

**PREVISÃO DE DIFICULDADE FINANCEIRA COM XGBOOST: Evidências no
Setor de Utilidade Pública da América Latina**

JOÃO PEDRO SILVA MENDES

UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

DANIEL VITOR TARTARI GARRUTI

UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

FLAVIO LUIZ DE MORAES BARBOZA

UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)

Agradecimento à órgão de fomento:

CAPES processo: 88887.823303/2023-00

PREVISÃO DE DIFICULDADE FINANCEIRA COM XGBOOST: Evidências no Setor de Utilidade Pública da América Latina

1 Introdução

O mercado financeiro latino-americano tem ganhado destaque global, com o Brasil possuindo um dos maiores mercados de derivativos do mundo e o México sendo uma das maiores economias globais (Barboza; Altman, 2024). No entanto, a região é caracterizada por alta volatilidade e riscos nos mercados de ações, com muitas de suas economias passando por períodos de instabilidade financeira e crises econômicas. Analisar o desempenho financeiro neste contexto é crucial para investidores e gestores, dada a importância de eventos negativos como inadimplência e Dificuldade Financeira (FD) (Cardona et al., 2017).

Dentro desse cenário, o setor de utilidade pública representa uma parcela importante, embora nem sempre a maior, do Produto Interno Bruto das nações, e se caracteriza pela essencialidade do serviço prestado para a população como distribuição de água, energia elétrica e gás natural (Kempenich, 2009). Assim, a relevância que o setor tem na produtividade da economia em geral torna a previsão da FD de particular importância.

A FD pode ser entendida como a incapacidade de uma empresa em saldar dívidas com seus credores, associando-se a problemas de inadimplência e fluxo de caixa (Beaver, 1966). Empresas em situação de estresse financeiro tendem a reduzir drasticamente seu valor de mercado, enfrentando perda de clientes e fornecedores, aumento nas taxas de juros, elevação dos custos de oportunidade de projetos e menor produtividade dos empregados (Canton et al., 2021). Esses eventos podem levar a crises financeiras, manifestando-se em problemas sociais como recessão econômica e aumento do desemprego (Huang; Yen, 2019).

Nesse contexto, previsão de FD em organizações é um tema de crescente relevância na literatura acadêmica, sendo tarefa impactante para diversos profissionais que precisam de ferramentas confiáveis que gerem alertas precoces.

Dentre esses profissionais, há: Analistas de Segurança, que precisam definir riscos de crédito; Instituições Reguladoras como bancos centrais para avaliar a qualidade de crédito dos portfólios bancários e monitorar as avaliações feitas por bancos individuais; Auditores que precisam avaliar a condição de continuidade de uma empresa; Advogados de Falência que precisam auxiliar na decisão e momento do pedido de falência ou reestruturação extrajudicial; Investidores que buscam evitar empresas que podem ir à falência, dentre outros (Altman, 2018).

Historicamente, a previsão de FD tem sido abordada por diversos modelos estatísticos e matemáticos, desde análises discriminantes múltiplas e regressões logísticas (LR) (Canton et al., 2021). Contudo, fenômenos financeiros são naturalmente complexos e muitas vezes apresentam aspectos de comportamento não lineares, especialmente na América Latina, o que pode diminuir a qualidade de técnicas tradicionais (Barboza; Altman, 2024).

Diante disso, a aplicação de modelos de *machine learning* (ML) tem-se mostrado promissora na previsão de FDs, superando frequentemente as abordagens tradicionais (Barboza e Altman, 2024). Entre esses modelos, o *Extreme Gradient Boosting* (XGBoost) tem-se destacado por sua capacidade de oferecer alta precisão na previsão de insolvência e é particularmente adequado para lidar com a complexidade e não-linearidade dos dados financeiros, o que o torna uma ferramenta robusta e valiosa para prever esse cenário em empresas (Huang; Yen, 2019; Yuan et al., 2021).

Além da capacidade preditiva, é igualmente importante entender quais variáveis financeiras são os principais impulsionadores desses eventos e como elas influenciam o desempenho de mercado. Assim, modelos baseados em árvores, como o XGBoost, são ferramentas úteis, pois permitem a avaliação da importância das variáveis, fornecendo explicações razoáveis e visíveis sobre as características de risco importantes e as formas como elas afetam a dificuldade financeira (Yuan et al., 2021).

