

## Avaliação do Risco de Crédito das Operações do BNDES

### Resumo:

O presente trabalho contribui para as avaliações de risco de crédito em financiamentos do BNDES, ao fazer uso de fontes alternativas de dados e utilizar métodos de aprendizado de máquina na mensuração da probabilidade de inadimplência nas operações do BNDES. Utilizamos uma marcação de status das empresas encontrada nos dados da Receita Federal como *proxy* para o *default* das empresas, e assim treinar os modelos de aprendizado de máquina supervisionados. Desta forma é possível mensurar os níveis de risco de crédito nas operações de financiamento do BNDES. Finalmente, utilizamos o conceito de inteligência artificial interpretável (eXplainable AI) para entender a relação entre cada *feature* e as previsões fornecidas pelo aprendizado de máquina, elucidando como estas técnicas podem aumentar o conforto e a confiança dos usuários, e afastar a designação de “caixa-preta” desses modelos.

**Palavras-Chave:** Risco de Crédito, Probabilidade de Inadimplência, Otimização de Hiperparâmetros, *XGboost*, ELM, *Catboost*.

### Abstract:

The present study contributes to credit risk assessments for BNDES financing by employing alternative data sources and using machine learning methods in measuring *default* rates. We use a company status marker found in the Federal Revenue data as a proxy for company *defaults*, thereby training supervised machine learning models. This allows for the measurement of credit risk levels in BNDES financing operations. Lastly, we utilize the concept of explainable AI (eXplainable Artificial Intelligence) to understand the relationship between each feature and the predictions provided by the machine learning, elucidating how these techniques can enhance user comfort and trust, and moving away from the "black-box" designation of these models.

**Key Words:** Credit Risk, Probability *Default*, Hyperparameters Tunning, *XGboost*, ELM, *Catboost*.

## Resumen:

El presente trabajo contribuye a las evaluaciones de riesgo de crédito en financiamientos del BNDES, al hacer uso de fuentes alternativas de datos y aplicar métodos de aprendizaje automático en la medición de la morosidad. Utilizamos una marca de estado de las empresas encontrada en los datos de la Receita Federal como proxy para el *default* de las empresas, y así entrenar los modelos de aprendizaje automático supervisados. De esta manera, es posible medir los niveles de riesgo de crédito en las operaciones de financiamiento del BNDES. Finalmente, utilizamos el concepto de inteligencia artificial explicativa (eXplainable AI) para entender la relación entre cada característica y las predicciones proporcionadas por el aprendizaje automático, elucidando cómo estas técnicas pueden aumentar la comodidad y la confianza de los usuarios, y alejarse de la designación de "caja negra" de estos modelos.

**Palabras-clave:** Riesgo de Crédito, Probabilidad de Incumplimiento, Optimización de Hiperparámetros, *XGboost*, ELM, *Catboost*.

## 1. Introdução

As imprecisões nas estimativas de risco de crédito das instituições financeiras, podem eventualmente ter contribuído para algumas crises financeiras sistêmicas, tal como a crise do *subprime* (acho que cabe uma nota de rodapé explicando) de 2008 (Altman, 1998.). As recentes experiências das instituições monitorando os efeitos de contágio ocorridos em algumas crises financeiras levaram os bancos a dedicar a alocação de crescentes recursos visando prever o risco de crédito com maior precisão. Este movimento ocorreu em direção à exploração de metodologias estatísticas mais sofisticadas, incorporação de fontes de dados alternativas para complementar os tradicionais bancos de dados e na incorporação de novas tecnologias de armazenamento e processamento de dados visando processar rapidamente conjuntos de dados cada vez maiores e com mais eficiência. Esse movimento de inovação nos procedimentos de classificação de risco crédito, juntamente com a busca por novas fontes de dados modificou alguns paradigmas de mercado na mensuração do risco de crédito. Consequentemente, em resposta às crises econômicas as instituições financeiras foram se afastando das análises subjetivas, em direção a metodologias mais objetivas e sofisticadas, e fazendo crescente o uso do armazenamento de dados em nuvem e explorando novas fontes de informações.

O crédito é necessário para financiar diversas iniciativas de negócios. Permitindo que agentes com “excesso de capital” possam alocar eficientemente seus recursos a agentes com necessidade de capital. A atuação dos bancos também possibilita que diversos correntistas

combinem seus excessos de recursos alocando seu capital coletivamente para grandes projetos. Mais recentemente, na tentativa de espelhar esta forma de atuação dos bancos, credores e empresas de tecnologia financeira (“fintech”) estão utilizando fontes alternativas de dados, e novas formas de análise de risco de crédito para alocação de crédito em plataformas de empréstimo P2P (González-Fernández & González-Velasco; 2020). Essas inovações podem ampliar o acesso ao crédito, especialmente para pessoas com históricos de crédito escasso. No mercado financeiro do consumidor, dados alternativos referem-se a informações usadas para avaliar a qualidade de crédito que geralmente não fazem parte de um relatório de crédito. Alguns exemplos incluem:

- Pagamentos de aluguel.
- Pagamentos por celular.
- Pagamentos de TV a cabo.
- Sites de reputação on-line.
- Métricas de análise de sentimento de mídias sociais.

O uso de dados alternativos pode expandir o acesso ao crédito, onde mesmo os clientes sem histórico de pagamentos podem ter uma avaliação de risco de crédito com precisão, desde que as fontes alternativas apresentem variabilidade. Outra vantagem é disponibilizar escores alternativos de risco de crédito, que podem ser combinados com escores de *bureaus* de crédito tradicionais, adicionando fontes de informação na gestão de risco.

Atualmente alguns credores adotam políticas contratuais que exigem escores de crédito acima de determinado limiar, mas sem restringir que sejam escores de *bureaus* de crédito tradicionais. Alguns desses credores podem estar dispostos a conceder o empréstimo, se encontrarem critérios objetivos confiáveis para determinar quais pessoas têm menos probabilidade de inadimplência no empréstimo, com base em fontes de dados alternativos (González-Fernández & González-Velasco; 2020).

Os dados tradicionalmente usados pelos credores e *bureaus* de crédito, podem não refletir todas as atividades que a empresa se envolve. Dados alternativos podem fornecer informações mais atualizadas e em tempo real. Por exemplo, podem existir cadastros desatualizados ou informações incompletas mesmo em dados tradicionais. Os escores de crédito tradicionais são fortemente influenciados pela própria conduta financeira da pessoa física ou jurídica. E alguns dados alternativos não são tão intimamente relacionados a conduta financeira propriamente dita, fazendo tais fontes de informação interessantes para concessão de crédito com objetivos

sociais ou para financiar projetos de maior risco, tais como linhas voltadas para inovação tecnologia. Tal como González-Fernández & González-Velasco (2020) que avalia risco de crédito utilizando dados do Google, e Giudici et al. (2020) que utiliza dados de plataformas de empréstimos peer-to-peer

Entretanto, o uso de fontes de dados alternativas pode penalizar ou recompensar certos grupos ou comportamentos de maneiras difíceis de prever. Por exemplo, empresas cujos processos de produção utilizam solução em nuvem e que adotaram trabalho remoto universal. Essas empresas podem mudar o endereço de seu estabelecimento sem impactar suas atividades produtivas, e isso pode dar uma falsa impressão de instabilidade que pode afetar o acesso ao crédito.

