evitar o viés.	Indicado para modelos com dificuldades em ajuste excessivo.
AdaBoost, Gradient Boosting, XGBoost, Light GBM e CatBoost.	Utiliza os algoritmos CART.

Fonte: Elaborado pelo autor

- **Floresta Aleatória (Random Forast)**

Algoritmo robusto baseado no método bagging, capaz de reduzir a correlação entre as decisões das árvores, ao modificar a forma de separação dos conjuntos de dados para torná-los diferentes. Seu funcionamento se assemelha ao método bagging, diferenciando-se por aleatorizar as várias árvores criadas a partir da separação dos conjuntos de dados e utilizar apenas uma vez os subconjunto de preditores, para garantir que as árvores sejam diferentes umas das outras em sua predição [Mishina et al., 2015].

Segundo, [Breiman, 2001], são algoritmos que mapeiam amostras em vetores aleatórios para produzir Árvores de Decisão independentes com a mesma distribuição. A medida

que a Floresta de Decisão cresce em tamanho, seu resultado converge para correlação de média ou frequência de todos os resultados. A Figura 3.20 ilustra o funcionamento do Random Forast.

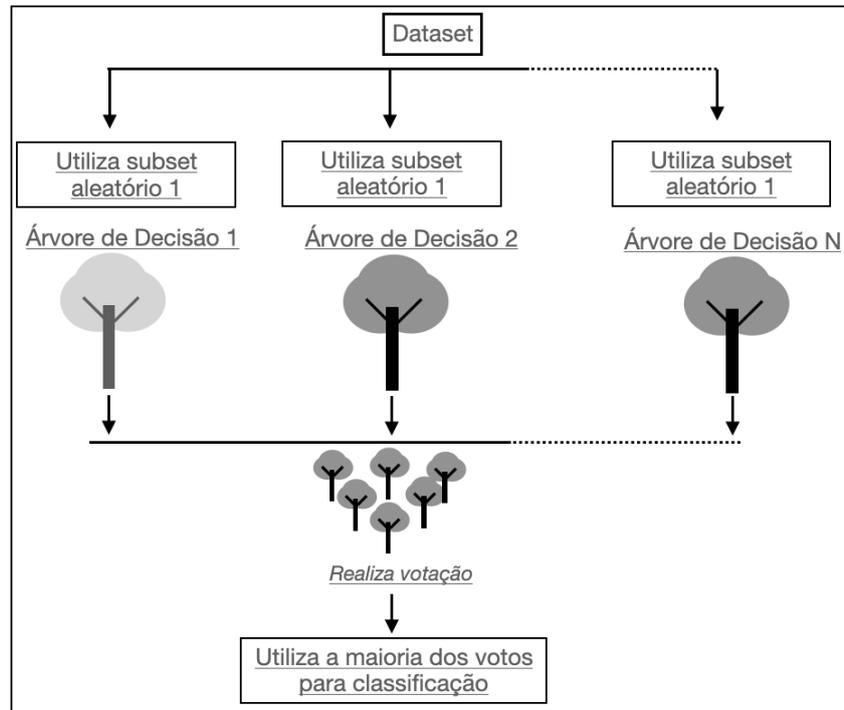


Figura 3.20: Random Forast  
 Fonte: Elaborado pelo autor

A fórmula normalmente utilizada para separar os dados de forma aleatória que irão compor o conjunto de dados de cada árvore pode ser representada da seguinte forma:

$$m = \sqrt{p} \text{ ou } \log_2 p \quad (3.22)$$

$m$  = covariáveis escolhidas de forma aleatória do total. A cada nó, um novo subconjunto é sorteado.

$p$  = representa o número de observações.

- **AdaBoost**

Algoritmo baseado no método de impulsionamento com a finalidade de utilizar a combinação de classificadores fracos para gerar classificadores fortes. Foi o primeiro algoritmo de boosting prático e continua sendo largamente utilizado [Schapire, 2013].

- **Gradient Boosting**

Introduzido por [Breiman, 1997], este algoritmo é baseado na técnica de impulsionamento e seu objetivo é minimizar a função de perda (o erro representado pela diferença do valor esperado e o valor previsto).

Ele calcula a derivada (gradiente) da função de perda (erro) visando descrever sua inclinação. Assim como representado na Figura 3.21.

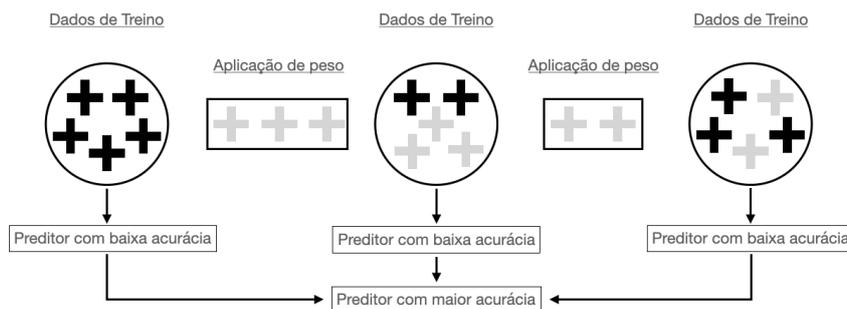


Figura 3.21: Gradient Boosting  
 Fonte: Elaborado pelo autor

Neste modelo, novos parâmetros são produzidos com base no gradiente da função de erro calculado para cada parâmetro do modelo. Esse processo deverá ser realizado para cada passo subsequente, com o intuito de reduzir o erro em cada rodada.

Também pode ser utilizado para informar a direção de alteração dos parâmetros do modelo no intuito de reduzir o erro nas rodadas de treinamento [Li, 2016].

Sua fórmula é representada da seguinte forma:

$$h_m(x) = y - F_m(x) \tag{3.23}$$

Sua principal diferença para o AdaBoost está em como os algoritmos identificam as deficiências dos classificadores fracos. Enquanto o AdaBoost atribui e eleva os pesos dos classificadores, o Gradient Boosting realiza o mesmo processo, porém, utilizando gradientes em função de perda.

- **Extreme Gradient Boosting**

Segundo [Chen & Guestrin, 2016], criador do algoritmo, o XGBoost, assim como o Gradient Boosting, também utiliza do método de impulsionamento e cálculo da derivada da função de erro. No entanto, o XGBoost utiliza as técnicas de regularização L1 e L2, calcula gradientes de segunda ordem que fornecem mais informações sobre a direção do gradiente e lida com valores ausentes, dessa forma permitindo a otimização dos modelos

ao controlar o overfitting. Também é capaz de paralelizar e distribuir os cálculos em clusters, dando maior poder computacional e resolver problemas do mundo real usando quantidade mínima de dados.

Projetado para melhorar a velocidade da árvore de decisão Gradient Boosting, exigem mais conhecimento de ajustes que modelos como Random Forest. Esse algoritmo tem dominado as competições da Kaggle, sendo líder em trabalhar com dados tabulares padrão em problemas de classificação [Kaggle, 2020].

Apesar de ter sido considerado um dos modelos mais poderosos para problemas de classificação com dados tabulares, também pode ser usado para problemas de regressão.

Sua fórmula pode ser representada da seguinte forma:

$$\hat{f}_{(0)}(x) = \arg \min_{\theta} \sum_{i=1}^N L(y_i, \theta) \quad (3.24)$$

## Medidas de ajuste

Ao realizarmos a avaliação de um modelo, este pode apresentar um baixo viés e alta variância, sugerindo um sinal indicador de Overfitting, ou apresentar baixa variância e alto viés, sugerindo um sinal indicador de Underfitting. Também é importante ressaltar que em machine learning, trabalhamos com os termos acurácia e precisão, em que uma alta acurácia corresponde a uma baixa variância e uma alta precisão corresponde a um baixo viés [Koehrsen, 2018].

Compreender como esses termos se relacionam ao realizarmos o ajuste do modelo é importante, para isso, primeiro precisamos entender o conceito de Matriz de Confusão.

Segundo [Visa et al., 2011], a Matriz de Confusão é composta por quatro valores:

- Verdadeiro positivo (VP): quantidade de classificações corretamente positivas. O modelo acertou na classificação positiva.
- Falso positivo (FP): quantidade de classificações erroneamente positivas. O modelo errou ao classificar de forma positiva um registro negativo.
- Verdadeiro negativo (VN): quantidade de classificações corretamente negativas. O modelo acertou em classificar negativamente.
- Falso negativo (FN): quantidade de classificações erroneamente negativas. O modelo errou ao classificar de forma negativa um registro positivo.

A Tabela 3.5 representa como os dados se comportam na Matriz de Confusão.

Tabela 3.5: Matriz de Confusão

Valor Verdadeiro	Valor Previsto	
	Positivo	Negativo
Positivo	<i>Verdadeiro Positivo (VP)</i>	<i>Falso Positivo (FP)</i>
Negativo	<i>Falso Negativo (FN)</i>	<i>Verdadeiro Negativo (VN)</i>

Fonte: Elaborado pelo autor

Por meio da avaliação dos valores apresentados na Matriz de Confusão são calculadas as medidas de avaliação: Acurácia, Precisão, Especificidade, Revocação, F1- Score, Curva ROC. Que serão detalhadas a seguir segundo [Powers, 2020] e [Sokolova et al., 2006].

- **Acurácia**

Medida básica para calcular o desempenho do modelo. Quanto maior for o valor da acurácia, maior será o número de acertos do modelo e menos erros serão cometidos. Representa quão próximos um conjunto de medições está do valor real. A Figura 3.22 ilustra o foco pretendido da Acurácia.

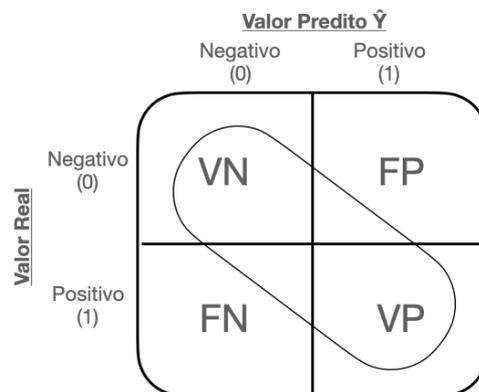


Figura 3.22: Acurácia  
Fonte: Elaborado pelo autor

É considerada um indicador de cálculo simples, obtida pela divisão de todos os acertos pelo total. Sua fórmula pode ser representada da seguinte forma:

$$\text{Acurácia} = \frac{vp + vn}{vp + vn + fp + fn} \quad (3.25)$$

Após a obtenção da Acurácia, calcula-se o Erro do modelo com a fórmula a seguir:

$$\text{Erro} = 1 - \text{Acurácia} \quad (3.26)$$

É importante ressaltar que a Acurácia fornece valores globais e não é suficiente para elucidar o que ocorre em determinados modelos. Pois, podemos identificar uma acurácia alta, porém, devido ao baixo balanceamento de dados. Ou seja, se tivermos conjunto de dados com classes predominantes, o modelo ao acertar os dados predominantes irá obter uma acurácia alta, mesmo que tenha falhado em acertar todos os dados não predominantes.

- **Precisão**

É a proporção das predições corretas de um conjunto em relação a todas as previsões feitas desse conjunto. Representa a fração recuperada que é considerada relevante, ou seja, seu foco é em quantificar a proporção de acertos dos dados previstos como positivos, assim como é representado na Figura 3.23.

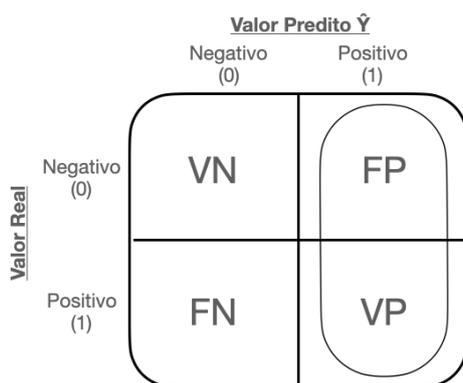


Figura 3.23: Precisão  
Fonte: Elaborado pelo autor

Sua fórmula pode ser apresentada da seguinte forma:

$$\text{Precisão} = \frac{vp}{vp + fp} \quad (3.27)$$

Assim como na Acurácia, o desbalanceamento dos dados também causam anomalias na avaliação dos dados por meio da Precisão, ao não considerar os falsos negativos (fn), um modelo terá uma precisão alta mesmo que sua taxa de verdadeiros positivos (tp) seja baixa.

- **Especificidade**

Visa identificar a capacidade do modelo de prever a classificação negativa ao apresentar a proporção dos dados classificados como verdadeiros negativos retornados pelo modelo frente ao total de dados realmente negativos. Assim como é representado na Figura 3.24 a seguir.

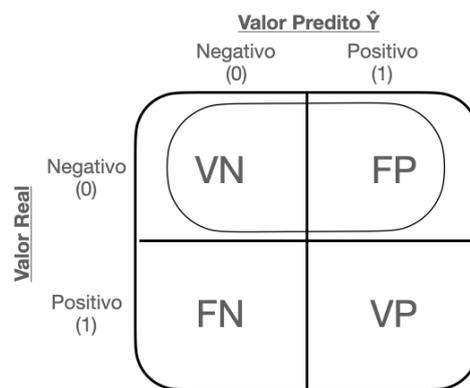


Figura 3.24: Especificidade  
Fonte: Elaborado pelo autor

Sua fórmula pode ser representada da seguinte forma:

$$\text{taxa de verdadeiros negativos} = \frac{vn}{vn + fp} \quad (3.28)$$

- **Revocação (Recall) ou Sensibilidade**

Visando mitigar os efeitos apresentados pela Acurácia e Precisão que foram causados pelo desbalanceamento dos dados, esta métrica é utilizada para avaliar os valores recuperados dentre os valores classificados como positivos reais. Seu foco visa quantificar quanto o modelo está capturando do que realmente ele deveria capturando. Ou seja, sua proporção de acertos. Indicado quando o objetivo é otimizar os verdadeiros positivos, mesmo que prejudique os falsos positivos. A Figura 3.25 ilustra o foco pretendido da Revocação.

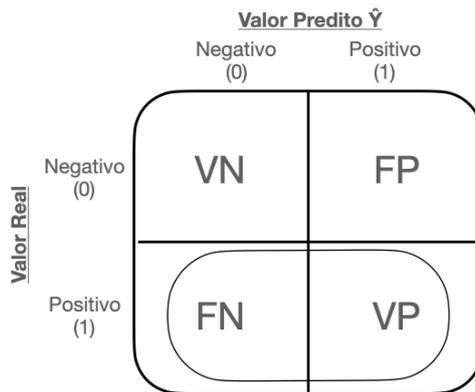


Figura 3.25: Revocação (Recall)  
Fonte: Elaborado pelo autor

Sua fórmula é expressa da seguinte forma:

$$\text{Revocação} = \frac{vp}{vp + fn} \quad (3.29)$$

- **F1-Score**

Utilizada como boa sugestão para mitigar problemas de conjunto de dados desproporcionais. É considerada como sendo uma média harmônica entre a precisão e o recall. Ou seja, é uma forma de observar em um único número a relação entre a precisão e o recall trazendo uma forma de observar a qualidade geral do modelo.

Sua fórmula pode ser representada da seguinte forma:

$$F1score = 2 \cdot \frac{\text{precis} * \text{revoc}}{\text{precis} + \text{revoc}} \quad (3.30)$$

- **Curva ROC**

Utilizada para problemas de classificação binária, é gerada a partir da medição da Sensibilidade (Recall) e Especificidade. Permite acompanhar de forma gráfica a variação da Especificidade e Sensibilidade da performance do modelo, desse modo, auxiliando na escolha do melhor ponto de corte para otimização do desempenho. A Figura 3.26 seguir descreve de forma gráfica como ficam as curvas ROC para um modelo de casos hipotéticos.

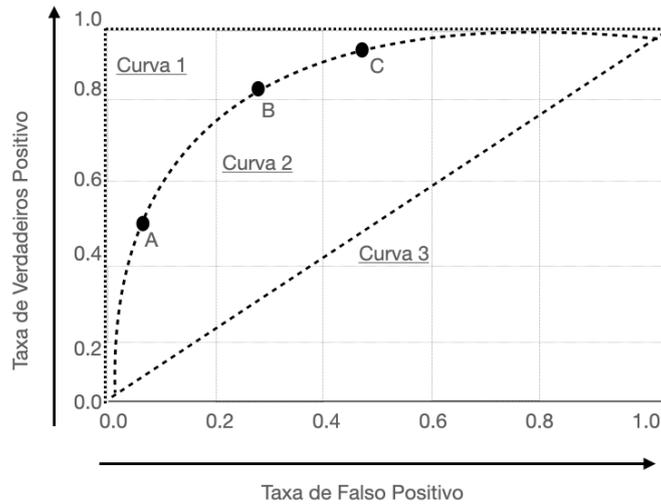


Figura 3.26: Ponto de ótimo Corte  
 Fonte: Elaborado pelo autor

A Figura 3.27 ilustra graficamente a distribuição dos dados presentes na área da Curva 1, em que não há falsos positivos e possui cem por cento de verdadeiros positivos.

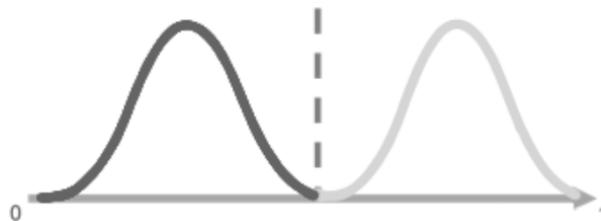


Figura 3.27: Curva 1  
 Fonte: Elaborado pelo autor

A Figura 3.28 ilustra graficamente a distribuição dos dados presentes na área da Curva 2, em que há possibilidade de falsos positivos e falsos negativos.

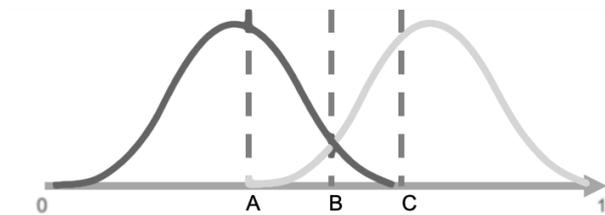


Figura 3.28: Curva 2  
 Fonte: Elaborado pelo autor

A Figura 3.29 ilustra graficamente a distribuição dos dados presentes na área da Curva 3, em os valores são aleatórios e não agregam ao modelo.

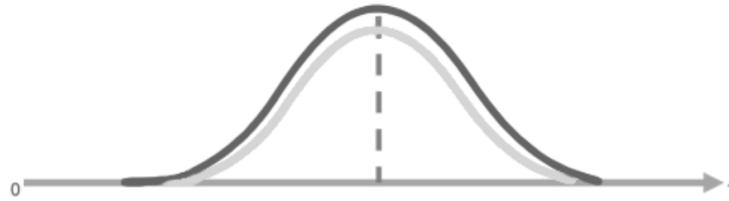


Figura 3.29: Curva 3  
Fonte: Elaborado pelo autor

A curva ROC permite discriminar o ponto de corte ideal, ou seja, quantificar a proporção dos resultados presentes na área da curva. Dessa forma, permite escolher a melhor distribuição para o problema proposto, assim como mostram os pontos A, B e C.

Em alguns casos, a curva ROC pode não ser facilmente interpretada, necessitando da aplicação do método AUC (Area Under the Curve) que apresenta valores entre 0 e 1, de modo que quanto mais próximo de 1, melhor. Ou seja, em um modelo com previsões 100% corretas, terá AUC igual a 1. No entanto, para previsão 100% erradas, terá AUC igual a 0. Esse método é utilizado para simplificar a análise da curva ROC ao computar todos os pontos limiares da curva ROC e apresentar uma área sob a curva.