Assim, este artigo tem como objetivo contribuir para a literatura ao explorar o poder preditivo da técnica XGBoost na previsão de FD em empresas latino-americanas do setor de utilidade pública negociadas em bolsa, além de identificar e interpretar as principais variáveis financeiras que sinalizam esse evento. Com isso, buscamos promover uma compreensão mais refinada dos determinantes contábeis e de mercado do risco financeiro, ampliando o entendimento sobre a aplicabilidade de modelos de ML em ambientes econômicos instáveis e em setores específicos. Do ponto de vista prático, os resultados oferecem suporte para analistas financeiros, investidores, reguladores e profissionais de auditoria, ao disponibilizar uma ferramenta preditiva robusta e interpretável, capaz de gerar alertas antecipados sobre deterioração financeira e auxiliar na tomada de decisão estratégica.

Além desta introdução, o presente artigo foi estruturado da seguinte forma: a seção 2 descreve os principais trabalhos relacionados; a seção 3 descreve os dados e a metodologia usada; a seção 4 discute os resultados e a seção 5 traz a conclusão.

2 Referencial Teórico

A dificuldade financeira (do inglês *financial distress*) pode ser compreendida como uma situação em que uma empresa enfrenta sérios problemas para cumprir suas obrigações financeiras, como o pagamento de dívidas, salários e fornecedores. Trata-se de um estágio anterior à falência, caracterizado pela deterioração da saúde financeira da organização (Altman, 1968).

Altman (1968) foi um dos pioneiros na modelagem da previsão de falência por meio de indicadores contábeis, ao propor o modelo Z-score. Este modelo combina cinco índices financeiros para estimar a probabilidade de insolvência: capital de giro sobre ativos totais (X1A), lucros retidos sobre ativos totais (X2A), lucros antes de juros e impostos sobre ativos totais (X3A), valor de mercado do capital próprio sobre a dívida total (X4A) e vendas sobre ativos totais (X5A). Tais índices são utilizados no presente trabalho e mostraram ter alta relevância preditiva.

Outro método clássico é a Regressão Logística, proposta por Ohlson (1980). Este modelo é baseado em uma função logística que estima a probabilidade de uma empresa entrar em falência com base em indicadores como alavancagem, liquidez e desempenho operacional. Apesar do avanço que esse método representa, o modelo possui limitações no tratamento de padrões não-lineares (e.g., Chen, 2011; Geng et al., 2015; Gepp; Kumar, 2015; Huang; Yen, 2019).

Com os avanços em *machine learning* (ML), novos métodos passaram a ser aplicados na previsão de dificuldade financeira, apresentando desempenho superior aos modelos tradicionais. Técnicas como *Random Forest* (RF) e *Extreme Gradient Boosting* (XGBoost) destacam-se pela capacidade de capturar relações não-lineares, lidar com bases de dados desbalanceadas e identificar a importância relativa das variáveis preditoras (Barboza; Altman, 2024; Yang et al., 2021).

Neste contexto, Huang e Yen (2019) buscaram rever a vasta literatura recente sobre abordagens de ML para previsão de dificuldade financeira, comparando seis algoritmos distintos em empresas taiwanesas de capital aberto. Dentre os modelos supervisionados, o XGBoost apresentou o melhor desempenho preditivo, com acurácia acima de 90% usando dados de 2 ou 4 trimestres antes do evento.

Yang et al. (2021), por sua vez, destacaram a importância da explicabilidade nos modelos preditivos. Utilizando o XGBoost em conjunto com o *framework* SHAP (*SHapley Additive exPlanations*), os autores conseguiram interpretar como cada variável influencia na probabilidade de FD, agregando valor diagnóstico ao processo de previsão.

Diversos estudos buscaram adaptar e testar modelos de previsão em contextos específicos, como o latino-americano, cujas economias apresentam alta volatilidade e

instabilidade institucional. O estudo de Barboza e Altman (2024), por exemplo, comparou a performance de modelos de LR com RF na predição de FD em empresas da América Latina. Os resultados mostraram que o RF apresentou menor taxa de erro tipo II e maior capacidade preditiva, sobretudo em horizontes de um a dois anos antes da ocorrência da dificuldade financeira.

Por sua vez, Canton et al. (2021) investigaram os efeitos de eventos de dificuldade financeira no desempenho dos mercados latino-americanos, destacando a relevância de indicadores regionais e contextuais. Ambos os estudos reforçam a necessidade de considerar as particularidades socioeconômicas da região ao aplicar modelos preditivos.

Modelos baseados em árvores de decisão têm ganhado destaque tanto pelo desempenho quanto pela capacidade de interpretação. Chen (2011) comparou algoritmos de árvores de decisão com LR em empresas taiwanesas e observou melhor desempenho das árvores no curto prazo, enquanto a regressão apresentou melhores resultados no longo prazo.