### **1.1. Escores de Risco de Crédito Contábeis**

Os escores de risco de crédito contábeis (ERCC) são fortemente baseados na comparação de vários índices contábeis de potenciais credores, com os valores médios observados na indústria ou em um cluster de empresas semelhantes. Esses indicadores são combinados em um escore de risco de crédito ou probabilidade de inadimplência. Assim, se o escore ou probabilidade de risco ultrapassar certo limiar de referência, a operação de crédito é rejeitada. Altman *et al.* (1998) identifica quatro abordagens metodológicas no desenvolvimento dos ERCC: (i) o modelo de probabilidade linear, (ii) o modelo *logit*, (iii) o modelo *probit*, e (iv) o modelo de análise discriminante. Entre essas abordagens, destacando-se a análise discriminante e o modelo *logit*. Essas as abordagens utilizam uma função de ligação entre indicadores contábeis e os tomadores de empréstimo inadimplentes e adimplentes.

Os chamados modelos de “risco de ruína” são uma classe de modelos de falência com forte sustentação e utilizados em risco de crédito (Altman *et al.*, 1998). Em sua definição mais simples, a falência ocorre quando o valor de mercado dos ativos é inferior às obrigações de dívidas. A literatura dos modelos de risco de ruína afirma que, se o valor de mercado dos ativos de uma empresa encontra-se abaixo das suas obrigações com dívidas, então em algum momento a empresa irá frustrar tais obrigações.

Os modelos de risco de ruína são encontrados em Wilcox (1973) e Scott (1981). Ainda em Scott (1983) são encontradas muitas semelhanças entre o modelo de risco de ruína e os modelos de precificação de opções (OPM) de Black & Scholes (1973), cuja volatilidade do preço das ações de uma empresa é usada como um *proxy* para se obter variabilidade esperada ou implícita nos valores dos ativos, ou seja, o risco dos ativos.

Uma classe de modelos de risco com forte sustentação teórica utilizam estruturas a termo de *spreads* de rendimentos de títulos corporativos para obter probabilidades implícitas de inadimplência (Jonkhart, 1979; Iben e Litterman, 1989). Esses modelos derivam taxas a termo implícitas em títulos com e sem risco, extraíndo expectativa de inadimplência em diferentes momentos no futuro.

Os chamados modelos de inadimplência e taxa mortalidade (Altman; 1988, 1989) ou modelos de envelhecimento (Asquith *et al.*, 1989), têm as probabilidades de inadimplência derivadas a partir de dados anteriores de inadimplência de títulos de crédito e do tempo até o vencimento. As agências de classificação *Moody's* (1990) e *Standard and Poor's* (1991) adotaram esta abordagem de mortalidade em análises de instrumentos financeiros.

Uma tendência mais recente em risco de crédito é a utilização de aprendizado de máquina. Esta abordagem faz uso de uma rede de relações não-lineares entre as variáveis explicativas, explorando relações latentes que expliquem o risco de crédito. As aplicações mais antigas incluem Altman *et al.* (1994), Coats & Fant (1993) e Turban (1996).

A maior crítica à utilização de inteligência artificial continua sendo a natureza obscura dos algoritmos. Devido à dificuldade de interpretação dos parâmetros e *features* (comumente chamadas de variáveis explicativas), são também referidos como modelos *black box*. Entretanto, recentemente surgiram diversas ferramentas que possibilitam interpretar esses algoritmos, tais como, *Partial Dependence Plot* (Molnar, 2023), valor de SHAP (Lundberg & Lee, 2017), gráfico de Efeitos Acumulados Locais (Apley & Zhu, 2020), *Feature Importance* (Janzing, *et al.*, 2020), e LIME (Ribeiro, *et al.*, 2016).

Avanços recentes no poder de computação, disponibilidade de grandes bancos de dados, e armazenagem em nuvem abriram caminho para expansão da mensuração do risco de crédito orientado por IA. Shi, *et al.* (2022) classificam os algoritmos de aprendizado entre técnicas convencionais, tais como, *k-Nearest Neighbor*, *Random Forest* e *Support Vector Machines*; e redes neurais de aprendizado profundo (*Deep Learning*). A partir da revisão de setenta e seis artigos científicos é mostrado que as abordagens baseadas em aprendizado profundo (*Deep Learning*) superam as técnicas mais convencionais de aprendizado de máquina (SVM, *k-Nearest Neighbor*, *Random Forest*). Por sua vez as técnicas tradicionais de aprendizado de máquina vez superam as técnicas estatísticas na previsão de risco de crédito, tanto em precisão quanto em eficiência.

González-Fernández & González-Velasco (2020) utilizam análise de sentimento para medir o risco de crédito bancário em países europeus, utilizando para isto dados do Google e avaliando um conjunto de palavras-chave relacionadas ao risco de crédito. As classificações de

crédito desse estudo apresentam grande semelhança com escores tradicionais bancários. Especialmente em tempos de instabilidade financeira, o desempenho do escore baseado em análise de sentimento ultrapassa o risco de crédito baseado em escores tradicionais.

Por sua vez, Giudici *et al.* (2020) realizam estimação de risco de crédito em plataformas de empréstimo *Peer-to-Peer* utilizando medidas de centralidade e similaridade derivadas da análise de redes. Os autores indicam indicadores alternativos juntamente com indicadores financeiros tradicionais. Os resultados encontrados corrigem erros de mensuração típicos encontrados em tais plataformas, melhorando a experiência do usuário.

## **2. Fontes de Dados**

A demanda por maior *accountability* nas finanças públicas tem estimulado algumas iniciativas do governo federal em disponibilizar bases de dados abertos, contendo microdados de interesse público. Entre essas iniciativas, podemos citar o Portal do Cadastro Nacional de Pessoas Jurídicas da Receita Federal e o Portal Transparência do BNDES, contendo microdados de empresas. A unidade observacional dos dados disponibilizada é a operação de crédito, contendo o CNPJ censurado das empresas e as condições da operação de crédito (taxa de juros, prazo, amortização, valor do contrato). A unidade observacional das informações da Receita Federal é a matriz empresa, a empresa final e o sócio da empresa, dependendo da tabela relacional. O foco está na concatenação desses dados para que seja desenvolvida uma proxy do risco de crédito das operações financeiras, relacionando esta medida de risco às operações do BNDES.