Sua fórmula pode ser representada da seguinte forma:

$$AUC(f) = \frac{\sum_{t_0 \in \mathcal{D}^0} \sum_{t_1 \in \mathcal{D}^1} \mathbf{1}[f(t_0) < f(t_1)]}{|\mathcal{D}^0| \cdot |\mathcal{D}^1|} \quad (3.31)$$

A curva ROC é considerada uma ferramenta comum de avaliação de modelos de classificação, por sua capacidade de avaliar a relação entre a sensibilidade e especificidade. Desse modo, é possível avaliar o ponto de corte que melhor atenda a solução do problema proposto [Brown & Mues, 2012].

A AUC pode ser entendida como a eficiência do modelo, por seu poder de discriminar os classificadores em todos os pontos de corte possíveis sem considerar a distribuição de classes ou o custo de classificação errada [Baesens et al., 2003].

### Comparação entre as medidas de avaliação

Cada métrica apresentada possui suas próprias peculiaridades que devem ser consideradas ao se escolher o modelo de classificação que será utilizado para resolver o problema concreto. Para isso, deve-se avaliar o problema e identificar o conjunto de métricas que mais se adéque a solução, no intuito de definir o melhor ponto de corte que reflita a necessitam de maior otimização para os resultados tanto de sensibilidade como especificidade, pois

em alguns casos, haverá a necessidade que o modelo acerte todos os verdadeiros positivos, mesmo que realize falsos positivos, como é o caso de diagnóstico de doenças. Por isso, encontrar o equilíbrio é um processo iterativo, necessitando de processo de treino com diferentes combinações, parâmetros e conjunto de dados.

### Medidas de explicabilidade:

Segundo [Lundberg & Lee, 2017], compreender como modelos realizam determinadas previsões pode tão crucial quanto a precisão com que as previsões são realizadas. A tensão criada entre a precisão dos modelos e sua interpretabilidade é um desafio. Existe vários métodos para auxiliar nessa questão, mas que não deixam claros os mecanismos utilizados. Por isso, o método SHAP ( Shapley Additive exPlanations) foi apresentado como uma estrutura unificada de interpretação de previsões.

Seu funcionamento parte do princípio da atribuição valores de importância para uma previsão específica.

- **Shapley Value**

Baseado na teoria do jogos e capaz de explica decisões de modelos de forma intuitiva e direta. Segundo [Štrumbelj & Kononenko, 2014], esta medida atribui valores de importância para influência de cada variável ao resultado geral esperado. Em outras palavras, os valores Shapley correspondem à contribuição de cada recurso para para obtenção do valor esperado.

Sua função pode ser representada da seguinte forma:

$$\varphi_i(v) = \frac{1}{n} \sum_{S \subseteq N \setminus \{i\}} \binom{n-1}{|S|}^{-1} (v(S \cup \{i\}) - v(S)) \quad (3.32)$$

$S$  - subconjunto aleatório de recursos.

$v(S \cup \{i\}) - v(S)$  - medição da contribuição marginal.

Os valores de Shapley simulam a exclusão de recursos em previsões de forma amostral, sendo obtidos por meio da realização da média ponderada de todas as contribuições marginais, que contabilizam as contribuições das observações para o modelo.

Sua função também pode ser representada da seguinte forma:

$$\varphi_i(v) = \frac{1}{\text{number of players}} \sum_{\text{coalitions excluding } i} \frac{\text{marginal contribution of } i \text{ to coalition}}{\text{number of coalitions excluding } i \text{ of this size}} \quad (3.33)$$

A diferença da utilização deste método está em ser considerado uma técnica independente do modelo, por requerer apenas o conhecimento dos dados de entrada e saída, não necessitando conhecimento interno do modelo.

Os valores SHAP estimam o impacto de um recurso nas previsões, enquanto as importâncias do recurso estimam o impacto de um recurso no ajuste do modelo.

Em Python, a biblioteca SHAP possui uma função para auxiliar na explicabilidade dos modelos, também possuindo funções aperfeiçoadas para utilização em modelos de árvores de decisão.

### 3.1.6 Execução

A execução dos passos que compuseram as etapas citadas no item 3.1 incluíram as atividades referentes ao carregamento e ajuste de base de dados, inserção e medição das variáveis que representam o risco climático, formatação dos dados, exploração dos resultados, avaliação e tratamento dos riscos, submissão dos dados aos classificadores: Regressão Logística, Árvore de Decisão, AdaBoost, Random Forest, Gradient Boosting e XGBoost, e pela explicabilidade dos resultados previstos pelo método Shapley Value. Todos os passos citados utilizaram a linguagem Python.

Essas ferramentas foram escolhidas por sua popularidade e por serem comumente citadas na literatura.

Os procedimentos necessários para utilização das ferramentas citadas e consequente realização das etapas desta pesquisa foram agrupados por tópicos e estão detalhados a seguir:

#### Base de Dados

- A base foi manipulada para manter os dados somente de operações realizadas por pessoas jurídicas, devido a necessidade de ser utilizada a classificação CNAE nas etapas de medição e incorporação do risco climático. O resultado totalizou 115.348 operações e 32 observações das 152.936 operações e 24 observações fornecidas.
- As informações de UF e município foram retiradas da base de dados do **Instituto Brasileiro de Geografia e Estatística (IBGE)**<sup>3</sup> em julho de 2021.
- Os dados da Taxonomia Verde foram inseridos.
- Os dados referentes ao **IVDNS** foram inseridos.
- Os valores **IVDNS** foram formatados para Baixo, Médio e Alto.

---

<sup>3</sup><https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/15778-divisoes-regionais-do-brasil.html?=&t=downloads>

- A medição da exposição ao risco climático referente à **RSRC** foi realizada.
- Os resultados foram formatados e os itens duplicados retirados, resultando em uma base com 109.017 linhas e 39 observações, que foram utilizadas para exploração dos dados.
- Os dados também foram filtrados somente em operações de agronegócio para melhor detalhamento das classificações, resultando em 606 operações.
- Para aprofundamento das investigações, os dados foram separados em operações de cultivo e criação de bovinos, resultando em 488 operações.

## Risco climático

O projeto foi composto por três indicadores de risco climático e indicadores presentes no escopo **Ambiental, Social e Governança (ASG)**, assim como detalhado a seguir:

- A Taxonomia Verde <sup>4</sup> forneceu as variáveis avaliadas: Risco Climático, Economia Verde e Risco Ambiental.
- A **RSRC** forneceu a metodologia de cálculo do risco climático que foi utilizada para gerar as classificações Risco Climático Hadgem e Risco Climático Miroc, que foram usadas como uma alternativa de comparação na execução dos modelos. Essas classificações utilizaram o **Índice de vulnerabilidade aos desastres naturais relacionados às secas no contexto da mudança do clima (IVDNS)**<sup>5</sup> para compor a variável de Risco Locacional. Esse procedimento foi necessário por não haver metodologia pacificada na instituição sobre essa variável. A escolha do **IVDNS** se deu por meio das investigações feitas no processo de desenvolvimento deste indicador, em que foi constatado que o risco **IVDNS** apresentou alta relevância em todas as metodologias em construção observadas. Desse modo, foram obtidos:
  - O indicador de Risco Climático Hadgem calculado utilizando como Risco Locacional a variável **IVDNS** HadGEM 8.5.
  - O indicador de Risco Climático Miroc calculado utilizando como Risco Locacional a variável **IVDNS** Miroc 8.5.

Os documentos citados produziram novas variáveis ao escopo do projeto. A Taxonomia Verde foi escolhida como indicador principal por possuir variáveis **ASG** em seu escopo e ser um documento robusto e de maior abrangência e qualidade de informações disponível

<sup>4</sup><https://portal.febraban.org.br/pagina/3292/1103/pt-br/consulta-publica>

<sup>5</sup>[https://wwfbrnew.awsassets.panda.org/downloads/tabela\\_estudo\\_ivdns.pdf](https://wwfbrnew.awsassets.panda.org/downloads/tabela_estudo_ivdns.pdf)

durante a realização do estudo, além de ter sido produzida considerando as informações disponibilizadas pelo **BACEN** a partir do **Sistema de Informações de Crédito** e ter 100% de abrangência dos créditos de pessoas jurídicas do sistema bancário no Brasil. Como metodologia de comparação, o **IVDNS** foi escolhido para compor o Risco Locacional na utilização da **RSRC**, tendo em vista que esta variável estava presente nas metodologias em desenvolvimento com um grande impacto, sua pontuação em 8.5 foi escolhida, pois, segundo o IPCC 2018, esse cenário corresponde a alta emissão das concentrações de **GEE** e está mais próximo das tendências observadas nas medições atuais das concentrações de gases de efeito estufa na atmosfera.

Ressalta-se que, por se tratar de volumes financeiros, deve-se optar pelos cenários mais conservadores.

Os detalhes deste passo encontram-se no item **3.1.1** e seus resultados no item **4.1**.

### **Formatação dos dados**

Os dados precisaram ter sua formação regularizada para utilização. Dessa modo, os seguintes passos foram realizados:

- Aplicação da formatação numérica para o padrão americano.
- Remoção de caracteres especiais e acentuações.
- Padronização das classificações ALTO, MEDIO e BAIXO.
- Os valores também foram padronizados em baixa exposição para os casos de baixa exposição ao risco climático ou risco ambiental e em baixa contribuição para os casos de baixa contribuição da economia verde.
- Padronização do risco da operação e risco do cliente para seguir a resolução CMN 2.682<sup>6</sup>.
- Ajuste nos tipos das variáveis.
- Substituição da variável que indica a localidade física das operações pela localidade onde a operação foi contratada, nos casos de ausência da variável localidade.
- Retirada de dados duplicados.
- Tratamento das ocorrências de ausência de dados.
- Técnicas de anonimização dos dados sensíveis foram utilizadas para garantir o sigilo das operações e cumprimento da **LGPD**.

---

<sup>6</sup>[https://www.bcb.gov.br/pre/normativos/res/1999/pdf/res\\_2682\\_v2\\_L.pdf](https://www.bcb.gov.br/pre/normativos/res/1999/pdf/res_2682_v2_L.pdf)

Esses passos possibilitaram investigar os dados para identificar e corrigir erros, lacunas e necessidades de alterações a fim de possibilitar a produção das variáveis necessárias para utilização da metodologia de classificação do risco climático.

### **Análise exploratória**

Para realizar a exploração dos dados e buscar identificar como as variáveis de risco climático se comportam, as variáveis incorporadas pelo material da Taxonomia Verde foram investigadas, por ser um indicador mais robusto. Logo, nesta etapa, não foram utilizados os resultados produzidos pela **RSRC**, devido à limitação encontrada no indicador de Risco Locacional.

As operações foram somatizadas e agrupadas pelo indicador de risco climático visando observar seu comportamento ao ser associado à quantidade de operações, ao volume financeiro e inadimplência agregada às classificações. Também foram feitas investigações sobre o indicador de economia verde, em que foi somatizado para contabilizar a quantidade de operações, o volume financeiro e inadimplência associados. Em seguida, os resultados da economia verde também foram agrupados pela variável de risco climático.

Para melhor detalhamento das informações foram escolhidos filtros baseados no estudo feito pela **Secretaria de Inteligência e Relações Estratégicas da Embrapa (Sire)**<sup>7</sup>. No estudo citado, foi constatado que o Brasil ganhou posições no mercado internacional na produção e exportação agropecuária entre os períodos de 2000 a 2020, apontando o país como sendo o quarto maior produtor de grãos do mundo. O estudo também destaca a posição do cultivo de soja que responde por 50% do comércio mundial, o rebanho de bovinos como sendo o maior do mundo e a liderança na produção de café.

Tendo o exposto, para melhor detalhamento, as operações foram separadas somente em operações de agronegócio, resultando em 606 observações e avaliadas novamente.

Por fim, com base nos resultados encontrados, para realizar uma investigação com mais detalhes, os dados foram novamente filtrados para manter somente operações de cultivos e bovinos. Desse modo, o filtro resultante de 488 operações também teve seus indicadores de risco climático comparados com a quantidade de operações, volume financeiro dos contratos, risco da operação, risco do cliente e inadimplência.

Essas comparações tiveram como finalidade verificar o potencial do indicador ao ser utilizado na tomada de decisões de crédito da instituição financeiras.

Os resultados estão descritos no item 4.1.

---

<sup>7</sup><https://www.embrapa.br/busca-de-noticias/-/noticia/62619259/brasil-e-o-quarto-maior-produtor-de-graos-e-o-maior-exportador-de-carne-bovina-do-mundo-diz-estudo>

## Processo de avaliação e tratamento dos riscos

Esta etapa contou com dez entrevistas com os funcionários da instituição e envolvidos no processo, sendo:

- Cinco com a área de riscos da instituição, sendo uma com o gestor, duas com o assessor responsável, e duas com o responsável técnico dos dados.
- Cinco entrevistas com a área de sustentabilidade da instituição, sendo duas com o gestor e três com o assessor responsável.

Foram produzidas tabelas contendo a categorização, identificação, análise, avaliação e propostas de tratamento para os riscos levantados no processo de medição do risco climático. Ao final, foram incluídas considerações para cada item avaliado.

Os detalhes deste passo encontram-se no item 3.1.2 e seus resultados no item 4.2.

## Análise dos dados

Por sua larga utilização e capacidade de fornecer ferramentas para aprendizado de máquina, ajustes de modelo, pré-processamento de dados, seleção de modelo, avaliação de modelo, dentre outras. A biblioteca de código aberto scikit-learn<sup>8</sup> foi utilizada para obter as ferramentas capazes de realizar os passos necessários para as predições e análises necessárias.

A biblioteca python SHAP<sup>9</sup> foi utilizada para explicar a interpretabilidade dos modelos de aprendizado de máquina utilizados.

O processo de avaliação busca responder se a inclusão do indicador de Risco Climático seria capaz de influenciar a composição do Risco de Crédito das operações e consequentemente o favorecimento de operações nas tomadas de decisões.

## Tratamento de dados ausentes

Os passos a seguir foram necessários para utilização da ferramenta KNNImputer para preencher os dados ausentes.

- As colunas consideradas sem importância foram retiradas.
- A variável dependente foi transformada em 0 e 1.
- As variáveis categóricas foram transformadas em numéricas binárias, com inclusão de variáveis ausentes e retirada da primeira ocorrência.

---

<sup>8</sup><https://scikit-learn.org/stable/about.html>

<sup>9</sup><https://shap.readthedocs.io/en/latest/index.html>

- As variáveis NaN (dados ausentes) criadas com o passo anterior foram retiradas.
- O algoritmo KNNImputer foi aplicado somente sobre as variáveis numéricas.

Os detalhes deste passo encontram-se no item [3.1.4](#).

## **Balanceamento**

Por se tratar de dados desbalanceados, em que um objeto de classificação ocorre em maior escala que os demais. Os dados precisaram ser balanceados utilizando a técnica SMOTE-ENN, que é uma variação da técnica SMOOTE, presente na biblioteca `imblearn.combine import SMOTEENN`.

Os detalhes deste passo encontram-se no item [3.1.4](#).

## **Regularização e Seleção**

Por se tratar de dados em série temporal, após serem ordenados por data de contrato, a data de contratação foi transformada em índice e os passos a seguir foram realizados:

- Os dados considerados redundantes ou que refletiam uma mesma variável foram retirados.
- A variável de inadimplência foi separada como dependente e as demais como independentes.
- Os dados de treino e teste foram separados com a função `TimeSeriesSplit`.
- Foram utilizadas as metodologias LASSO e `ElasticNet` para seleção de variáveis e redução de dimensionalidade.
- Após a avaliação das seleções realizadas pelas técnicas, as variáveis que foram consideradas sem importância para o modelo foram retiradas e as metodologias rodadas novamente no intuito de investigar se os resultados poderiam estar sendo afetados por alguma das variáveis retiradas.
- Os coeficientes apresentados pelas metodologias também foram investigados. Seus valores foram separados e apresentados utilizando a normalização máximos e mínimos.

Os detalhes deste passo encontram-se no item [3.1.5](#).

## Modelos

Os algoritmos utilizados no modelo foram: Regressão Logística, Árvore de Decisão, Ada-Boost, Random Forast, Gradient Boosting e XGBoost.

Os detalhes das funções utilizadas estão no [3.1.5](#).

## Avaliando o modelo

As métricas utilizadas para avaliação do modelo foram a acurácia, precisão, recall, f1-score, a matriz de confusão e a curva ROC e AUC que foram amplamente utilizadas na literatura de referência.

Os detalhes das funções utilizadas estão no [3.1.5](#).

## Explicabilidade do modelo

As ferramentas de explicabilidade foram obtidas da biblioteca Python SHAP. Com as funções presentes nesta biblioteca foi possível obter os valores de shap (shapley values) e investigar graficamente o potencial de impacto das variáveis em relação a variável dependente.

Os detalhes das funções utilizadas estão no [3.1.5](#).

# Capítulo 4

## Resultados e Discussão

### 4.1 Classificação da Exposição ao Risco Climático

A medição da exposição ao risco climático e obtenção dos indicadores foi realizada seguindo o exposto no item 3.1.6, resultando no acréscimo de três indicadores de exposição ao risco climático e nas variáveis representativas do escopo ASG (economia verde e risco ambiental) presentes na Taxonomia Verde.

Como citado no 3.1.6, a exploração dos dados se concentrou nas variáveis fornecidas pela taxonomia verde, por ser um indicador mais robusto e se relacionar com os demais indicadores do escopo ASG. Desse modo, os resultados das medições realizadas utilizando a RSRC com o IVDNS foram utilizados somente na construção do modelo como indicador alternativo.

As operações foram somatizadas e agrupadas pela da exposição ao risco climático da taxonomia verde, para ilustrar a quantidade de operações, o volume financeiro e a inadimplência agregado a sua classificação.

A Tabela 4.1 a seguir apresenta o agrupamento realizado com todas as operações, utilizando a medição fornecida pela Taxonomia Verde. Do total de valores, foram classificadas 109.017 operações.

Tabela 4.1: Taxonomia Verde: Exposição ao Risco climático

Risco Climático	Quantidade	Total Contratado	Inadimplência
Alto	16.656	76%	17,7%
Médio	4.956	3,4%	2,4%
Baixo	87.405	20,6%	4,3%

Fonte: Elaborado pelo autor

## Resultados observados

- Todas as operações foram classificadas pela classificação da Taxonomia Verde.
- Os resultados indicam que o volume financeiro está concentrado nas operações classificadas com Alta Exposição ao Risco Climático, totalizando 76% do valor total de todas as operações, contra 3,4% classificadas como Média Exposição ao Risco Climático e 20,6% classificadas como Baixa Exposição ao Risco Climático.
- Ao comparar esses resultados com a inadimplência de sua respectiva categoria, as operações classificadas como sendo de Alta Exposição ao Risco Climático obtiveram 17,7% de inadimplência, contra 2,4% das operações classificadas como Média Exposição ao Risco Climático e 4,3% das operações classificadas como sendo de Baixa Exposição ao Risco Climático.
- Ao comparar os resultados com a inadimplência somada total, as operações classificadas como Alta Exposição ao Risco Climático acumularam 93,3% da inadimplência, contra 0,6% das operações classificadas como Média Exposição ao Risco Climático e 6,2% das operações classificadas como Baixa Exposição ao Risco Climático.
- Os resultados também indicam que a maior quantidade de operações está classificada como Baixa Exposição ao Risco Climático, totalizando 80,2% das operações, contra 4,5% classificadas como Média Exposição ao Risco Climático e 15,3% classificadas como Alta Exposição ao Risco Climático.

Os resultados mostram que, apesar de 80,2% das operações serem classificadas como Baixa Exposição ao Risco Climático, há uma concentração de 76% do valor financeiro nas operações com o indicador de Alta Exposição ao Risco Climático, mesmo que essas operações tenham apresentado valores muito superiores de inadimplência.

Para entender como as operações estão se relacionando com as variáveis ambientais e sociais, elas foram somatizadas e agrupadas pelo indicador de economia verde para ilustrar a quantidade de operações, o volume financeiro e a inadimplência agregado a cada classificação. Os resultados estão apresentados na Tabela 4.2.