Gepp e Kumar (2015) realizaram uma comparação entre árvores de decisão (CART) e abordagens tradicionais amplamente utilizadas na literatura, como a Análise Discriminante (DA) e a LR, concluindo que as árvores apresentaram superioridade em termos de acurácia de classificação na maioria dos cenários. Os autores também identificaram que os indicadores de lucratividade são os preditores mais relevantes, seguidos por variações anuais de indicadores financeiros, evidenciando que tanto os valores atuais quanto suas mudanças ao longo do tempo agregam valor preditivo.

Esses resultados reforçam a utilidade de modelos baseados em árvores para identificar e interpretar sinais precoces de deterioração financeira, principalmente em contextos com alta variabilidade e múltiplas fontes de risco.

3 Dados e Procedimentos Metodológicos

Os dados coletados, assim como os critérios para definir se uma empresa é saudável ou se está em estresse financeiro, são os mesmos usados por Barboza e Altman (2024). Porém, para o presente estudo, a base de dados foi filtrada para incluir exclusivamente empresas do setor de utilidade pública, utilizando a classificação setorial da Thomson Reuters. Estudos prévios, como Barboza e Altman (2024), demonstraram a eficácia de modelos preditivos de dificuldade financeira em amostras amplas e setorialmente diversificadas. Contudo, ainda se observa uma lacuna quanto à avaliação isolada de setores específicos, como o de utilidade pública. Mais que isso, a literatura aponta que diferentes setores econômicos reagem de modo distinto a choques de mercado (Cardona et al., 2017), o que reforça a importância de investigar setores individualmente.

Os dados foram coletados no repositório “Data and Code (in R) for predicting Financial Distress in Latin America” no Harvard Dataverse (Barboza, 2023) e incluem dados oriundos de demonstrações financeiras (Balanço Patrimonial e Resultados do Exercício), entre 2000 e 2020, de 808 empresas em 6 países latino-americanos, quais sejam: Argentina, Brasil, Chile, Colômbia, México e Peru, totalizando 10.118 observações (empresa-ano) na base.

Ao selecionar apenas as empresas do setor de utilidade pública, temos 1.498 observações (empresa-ano), entre 2000 e 2019, de 103 empresas dos mesmos 6 países, conforme apresentado na Tabela 1. Outros países latino-americanos não foram incluídos devido à falta de dados suficientes.

Pindado et al. (2008) propõe uma definição de FD baseada em mercado e outra baseada em dados financeiros. Para os autores, uma empresa é considerada em dificuldade financeira se pelo menos uma das seguintes condições for atendida: seu Lucro Antes de Juros, Impostos, Depreciação e Amortização (EBITDA) for inferior às suas despesas financeiras por dois anos consecutivos; o valor de mercado da empresa diminuir dentro de um intervalo de dois anos.

TABELA 1 - Total de empresas do setor de utilidade pública e observações por país (2000 – 2019)

País	Número de empresas	Observações
Argentina	16	232
Brasil	44	673
Chile	19	276
Colômbia	7	84
México	2	29
Peru	15	204
TOTAL	103	1.498

Fonte: Dados da pesquisa.

Seguindo Barboza e Altman (2024), no presente estudo aplicou-se essas duas definições e 4 tipos de FD, sendo eles: o critério financeiro (FD1) e o de mercado (FD2) propostos por Pindado et al. (2008), ao menos um dos 2 critérios satisfeitos (FD3) e a intersecção entre os 2 (FD4). Estas definições estão expostas na Tabela 2.

TABELA 2 - Tipos de Dificuldade Financeira

Tipo de FD	Definição	Empresas-ano
FD1	EBITDA < despesas financeiras por 2 anos consecutivos	44
FD2	Valor de mercado cai entre 2 anos	516
FD3	FD1 ou FD2	540
FD4	FD1 e FD2	20

Fonte: Dados da pesquisa.

Foram usadas as mesmas variáveis independentes do estudo de Barboza e Altman (2024), com exceção da variável Sector (Setor, justamente pela escolha metodológica), que foram selecionadas considerando a relevância em estudos anteriores e sua disponibilidade na base de dados. A Tabela 3 apresenta as variáveis usadas e suas definições.

TABELA 3 - Variáveis independentes e suas descrições

Variável	Descrição
Country	País onde localiza-se a sede da empresa
CPB	Variação do Preço/Patrimônio Líquido entre dois anos consecutivos
CROE	Variação do Retorno sobre o Patrimônio Líquido entre dois anos consecutivos
DTE	Dívida Total sobre o Patrimônio Total
EPS	Lucro por Ação
GSA	Crescimento das Vendas entre dois anos consecutivos
LIQ	Capital de Giro sobre Passivos Atuais
LVC	Composição da Alavancagem: Passivos Atuais sobre Dívida Total
NICA	Lucro Líquido sobre Ativos Atuais Circulantes
NIE	Lucro Líquido sobre Patrimônio Líquido Total
OPM	Margem Operacional
X1A	Medida de Liquidez de Altman (1968)
X2A	Medida de Rentabilidade de Altman (1968)
X3A	Medida de Eficiência Operacional de Altman (1968)
X4A	Medida de Mercado de Altman (1968)
X5A	Medida de Giro de Ativos de Altman (1968)

Fonte: Dados da pesquisa.