A Receita Federal disponibiliza 03 tabelas de microdados: Empresa (1), Filiais(2) e Sócios(3). A tabela (1) possui em torno de 40 milhões de registros, já a (2) possui aproximadamente 42 milhões de observações, e a última aproximadamente 18 milhões de observações. As informações de empresas, filiais, sócios, bem como as operações de financiamento do BNDES possuem como chave de ligação o CNPJ da empresa.

### **2.2 Variável Dependente**

Com a publicação das diretrizes de conformidade de Basel II e Basel III, e a necessidade de maior acurácia nas mensurações de risco de crédito, os modelos de análise de sobrevivência (MAS) ganharam maior importância ao longo dos anos (Dirick *et al.*, 2017). Nesse sentido é estimulada a introdução de novas técnicas para mensuração de risco de crédito e a exploração

de novas fontes de dados. Narain (1992) foi a primeira pessoa que sugeriu a utilização de MAS no contexto de risco de crédito. Posteriormente, Dirick *et al.* (2017) avaliam o desempenho destas técnicas para mensuração do risco de crédito. Nesse contexto, o interesse é modelar o período até a inadimplência ocorrer (Thomas *et al.*, 2002). Desde então diversos autores passaram a utilizar os MAS para mensuração de risco de crédito.

Apesar do presente trabalho não utilizar os métodos de análise de sobrevivência, o mesmo se aproximou dessa literatura pela característica da variável dependente. Foi utilizada a variável *status* como uma proxy de inadimplência das empresas com três níveis apenas: 3 – Suspensa\*, 4 – Inapta\* e 08 – Baixada\*, (apresentados na Tabela 1). Esta classificação encontra-se no campo “Situação Cadastral” (SIT\_CAD) da tabela de Filiais da Receita Federal.

**Tabela 1 –Situação Cadastral (SIT\_CAD) da Receita Federal (COD\_MOTIVO)**

<b>Campo</b>	<b>Descrição</b>
SITUAÇÃO CADASTRAL: (SIT_CAD)	01 – NULA;
	02 – ATIVA;
	03 – SUSPENSA*;
	04 – INAPTA*;
	08 – BAIXADA*

As situações cadastrais 03, 04 e 08 foram identificadas como empresas sob suspensão, inaptas ou baixadas na Receita Federal aquelas que apresentavam uma probabilidade de inadimplência dentre aquelas financiadas pelo BNDES.

A relação com a análise de sobrevivência ocorre por meio da possibilidade de falência dessas empresas, e a partir da referência aos “Risco Ruína”, citados na seção 1.1.

**Tabela 2 – Variáveis Explicativas (*Features*) e Dependentes (*Target*) do Modelo de Aprendizado de Máquina para Risco de Crédito**

	Variáveis Explicativas BNDES	Variáveis Explicativas Receita Federal
<p>Situações Cadastrais (SIT_CAD):</p> <p>3 – SUSPENSA</p> <p>4 – INAPTA</p> <p>8 – BAIXADA</p> <p>Em que a data da situação é posterior à data de financiamento, mas não é posterior à data do fim do contrato</p>	<ul style="list-style-type: none"> <li>• Porte Receita</li> <li>• Idade da Firma</li> <li>• Idade da Situação Cadastral</li> <li>• Natureza Jurídica</li> <li>• Número de Sócios</li> <li>• Idade média dos sócios</li> <li>• Estrangeiro Sócio (Dummy)</li> <li>• Qualificação do Sócio Responsável</li> </ul>	<ul style="list-style-type: none"> <li>• Taxa de Juros</li> <li>• Valor Contratado</li> <li>• Valor Desembolsado</li> <li>• Prazo de Amortização</li> <li>• Prazo de Carência</li> <li>• Modalidade de Apoio</li> <li>• Produto</li> <li>• Inovação (Dummy)</li> <li>• Porte BNDES (P, M,G)</li> <li>• UF</li> <li>• CNAE</li> </ul>

### 2.3 Interpretação de Algoritmos de Aprendizado de Máquina

Os modelos “black box” são metodologias em que é possível identificar as suas entradas (*inputs*) e saídas (*outputs*), sendo “obscuros” os processos pelos quais os *inputs* foram transformados em *outputs*. Tal obscuridade está intimamente ligada à complexidade do entendimento da conexão entre cada variável explicativa (*feature*) e as saídas do modelo, sejam elas probabilidades, classificações ou valores preditos.

O grande poder preditivo dos algoritmos de aprendizado de máquina vem acompanhado de maiores dificuldades na sua interpretação, devido a maior complexidade de tais algoritmos, e pela não-linearidade dos vínculos entre as *features* e o valor predito. A interpretabilidade é especialmente importante em instituições financeiras, a qual precisa atender diferentes regulações, incluindo a explicabilidade das metodologias de classificação de risco de crédito. Tal explicabilidade dos modelos de aprendizado de máquina também esbarra em questões éticas e



regulamentares, cujos desdobramentos já começam a ser abordadas, por exemplo, pela Comissão Europeia (2020).

Diante desses desafios, surgiram diferentes metodologias visando esclarecer ao usuário de soluções como uma determinada entrada gerou uma determinada saída, denominadas *eXplainable AI (XAI)*. Diz-se que esse conceito traz interpretabilidade aos resultados gerados do modelo aumentando a confiança das diversas instâncias de decisão das instituições quanto a esses modelos. A *Feature Importance* (Kuhn & Johnson, 2013) é uma metodologia mais intuitiva para interpretação de algoritmos de aprendizado de máquina. Entretanto, também pode ser citada a abordagem LIME e valor-SHAP que têm sido mais reconhecidos como o estado da arte.

Em finanças, a interpretabilidade é também importante para estabelecer a confiança do consumidor na metodologia de avaliação de risco de crédito utilizada. Nesse sentido, Bussmann (2020) e Ariza-Garzón et al. (2020) propõem um modelo XAI baseado em valores de Shapley aplicado no contexto de decisões de empréstimo para pequenas e médias empresas em plataformas de financiamento P2P. Por sua vez, HadjiMisheva et al. (2021) também exploram as estruturas do SHAP e LIME no contexto do gerenciamento de risco de crédito, identificando obstáculos práticos na aplicação dessas técnicas a vários tipos de algoritmos de aprendizado de máquina e propondo algumas soluções.

## LIME

A técnica denominada modelo agnóstico interpretado localmente (LIME) procura aproximar um modelo de *black box* por um modelo local interpretável em cada previsão individual. Ribeiro *et al.* (2016) sugerem que essa abordagem é aplicável em qualquer algoritmo de classificação, uma vez que o LIME é independente do classificador original. O LIME procura ajustar um modelo local usando amostra de dados semelhantes à observação que está sendo explicada, fornecendo explicações para a previsão relativa a cada observação. As explicações fornecidas pelo LIME para cada observação  $X$  são obtidas da seguinte forma:

$$\Phi(x) = \operatorname{argmin}_{g \in G} L(f, g, \pi_x(z)) + \Omega(g)$$

onde,  $G$  é a classe de modelos potencialmente interpretáveis, tais como modelos lineares e árvores de decisão:  $g \in G$ .