Tabela 4.2: Taxonomia Verde: Economia Verde

<b>Economia Verde</b>	<b>Quantidade</b>	<b>Total Contratado</b>	<b>Inadimplência</b>
Alta contribuição [Social+Ambiental]	542	1,3%	6,2%
Alta contribuição [Ambiental]	3.956	1,5%	3,5%
Alta contribuição [Social]	85.782	28,3%	17,8%
Moderada contribuição [Social + Ambiental]	8.481	38,3%	2,6%
Moderada contribuição [Ambiental]	7.903	29,1%	28,5%
Moderada contribuição [Social]	944	0,9%	2,4%
baixa contribuição	1409	0,5%	2,7%

Fonte: Elaborado pelo autor

### Resultados observados

- As operações se concentraram na classificação de Alta contribuição [Social] representam 78.7% das operações. No entanto, estão na terceira posição de valores financeiros e segunda de inadimplência.
- Moderada contribuição [Ambiental] representa a segunda posição de valores financeiros, com a primeira posição de inadimplência.
- Moderada contribuição [Social + Ambiental] possui o maior volume financeiro e penúltimo lugar em inadimplência.
- A classificação baixa contribuição representa operações que possuem baixa ou nenhuma contribuição para os itens citados.
- O valor financeiro mais baixo está com as operações classificadas como baixa contribuição.

As operações foram agrupadas para ilustrar como o comportamento da exposição ao risco climático está associado com os resultados apresentados, assim como é apresentado na Tabela 4.3.

Tabela 4.3: Economia Verde: Exposição ao Risco Climático

<b>Economia Verde</b>	<b>Quantidade</b>	<b>Total Contratado</b>	<b>Inadimplência</b>
Alta Exposição ao Risco Climático			
Alta contribuição [Social + Ambiental]	255	1.54%	0.86%
Alta contribuição [Ambiental]	488	0.82%	3.54%
Alta contribuição [Social]	1335	11.87%	48.38%
Moderada contribuição [Social + Ambiental]	7522	48.00%	2.22%
Moderada contribuição [Ambiental]	6623	37.49%	28.97%
Moderada contribuição [Social]	189	0.22%	5.34%
baixa contribuição	244	0.07%	2.97%
Média Exposição ao Risco Climático			
Alta contribuição [Social + Ambiental]	40	0.1%	5.3%
Alta contribuição [Ambiental]	244	9.4%	2.3%
Alta contribuição [Social]	1821	19.4%	7.0%
Moderada contribuição [Social + Ambiental]	886	33.1%	0.3%
Moderada contribuição [Ambiental]	1069	14.0%	2.0%
Moderada contribuição [Social]	755	21.7%	1.8%
baixa contribuição	141	2.4%	1.3%
Baixa Exposição ao Risco Climático			
Alta contribuição [Social+Ambiental]	247	0.6%	35.7%
Alta contribuição [Ambiental]	3224	2.9%	4.2%

Alta contribuição [Social]	82626	90.7%	3.4%
Moderada contribuição [Social + Ambiental]	73	3.4%	26.0%
Moderada contribuição [Ambiental]	211	0.7%	21.6%
Moderada contribuição [Social]	0	0%	0%
baixa contribuição	1024	1.6%	3.0%

Fonte: Elaborado pelo autor

## Resultados observados

- Alta Exposição ao Risco Climático:
  - Moderada contribuição [Social + Ambiental] concentra a maioria das operações, o maior volume financeiro e a segunda menor inadimplência.
  - Alta contribuição [Social + Ambiental] possui a menor inadimplência.
  - Alta contribuição [Social] concentra a maior inadimplência.
- Média Exposição ao Risco Climático:
  - Moderada contribuição [Social + Ambiental] concentra o maior volume financeiro e a menor inadimplência.
  - Alta contribuição [Social] possui o maior número de operações, maior inadimplência e terceira posição de representatividade em seu volume financeiro.
- Baixa Exposição ao Risco Climático:
  - Alta contribuição [Social + Ambiental] está em último lugar de volume financeiro, com o primeiro lugar de inadimplência.
  - Alta contribuição [Social] possui o maior número de operações, maior volume financeiro, representando 90,7% do total e menor inadimplência.

As análises trouxeram resultados interessantes. Foi observado a prevalência de algumas comportamentos nas classificações de Alta e Média exposição ao risco climático que não se refletiram nas classificações de Baixa exposição ao risco climático.

Por esse motivo, visando o fornecimento de informações mais relevantes para as instituições, as operações foram separadas somente por agronegócio, assim como detalhado no item 3.1.6. Desse modo, a base fornecida de 109.017 resultou em um filtro de 606 operações. A Tabela 4.4 apresenta a quantidade de operações, seu valor financeiro e inadimplência por classificação das operações de agronegócio filtradas da base recebida.

Tabela 4.4: Exposição ao Risco Climático: Agronegócio

Risco Climático	Quantidade	Total Contratado	Inadimplência
Alto	593	97.30%	0.96%
Médio	5	1.89%	0.01%
Baixo	8	0.81%	1.74%

Fonte: Elaborado pelo autor

#### Resultados observados:

- A Tabela 4.4 mostra que há uma predominância de operações classificadas como sendo de Alta Exposição ao Risco Climático.
- As operações consideradas de Alta Exposição ao Risco Climático possuem 0,96% de inadimplência.
- As operações consideradas de Média Exposição ao Risco Climático possuem 0,01% de inadimplência.
- As operações consideradas de Baixa Exposição ao Risco Climático possuem a maior taxa de inadimplência, sendo de 1,74%.
- Os valores médios das operações classificadas como Média Exposição ao Risco Climático são o dobro das operações classificadas como Alta Exposição ao Risco Climático.
- Os valores médios das operações classificadas como Baixa Exposição ao Risco Climático foram os menores.

As taxas gerais de inadimplência são relativamente baixas. Mas as operações de Média Exposição ao Risco Climático merecem atenção por quase não possuírem inadimplência, mesmo com valores de operação mais elevados.

Visando o aprofundamento das investigações e fornecimento de dados relevantes, as operações de cultivos e criação de bovinos foram selecionadas para serem investigadas por representarem 80% das operações de agronegócio, totalizando 488 operações.

Os resultados indicaram que as operações de cultivo de soja e criação de bovinos representaram 56% do volume de operações e acumularam mais de 29% do volume financeiro de inadimplência.

- Cultivo de soja:
  - Representa 20,08% do total de operações do filtro realizado.
  - Acumula 28,26% do volume financeiro total de operações do filtro realizado.
  - Possuem 10,46% do volume financeiro total de inadimplência.
  - Apesar de possuir um alto volume financeiro de inadimplência, a inadimplência acumulou 0,41%, quando considerado somente a inadimplência presente nas operações de cultivo e soja.
  
- Criação de bovinos:
  - Representa 36,07% do total de operações do filtro realizado.
  - Acumula 31,83% do volume financeiro total de operações do filtro realizado.
  - Possui 18,89% do volume financeiro total de inadimplência.
  - Apesar de possuir um alto volume financeiro de inadimplência, a inadimplência acumulou 0,66%, quando considerado somente a inadimplência presente nas operações de criação de bovinos,

Esses resultados vão ao encontro do exposto no item 3.1.6, em que a representatividade das operações relacionadas ao cultivo de soja e criação de bovinos são de grande importância para economia do Brasil. Também foi possível observar que a porcentagem total de inadimplência se concentrou nessas operações. No entanto, não apresentaram uma alta taxa de inadimplência, ficando bem abaixo de operações como cultivo de mamão que apresentou a maior taxa com 22,09% de inadimplência.

Visando aprofundar as investigações para entender o comportamento das operações filtradas frente aos indicadores de risco da operação e do cliente, as operações foram agrupadas pela classificação de exposição ao risco climático e assim comparadas com o risco das operações e risco do cliente.

As classificações de risco do cliente e da operação seguem a metodologia definida pela instituição e seguem ordem alfabética do maior para o menor, ou seja, de AA até H, sendo H a pior classificação. Ressalta-se que, essa classificação segue a resolução CMN 2.682<sup>1</sup>.

A classificação encontrada nas operações filtradas do rating de risco das operações é ilustrada na Tabela 4.5 e a classificação do rating de risco do Cliente encontrada nas operações filtradas é ilustrado conforme a Tabela 4.6.

---

<sup>1</sup>[https://www.bcb.gov.br/pre/normativos/res/1999/pdf/res\\_2682\\_v2\\_L.pdf](https://www.bcb.gov.br/pre/normativos/res/1999/pdf/res_2682_v2_L.pdf)

Tabela 4.5: Classificação do Rating das Operações

AA	A	B	C	D	E	F	G	H
----	---	---	---	---	---	---	---	---

Fonte: Elaborado pelo autor

Tabela 4.6: Classificação do Rating dos Clientes

A	B	C	D	E
---	---	---	---	---

Fonte: Elaborado pelo autor

Foram encontrados cinco campos sem informações no filtro realizado, assim como ilustra a Tabela 4.7.

Tabela 4.7: Agrupamento dos dados

Variáveis	Branco	Total de dados úteis
Risco do Cliente	5	483
Risco da Operação	0	488

Fonte: Elaborado pelo autor

A Tabela 4.8 apresenta a quantidade das operações de cultivo e bovinos com seu valor financeiro, agrupado pela classificação de exposição ao risco climático.

Tabela 4.8: Exposição ao Risco Climático: Cultivos e Bovinos

Risco Climático	Quantidade	Total Contratado	Inadimplência
Alto	486	99,92%	1,12%
Médio	2	0,08%	0,23%

Fonte: Elaborado pelo autor

### Resultados observados:

- A Tabela 4.8 mostra que há uma predominância de operações classificadas como sendo de Alta Exposição ao Risco Climático.
- As operações consideradas de Alta Exposição ao Risco Climático possuem 1,12% de inadimplência.

- As operações consideradas de Média Exposição ao Risco Climático possuem 0,23% de inadimplência.

As taxas gerais de inadimplência são relativamente baixas. Mas as operações de Média Exposição ao Risco Climático merecem atenção por quase não possuírem inadimplência, mesmo com valores de operação mais elevados.

A Tabela 4.9 comparou a *Exposição ao Risco Climático* com o *Risco da Operação* ao quantificar o volume de operações e seu valor financeiro agregado.

Tabela 4.9: Exposição ao Risco Climático e Risco da Operação

Risco da Operação	Risco Climático	Quantidade	Valor	Inadimplência
AA	Alto	80	40.58%	0.00%
AA	Médio	1	0.04%	0.00%
AA	Baixo	0	0.00%	0.00%
A	Alto	96	15.23%	0.37%
A	Médio	1	0.04%	0.47%
A	Baixo	0	0.00%	0.00%
B	Alto	186	34.43%	1.18%
B	Médio	0	0,00%	0,00%
B	Baixo	0	0,00%	0,00%
C	Alto	104	7.17%	1.67%
C	Médio	0	0,00%	0,00%
C	Baixo	0	0,00%	0,00%
D	Alto	7	0.78%	3.88%
D	Médio	0	0,00%	0,00%
D	Baixo	0	0,00%	0,00%
E	Alto	2	0.27%	6.57%
E	Médio	0	0,00%	0,00%
E	Baixo	0	0,00%	0,00%
F	Alto	5	1.22%	29.92%
F	Médio	0	0,00%	0,00%
F	Baixo	0	0,00%	0,00%
H	Alto	6	0.23%	51.56%
H	Médio	0	0,00%	0,00%
H	Baixo	0	0,00%	0,00%

Fonte: Elaborado pelo autor

A Tabela 4.10 comparou a *Exposição ao Risco Climático* com o *Risco do Cliente* ao quantificar o volume de operações e seu valor financeiro agregado.

Tabela 4.10: Exposição ao Risco Climático e Risco do Cliente

Risco do Cliente	Risco Climático	Quantidade	Valor	Inadimplência
A	Alto	76	21.0%	0.2%
A	Médio	2	0.1%	0.2%
A	Baixo	0	0,00%	0,00%
B	Alto	306	59.9%	0.8%
B	Médio	0	0,00%	0,00%
B	Baixo	0	0,00%	0,00%
C	Alto	67	15.2%	0.6%
C	Médio	0	0,00%	0,00%
C	Baixo	0	0,00%	0,00%
D	Alto	9	0.2%	1.3%
D	Médio	0	0,00%	0,00%
D	Baixo	0	0,00%	0,00%
E	Alto	23	3.1%	15.6%
E	Médio	0	0,00%	0,00%
E	Baixo	0	0,00%	0,00%
Sem Classificação	Alto	5	0.4%	0.7%

Fonte: Elaborado pelo autor

Com a análise realizada na Tabela 4.9 e Tabela 4.10 é possível observar que:

#### Resultados observados:

- Em todos os casos a classificação de Alta Exposição ao Risco Climático permaneceu com maior volume financeiro e quantidade de operações.
- O valor da inadimplência aumentou de acordo com a piora do risco da operação.
- A quantidade de operações e volume financeiro estão concentrados majoritariamente nas categorias AA, A, B e C.

Os resultados indicam que, apesar do volume expressivo dentro da classificação de Alta Exposição ao Risco Climático, a instituição possui excelência em sua gestão, pois suas operações estão concentradas nas categorias de risco AA, A, B e C. Por isso, é importante ressaltar que as operações compreendidas entre AA a C recebem a classificação de curso normal, segundo a resolução CMN 2.682.

O volume expressivo nas classificações de Alta Exposição ao Risco Climático pode ser explicado pelas operações de cultivo de soja e criação de bovinos serem majoritariamente classificadas como sendo de alta emissão de carbono e consumo intensivo.

Os resultados forneceram uma base geral de comportamento das operações ao serem expostas às variáveis de Exposição ao Risco Climático e Economia Verde, em que possível observar um claro potencial de impactar o risco de crédito das operações, pois foi possível identificar alguns padrões de comportamento em relação às classificações.

Ressalta-se que apesar das operações classificadas como Alta Exposição ao Risco Climático apresentarem melhores resultados em sua inadimplência, o volume financeiro de operações classificadas como Alta Exposição traz um alerta, pois, em cenário de mudanças climáticas e aumento de desastres relacionados às secas no Brasil, viabilizar a redução da exposição ao risco climático torna-se essencial, e planejar meios de contingência a eventos extremos é indispensável.

Segundo [Kemp et al., 2022], as mudanças climáticas têm o potencial de colapsar sistemas, por isso, produzir informações sobre seus impactos é essencial para enfrentar os piores cenários. Os resultados também vão ao encontro do indicado pelo IVDNS detalhado no item 2.2.2, que alerta para os cuidados com o aumento de desastres em todo território nacional.

No que tange às contribuições sociais e ambientais classificadas pelo indicador de economia verde, foi possível observar que as operações que contemplavam contribuição social e ambiental prevaleceram com maior quantidade de operações e menor inadimplência que as demais e as operações que contemplavam somente contribuições sociais tiveram uma incidência de maior inadimplência.

Tendo o exposto, com os altos valores de retorno envolvidos nas operações investigadas, analisar a relevância da Exposição ao Risco Climático frente aos demais riscos de crédito se torna necessário, pois os impactos produzidos pelos indicadores incluídos podem ser considerados irrisórios frente aos altos retornos financeiros e demais riscos avaliados ou ter o potencial de gerar perdas sistêmicas.

Por isso, com o intuito de gerar resultados que possibilitem avaliar o impacto de cada variável no risco de crédito e viabilizar a manipulação dos fatores para se obter uma melhor classificação de exposição ao risco climático sem perder os benefícios ambientais e sociais, o comportamento das variáveis foi analisado e explicado, assim como foi detalhado nos itens 4.3 e 4.4.

## 4.2 Processo de avaliação e tratamento dos riscos

Mapear os riscos é um processo sistemático que deve investigar diferentes pontos de vista e usar a melhor informação disponível. Seu propósito é encontrar, reconhecer e descrever os riscos de forma detalhada, para identificar incertezas que podem afetar os objetivos da instituição. Este estudo investigou os principais fatores de risco envolvidos na criação dos

elementos que compõem a utilização do indicador de exposição ao risco climático, assim como demais elementos que pudessem interferir no seu desenvolvimento. O processo de avaliação dos riscos foi acompanhado durante as reuniões com a equipe de riscos e sustentabilidade. Foram totalizadas 10 reuniões:

- Cinco com a área de riscos da instituição, sendo uma com o gestor, duas com o assessor responsável, e duas com o responsável técnico dos dados.
- Cinco entrevistas com a área de sustentabilidade da instituição, sendo duas com o gestor e três com o assessor responsável, incluindo a participação de entidades parceiras.

Os riscos foram categorizados segundo sua capacidade de incapacitar ou impactar a medição do indicador final. Esses impactos foram descritos seguindo as orientações da equipe de risco e norteados pela A ISO 31000. Em seguida, foram feitas sugestões de melhoria para mitigar e oferecer soluções de contorno.

Desse modo, a classificação Alta, Média e Baixa foi efetuada sobre a possibilidade de produção do indicador de exposição ao risco climático. Ressalta-se que as classificações Alta e Média possuem um grande potencial de alterar ou inviabilizar a classificação das operação.

Também é importante ressaltar que a instituição trabalha com retornos financeiros e, por esse motivo, deve adotar a posição mais conservadora em todos os fatores de mitigação propostos.

A Tabela 4.11 ilustra a categorização dos riscos em alto, médio e baixo e sua descrição.

Tabela 4.11: Classificação do Risco

<b>Impacto</b>	<b>Descrição</b>
<b>Alto</b>	Impacta uma ou mais etapas capazes de alterar o indicador consideravelmente.
<b>Médio</b>	Possibilidade de alterar o resultado do indicador de forma considerável.
<b>Baixo</b>	Possibilidade de alterar o indicador final.

Fonte: Elaborado pelo autor

As Tabelas 4.12, 4.13, 4.14, 4.15, 4.16, 4.17 e 4.18 descrevem os principais fatores de risco envolvidos no desenvolvimento do indicador e as sugestões para tratamento.

Tabela 4.12: Rating da operação

<b>Risco</b>	<b>Rating da operação</b>
<b>Identificação do Risco</b>	Informações desatualizadas para identificar o rating de crédito.
<b>Análise do Risco</b>	A metodologia de crédito segue aspectos que incluem as informações cadastrais do cliente. Caso estejam desatualizadas, a classificação será equivocada.
<b>Avaliação do Risco</b>	Falha no cadastro: clientes podem não possuir código vinculado a seu cadastrado no sistema, isso pode ocorrer por falhas no sistema, cessão de dívidas, dentre outros motivos. Títulos internalizados com o clientes: isso faz com que ele não tenha operações de crédito ligadas a ele, mas o risco de crédito ainda está com ele.
<b>Impacto</b>	Médio.
<b>Tratamento do Risco</b>	Rotinas de conformidade cadastral devem ser aplicadas.
<b>Considerações</b>	Essas rotinas devem seguir a ordem de prioridade de impacto das operações na instituição.

Fonte: Elaborado pelo autor

Tabela 4.13: Locacional

<b>Risco</b>	<b>Locacional</b>
<b>Identificação do Risco</b>	Podem haver diversas unidades atreladas à operação.
<b>Análise do Risco</b>	As unidades produtivas devem ser localizadas, podendo ocorrer de várias estarem atreladas a uma operação.
<b>Avaliação do Risco</b>	A representatividade do risco das unidades produtivas pode ter grau elevado de diferença e afetar a régua final.
<b>Impacto</b>	Baixo.
<b>Tratamento do Risco</b>	Deve-se avaliar a operação com maior grau de risco ou atribuir uma média ponderada do grau de risco locacional das unidades identificadas, considerando a representatividade de cada uma delas.
<b>Considerações</b>	Ação de contorno prevista no guia.