$f: \mathbb{R}^d \rightarrow \mathbb{R}$  é a explicação do modelo.

$\pi_x(Z)$ : Medida de proximidade de uma instância  $z$  de  $x$ .

$\Omega(g)$ : Medida de complexidade da explicação de  $g \in G$ .

O objetivo é minimizar a função de perda  $L$  localmente, sem lançar mão de pressupostos a respeito de  $f$ , assegurando a natureza agnóstica de  $L$ . Por sua vez,  $L$  é a medida agnóstica de quão infiel  $g$  é na aproximação de  $f$  na localidade definida por  $\pi_x$ .

## SHAP

Lundberg & Lee (2017) se inspiraram em conceitos de teoria dos jogos ao propor a metodologia SHAP de interpretação de modelos de aprendizado de máquina. A variabilidade das previsões é dividida entre as covariáveis disponíveis, e a contribuição de cada variável explicativa para cada previsão de ponto pode ser avaliada independentemente do modelo subjacente (Joseph, 2019).

A metodologia SHAP (SHapley Additive exPlanation) procura expressar as previsões como se fossem combinações lineares de variáveis binárias, as quais descrevem se determinada covariável está ou não presente no modelo. O SHAP aproxima cada previsão  $f(x)$  por uma função linear das variáveis binárias  $g(x)$ , sendo  $z' \in \{0,1\}$  as variáveis binárias e  $\phi_i \in \mathbb{R}$  as quantidades de explicação, conforme:

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i$$

onde  $M$  é o número de variáveis explicativas.

Somente modelos aditivos que satisfazem as propriedades de precisão local, *missingness* e consistência podem ser obtidos ao atribuir para cada variável  $x'_i$  um efeito  $\phi_i$ , isto é, o valor de Shapley (Scott et al., 2018):

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|! (M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \setminus i)]$$

onde  $f$  é a explicação do modelo,  $x$  são as variáveis disponíveis e  $x'$  são as variáveis selecionadas. A expressão  $f_x(z') - f_x(z' \setminus i)$  demonstrada, para cada predição individual, o desvio dos valores de Shapley de sua média, ou seja, a contribuição da  $i$ -ésima variável.

Intuitivamente, os valores de Shapley são um modelo explicativo que se aproxima localmente do modelo original. Sempre que  $x$  for diferente de zero, o valor de Shapley também será (propriedade *missingness*). Se para um modelo diferente, a contribuição de uma variável para a predição for maior, então o valor de Shapley correspondente também será (propriedade de consistência).

Ambas as metodologias, LIME e SHAP, obtêm os parâmetros para contribuição de cada *feature* em cada observação (explicação local). Entretanto, eles diferem -na construção do algoritmo. A fim de ver qual abordagem é melhor para detectar a contribuição das variáveis no nível local, foi aplicada uma abordagem não supervisionada e foi verificada a possibilidade de agrupamento das observações empregando uma matriz de dissimilaridade construída sobre pesos LIME e valores SHAP, empregando a distância euclidiana padronizada como o base para tal agrupamento.

### ***Feature Importance***

A importância das entradas do modelo (*feature importance*) atribui um escore cada uma das entradas (inputs), com base em sua importância para prever as saídas (outputs). Uma das formas de obter a *feature importance* é por meio do método de permutação de variáveis.

- **Passo 1:** Escolher aleatoriamente uma variável (entrada), ordenar aleatoriamente os valores desta variável, e realizar as previsões utilizando todas as demais *features* originais juntamente com a *feature* embaralhada.
- **Passo 2:** Utilizar uma função de perda para comparar a variabilidade entre as saídas previstas utilizando a *feature* embaralhada e a *feature* original.
- **Passo 3:** Retornar a *feature* para sua forma original, e selecionar aleatoriamente uma nova *feature* para ser embaralhada.
- **Passo 4:** Repete os passos 1 a 3 para todas as entradas do algoritmo.
- **Passo 5:** As importâncias das entradas (*feature importance*) são calculadas comparando o escore da *feature* original em cada um dos sorteios das demais variáveis com o escore médio desta mesma *feature* aleatorizada.

Uma *feature* é dita “sem importância” quando a previsão do modelo permanece inalterada mesmo quando os valores desta *feature* são embaralhados. A *feature importance* foi introduzida inicialmente por Breiman (2001) para o algoritmo *random forest*, sendo posteriormente generalizada por Aaron et al. (2018), os quais elaboraram uma versão “agnóstica” da *feature importance*, a qual os autores denominaram de “confiança no modelo” (*model reliance*).

### **3. Spread Bancário**

Visando o monitoramento de tendências de mercado e da eficiência dos bancos a literatura tem se ocupado em entender quais são os determinantes do *spread* bancário. *Spreads*

bancários muito altos podem refletir um ambiente bancário regulatório inadequado. Uma preocupação é quanto à cobrança de taxas de juros injustas e inacessíveis aos consumidores e empresas (Hawtrey & Liang, 2008). Desta forma, o Banco Nacional de Desenvolvimento Econômico e Social (BNDES) tem manifestado sua preocupação em relação à adequação dos *spreads* cobrados em operações que utilizam recursos subsidiados, seja elas diretas ou indiretas (BNDES, 2018). Da mesma forma, algumas iniciativas do Banco Central do Brasil têm procurado incentivar diminuição dos *spreads* bancários por meio do aumento da competição de mercado (Dantas, *et al.*, 2011).

As equações de formação do *spread* no nível macro de Hawtrey & Liang (2008), consideram a concentração de mercado, aversão ao risco e tamanho do mercado como os principais determinantes do *spread* bancário:

$$s = \frac{\alpha}{\beta} + \frac{1}{2} R \sigma_I^2 Q$$

onde,

$s$ : *spread*

$\frac{\alpha}{\beta}$ : concentração de mercado

$R$ : taxa de aversão ao risco

$\sigma_I^2$ : variância da taxa de juros da economia

$Q$ : Tamanho das transações bancárias.

O aumento da concentração de mercado estaria associado ao aumento do *spread* bancário. O efeito da aversão ao risco ( $R$ ) sobre o *spread* dependeria também dos movimentos simultâneos na volatilidade da taxa de juros da economia ( $\sigma_I^2$ ) e do tamanho do mercado ( $Q$ ). Entretanto, de forma geral, uma maior aversão ao risco resultaria em maiores *spreads* bancários.