Fonte: Elaborado pelo autor

Tabela 4.14: Clientes sem CNAE

<b>Risco</b>	<b>Clientes sem CNAE</b>
<b>Identificação do Risco</b>	A metodologia CNAE é uma padronização brasileira, por esse motivo, pode não seguir os mesmos padrões do exterior ou outros sistemas.
<b>Análise do Risco</b>	Entidades no exterior ou entidades ligadas podem seguir padrões de classificação diferentes. Como por exemplo, o padrão Europeu.
<b>Avaliação do Risco</b>	Classificações equivocadas ou falta de classificação das operações impactam todo o processo de medição do risco.
<b>Impacto</b>	Alto.
<b>Tratamento do Risco</b>	Atualizar a base de dados com relação direta entre os códigos CNAE brasileiros e os códigos de natureza de atividades mais utilizados no mercado. Decidir a prioridade dos códigos a serem atualizados mediante análise de volume de operações da instituição. Ou seja, dar prioridade a operações realizadas com mais volume e/ou mais frequência.
<b>Considerações</b>	Após o tratamento, as operações remanescentes representariam um impacto não significativo na visão institucional. Pois seu volume não representaria uma exposição ao risco climático que devesse ter atenção.

Fonte: Elaborado pelo autor

Tabela 4.15: CNAE sem subtipo de operação

<b>Risco</b>	<b>CNAE sem subtipo de operação</b>
<b>Identificação do Risco</b>	Os códigos CNAE não possuem subtipo de operação.
<b>Análise do Risco</b>	Não diferenciam operações da mesma categoria com atividades diferentes. Por exemplo: Operações classificadas como sendo de Energia pelo código CNAE e fontes diferentes recebem a mesma classificação. Ou seja, uma empresa A que trabalha com fonte tradicional de energia receberá a classificação de CNAE de energia, no entanto, a empresa B que também trabalha com energia, porém, com fonte eólica, receberá a mesma classificação CNAE.
<b>Avaliação do Risco</b>	Classificação padrão para operações com riscos evidentemente diferentes.
<b>Impacto</b>	Alto.
<b>Tratamento do Risco</b>	Questionários aplicados aos clientes que gerem insumos capazes de produzir fatores de mitigação para as operações no momento da medição.
<b>Considerações</b>	Os questionários devem ter rastreabilidade, ou seja, ao se fazer um batimento das informações com o sistema, deverá ser possível avaliar o grau de fidedignidade das informações prestas.

Fonte: Elaborado pelo autor

Tabela 4.16: Volume de operações

<b>Risco</b>	<b>Volume de operações</b>
<b>Identificação do Risco</b>	O volume de operações é dado pelo valor do contrato, esse valor pode não ser liberado de uma única vez e podem ocorrer abatimentos durante o período.
<b>Análise do Risco</b>	O valor total não é liberado de uma única vez e podem ocorrer abatimentos durante o período modificando o valor total exposto.
<b>Avaliação do Risco</b>	Com o saldo não sendo liberado de uma única vez e abatimentos ocorrendo durante o período, isso reduz a exposição total do cliente, e não o volume cadastrado de sua operação. Ou seja, um cliente A que realizar uma operação de R\$ 50.000.000,00, poderá ter um valor de R\$ 20.000.000,00 liberado e quitar esse valor antes que o restante seja liberado. Dessa forma, o saldo exposto total real seria de R\$ 30.000.000,00. Em contra partida, um cliente B que realize uma operação de R\$ 30.000.000,00 terá o mesmo saldo exposto total real. Desse modo, operações com valores de exposição total diferentes podem representar o mesmo saldo exposto real para instituição. Também pode ocorrer de as operações não terem seus prazos finais definidos.
<b>Impacto</b>	Médio.
<b>Tratamento do Risco</b>	Técnicas de proporcionalidade devem ser aplicadas, ou seja, deve-se manter o foco na data atual para se obter os saldos restantes expostos até o término das operações. Com esses valores, um peso deve ser aplicado às operações sem desconsiderar o valor total do contrato.
<b>Considerações</b>	Os valores já pagos das operações não estão expostos ao risco climático, por isso, há a necessidade de utilização do valor do saldo atual. Apesar desse ajuste, o valor empregado total do contrato deve ser considerado com um fator de ajuste. Tendo em vista que, mesmo que existam operações com prazos e saldos restantes semelhantes, o valor empregado total no investimento do contrato é diferente e conseqüentemente seu risco, por isso, deve ser considerado.

Fonte: Elaborado pelo autor

Tabela 4.17: Relevância do Cliente

<b>Risco</b>	<b>Cliente</b>
<b>Identificação do Risco</b>	A relevância do cliente deve ser considerada.
<b>Análise do Risco</b>	Existem clientes com alta relevância em suas operações, quer seja pelo tipo de operação ou pelo volume de operações que possui. E isso pode representar uma alta porcentagem sobre os ativos totais da instituição.
<b>Avaliação do Risco</b>	Devido ao tipo de movimentação que possuem, os clientes com alta relevância financeira podem afetar o valor arrecadado pela instituição ou impactar diretamente em sua reputação: <ul style="list-style-type: none"> <li>• Operações com alto valor concentrado em um único ativo possuem relevância maior nessas operações do que clientes com valores pulverizados em diversos ativos.</li> <li>• Operações consideradas críticas por diversos fatores com o potencial de impactar a instituição. Exemplo: a indústria de petróleo, tabaco... Os impactos sofridos por essas operações possuem um grande potencial de risco de reputação.</li> </ul>
<b>Impacto</b>	Alto.
<b>Tratamento do Risco</b>	Deve haver uma ponderação atribuída, tanto pelo volume movimento e concentrado em um único cliente, como pelo tipo de movimentação capaz de afetar a reputação da instituição. Esses clientes não devem ter o peso padrão de clientes que tem suas operações pulverizadas. As operações consideradas mais arriscadas também devem ser analisadas com peso diferente.
<b>Considerações</b>	Há a necessidade de identificação dessas operações e atribuição de pesos, pois esse risco tem um grande potencial e não está previsto nos documentos de referência.

Fonte: Elaborado pelo autor

Tabela 4.18: Melhorias operacionais

<b>Risco</b>	<b>Melhorias operacionais</b>
<b>Identificação do Risco</b>	Não há variável de ponderação para melhorias no maquinário com o intuito de reduzir emissões.
<b>Análise do Risco</b>	Uma empresa que investir em melhorias de seu maquinário, como por exemplo, carros elétricos ou motores que consomem menos combustível e produzem menos emissões terá a mesma classificação que outra empresa que não investiu. Não há como fazer comparativo linear entre as empresas.
<b>Avaliação do Risco</b>	A exposição ao risco climático atrelada a uma operação que não fez investimento em renovação de maquinário será igual a de uma empresa que fez. Isso, além de trazer uma variável equivocada, pode desestimular empresas a fazer renovação de seu maquinário buscando o desenvolvimento sustentável.
<b>Impacto</b>	Medio.
<b>Tratamento do Risco</b>	Deve haver variável de mitigação de risco, para incentivar esse tipo de prática. Em muitas situações, é possível reduzir o valor da exposição ao risco climático de forma exponencial apenas renovando o maquinário para equipamentos mais avançados ou considerados verdes.
<b>Considerações</b>	Indicadores capazes de favorecer empresas que investem em renovação de maquinário são eficazes para estimular a prática dessa renovação.

Fonte: Elaborado pelo autor

### 4.3 Modelos

As técnicas de regularização e seleção de variáveis Ridge, Lasso e Elastic Net referenciadas no item 3.1.5 foram utilizadas para produzir resultados que auxiliassem na redução de dimensionalidade e mitigassem os fenômenos de overfitting e underfitting citados no 3.1.5. Como resultado das investigações, foram escolhidas as variáveis selecionadas pela técnica Elastic Net. Desse modo, das 3.708 observações, foram selecionadas as observações: *Valor Contrato, Saldo Devedor, Prazo Médio do contrato: Longo, Risco da Operação: AA*.

O resultado produzido pela técnica Elastic Net não selecionou as variáveis de exposição ao risco climático geradas pelas metodologias da Taxonomia Verde e RSRC. Portanto, elas foram incluídas manualmente ao conjunto de dados selecionado.

Tendo o exposto, Os dados foram submetidos aos algoritmos Regressão Lógica, Árvore de Decisão, Random Forast, Gradient Boosting, AdaBoosting e XGBoost e avaliados pelas métricas informadas no item 3.1.5. Os resultados obtidos estão apresentados na Tabela 4.19

Tabela 4.19: Avaliação de desempenho

Modelo	Acurácia	Precisão	Recall	F1-Score	ROC	t/s
<i>Regressão logística</i>	0,35	0,84	0,35	0,40	0,77	23
Árvore de Decisão	0,91	0,94	0,91	0,91	0,95	0,51
<i>Random Forast</i>	0,93	0,95	0,93	0,94	0,95	8,5
<i>Gradient Boosting</i>	0,98	0,98	0,98	0,98	0,88	7,68
<i>AdaBoost</i>	0,90	0,94	0,90	0,91	0,96	11
<i>XGBoost</i>	0,94	0,95	0,94	0,94	0,97	6,53

Fonte: Elaborado pelo autor

- O Gradient Boosting apresentou melhores resultados de acurácia, precisão, recall e f1-score.
- O modelo XGBoost apresentou a maior área sob a curva roc seguido pelo AdaBoost em segundo.

As figuras Figuras 4.1 a 4.6 ilustram a matriz de confusão e a curva roc dos modelos. A matriz de confusão apresenta os resultados referentes a taxa de predição dos algoritmos que foi utilizada para realização das métricas, assim como referenciado no item 3.1.5, a curva roc apresenta a eficiência do modelo, em que quanto maior a área abaixo da curva, melhor, assim como foi referenciado no item 3.1.5.

roc\_auc: 0.7683404281343056

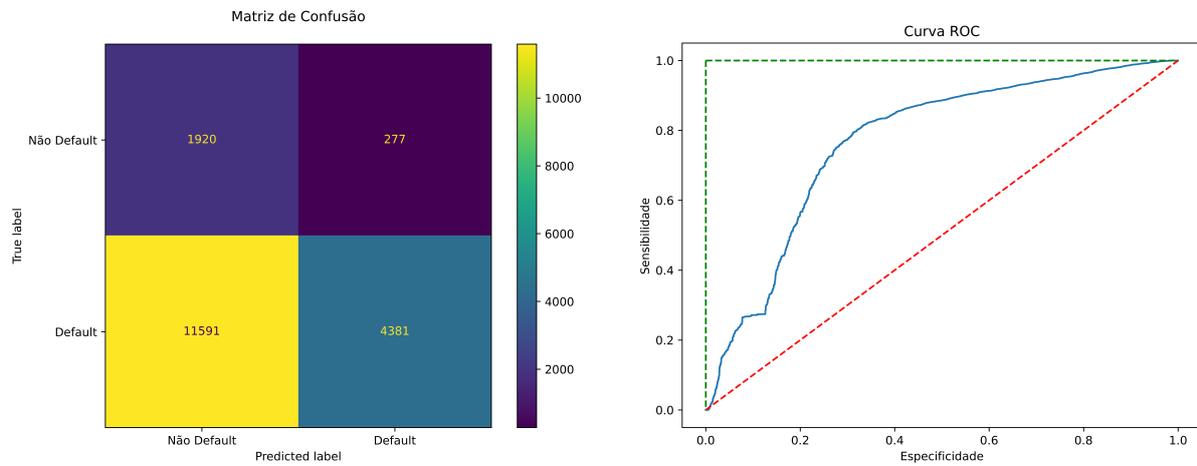


Figura 4.1: Matriz de Confusão e Curva ROC: Regressão Logística  
Fonte: Elaborado pelo autor

roc\_auc: 0.954289530460737

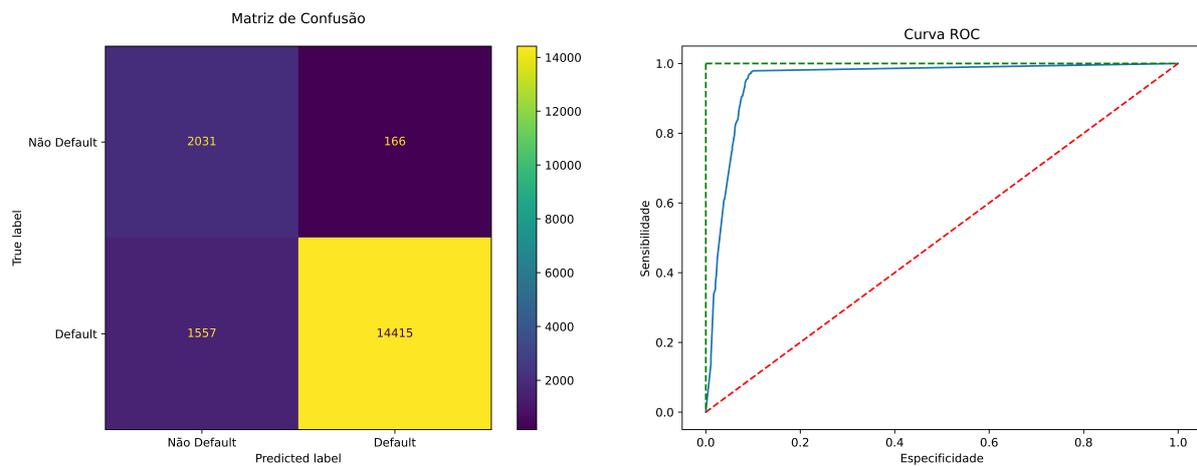


Figura 4.2: Matriz de Confusão e Curva ROC: Árvore de Decisão  
Fonte: Elaborado pelo autor

roc\_auc: 0.954833353110775

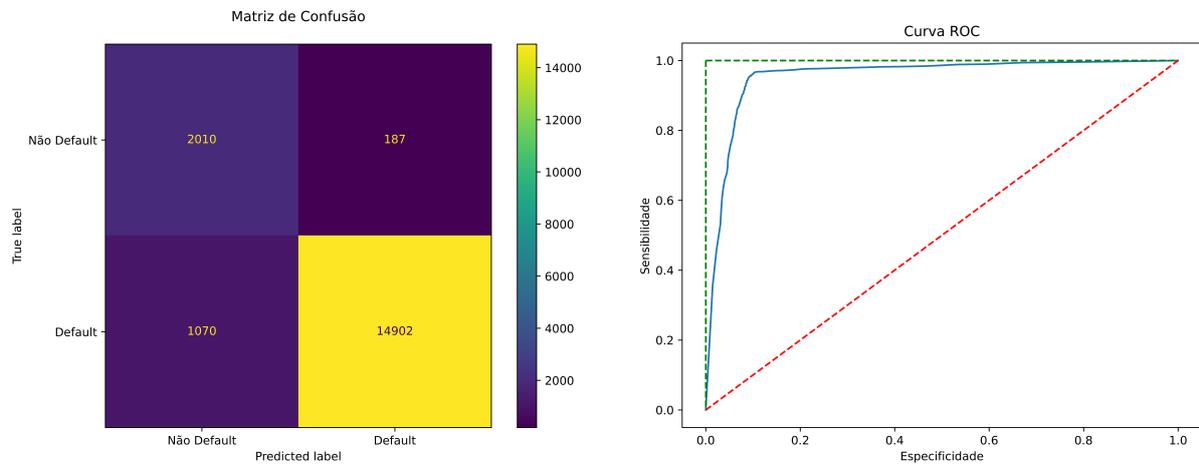


Figura 4.3: Matriz de Confusão e Curva ROC: Random Forast  
Fonte: Elaborado pelo autor

roc\_auc: 0.8753653554621817

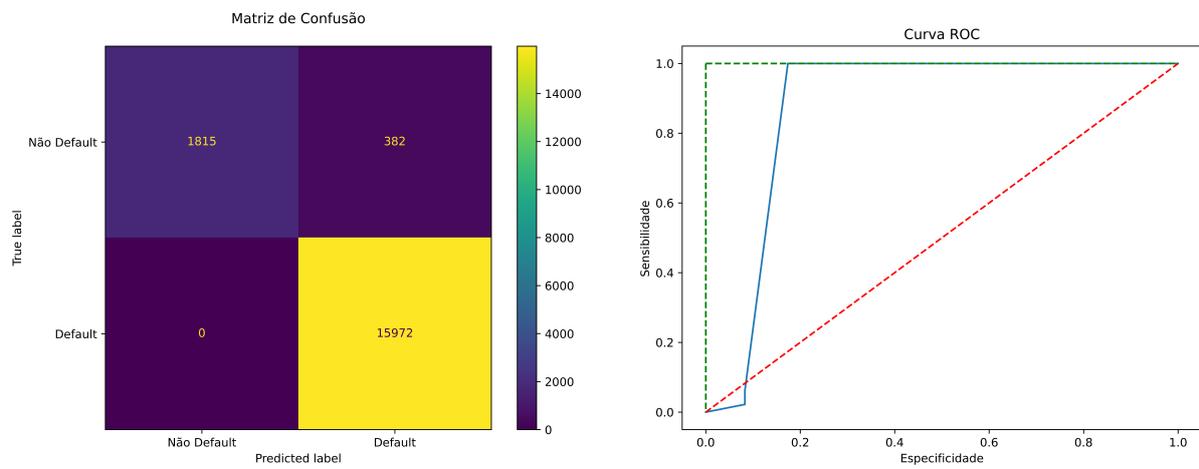


Figura 4.4: Matriz de Confusão e Curva ROC: Gradient Boosting  
Fonte: Elaborado pelo autor

roc\_auc: 0.963856212413599

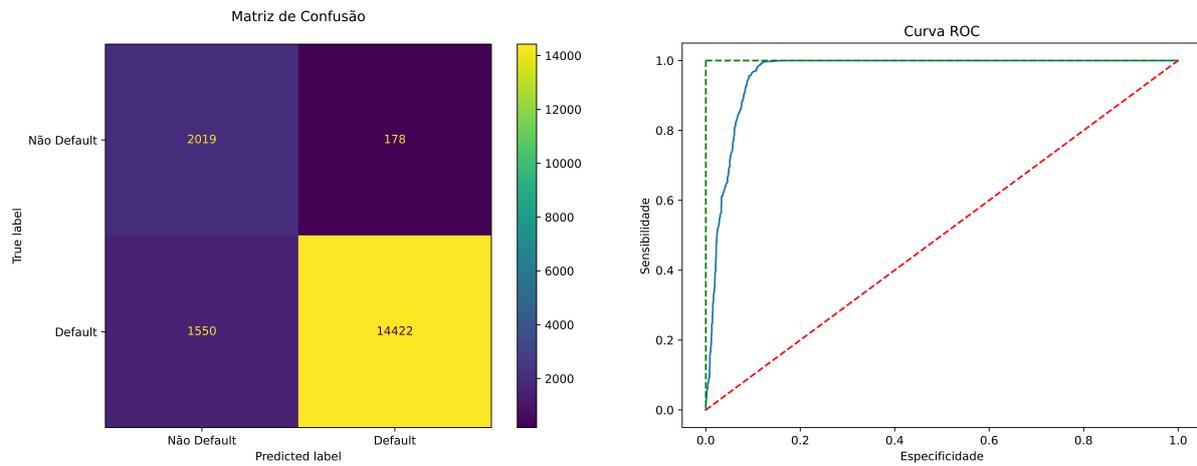


Figura 4.5: Matriz de Confusão e Curva ROC: AdaBoost  
Fonte: Elaborado pelo autor

roc\_auc: 0.969695886782297

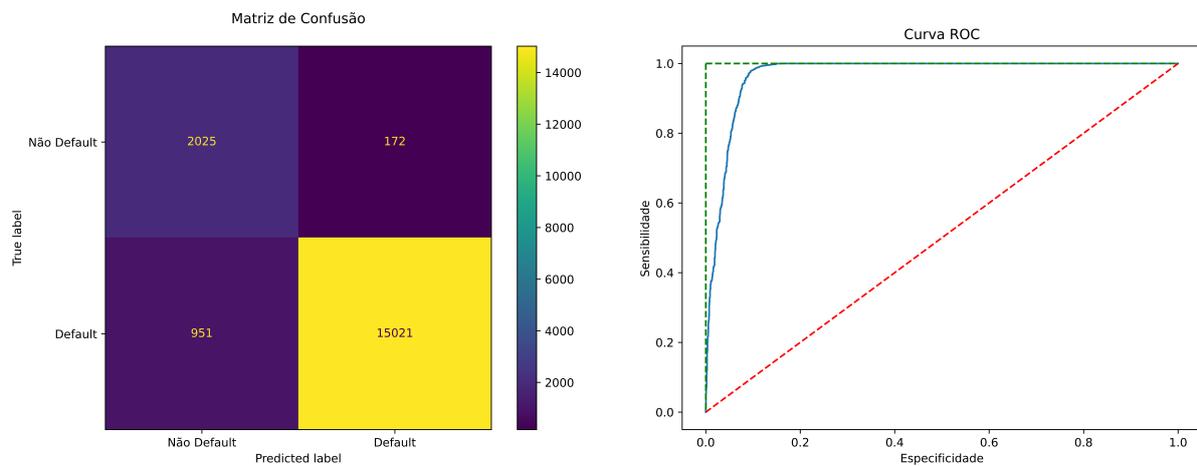


Figura 4.6: Matriz de Confusão e Curva ROC: XGBoost  
Fonte: Elaborado pelo autor

Todos os modelos tiveram seu desempenho melhorado após a redução de dimensionalidade.