Subtraindo a concentração de mercado dos dois lados da equação e aplicando logaritmo natural obtém-se:

$$s - \frac{\alpha}{\beta} = \frac{1}{2} R \sigma_I^2 Q$$

$$\ln\left(s - \frac{\alpha}{\beta}\right) = \ln\left(\frac{1}{2}\right) + \ln(R) + 2 \ln(\sigma_I) + \ln(Q)$$

Considerando a concentração de mercado como um fator fixo, do ponto de vista macroeconômico é possível expressar o *spread* como uma relação linear entre taxa de aversão ao risco, variabilidade da taxa de juros e tamanho do mercado.

$$\ln(s) = \beta_0 + \beta_1 \ln(R) + \beta_2 \ln(\sigma_I) + \beta_3 \ln(Q)$$

A equação acima é uma relação macroeconômica, e mesmo assim já representa uma relativa simplificação dos determinantes do *spread* bancário. Dias & Ichikava (2011) testam muitos outros determinantes, inclusive os índices de bolsa de valores no mercado americano. Dentre todos os indicadores verificados, destaca-se a taxa de inadimplência, que seria capturada pelo índice de aversão ao risco.

Hawtrea & Liang (2008) defendem também que a complexidade no entendimento dos determinantes do *spread* aumentará conforme distancia-se da representação do nível macroeconômico, para o nível mesoeconômico (setorial) e microeconômico (firmas e consumidores).

#### 4. Resultados

Foi apresentada a comparação do treino de três algoritmos de aprendizado de máquina: *Extreme Gradient Boosting (XGboost)*, *Categorical Boosting (Catboost)* e *Extreme Learning Machine (EML)*.

O *Extreme Gradient Boosting (XGboost)* é um algoritmo de aprendizado de reforço de gradiente escalável que ganhou muita projeção ao ser utilizado por muitas equipes vencedoras nas competições de aprendizagem de máquina da comunidade *Kaggle*. Essa biblioteca foi integrada com o *scikit-learn* no *Python*, tornando-se uma implementação eficiente e escalável, utilizada em várias instituições.

*Catboost* é um algoritmo de gradiente *boosting* com tratamento especial para *features* categóricas. Também é amplamente utilizada pela comunidade *Kaggle*, tendo sido listada entre as abordagens mais utilizadas nas competições. Além do tratamento especializado para variáveis categóricas, alta performance em treinamento utilizando GPU, e implementa árvores (*Oblivious*) como forma de execução mais rápida, e para evitar *overfitting*.

*Extreme Learning Machines* são redes neurais retroalimentadas utilizadas em problemas de classificação e regressão. Huang *et al.* (2005) demonstrou que a utilização de uma única camada (*hidden layer*) com uma grande quantidade de nós (*hidden nodes*) apresentava performance mais elevada que redes neurais complexas, com muitas camadas. Entre as vantagens de haver uma única camada, está a capacidade de bom desempenho e generalização. Tem sido

demonstrado que esses modelos podem superar as máquinas de vetores de suporte em performance e velocidade (Huang, *et al.*, 2012).

O sobreajuste (*overfitting*) ocorre quando um algoritmo se ajusta muito bem aos dados de treino, mas a performance observada cai drasticamente em novos conjuntos de dados. O *overfitting* indica que o algoritmo incorporou inclusive os erros de medição e outros fatores aleatórios. Uma forma de se evitar o sobreajuste é ajustando os hiperparâmetros do modelo, mais especificamente os hiperparâmetros relacionados a regularização. Validação cruzada (*cross-validation*) é a forma de verificação de evitar a ocorrência de *overfitting*, onde em cada iteração do treino a métrica de ajuste é maximizada em um conjunto de dados de teste. Tal como demonstrado na Tabela 3, também foram comparadas as métricas de ajuste no conjunto de dados de treino, nos dados de teste (*test data*) e um conjunto de dados que não visto pelos algoritmos em nenhum momento (*Out-of-sample*). Quando as métricas de ajuste se encontram muito próximas nos conjuntos de dados de treino, teste e validação, então conclui-se que não existem problemas de *overfitting*.

*Catboost* e *XGboost* foram os modelos que apresentaram os melhores desempenhos. Os dois algoritmos apresentaram também desempenhos muito parecidos. Dependendo da métrica avaliada, pode-se considerar que o algoritmo *XGboost* apresentou problemas de *overfitting*.

O valor das métricas Acurácia, Escore F1, Recall e Precisão dependem do limiar de ajuste para se definir que a classificação predita é igual a 1, enquanto as métricas KS e AUC não dependem desse limiar. Por padrão da biblioteca scikit-learn, esse limiar é de 50%, sendo possível alterar esse valor para um limiar que melhor discrimine a variável dependente. Esse limiar pode alterar o valor das métricas de ajuste, e idealmente é definido avaliando os percentis de probabilidade nos grupos 0 e 1, conforme avaliação da Tabela 4.

**Tabela 3 – Métricas de Precisão: Validação Cruzada**

		<i>Catboost</i>	ELM	<i>XGboost</i>
Treino	Acurácia	0,9574	0,9285	0,8587
	Escore F1	0,5806	0,1135	0,4926
	Recall	0,4175	0,0644	0,9643
	Precisão	0,9528	0,4785	0,3308
	Escore KS	0,7080	0,7790	0,8176
	AUC	0,9574	0,5295	0,9409

Teste	Acurácia	0,9325	0,9284	0,8563
	Escore F1	0,2455	0,1098	0,4870
	Recall	0,1556	0,0628	0,9707
	Precisão	0,5817	0,4332	0,3250
	Escore KS	0,5736	0,7795	0,8186
	AUC	0,9325	0,5283	0,9387
Out-of-sample (Validação)	Acurácia	0,9313	0,9286	0,8587
	Recall	0,1446	0,0662	0,9683
	Precisão	0,5526	0,4035	0,3250
	Escore KS	0,5678	0,7712	0,8199
	AUC	0,9313	0,9184	0,9391

Conforme apresentado na Tabela 3, todos os modelos apresentam bons resultados. Particularmente, o algoritmo *XGboost* apresentou performance superior em todas as métricas de ajuste. As métricas AUC e KS utilizam a própria distribuição de probabilidade de inadimplência. Portanto, são as duas métricas que não dependem da definição de um limiar.