## 4.4 Explicabilidade

Os algoritmos foram submetidos às ferramentas de explicabilidade da biblioteca shapley value. Os algoritmos Gradient Boosting e Extreme Gradient Boosting (XGBoost) foram avaliados por terem apresentado os melhores resultados. Após a avaliação do melhor modelo, os valores SHAP foram calculados para ser possível identificar o impacto das variáveis nos modelos e extrair conclusões de seus relacionamentos, assim como foi explicado nos itens 3.1.5 e 3.1.6

Apesar do algoritmo Gradient Boosting ter apresentado melhores resultados nas métricas de avaliação, os resultados produzidos pelo algoritmo de explicabilidade não apresentaram interpretabilidade satisfatória. Desse modo, somente o algoritmo XGBoost foi utilizado por apresentar o segundo melhor resultado nas métricas de avaliação e a maior área sob a curva ROC.

Os resultados sumarizados podem ser visualizados na Figura 4.7 e interpretados da seguinte forma:

- O Eixo Y apresenta as variáveis comparadas com os casos de inadimplência que possuem os valores maiores que 0 no eixo X (variável em *default*) e os casos de não inadimplência que possuem os valores iguais ou menores que 0 no eixo X (variável não está em *default*).
- O Eixo X apresenta a numeração que possibilita identificar o quanto a variável pode influenciar os casos de *default* ou não. Ou seja, Quanto maior for a numeração positiva ou negativa do eixo X, maior o potencial de influência da variável para o respectivo caso.
- O gradiente de cores entre azul e vermelho representam a incidência da presença da variável. Em que quanto mais perto da cor vermelha, mais forte é a presença da variável e quanto mais próxima da cor azul, mais fraca é a presença da variável.
- Para os casos de variáveis numéricas, as cores representam a distribuição dos valores. Ou seja, quanto mais próximos da cor vermelha, maiores serão os valores da variável e quanto mais próximas da cor azul, menores serão os valores.
- Para os casos de variáveis categóricas, as cores representam a presença da variável ou sua ausência. Ou seja, quanto mais próximos da cor vermelha, maior a presença da variável e quanto mais próximas da cor azul, menor será a presença da variável.

Dessa forma é possível identificar o comportamento da inadimplência em relação aos valores assumidos pelas variáveis de forma individual ou comparada.

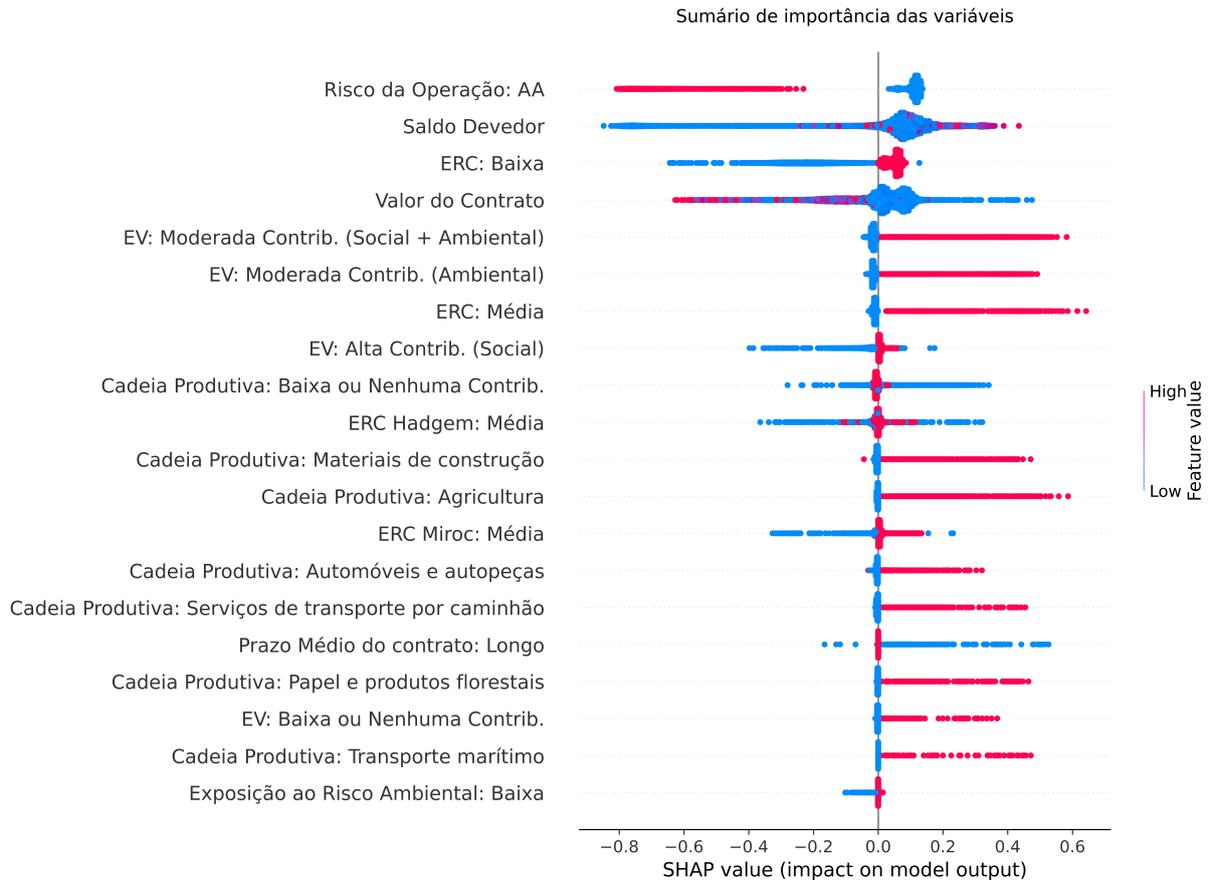


Figura 4.7: SHAP value: Sumário  
 Fonte: Elaborado pelo autor

**Resultados Observados:**

- *Risco da Operação AA:*
  - A presença da variável apresentou um claro potencial de redução de *default*.
  - A ausência da variável apresentou tendência de aumento de *default*.
  
- *Saldo Devedor:*
  - É possível identificar um aumento de pontos vermelhos que indicam maiores valores ao lado direito, indicando uma tendência de aumento de *default* a medida que os valores da variável sobem.
  - A redução dos valores da variável possui uma clara tendência de redução de *default*.
  - Recomenda-se a investigação dos pontos vermelhos que se destacam ao lado esquerdo e indicam redução de *default*.

- *ERC: Baixa:*
  - A presença da variável apresenta tendência de aumento de *default*.
  - A ausência da variável apresentou uma clara tendência de redução de *default*.
- *Valor do Contrato:*
  - É possível identificar um aumento de pontos vermelhos que indicam maiores valores ao lado esquerdo, indicando tendência de redução de *default* a medida que os valores da variável sobem.
  - A ausência da variável apresentou tendência de aumento de *default*.
  - Os pontos que indicam a redução dos valores da variável estão presentes em ambos os lados, mas se concentram ao lado direito. Recomenda-se investigar os pontos azuis do lado esquerdo.
- *EV: Moderada Contrib. (Social + Ambiental):*
  - A presença da variável apresentou uma clara tendência de aumento de *default*.
  - A ausência da variável não sugeriu significância.
- *EV: Moderada Contrib. (Ambiental):*
  - A presença da variável apresentou uma clara tendência de aumento de *default*.
  - A ausência da variável não sugeriu significância.
- *ERC: Média:*
  - A presença da variável apresentou uma clara tendência de aumento de *default*.
  - A ausência da variável não sugeriu significância.
- *EV: Alta Contrib. (Social):*
  - A presença da variável sugeriu aumento de *default*.
  - A ausência da variável apresentou uma clara tendência de redução de *default*.
- *Cadeia Produtiva: Baixa ou Nenhuma Contrib.:*
  - A presença da variável não apresentou significância.
  - Os pontos que indicam a redução da presença da variável estão presentes em ambos os lados, mas se concentram ao lado direito, indicando uma maior tendência de aumento de *default* a medida que não há a presença da variável.

- Recomenda-se investigar os limiares que geram a mudança do indicador de *default*.
- *ERC Hadgem: Média:*
  - Os pontos que indicam o aumento da presença da variável estão presentes em ambos os lados, mas se concentram ao lado direito, indicando uma maior tendência de aumento de *default* a medida que aumenta a presença da variável.
  - Os pontos que indicam a redução da presença da variável estão presentes em ambos os lados, mas se concentram ao lado esquerdo, indicando uma maior tendência de redução de *default* a medida que reduz a presença da variável.
  - Recomenda-se a investigação dos pontos vermelhos que se destacam de forma isolada ao lado esquerdo.
- *Cadeia Produtiva: Materiais de construção:*
  - A presença da variável apresentou uma clara tendência de aumento de *default*.
  - A ausência da variável não apresentou significância.
- *Cadeia Produtiva: Agricultura:*
  - A presença da variável indica uma clara tendência de aumento de *default*.
  - A ausência da variável não apresentou significância.
- *ERC Miroc: Média:*
  - A presença da variável indica uma clara tendência de aumento de *default*.
  - Os pontos que indicam a redução da presença da variável estão presentes em ambos os lados, mas se concentram ao lado esquerdo, indicando uma maior tendência de redução de *default* a medida que reduz a presença da variável.
  - Recomenda-se a investigação do ponto azul isolado ao lado direito.
- *Cadeia Produtiva: Automóveis e autopeças:*
  - A presença da variável indica uma clara tendência de aumento de *default*.
  - A ausência da variável não apresentou significância.
- *Cadeia Produtiva Serviços de transporte por caminhão:*
  - A presença da variável indica uma clara tendência de aumento de *default*.
  - A ausência da variável não apresentou significância.

- *Prazo Médio do contrato: Longo:*
  - A presença da variável não apresentou significância.
  - A ausência da variável indica uma clara tendência de aumento de *default*.
  - Recomenda-se investigar os pontos azuis ao lado esquerdo.
- *Cadeia Produtiva: Papel e produtos florestais:*
  - A presença da variável indica uma clara tendência de aumento de *default*.
  - A ausência da variável não apresentou significância.
- *EV: Baixa ou Nenhuma Contrib.:*
  - A presença da variável indica uma clara tendência de aumento de *default*.
  - A ausência da variável não apresentou significância.
- *Cadeia Produtiva: Transporte marítimo:*
  - A presença da variável indica uma clara tendência de aumento de *default*.
  - A ausência da variável não apresentou significância.
- *Exposição ao Risco Ambiental: Baixa:*
  - A presença da variável não apresentou significância.
  - A ausência da variável indica tendência de redução de *default*.

Os resultados produzidos mostraram de forma sumarizada a capacidade das variáveis selecionadas de influenciar as operações aumentando ou reduzindo o risco quando entram ou não em *default*.

Com essas informações, os resultados observados foram gerados dando preferência aos agrupamentos com a maior concentração de pontos das respectivas cores.

Também foi possível identificar situações em que haviam consideráveis agrupamentos de pontos vermelhos e azuis em ambos os lados do gráfico, mesmo que estivessem majoritariamente em uma única direção. Investigar essas situações para entender o que influenciou o indicador *default* a sofrer alteração em sua distribuição é importante.

Para isso, análises comparativas foram realizadas para identificar, nos casos de valores numéricos, os limiares dos gradientes de valores responsáveis pelo indicador entrar em *default* e, nos casos de valores categóricos, os fatores adjacentes que possam ter influenciado a mudança do indicativo de *default*. Essas análises tiveram como foco a produção de insumos que auxiliassem a responder o problema proposto pela pesquisa.

Tendo o exposto, as três variáveis de exposição ao risco climático foram analisadas por gráficos de dispersão, assim como ilustra a Figura 4.8. Sua interpretação é feita da seguinte forma:

- O nome presente no eixo esquerdo Y é a variável investigada.
- O valor 0 do eixo X representa a situação de não de presença inadimplência, ou seja, a variável não está em *default*.
- O valor 1 do eixo X representa a situação de presença de inadimplência, ou seja, a variável está em *default*.
- Os valores apresentados no eixo Y representam o potencial de impacto da variável em uma das duas situações citadas, ou seja, a influência da variável em *default* ou não em *default*.

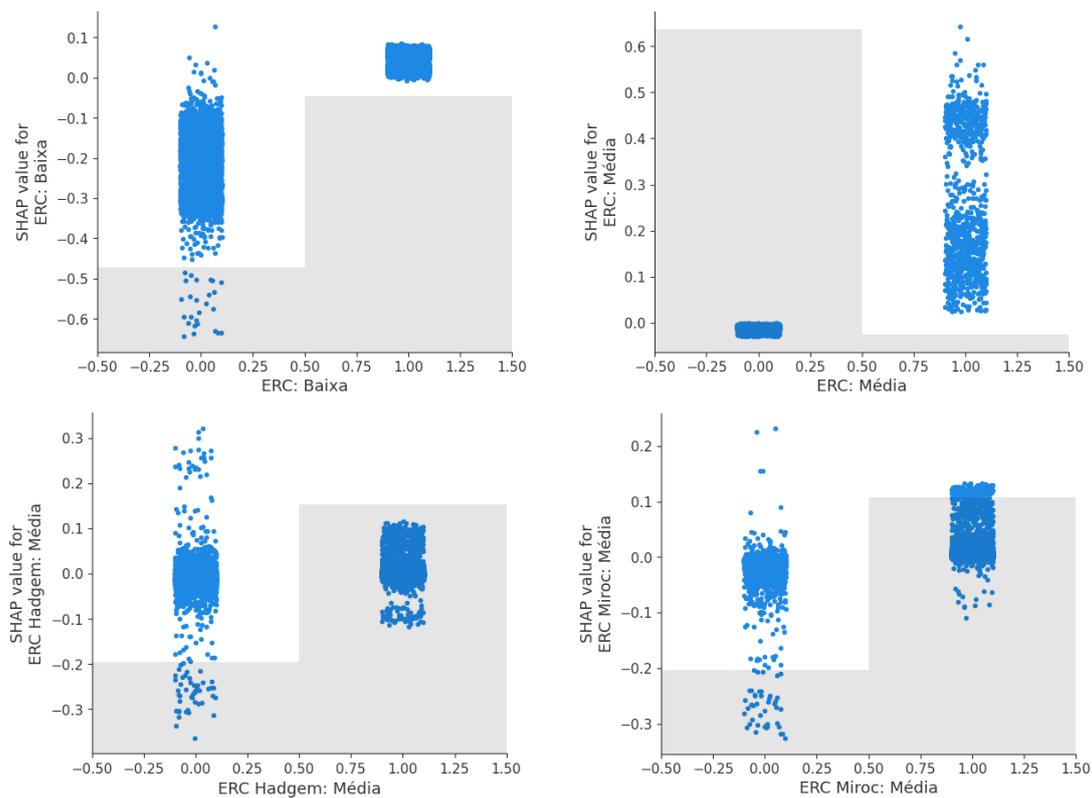


Figura 4.8: Shapley Value: Exposição ao Risco Climático e Risco de Crédito  
 Fonte: Elaborado pelo autor

## Resultados Observados

- A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 70%.
- A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10%.
- A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* próximo de 0%.
- A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 70%.
- A média exposição ao risco climático hadgem tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 40% e aumentar o risco de até 30%.
- A média exposição ao risco climático hadgem tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e reduzir o risco de até 10%.
- A média exposição ao risco climático miroc tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30% e aumentar o risco de até 20%.
- A média exposição ao risco climático miroc tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e reduzir o risco de até 20%.

Os resultados indicam que a exposição ao risco climático possui um grande potencial de impactar o risco de inadimplência das operações de crédito.

Ressalta-se que os indicadores alternativos realizaram classificações mais brandas que o principal (taxonomia verde) e seus resultados foram mais esparsos, porém semelhantes, apresentando maiores concentrações em direções que sugerem interpretação semelhante aos do indicador principal.

Em seguida, essas variáveis foram combinadas, em dupla, com as demais variáveis. Essas combinações produziram um resultado de 400 gráficos no total. Desse total, foram retiradas as combinações das variáveis com ela mesma e as combinações das três variáveis de exposição ao risco climático entre si. Por fim, foram selecionadas as variáveis identificadas como tendo maior potencial de oferecer detalhes para responder o problema proposto desta pesquisa. Essas variáveis foram divididas em três grupos, assim como é detalhado a seguir.

- **Variáveis de exposição ao risco climático:** *ERC: Baixa; ERC: Média; ERC Hadgem: Média; ERC Miroc: Média.*
- **Variáveis de Crédito:** *Risco da Operação: AA; Valor do Contrato.*
- **Variáveis de Economia verde:** *EV: Moderada Contribuição (Social + Ambiental); EV: Moderada Contribuição (Ambiental); EV: Baixa ou Nenhuma Contribuição; EV: Alta Contribuição (Social).;*
- **Variáveis de Risco Ambiental:** *Exposição ao Risco Ambiental: Baixa.*

Todas as combinações selecionadas possuem como variável foco as variáveis de exposição ao risco climático na seguinte ordem: *ERC: Baixa; ERC: Média; ERC Hadgem: Média; ERC Miroc: Média.*

As combinações apresentadas nas Figuras 4.9 a 4.15 ilustram de forma visual o potencial de influência das variáveis de exposição ao risco climático sobre o indicador de inadimplência (*default*). As imagens são interpretadas da seguinte forma:

- O nome presente no eixo esquerdo Y é a variável investigada.
- O valor 0 do eixo X representa a situação de não presença de inadimplência, ou seja, a variável não está em *default*.
- O valor 1 do eixo X representa a situação de presença de inadimplência, ou seja, a variável está em *default*.
- Os valores apresentados no eixo Y representam o potencial de impacto da variável em uma das duas situações citadas, ou seja, influenciar a variável em *default* ou não em *default*.
- Ao lado direito é apresentada a variável que foi utilizada como comparação, em que o gradiente de cores entre azul e vermelho representam a incidência da presença da variável. Quanto mais perto da cor vermelha, mais forte é a presença da variável e quanto mais próxima da cor azul, mais fraca é a presença da variável.
- Para os casos de variáveis numéricas, as cores representam a distribuição dos valores. Ou seja, quanto mais próximos da cor vermelha, maiores serão os valores da variável e quanto mais próximas da cor azul, menores serão os valores.
- Para os casos de variáveis categóricas, as cores representam a presença da variável ou sua ausência. Ou seja, quanto mais próximos da cor vermelha, maior a presença da variável e quanto mais próximas da cor azul, menor será a presença da variável.