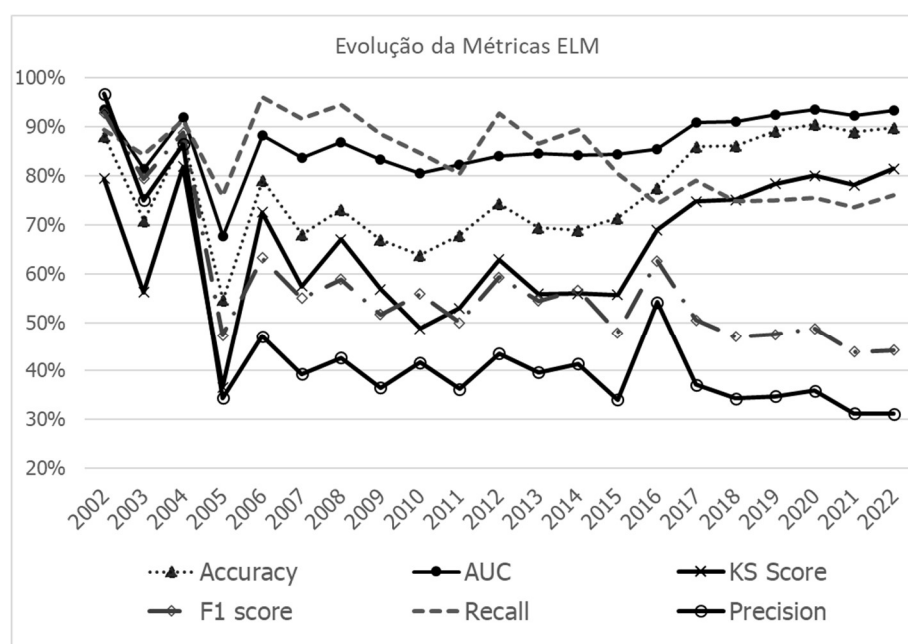
Os algoritmos ELM e *XGboost* não apresentam sobreajuste aos dados. As métricas de performance são muito semelhantes nas três partições do conjunto de dados: treino, teste e validação. Em relação ao sobreajuste, o algoritmo *Catboost* apresentou boa performance quando foram avaliadas as métricas Acurácia e AUC. Mas quando as métricas Recall, Precisão e KS foram avaliadas, verificou-se que a performance do modelo diminui da partição de teste dos dados em relação às partições de treino e validação. Uma forma de eliminar o sobreajuste do *Catboost* é realizar ajustes limitando o espaço de hiperparâmetros, e adicionalmente modificar a métrica de avaliação do modelo em cada interação (não seria interação???) da etapa de otimização dos parâmetros. No algoritmo *Catboost* a métrica utilizada na etapa de otimização de hiperparâmetros bayesiana foi a métrica AUC. Curiosamente, foi exatamente a mesma métrica que não apresentou problemas de sobreajuste no *Catboost*. Assim, entende-se que utilizando uma abordagem de otimização bayesiana utilizando multi-métrica, poder-se-ia eliminar os problemas de sobreajuste em todas as métricas. Entretanto, para o presente exercício considerou-se o esforço de ajuste suficiente. Já foram obtidos os resultados dos algoritmos ELM e *XGboost*, os quais possuem boa performance e não apresentam sobreajuste.

**Tabela 4 – Distribuição das Probabilidade de Inadimplência**

	Target	Contagem	Média	Desvio- Padrão	Percentil 25%	Percentil 50%	Percentil 75%
<i>Catboost</i>	0	116.654	0,036	0,093	0,0001	0,001	0,005
	1	8.862	0,415	0,197	0,2828	0,404	0,540
ELM	0	116.654	0,130	0,081	0,098	0,101	0,126
	1	8.862	0,292	0,248	0,203	0,248	0,318
<i>XGboost</i>	0	116.654	0,151	0,197	0,007	0,012	0,063
	1	8862	0,844	0,470	0,843	0,879	0,903

A Tabela 4 permite avaliar qual limiar de probabilidade é ideal para discriminação entre os grupos 0 e 1. Em geral, uma boa prática é considerar o percentil 25% para os grupos com maior probabilidade média, e o percentil 75% para o grupo com menor probabilidade média. Não existe nenhum argumento teórico ou prático que determine que este limiar idealmente é 50%. Também não há nenhum argumento teórico que defina que os percentis das distribuições de probabilidades devam ser parecidos. O que estamos procurando é determinar limiares de probabilidades onde o poder discriminatório dos algoritmos seja máximo.

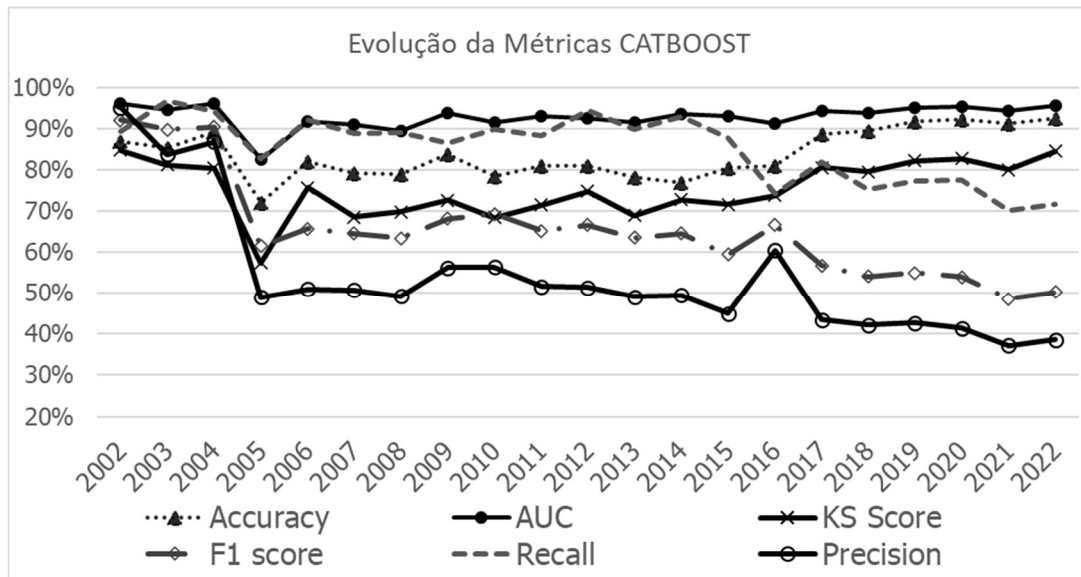
**Figura 1 – Evolução das Métricas do *Extreme Learning Machine* (ELM)**