As Figura 4.9 ilustra a combinação das variáveis de exposição ao risco climático com a variável de crédito *Risco da Operação: AA*.

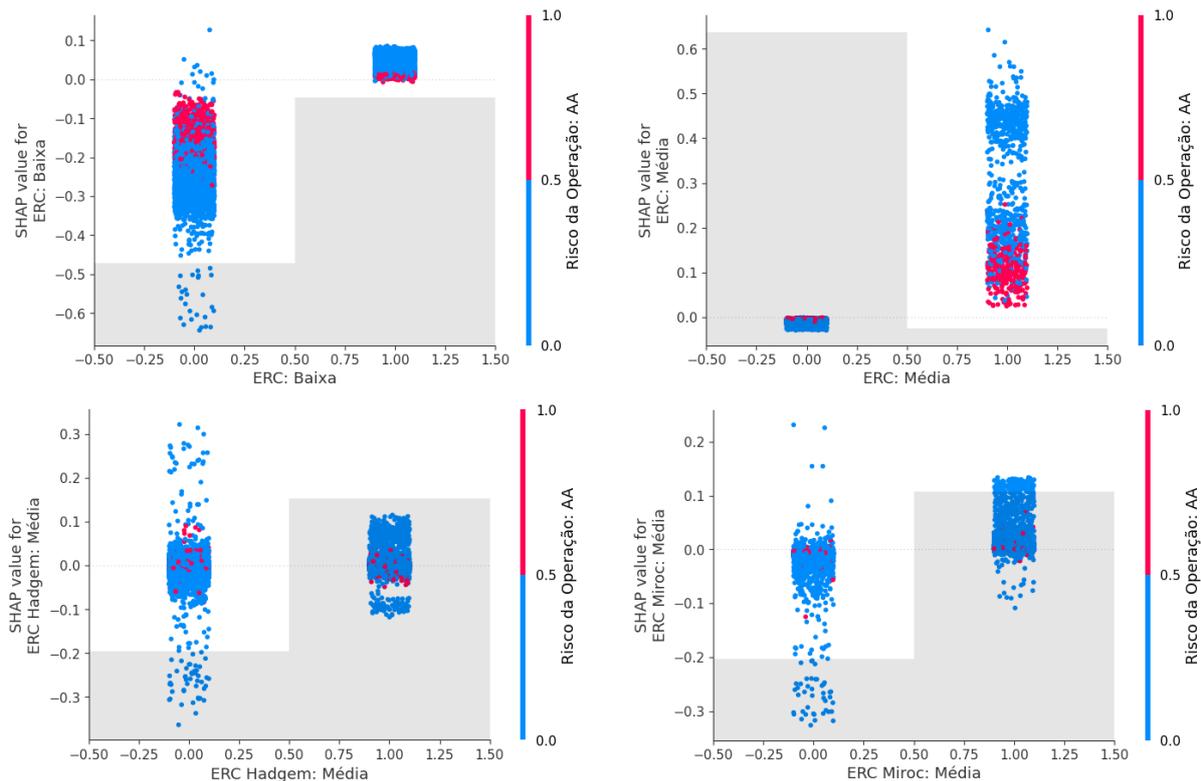


Figura 4.9: Shapley Value: Exposição ao Risco Climático e Risco da Operação AA  
 Fonte: Elaborado pelo autor

## Resultados Observados

- *ERC: Baixa e Risco da Operação: AA:*
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30%, se concentrando em até 20%.
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* próximo de 0%. A variável reduz o risco de *default*.
- *ERC: Média e Risco da Operação: AA:*
  - A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de 0%.
  - A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 30%. A variável reduz o risco de *default*.

- *ERC Hadgem: Média e Risco da Operação: AA:*
  - A média exposição ao risco climático hadgem tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 10% e de aumentar o risco de até 10%.
  - A média exposição ao risco climático hadgem tem o potencial de influenciar aumentando o risco quando a variável está em *default* de 0%.
- *ERC Miroc: Média e Risco da Operação: AA:*
  - A média exposição ao risco climático miroc tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 10%.
  - A média exposição ao risco climático miroc tem o potencial de influenciar aumentando o risco quando a variável está em *default* próximo de 0%.

Os resultados indicam que a classificação de baixa exposição ao risco climático da taxonomia verde associada ao risco de crédito mostrou grande potencial de auxiliar na redução do risco de *default*, ao mostrar grandes concentrações da presença da variável ao reduzir o risco quando a variável não está em *default* e capacidade de neutralizar o risco ao entrar em *default*. A classificação de média exposição ao risco climático não mostrou potencial de reduzir o risco quando a variável não está em *default*, mas mostrou um grande potencial ao reduzir o risco ao entrar em *default*.

A Figura 4.9 ilustrou a comparação direta da exposição ao risco climático com a variável de risco de crédito. A seguir, as Figuras 4.10 a 4.15 apresentam as demais combinações consideradas importantes para o entendimento do comportamento da exposição ao risco climático.

As Figura 4.10 ilustra a combinação das variáveis de exposição ao risco climático com a variável de crédito *Valor do Contrato*.

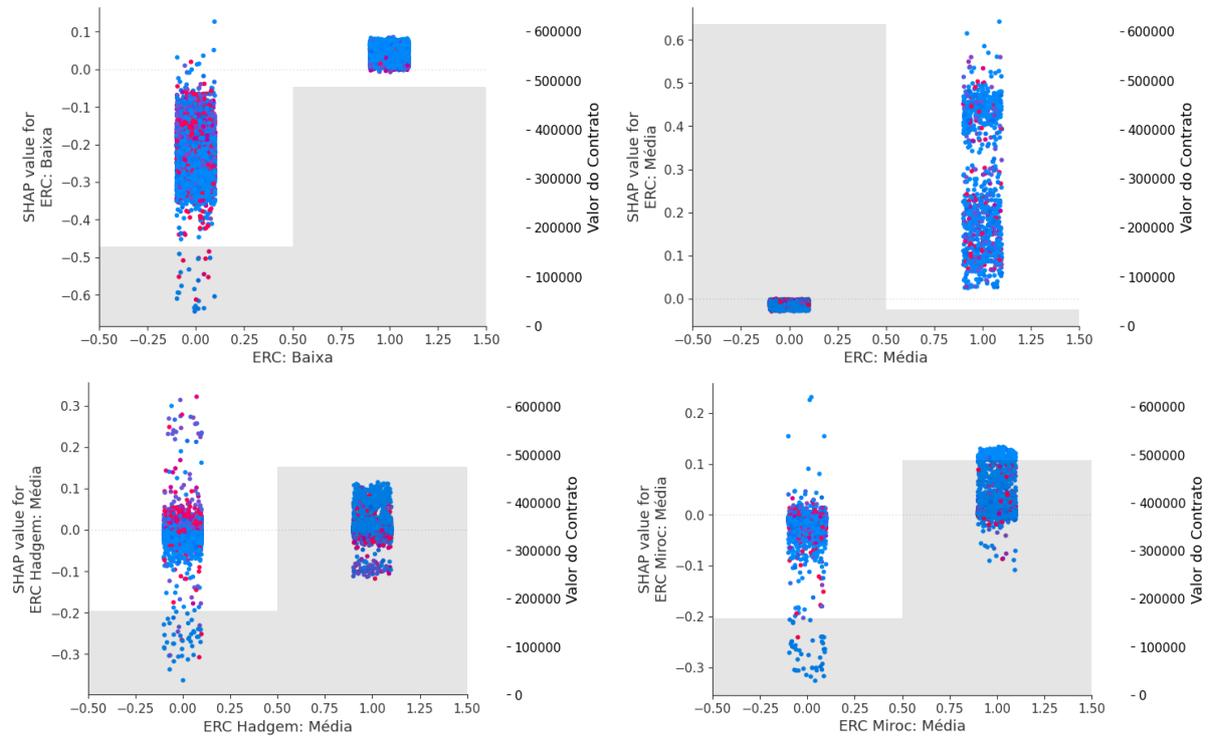


Figura 4.10: Shapley Value: Exposição ao Risco Climático e Valor do Contrato  
Fonte: Elaborado pelo autor

## Resultados Observados

- *ERC: Baixa e Valor do Contrato:*
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 60%, quando há aumento dos valores da variável.
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10%, com sua concentração próxima de 0%, quando há aumento dos valores da variável.
- *ERC: Média e Valor do Contrato:*
  - A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* próximo de 0%, quando há aumento dos valores da variável.

- A exposição ao risco climático tem o potencial de influenciar o risco da variável a entrar em inadimplência de até 60%, quando há a presença de contratos com maiores valores.
- *ERC Hadgem: Média e Valor do Contrato:*
  - A média exposição ao risco climático hadgem tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30% e aumentar o risco de até 30%, quando há a presença de contratos com maiores valores. Sua maior concentração de maiores valores indica redução do risco.
  - A média exposição ao risco climático hadgem tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e reduzir o risco de até 10%, quando há a presença de contratos com maiores valores. Sua maior concentração de maiores valores indica aumento de risco.
- *ERC Miroc: Média e Valor do Contrato:*
  - A média exposição ao risco climático miroc tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30% e aumentar o risco de até 10%, quando há a presença de contratos com maiores valores. Sua maior concentração de maiores valores indicam redução do risco.
  - A média exposição ao risco climático miroc tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e reduzir o risco de até 10%, quando há a presença de contratos com maiores valores. Sua maior concentração de maiores valores indica aumento de risco.

Os resultados indicam que a classificação de baixa exposição ao risco climático da taxonomia verde associada ao valor do contrato mostrou potencial de auxiliar na redução do risco de *default*. Mostrando grandes concentrações de maiores valores da variável ao reduzir o risco quando a variável não está em *default* e também mostrou potencial ao reduzir o risco ao entrar em *default*. A classificação de média exposição ao risco climático não mostrou potencial ao reduzir o risco quando a variável não está em *default*, nem mostrou potencial ao reduzir o risco ao entrar em *default*.

As Figura 4.11 detalha ilustra a combinação das variáveis de exposição ao risco climático com a variável de Economia Verde Moderada Contribuição (Social + Ambiental).

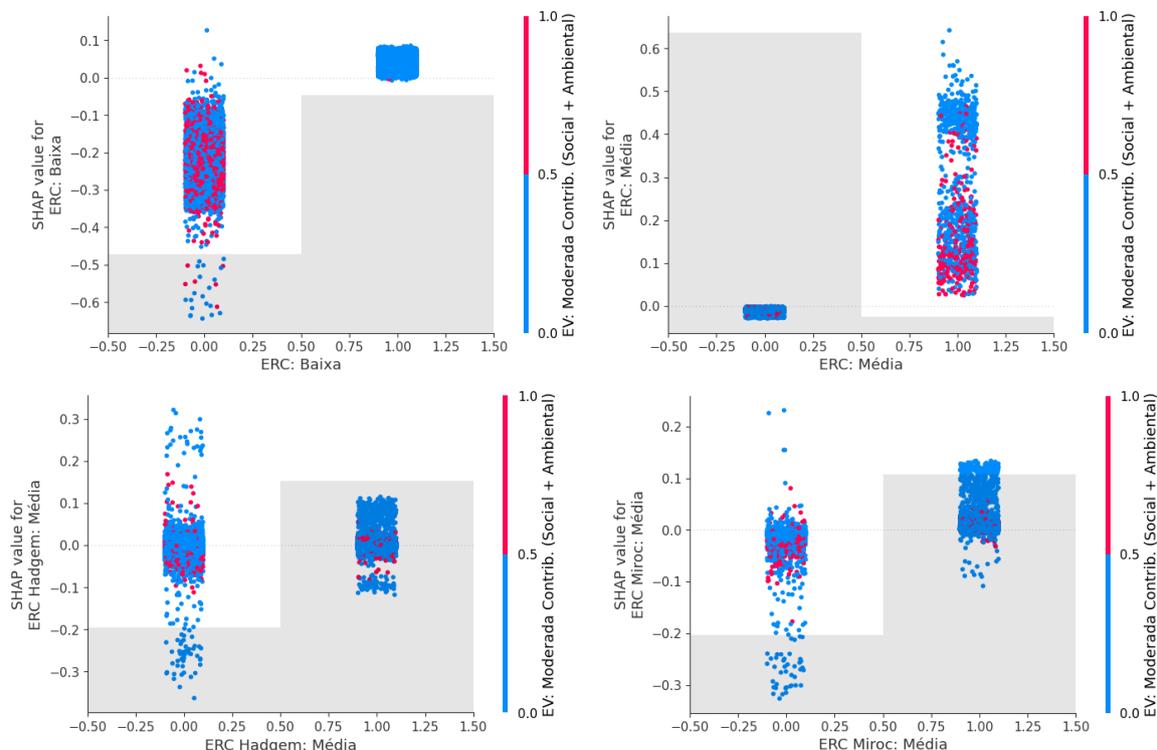


Figura 4.11: Shapley Value: Exposição ao Risco Climático e Economia Verde Moderada Contribuição (Social + Ambiental)

Fonte: Elaborado pelo autor

## Resultados Observados

- *ERC: Baixa e EV: Moderada Contribuição (Social + Ambiental):*
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 60%.
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável está em *default* de até 0%.
- *ERC: Média e EV: Moderada Contribuição (Social + Ambiental):*
  - A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* próximo de 0%.

- A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 50%.
- *ERC Hadgem: Média e EV: Moderada Contribuição (Social + Ambiental):*
  - A média exposição ao risco climático hadgem tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 10% e aumentar o risco de até 20%. Sua concentração está em reduzir o risco até 10%.
  - A média exposição ao risco climático hadgem tem o potencial de influenciar aumentando o risco quando a variável está em *default* próximo de até 10% e reduzir o risco de até 10%. Sua concentração está em reduzir o risco até 10%
- *ERC Miroc: Média e EV: Moderada Contribuição (Social + Ambiental):*
  - A média exposição ao risco climático miroc tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 20% e aumentar o risco de até 10%. Sua concentração está em reduzir o risco até 10%.
  - A média exposição ao risco climático miroc tem o potencial de influenciar aumentando o risco quando a variável está em *default* próximo de 0%.

Os resultados indicam que a classificação de baixa exposição ao risco climático da taxonomia verde associada à variável de economia verde mostrou um grande potencial de auxiliar na redução do risco de *default*. Mostrando grandes concentrações da presença da variável ao reduzir o risco quando a variável não está em *default* mas não apresentou casos quando a variável entra em *default*. A classificação de média exposição ao risco climático não mostrou potencial ao reduzir o risco quando a variável não está em *default*, mas mostrou potencial de reduzir o risco quando a variável entra em *default*.

As Figura 4.12 ilustra a combinação das variáveis de exposição ao risco climático com a variável de *Economia Verde Moderada Contribuição (Ambiental)*.

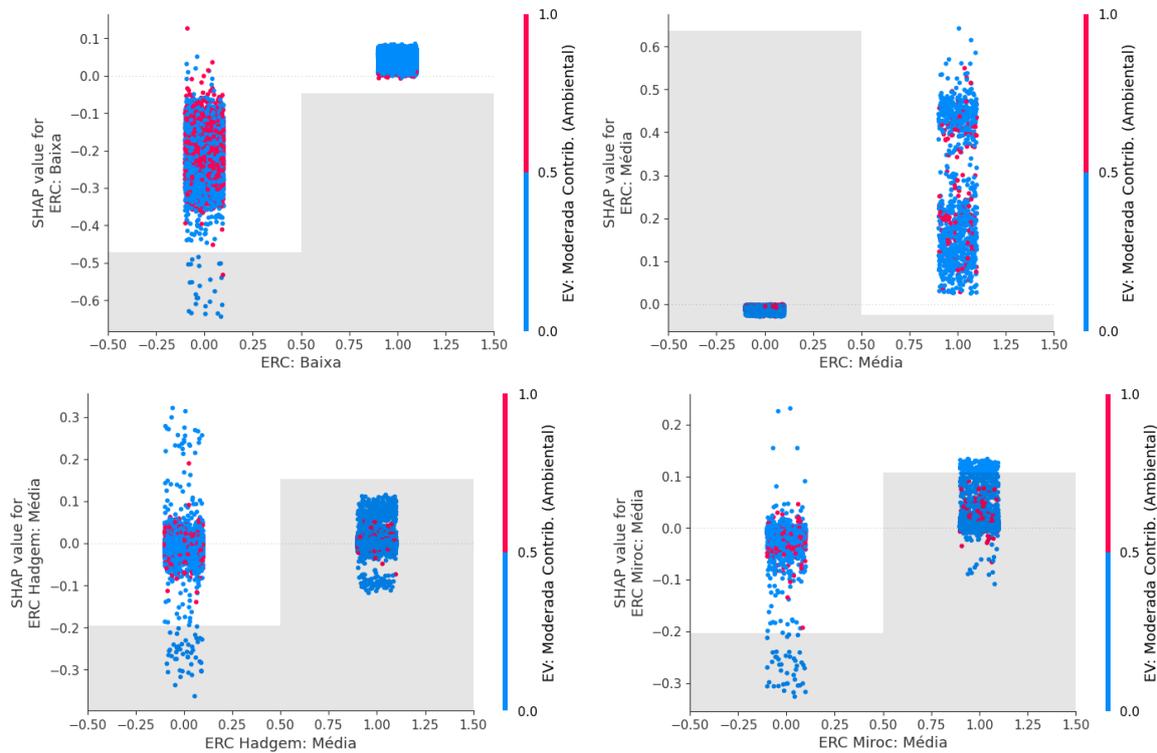


Figura 4.12: Shapley Value: Exposição ao Risco Climático e Economia Verde Moderada Contribuição (Ambiental)

Fonte: Elaborado pelo autor

## Resultados Observados

- *ERC: Baixa e EV: Moderada contribuição (Ambiental)*:
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 50%.
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 0%.
- *ERC: Média e EV: Moderada contribuição (Ambiental)*:
  - A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* próximo de 0%.

- A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 70%.
- *ERC Hadgem: Média e EV: Moderada contribuição (Ambiental):*
  - A média exposição ao risco climático hadgem tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 20% e aumentar o risco de até 20%. Com concentrações em até 10% de redução do risco.
  - A média exposição ao risco climático hadgem tem o potencial de influenciar aumentando o risco quando a variável está em *default* próximo de 0%.
- *ERC Miroc: Média e EV: Moderada contribuição (Ambiental):*
  - A média exposição ao risco climático miroc tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 20%.
  - A média exposição ao risco climático miroc tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e reduzir o risco de até 10%. Com maiores concentrações em até 10% de aumento do risco.

Os resultados indicam que a classificação de baixa exposição ao risco climático da taxonomia verde associada à variável de economia verde mostrou potencial de auxiliar na redução do risco de *default*. Mostrando grandes concentrações da presença da variável ao reduzir o risco quando a variável não está em *default* e também mostrou capacidade de reduzir o risco quando a variável está em *default*. A classificação de média exposição ao risco climático não mostrou potencial ao reduzir o risco quando a variável está ou não em *default*.

As Figura 4.13 ilustra a combinação das variáveis de risco climático com a variável Economia Verde Alta Contribuição (Social).