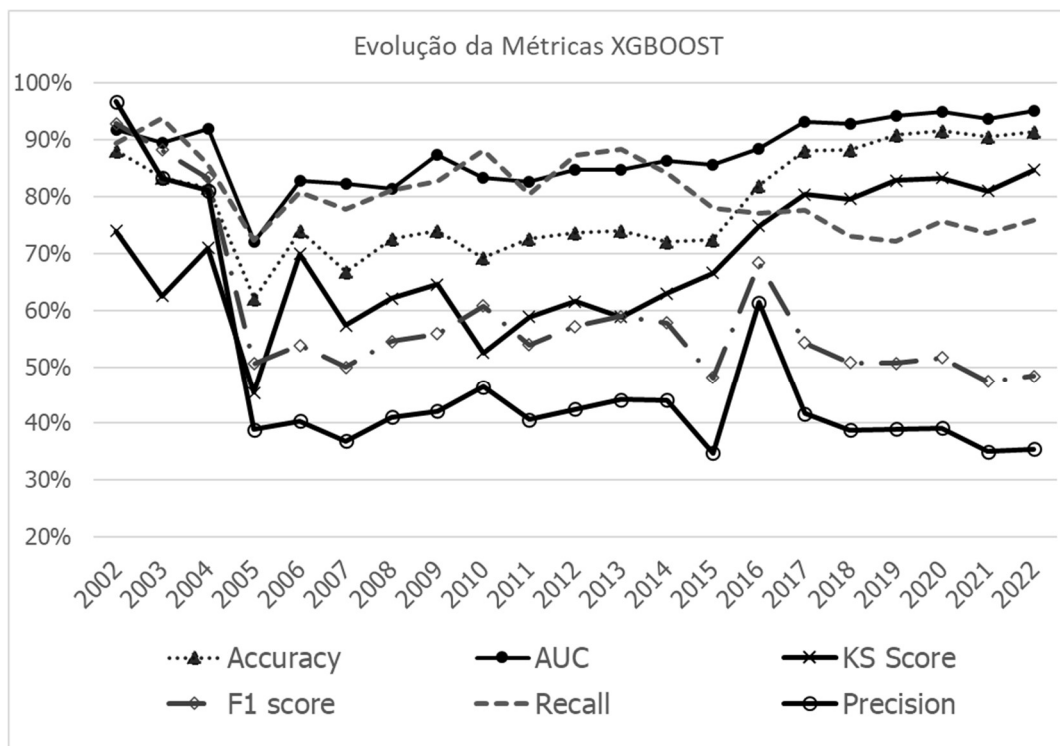


Para o algoritmo *Catboost* , algum valor abaixo de 28% discrimina 75% dos casos em que a inadimplência é zero de 75% dos casos em que a inadimplência. Para o algoritmo ELM esse limiar seria 20%, enquanto para o algoritmo *XGboost* este limiar seria 84%.

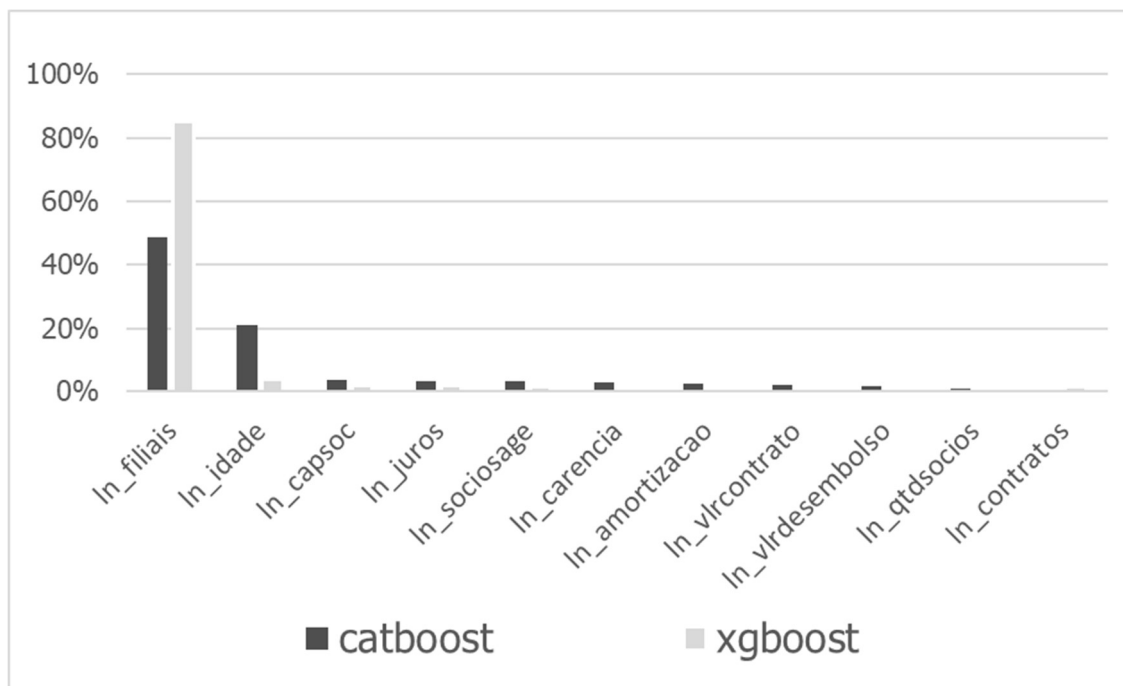
**Figura 2 – Evolução das Métricas do Categorical Boosting (*CATBOOST*)**



**Figura 3 – Evolução das Métricas do Categorical Boosting (*XGBOOST*)**



**Figura 4 – Importância das Características do *Catboost* e *XGboost***



Avaliando a interpretabilidade dos algoritmos *Catboost* e *XGboost*, verifica-se que o número de filiais é a *feature* (característica) mais importante dos dois algoritmos, seguida da idade da empresa, capital social e a taxa de juros cobrada. Isto não significa que essas características afetem a probabilidade de *default* de forma linear, pois esses modelos exploram características de não-linearidade presentes nos dados. Como não há o pressuposto de linearidade, não se pode afirmar que um aumento do número de filiais leve a um aumento ou diminuição na probabilidade de inadimplência, pois uma modificação no número de filiais iria interagir com todas as demais variáveis do modelo ao produzir uma probabilidade de inadimplência.

**Tabela 5 – Classificação de Risco utilizando as Probabilidade de Inadimplência do *XGboost***

Classificação de Risco	Percentil	Prob.	Limite Inferior	Limite Superior	Freq.	Freq. Acumulada	% Acumulado
AA	1,0%	0,27%	0,00%	0,27%	1.050	1.050	0,8%
A	2,5%	0,30%	0,27%	0,30%	3.010	4.060	3,2%
B	5,0%	0,41%	0,30%	0,41%	1.968	6.028	4,8%
C	10,0%	0,48%	0,41%	0,48%	25.266	31.294	24,9%
D	25,0%	0,74%	0,48%	0,74%	13.308	44.602	35,5%
E	35,0%	0,87%	0,74%	0,87%	18.384	62.986	50,2%
F	50,0%	1,31%	0,87%	1,31%	31.151	94.137	75,0%
G	75,0%	13,70%	1,31%	13,70%	12.552	106.689	85,0%
H	85,0%	82,61%	100,00%	82,61%	18.827	125.516	100,0%

Pode-se utilizar os percentis das probabilidades preditas para reproduzir a classificação de risco do Banco Central. Como há o conhecimento que somente um pequeno percentual das empresas possui classificação de risco AA, e aproximadamente 5% das empresas possuem classificação de risco até 5%, pode-se intuir para todas as categorias de risco do Banco Central, conforme é apresentado na Tabela 5.