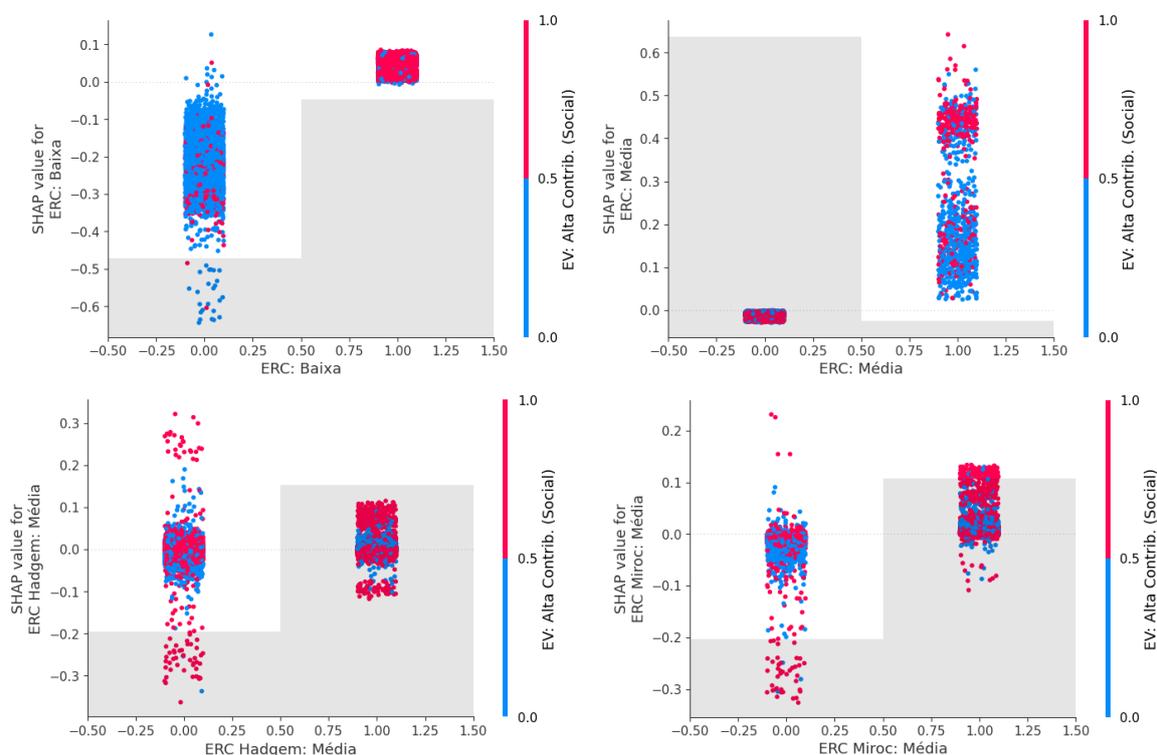


Figura 4.13: Shapley Value: Exposição ao Risco Climático e Economia Verde Alta Contribuição (Social)

Fonte: Elaborado pelo autor

## Resultados Observados

- *ERC: Baixa e EV: Alta Contribuição (Social):*
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 60%.
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10%
- *ERC: Média e EV: Alta Contribuição (Social):*
  - A média exposição ao risco climático da taxonomia verde apresentou potencial de reduzir o risco quando a variável não está em *default* próximo de 0%.
  - A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 60%. Com suas concentrações permanecendo nos níveis mais altos.

- *ERC Hadgem: Média e EV: Alta Contribuição (Social):*
  - A média exposição ao risco climático hadgem tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30% e de aumentar o risco em até 30%. Com suas concentrações indicando redução do risco.
  - A média exposição ao risco climático hadgem tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e de reduzir o risco de até 10%. Com suas concentrações indicando aumento do risco.
  
- *ERC Miroc: Média e EV: Alta Contribuição (Social):*
  - A média exposição ao risco climático miroc tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30% e aumentar o risco em até 20%. Com suas concentrações indicando redução do risco.
  - A média exposição ao risco climático miroc tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e reduzir o risco em até 10%. Com suas concentrações indicando aumento do risco.

Os resultados indicam que a classificação de baixa exposição ao risco climático da taxonomia verde associada à variável de economia verde mostrou potencial de auxiliar na redução do risco quando a variável não está em *default*, mostrando maiores concentrações da presença da variável nos pontos de maior redução de risco, mas apresentou maiores concentrações nos níveis mais elevados de risco nos casos em que a variável entra em *default*.

As Figura 4.14 ilustra a combinação das variáveis de risco climático com a variável Economia Verde Baixa Contribuição.

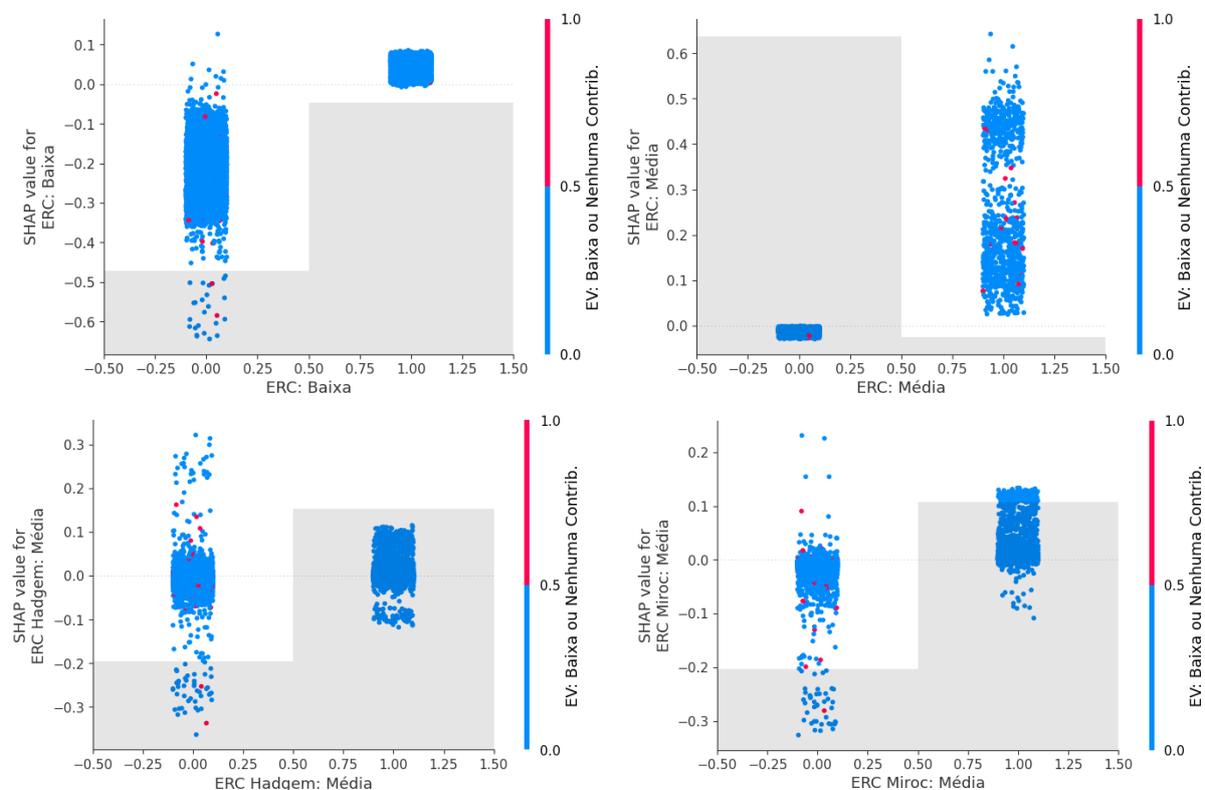


Figura 4.14: Shapley Value: Exposição ao Risco Climático e Economia Verde (Baixa ou Nenhuma Contribuição)

Fonte: Elaborado pelo autor

## Resultados Observados

- *ERC: Baixa e EV: Baixa Contrib.:*
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 60%.
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* próximo de 0%.
- *ERC: Média e EV: Baixa Contrib.:*
  - A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* próximo de 0%.

- A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 50%.
- *ERC Hadgem: Média e EV: Baixa Contrib.:*
  - A média exposição ao risco climático hadgem tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30% e aumentar o risco de até 10%, com maiores concentrações na redução do risco.
  - A média exposição ao risco climático hadgem não apresentou casos quando a variável está em *default*.
- *ERC Miroc: Média e EV: Baixa Contrib.:*
  - A média exposição ao risco climático miroc tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30%.
  - A média exposição ao risco climático hadgem não apresentou casos quando a variável está em *default*.

Os resultados da taxonomia verde foram esparsos. A classificação de baixa exposição ao risco climático da taxonomia verde sugeriu um potencial de influenciar na redução da inadimplência quando a variável não está em *default*. A classificação de média exposição ao risco climático mostrou que a variável de economia verde não impactou significativamente na redução do risco quando a variável entra em *default*. Mesmo com resultados esparsos, há uma indicação de redução de inadimplência.

As Figura 4.15 ilustra a combinação das variáveis de exposição ao risco climático com a variável de Risco Ambiental Baixa Exposição.

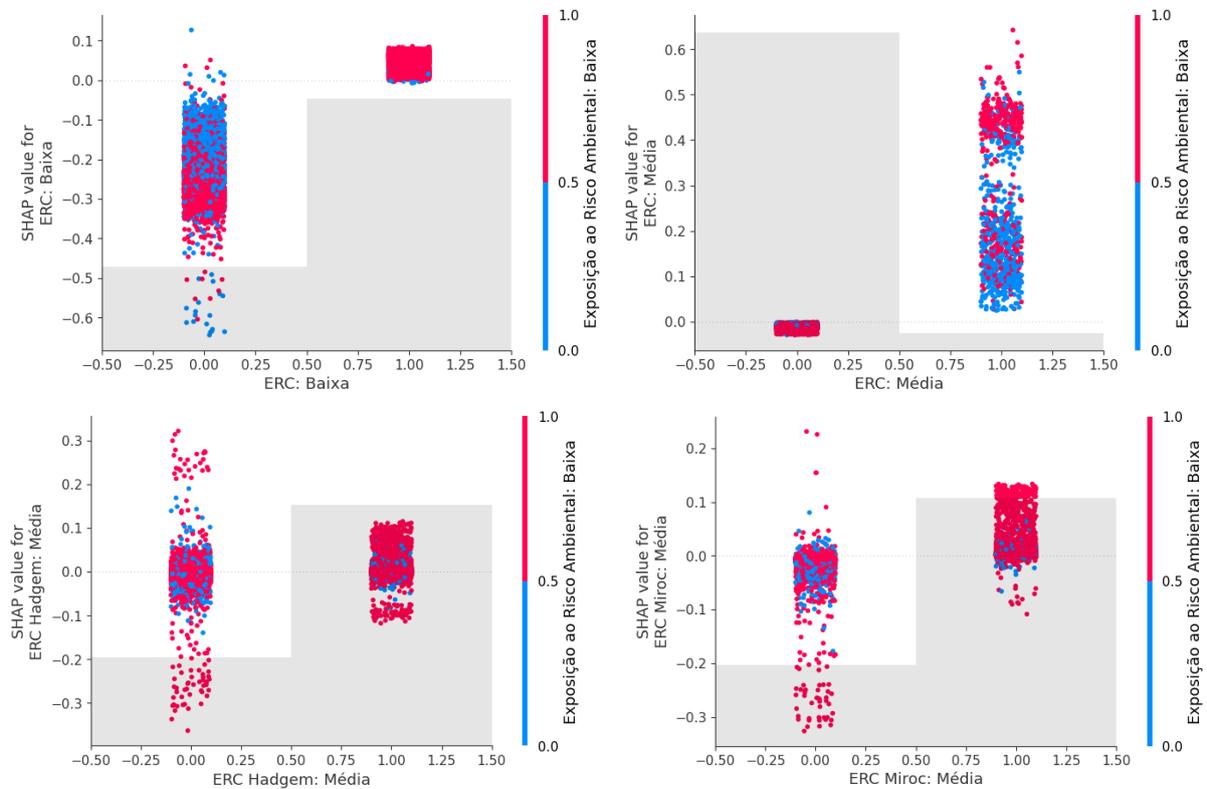


Figura 4.15: Shapley Value: Exposição ao Risco Climático e Exposição ao Risco Ambiental (Baixa ou Nenhuma Exposição)

Fonte: Elaborado pelo autor

## Resultados Observados

- *ERC: Baixa e Exposição ao Risco Ambiental: Baixa:*
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 60%, com maiores concentrações em até 40%.
  - A baixa exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10%.
- *ERC: Média e Exposição ao Risco Ambiental: Baixa*
  - A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* próximo de 0%.

- A média exposição ao risco climático da taxonomia verde tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 60%, com maiores concentrações em impactos mais altos.
- *ERC Hadgem: Média e Exposição ao Risco Ambiental: Baixa:*
  - A média exposição ao risco climático hadgem tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30% e aumentar o risco de até 30%, com maiores concentrações próximas de 0%.
  - A média exposição ao risco climático hadgem tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e reduzir o risco de até 10%, com maiores concentrações indicando aumento.
- *ERC Miroc: Média e Exposição ao Risco Ambiental: Baixa:*
  - A média exposição ao risco climático miroc tem o potencial de influenciar reduzindo o risco quando a variável não está em *default* de até 30% e aumentar o risco de até 20%. Com maiores concentrações de redução.
  - A média exposição ao risco climático miroc tem o potencial de influenciar aumentando o risco quando a variável está em *default* de até 10% e reduzir o risco de até 10%. Com maiores concentrações de aumento.

Os resultados indicam que a classificação de baixa exposição ao risco climático da taxonomia verde associada a variável de risco ambiental possui um claro potencial de auxiliar na redução do risco quando a variável não está em *default*, mostrando grandes concentrações da presença da variável ao reduzir o risco quando a variável não está em *default*, mas não mostrou capacidade de reduzir o risco quando a variável entra em *default*. A classificação de média exposição ao risco climático não mostrou potencial ao reduzir o risco quando a variável entra em *default* e manteve suas concentrações nos valores mais altos de risco.

# Capítulo 5

## Considerações Finais

### 5.1 Limitações

#### Base de dados

Os dados não podem ser divulgados em sua composição original por questões de sigilo total das operações. Dessa forma, receberam tratamento com técnicas de processamento para remover ou modificar informações que possam identificar clientes e operações.

Foram encontradas ausências no preenchimento de informação das operações.

Os dados do IBGE precisaram ser tratados para refletirem corretamente os códigos de localização atrelados às operações.

Os dados utilizados para gerar os indicadores precisaram ser tratados para refletirem o mesmo padrão.

#### Indicador de Risco climático

Durante a realização deste estudo, não foram encontradas metodologias pacificadas sobre a medição do risco climático em instituições financeiras, tendo como material disponível somente as metodologias disponibilizadas pela **Febraban**.

Para gerar o indicador de risco climático por meio da **Régua de Sensibilidade ao Risco Climático (RSRC)** era necessário possuir o indicador de risco locacional que é obtido por meio de metodologias baseadas nas informações fornecidas por sites de referência especializados. Porém, até a conclusão deste estudo, não havia ferramenta disponível direcionada para operações em instituições financeiras que realizasse a medição do risco locacional em operações realizadas no Brasil. Desse modo, o estudo utilizou o material disponibilizado **Índice de vulnerabilidade aos desastres naturais relacionados às secas no contexto da mudança do clima (IVDNS)** como indicador de risco locacional.

A metodologia disponibilizada pela Taxonomia Verde foi o único material de relevância focado em operações no Brasil utilizado.

## **Análise de Crédito**

O indicador não pode ser incorporado à metodologia de liberação de crédito. Essa metodologia possui maturidade elevada e não é possível fazer uso do indicador de risco climático como parâmetro para decidir se o crédito será ou não liberado.

## **5.2 Conclusões**

Neste estudo, fornecemos uma visão detalhada da abordagem utilizada para calcular a influência das variáveis de risco climático nas operações de crédito das instituições financeiras. Os resultados das medições de risco climático foram comparados quantitativamente e seus impactos foram estimados por modelos de machine learning. Essas análises foram capazes de fornecer fontes de dados valiosas para tomadas de decisão e fornecem às empresas e organizações uma visão detalhada do comportamento das operações financeiras em relação às variáveis de risco climático e demais variáveis do escopo **Ambiental, Social e Governança (ASG)** que foram utilizadas. O risco climático foi inserido utilizando duas metodologias: a Taxonomia Verde, que incluiu as classificações de risco climático, economia verde e risco ambiental; e a **Régua de Sensibilidade ao Risco Climático (RSRC)** que produziu indicadores de risco climático alternativos. Desse modo, foram incluídos três indicadores de exposição ao risco climático para avaliação e dois indicadores do escopo **ASG** referentes às contribuições ambiental e social, e à exposição ao risco ambiental. Os resultados se dividiram em três partes.

A primeira foi a análise quantitativa do comportamento das operações em relação aos indicadores de risco climático. Foi identificado que a concentração da inadimplência variou de acordo com a classificação de risco climático recebida, sendo que a classificação de alta exposição ao risco climático concentrou a pior inadimplência e o maior volume financeiro. Esse cenário foi observado mesmo que a maioria das operações tenha sido classificada como baixa exposição ao risco climático. Ao associar os resultados com as classificações de economia verde, foi observado que as classificações de alta e moderada contribuição social mais ambiental apresentaram os melhores resultados em quantidade de operações, volume financeiro e inadimplência, enquanto a classificação de alta contribuição somente social apresentou os piores. Esses resultados foram observados nas classificações de alta e média exposição ao risco climático, e não foram replicados na classificação de baixa exposição ao risco climático que apresentou o pior resultado para contribuições sociais mais ambientais e melhores resultados para contribuições somente sociais.

Também foi observada elevada representatividade de operações de soja e bovinos, totalizando 56,15% das operações com 60,09% do volume financeiro total das operações de agronegócio. Essas operações foram quase que em sua totalidade classificadas como sendo de alta exposição ao risco climático, porém, sua inadimplência foi considerada extremamente baixa, sendo de 0,41% em soja e 0,66% em bovinos. Ao serem expostas ao risco da operação e do cliente, foi observado que quanto pior o risco, maior a inadimplência e que as operações estão concentradas nas classificações de risco AA, A, B e C. Ressalta-se que ter o volume financeiro concentrado em operações de alto risco climático exige cautela, pois, apesar das operações de soja e bovinos terem apresentado uma baixa inadimplência, as operações totais com essa classificação apresentaram a maior inadimplência e é preciso lembrar que estamos em um cenário de mudanças climáticas com aumento de desastres relacionados às secas em todo território nacional e que, segundo, [Shackelford et al., 2010] e [Painter, 2020], a maioria dos impostos corretivos e aumento de custos na emissão de títulos está direcionado a externalidades ambientais.

A segunda foi a análise dos riscos envolvidos durante as classificações, em que os riscos foram mapeados e classificados segundo sua capacidade de impactar a produção do indicador. Nesta etapa, foi possível identificar que o indicador ainda precisa amadurecer, pois ainda existem muitas situações capazes de atrapalhar ou inviabilizar sua concepção. Tendo em vista esses desafios, o indicador da Taxonomia Verde foi considerado o mais robusto para as medições de risco climático.

Por fim, a terceira etapa visou identificar um modelo capaz de explicar qual a influência das variáveis de risco climático e demais variáveis de sustentabilidade inseridas ao prever a inadimplência. Foram utilizados os modelos de machine learning: Regressão Logística, Árvore de Decisão, AdaBoost, Random Forest, Gradient Boosting e XGBoost. Desses modelos, os modelos Gradient Boosting e XGBoost apresentaram melhores resultados, sendo que o modelo Gradient Boosting apresentou melhores métricas e o modelo XGBoost apresentou maior área sob a curva ROC. O modelo XGBoost foi escolhido para produzir a explicabilidade do modelo, tendo em vista que os resultados de explicabilidade produzidos pelo modelo Gradient Boosting não foram satisfatórios. Esses resultados mostram que os algoritmos podem apresentar desempenhos semelhantes, mas se diferenciarem em sua implementação, diversidade de funções ou interpretabilidade assim como foi relatado por [James et al., 2013] e [Kuhn et al., 2013].

O método Shapley Value foi utilizado para gerar os resultados de explicabilidade. Com os resultados produzidos pela técnica Shapley Value foi possível concluir que o risco climático é capaz de influenciar o risco das operações de crédito. Assim sendo, a classificação de baixa exposição ao risco climático possui um grande potencial de auxiliar na redução do risco de inadimplência, enquanto classificações mais elevadas mostraram

impactos de favorecimento da inadimplência. Esses resultados também foram associação com as demais variáveis de sustentabilidade inseridas e foi observado que a presença dessas variáveis pode ampliar a influência positiva e reduzir a influência negativa sobre o risco das operações de crédito das instituições financeiras.