## Conclusão

Neste trabalho foi utilizado o aprendizado de máquinas para classificar em níveis de risco crédito as empresas financiadas pelo BNDES. Algumas das ferramentas de aprendizado de máquina foram apresentadas, onde exemplificamos algumas possíveis contribuições dessas ferramentas na mensuração do risco de crédito. Também mostramos que diferentes fontes de dados abertas e públicas podem servir para mensuração de risco de crédito. Foi utilizada a classificação de status das empresas encontrada nos dados da Receita Federal como *proxy* para o *default* das empresas, e assim treinar os modelos de aprendizado de máquina supervisionados. Desta forma é possível mensurar os níveis de risco de crédito nas operações de financiamento do BNDES.

## Referências:

AARON, F.; CYNTHIA, R.; DOMINICI, F. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Disponível em: <<http://arxiv.org/abs/1801.01489>>. Acesso em: [20 de outubro de 2023].

ALTMAN, E.; SAUNDERS, A. Credit Risk Measurement: Developments over the last 20 years. *Journal of Banking & Finance*, v. 21, p. 1721-1742, 1998.

APLEY, D.; ZHU, J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *Journal of the Royal Statistical Society: Series B*, v. 82, p. 1059-1086, 2020.

BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. *Journal of Political Economy*. v. 8, p. 637-659, 1973. Disponível em: <<http://dx.doi.org/10.1086/260062>>. Acesso em: [20 de outubro de 2023].

BREIMAN, L. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5-32, Springer, 2001.

BROCHU, E.; VLAD, C.; FREITAS, N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. *CoRR*, 2010.

BNDES. BNDES promove nova redução de *spreads* e de custo dos financiamentos. 2018. Disponível em: <[www.bndes.gov.br/wps/portal/site/home/imprensa/noticias/conteudo/bndes-promove-nova-reducao-de-spreads-e-de-custo-dos-financiamentos](http://www.bndes.gov.br/wps/portal/site/home/imprensa/noticias/conteudo/bndes-promove-nova-reducao-de-spreads-e-de-custo-dos-financiamentos)>. Acesso em: [20 de outubro de 2023].

COATS, P.; FANT, L. Recognizing Financial Distress Patterns Using a Neural Network Tool. *Financial Management Association*, v. 22, n. 3, p. 142-155, 1993. Disponível em: <<https://ideas.repec.org/a/fma/fmanag/coats93.html>>. Acesso em: [20 de outubro de 2023].

DANTAS, J. et al. Determinantes do *Spread* Bancário Ex-Post no Mercado Brasileiro. Banco Central do Brasil, Texto para Discussão 242, 2011.

DIAS, V.; ICHIKAVA, E. Uma análise empírica da relação entre *spread* e risco. Revista do BNDES, n. 36, 2011.

DIRICK, L.; CLAESKENS, C.; BAESENS, B. Time to *Default* in Credit Scoring Using Survival Analysis: A benchmark study. Journal of The Operational Research Society, v. 68, p. 652-665.

EUROPEAN UNION. European Parliament resolution of 20 October 2020 on a framework of ethical aspects of artificial intelligence, robotics and related technologies. 2020/2012(INL), 2020.

GONZÁLEZ-FERNÁNDEZ, M.; GONZÁLEZ-VELASCO, C. An alternative approach to predicting bank credit risk in Europe with Google data. Finance Research Letters, v. 35, 101281, 2020. Disponível em: <<https://doi.org/10.1016/j.frl.2019.08.029>>. Acesso em: [data de acesso].

GIUDICI, P.; HADJI-MISHEVA, B.; SPELTA, A. Network based credit risk models. v. 32, n. 2, p. 199-211, 2020. Disponível em: <<https://doi.org/10.1080/08982112.2019.1655159>>. Acesso em: [data de acesso].

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning. Data Mining, Inference and Prediction. Springer Series in Statistics, 2.ed., 2001.

HAWTREY, K.; LIANG, H. Bank interest margins in OECD countries. The North American Journal of Economics and Finance, v. 19, n. 3, p. 249-260, 2008.

Huang, G.; Zhu, Q; Siew, C. *Extreme Learning Machine: Theory and Applications*. Neurocomputing, v. 70, n. 1, p. 489–501, 2006. doi:10.1016/j.neucom.2005.12.126.

HUANG, G.; ZHU, Q; SIEW, C. *Extreme Learning Machine* for Regression and Multiclass Classification. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, v. 42, n. 2, p. 513–529, 2012. doi:10.1109/tsmcb.2011.2168604.

IBEN, T.; LITTERMAN, R. Corporate bond valuation and the term structure of credit *spreads*. Journal of Portfolio Management, 1989.

JONKHART, M. On the term structure of interest rates and the risk of *default*. Journal of Banking and Finance, 1979.

JOSEPH, A. Shapley Regressions: A Framework for Statistical Inference on Machine Learning Models. 2019. Disponível em: <<https://www.kcl.ac.uk/business/assets/pdf/dafm-working-papers/2019-papers/shapley-regressions-a-framework-for-statistical-inference-on-machine-learning-models.pdf>>. Acesso em: [data de acesso].

LUNDBERG, M.; LEE, S-I. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems 30, p. 4766-4775, 2017. Disponível em: <<https://arxiv.org/abs/1705.07874>>. Acesso em: [20 de outubro de 2023].

MOODY'S SPECIAL REPORT. Corporate Bond *Defaults* and *Default* Rates, p. 1970-1989, April, 1990.

MOLNAR, C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Lulu.com, 2023. ISBN: 978-0244768522. Disponível em: <<https://christophm.github.io/interpretable-ml-book/lime.html#fn50>>. Acesso em: [data de acesso].

NARAIN, B. Survival analysis and the credit granting decision. In: THOMAS, L. C.; CROOK, J. N.; EDELMAN, D. B. (eds.). Credit Scoring and Credit Control. Clarendon Press: Oxford, 1992. p. 109-121.

RIBEIRO, M.; SAMEER, S.; CARLOS, G. Why Should I Trust You? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, 2016. Disponível em: <https://arxiv.org/abs/1602.04938>.

SCOTT, L.; GABRIEL, E.; SU-IN, L. Consistent Individualized Feature Attribution for Tree Ensembles, 2018. Disponível em: <https://arxiv.org/pdf/1802.03888.pdf>.

SHI, S. et al. Machine Learning-Driven Credit Risk: A systemic review. *Neural Computing and Applications*, v. 34, p. 14327–14339, 2022. Disponível em: <https://link.springer.com/article/10.1007/s00521-022-07472-2>.

SNOEK, J.; LAROCHELLE, H.; ADAMS, R. Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems*, v. 25, p. 2960-2968, 2012. Disponível em: <https://arxiv.org/abs/1206.2944>.

THOMAS, L.; EDELMAN, D.; CROOK, J. Credit Scoring and Its Applications. *Monographs on Mathematical Modeling and Computation*. Society for Industrial and Applied Mathematics: Philadelphia, 2002.

KUHN, M.; JOHNSON, K. *Applied Predictive Modeling*. 1ª ed., 2013.