Os resultados encontrados corroboram o relatado por [Chava, 2014], ao comentar que o perfil ambiental de uma empresa reflete sobre seu capital e seu risco de inadimplência, e que investidores esperam retornos mais elevados de operações não bem classificadas em pautas ambientais.

Os desafios encontrados para selecionar metodologias capazes de classificar o risco climático vão ao encontro do dito por [Eidelwein et al., 2018], quando relata a escassez de métodos de internalização das externalidades pelas demonstrações de resultado.

Apesar dos resultados observados nas operações agrícolas terem demonstrado majoritariamente classificações de alta exposição ao risco climático, também apresentaram retornos financeiros volumosos, com uma taxa extremamente baixa de inadimplência e altas contribuições sociais, indo ao encontro do dito por [Lv et al., 2010], quando informa que essas operações não possuem só efeitos negativos.

Indo ao encontro do relatado por [Munkhdalai et al., 2019a],[Guégan & Hassani, 2018] e [Dastile et al., 2020], os resultados mostraram a importância das avaliações de risco considerando metodologias atuais de machine learning, por sua capacidade de evoluir dinamicamente com os diversos cenários e por já ser possível encontrar na literatura modelos sofisticados de aprendizagem de máquina que podem substituir os modelos tradicionais como o de regressão logística, assim como foi evidenciado pelos resultados deste estudo.

Os resultados também mostraram que práticas de gestão de sustentabilidade podem ter um efeito positivo ambiental e financeiro para os bancos, ser capaz de mitigar impactos por meio de operações e experimentar um aumento de desempenho financeiro, assim como relatado também foi relatado por [Ali et al., 2020].

Tendo o exposto, conclui-se que o indicador da externalidade representada pelo risco climático possui um potencial capaz de influenciar significativamente a régua econômica das operações ao fornecer valiosas fontes de dados capazes de respaldar a tomada de decisão ao favorecer operações com menos exposição ao risco climático e melhores classificações em pautas ambientais dentro da carteira de crédito. Por isso, incorporar o indicador de exposição ao risco climático aos instrumentos de medição aprimorará os mecanismos de gestão e controle das instituições.

Por fim, devido ao cronograma e objetivos da pesquisa, mais análises não puderam ser feitas. Por isso, para investigar as particularidades encontradas e superar as limitações descritas, mais estudos são recomendados.

# Referências

- [Alexandre et al., 2021] Alexandre, M., Silva, T. C., Connaughton, C., & Rodrigues, F. A. (2021). The drivers of systemic risk in financial networks: a data-driven machine learning analysis. *Chaos, Solitons & Fractals*, 153, 111588. 15
- [Ali et al., 2020] Ali, Q., Salman, A., Yaacob, H., Zaini, Z., & ABDULLAH, R. (2020). Does big data analytics enhance sustainability and financial performance? the case of asean banks. *The Journal of Asian Finance, Economics, and Business*, 7(7), 1–13. 2, 122
- [Allara et al., 2012] Allara, M., Kugbei, S., Dusunceli, F., & Gbehounou, G. (2012). Coping with changes in cropping systems: plant pests and seeds. *Building resilience for adaptation to climate change in the agriculture sector*, 23, 91. 19
- [Antoci et al., 2015] Antoci, A., Borghesi, S., Russu, P., & Ticci, E. (2015). Foreign direct investments, environmental externalities and capital segmentation in a rural economy. *Ecological Economics*, 116, 341–353. 10
- [BACEN, 2021] BACEN (2021). Sistema financeiro nacional (sfm). 18, 19
- [Baesens et al., 2003] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, 54, 627–635. 66
- [Batista et al., 2004] Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1), 20–29. 36
- [Battiston et al., 2017] Battiston, S., Mandel, A., Monasterolo, I., Schütze, F., & Visentin, G. (2017). A climate stress-test of the financial system. *Nature Climate Change*, 7(4), 283–288. 4
- [BBC, 2017] BBC (2017). Mais de 3 décadas após ‘vale da morte’, cubatão volta a lutar contra alta na poluição. 3
- [Blatt, 1999] Blatt, A. (1999). Avaliação de risco e decisão de crédito: um enfoque prático. 22
- [Borges et al., 2009] Borges, L. A. C., de Rezende, J. L. P., & Pereira, J. A. A. (2009). Evolução da legislação ambiental no brasil. *Revista em Agronegócio e Meio Ambiente*, 2(3), 447–466. 2

- [BPBES, 2019] BPBES (2019). Crescem as leis ambientais, mas há graves falhas de implementação, afirma onu. **3**
- [Breiman, 1996] Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123–140. **56**
- [Breiman, 1997] Breiman, L. (1997). *Arcing the edge*. Technical report, Technical Report 486, Statistics Department, University of California at . . . . **59**
- [Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32. **57**
- [Breiman et al., 1984] Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and Regression Trees*. Taylor & Francis. **37**
- [Breiman et al., 2017] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge. **37**
- [Brito & Assaf Neto, 2008] Brito, G. A. S. & Assaf Neto, A. (2008). Modelo de classificação de risco de crédito de empresas. *Revista Contabilidade & Finanças*, 19(46), 18–29. **12, 23**
- [Brito, 2019] Brito, O. (2019). *Mercado financeiro*. Saraiva Educação SA. **23**
- [Brown & Mues, 2012] Brown, I. & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446–3453. **66**
- [Bussmann et al., 2021] Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, 57(1), 203–216. **14**
- [Capelletto & Corrar, 2008] Capelletto, L. R. & Corrar, L. J. (2008). Índices de risco sistêmico para o setor bancário. *Revista Contabilidade & Finanças*, 19, 6–18. **23**
- [Caruana & Niculescu-Mizil, 2006] Caruana, R. & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161–168). **49**
- [Chandrashekar & Sahin, 2014] Chandrashekar, G. & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16–28. **33**
- [Charbuty & Abdulazeez, 2021] Charbuty, B. & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. **48, 55**
- [Chava, 2014] Chava, S. (2014). Environmental externalities and cost of capital. *Management science*, 60(9), 2223–2247. **2, 9, 122**
- [Chawla et al., 2002] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321–357. **35, 36**

- [Chen & Guestrin, 2016] Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). 59
- [Chen et al., 2020] Chen, X., Li, S., Xu, X., Meng, F., & Cao, W. (2020). A novel gsci-based ensemble approach for credit scoring. *IEEE Access*, 8, 222449–222465. 13
- [Coutinho, 1994] Coutinho, G. Luciano; Ferraz, C. J. (1994). *Estudo da competitividade da indústria brasileira. 3. ed.* Editora da Universidade Estadual de Campinas, 1994. 22
- [Dastile et al., 2020] Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, 91, 106263. 14, 122
- [Dong & Peng, 2013] Dong, Y. & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1), 1–17. 34
- [Eckstein et al., 2021] Eckstein, D., Künzel, V., & Schäfer, L. (2021). Global climate risk index 2021. *Who Suffers Most from Extreme Weather Events*, (pp. 2000–2019). 2
- [Eidelwein et al., 2018] Eidelwein, F., Collatto, D. C., Rodrigues, L. H., Lacerda, D. P., & Piran, F. S. (2018). Internalization of environmental externalities: Development of a method for elaborating the statement of economic and environmental results. *Journal of Cleaner Production*, 170, 1316–1327. 9, 122
- [Feng et al., 2018] Feng, X., Xiao, Z., Zhong, B., Qiu, J., & Dong, Y. (2018). Dynamic ensemble classification for credit scoring using soft probability. *Applied Soft Computing*, 65, 139–151. 13
- [FieldView, 2022] FieldView, E. (2022). Qual a participação do agronegócio no pib brasileiro? 3
- [Gitman et al., 2010] Gitman, L. J. et al. (2010). Princípios de administração financeira. 22
- [Gonçalves et al., 2013] Gonçalves, E. B., Gouvêa, M. A., & Mantovani, D. M. N. (2013). Análise de risco de crédito com o uso de regressão logística. *Revista Contemporânea de Contabilidade*, 10(20), 139–160. 24
- [Gouveia Jr, 1999] Gouveia Jr, A. (1999). O conceito de modelo e sua utilização nas ciências do comportamento: breves notas introdutórias. *Estudos de Psicologia (Campinas)*, 16, 13–16. 36
- [Guegan & Hassani, 2018] Guegan, D. & Hassani, B. (2018). Regulatory learning: How to supervise machine learning models? an application to credit scoring. *The Journal of Finance and Data Science*, 4(3), 157–171. 24
- [Guégan & Hassani, 2018] Guégan, D. & Hassani, B. (2018). Regulatory learning: How to supervise machine learning models? an application to credit scoring. *The Journal of Finance and Data Science*, 4(3), 157–171. 12, 49, 122

- [Hoerl & Kennard, 1970] Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. 41
- [Holm & Rikhardsson, 2008] Holm, C. & Rikhardsson, P. (2008). Experienced and novice investors: does environmental information influence investment allocation decisions? *European Accounting Review*, 17(3), 537–557. 2
- [Hosmer Jr et al., 2013] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons. 45
- [Huang et al., 2018] Huang, H. H., Kerstein, J., & Wang, C. (2018). The impact of climate risk on firm performance and financing choices: An international comparison. *Journal of International Business Studies*, 49(5), 633–656. 2
- [James et al., 2013] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer. 33, 121
- [Jordan & Mitchell, 2015] Jordan, M. I. & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. 24
- [Kaggle, 2020] Kaggle (2020). Xgboost. 60
- [Kemp et al., 2022] Kemp, L., Xu, C., Depledge, J., Ebi, K. L., Gibbins, G., Kohler, T. A., Rockström, J., Scheffer, M., Schellnhuber, H. J., Steffen, W., et al. (2022). Climate endgame: Exploring catastrophic climate change scenarios. *Proceedings of the National Academy of Sciences*, 119(34), e2108146119. 85
- [Kitchenham & Charters, 2007] Kitchenham, B. & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. 7
- [Koehrsen, 2018] Koehrsen, W. (2018). Overfitting vs. underfitting: A complete example. *Towards Data Science*. 60
- [Krepesky & Scherer, 2019] Krepesky, G. M. & Scherer, K. R. (2019). O risco climático e o compromisso intergeracional-constitucional. *Revista Direito Ambiental e Sociedade*, 8(3). 2, 20
- [Kuhn et al., 2013] Kuhn, M., Johnson, K., et al. (2013). *Applied predictive modeling*, volume 26. Springer. 33, 121
- [Li, 2016] Li, C. (2016). A gentle introduction to gradient boosting. URL: [http://www.ccs.neu.edu/home/vip/teach/MLcourse/4\\_boosting/slides/gradient\\_boosting.pdf](http://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/slides/gradient_boosting.pdf). 59
- [Lundberg & Lee, 2017] Lundberg, S. M. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30. 67
- [Lv et al., 2010] Lv, Y., zhong Gu, S., & mei Guo, D. (2010). Valuing environmental externalities from rice–wheat farming in the lower reaches of the yangtze river. *Ecological Economics*, 69(7), 1436–1442. Special Section: Ecosystem Services Valuation in China. 9, 122

- [Marengo, 2008] Marengo, J. A. (2008). Água e mudanças climáticas. *Estudos avançados*, 22(63), 83–96. **3**
- [MARION, 2013] MARION, C. V. (2013). A questão ambiental e suas problemáticas atuais: uma visão sistêmica da crise ambiental. In *Congresso Internacional de Direito e Contemporaneidade*, volume 2 (pp. 5–4). **3**
- [Mattmann et al., 2016] Mattmann, M., Logar, I., & Brouwer, R. (2016). Wind power externalities: A meta-analysis. *Ecological Economics*, 127, 23–36. **10**
- [McDonald, 2009] McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 93–100. **41**
- [Mileo et al., 2013] Mileo, R., Kimura, H., & Kayo, E. K. (2013). Análise do modelo creditrisk+ em uma amostra de portfólio de crédito. **12, 23**
- [Mishina et al., 2015] Mishina, Y., Murata, R., Yamauchi, Y., Yamashita, T., & Fujiyoshi, H. (2015). Boosted random forest. *IEICE TRANSACTIONS on Information and Systems*, 98(9), 1630–1636. **57**
- [Mota, 2014] Mota, L. d. A. (2014). Capitalismo contemporâneo: olhares multidisciplinares. *Campina Grande: Eduepb*. **1**
- [Müller & Guido, 2016] Müller, A. C. & Guido, S. (2016). *Introduction to machine learning with Python: a guide for data scientists*. "O'Reilly Media, Inc.". **36**
- [Munkhdalai et al., 2019a] Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J. Y., & Ryu, K. H. (2019a). An empirical comparison of machine-learning methods on bank client credit assessments. *Sustainability*, 11(3). **12, 122**
- [Munkhdalai et al., 2019b] Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J. Y., & Ryu, K. H. (2019b). An empirical comparison of machine-learning methods on bank client credit assessments. *Sustainability*, 11(3), 699. **24, 50**
- [Müller et al., 2001] Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 12, 181–201. **38**
- [Neto & Lima, 2009] Neto, A. A. & Lima, F. G. (2009). *Curso de administração financeira*. Atlas. **23**
- [Onoda, 2001] Onoda, M. (2001). *Estudo sobre um algoritmo de árvores de decisão acoplado a um sistema de banco de dados relacional. 2001. 110p*. PhD thesis, Dissertação (Mestrado)-Universidade Federal do Rio de Janeiro, Rio de Janeiro. **50**
- [ONU, 2015] ONU (2015). Agenda 30. **17**
- [Opitz & Maclin, 1999] Opitz, D. & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, 11, 169–198. **55, 56**
- [Painter, 2020] Painter, M. (2020). An inconvenient cost: The effects of climate change on municipal bonds. *Journal of Financial Economics*, 135(2), 468–482. **1, 121**

- [Paiva, 2003] Paiva, P. R. d. (2003). Contabilidade ambiental: evidenciação dos gastos ambientais com transparência e focada na prevenção. In *Contabilidade ambiental: evidenciação dos gastos ambientais com transparência e focada na prevenção* (pp. 154–154). 1
- [Paul, 2006] Paul, R. K. (2006). Multicollinearity: Causes, effects and remedies. *IASRI, New Delhi*, 1(1), 58–65. 40
- [Phelan et al., 2017] Phelan, A. A., Dawes, L., Costanza, R., & Kubiszewski, I. (2017). Evaluation of social externalities in regional communities affected by coal seam gas projects: A case study from southeast queensland. *Ecological economics*, 131, 300–311. 11
- [Phua et al., 2004] Phua, C., Alahakoon, D., & Lee, V. (2004). Minority report in fraud detection: classification of skewed data. *Acm sigkdd explorations newsletter*, 6(1), 50–59. 35
- [Powers, 2020] Powers, D. M. (2020). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*. 61
- [Provost & Fawcett, 2013] Provost, F. & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. "O'Reilly Media, Inc.". 50, 51, 53
- [Roncoroni et al., 2021] Roncoroni, A., Battiston, S., Escobar-Farfán, L. O., & Martinez-Jaramillo, S. (2021). Climate risk and financial stability in the network of banks and investment funds. *Journal of Financial Stability*, 54, 100870. 4
- [Rubin, 1976] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. 34
- [Salvatore, 2000] Salvatore, D. (2000). *Economia internacional*. Rio de Janeiro: LTC, 2000. 1
- [Sanga & Mungatana, 2016] Sanga, G. & Mungatana, E. (2016). Integrating ecology and economics in understanding responses in securing land-use externalities internalization in water catchments. *Ecological Economics*, 121, 28–39. 10
- [Santos, 2007] Santos, J. O. D. F. R. (2007). Avaliação da aplicabilidade de um modelo de credit scoring com variáveis sistêmicas e não sistêmicas em carteiras de crédito bancário rotativo de pessoas físicas. *Contabilidade Finanças, São Paulo*, 2(44), 105–117. 2
- [Schapire, 2013] Schapire, R. E. (2013). Explaining adaboost. *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, (pp. 37–52). 58
- [Sellare et al., 2020] Sellare, J., Meemken, E.-M., & Qaim, M. (2020). Fairtrade, agro-chemical input use, and effects on human health and the environment. *Ecological economics*, 176, 106718. 11

- [Selvaraju, 2012] Selvaraju, R. (2012). Climate risk assessment and management in agriculture. *Building resilience for adaptation to climate change in the agriculture sector*, 23, 71. 20
- [Shackelford et al., 2010] Shackelford, D. A., Shaviro, D. N., & Slemrod, J. (2010). Taxation and the financial sector. *National Tax Journal*, 63(4), 781–806. 1, 121
- [Shrestha, 2020] Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, 8(2), 39–42. 40
- [Sokolova et al., 2006] Sokolova, M., Japkowicz, N., & Szpakowicz, S. (2006). Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence* (pp. 1015–1021).: Springer. 61
- [Stern, 2006] Stern, N. (2006). The economics of climate change: the stern review. *Cambridge University Press, UK*, (pp. 267). 1
- [Štrumbelj & Kononenko, 2014] Štrumbelj, E. & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41(3), 647–665. 67
- [Su & Zhang, 2006] Su, J. & Zhang, H. (2006). A fast decision tree learning algorithm. In *Aaai*, volume 6 (pp. 500–505). 47
- [TCFD, 2020] TCFD (2020). Climate change presents financial risk to the global economy. 18
- [Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. 42
- [Trueck & Rachev, 2009] Trueck, S. & Rachev, S. T. (2009). *Rating based modeling of credit risk: theory and application of migration matrices*. Academic press. 45
- [Van Dusen, 2016] Van Dusen, C. (2016). Methods to prevent overwriting and solve ill-posed problems in statistics: Ridge regression and lasso. *Preprint submitted to Colorado College Department of Mathematics September*, 16. 40
- [Vasconcellos & Garcia, 2014] Vasconcellos, M. A. S. D. & Garcia, M. E. (2014). *Fundamentos de economia 5. ed.* Editora Saraiva, São Paulo. 1
- [Verdiano et al., 2017] Verdiano, A. d. S., Aeroza, A., Jesus, J., Moura, N., & Machado, C. (2017). Problemáticas da legislação ambiental brasileira. *Jusbrasil, Jusbrasil*. <https://jemersonmoreira.jusbrasil.com.br/artigos/469080712/problematicas-da-legislacaoambiental-brasileira> (Page consultée le 18 mai 2018). 3
- [Visa et al., 2011] Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. *MAICS*, 710, 120–127. 60
- [Wang et al., 2006] Wang, L., Zhu, J., & Zou, H. (2006). The doubly regularized support vector machine. *Statistica Sinica*, (pp. 589–615). 44

- [Wei et al., 2010] Wei, Y., White, R., hu, K., & Willett, I. (2010). Valuing the environmental externalities of oasis farming in left banner, alxa, china. *Ecological Economics*, 69, 2151–2157. 10
- [Weston & Brigham, 2000] Weston, J. F. & Brigham, E. F. (2000). *Fundamentos da administração financeira*. 23
- [Winden et al., 2014] Winden, M., Cruze, N., Haab, T., & Bakshi, B. (2014). Integrating life-cycle assessment and choice analysis for alternative fuel valuation. *Ecological Economics*, 102, 83–93. 11
- [Young et al., 2012] Young, C., Soto, D., Bahri, T., & Brown, D. (2012). *Building resilience for adaptation to climate change in the fisheries and aquaculture sector*, (pp. 346). 4
- [Zhang et al., 2017] Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2017). Efficient knn classification with different numbers of nearest neighbors. *IEEE transactions on neural networks and learning systems*, 29(5), 1774–1785. 34
- [Zhou et al., 2020] Zhou, X. Y., Caldecott, B., Hoepner, A. G., & Wang, Y. (2020). Bank green lending and credit risk: an empirical analysis of china’s green credit policy. *Business Strategy and the Environment*. 15
- [Zhou, 2012] Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*. CRC press. 55
- [Érika Motoda, 2021] Érika Motoda (2021). O que falta para haver mais títulos verdes no mercado. 22