As amostras foram divididas, cronologicamente, em uma base de treino, que possui as observações de 2000 a 2014 (aproximadamente, 70%), e uma base de teste, que possui as observações de 2015 a 2019 (aproximadamente, 30%). Essa divisão é amplamente adotada na literatura (Barboza; Altman, 2024; Huang; Yen, 2019; Chen, 2011; Geng et al., 2015). A função

da base de treino é desenvolver e calibrar o modelo de previsões. Enquanto a base de teste serve para avaliar o modelo criado (Barboza; Altman, 2024).

Como as ocorrências de dificuldade financeira são significativamente menos frequentes que as de empresas saudáveis, foi necessário aplicar uma técnica de balanceamento na etapa de treinamento. O método utilizado foi o *Undersampling*, que consiste na remoção de instâncias da classe majoritária para reduzir a desproporção entre os grupos. Esse procedimento é relevante para evitar que o modelo se torne tendencioso à classe mais comum, o que comprometeria sua capacidade de identificar casos de FD (Barboza; Altman, 2024).

Para fins comparativos, o desempenho do modelo também foi testado em uma amostra desbalanceada e com a técnica de balanceamento *Synthetic Minority Over-sampling Technique* (SMOTE). Em ambos os cenários alternativos, o modelo apresentou sérias dificuldades para identificar a classe minoritária, chegando a registrar erro do tipo II de 100% na previsão de FD1. Por isso, esses resultados foram desconsiderados nesse estudo.

No que se refere a técnica, o XGBoost é um algoritmo de ML que se destaca por sua alta capacidade de reconhecimento de informações não-lineares e alta precisão de predição, sendo utilizado tanto para problemas de regressão quanto classificação. É um modelo de aprendizagem em conjunto (*Ensemble Learning*), ou seja, combina uma série de modelos e os une para conseguir previsões mais robustas (Yang et al, 2021; Huang; Yen, 2019).

Este algoritmo realiza treinamentos aditivos, em que árvores de decisão são, sequencialmente, adicionadas aos modelos, sempre buscando corrigir os erros da anterior (Huang; Yen, 2019). Cada uma dessas árvores faz suas previsões considerando a anterior e a previsão final é a soma ponderada de todas essas previsões (Yang et al., 2021).

O modelo foi desenvolvido em Python, por meio da ferramenta Colab, e foi treinado para prever FDs com 1 ano de antecedência. As principais bibliotecas usadas foram: pandas, sklearn e xgboost. Também foram usados os valores padrão dos hiperparâmetros.

As métricas adotadas para avaliar a predição do modelo seguiram as usadas por Barboza e Altman (2024) com o intuito de permitir a comparação de desempenho. São elas: Área Sob a Curva (AUC), Acurácia (ACC), Erro Tipo I, Erro Tipo II e *Brier Score* (BS).

As duas primeiras tratam de informar a respeito do desempenho. A AUC revela a capacidade do modelo em separar as classes positivas e negativas. ACC representa o percentual de classificações corretas em relação a todas as observações testadas. É uma métrica importante, porém, pode passar uma falsa ideia de sucesso do modelo no caso de dados desbalanceados (Barboza; Altman, 2024). O BS mede a diferença entre a probabilidade prevista e o resultado real. Portanto, quanto menor o BS, melhor o desempenho do modelo.

As demais métricas de avaliação do modelo estão relacionadas aos erros de classificação. O Erro Tipo I representa os falsos positivos, ou seja, a proporção de empresas que foram classificadas pelo modelo como insolventes (FD = 1), mas que na realidade eram solventes (FD = 0). Esse tipo de erro pode gerar alertas indevidos, levando a decisões equivocadas por parte de analistas e investidores, como a retirada de capital, aumento do custo de financiamento ou impactos na reputação da empresa. Já o erro do tipo II corresponde aos falsos negativos, indicando a proporção de empresas em situação real de dificuldade financeira (FD = 1) que foram erroneamente classificadas como saudáveis (FD = 0). Este é considerado o erro mais crítico, pois pode impedir a adoção de medidas preventivas ou de mitigação de risco, expondo *stakeholders* a perdas financeiras expressivas, inadimplência inesperada ou falhas em processos de auditoria e supervisão regulatória.

Além da avaliação da qualidade preditiva, foi realizada uma análise da importância relativa das variáveis para o modelo. Para isso, utilizou-se o atributo “*feature_importances_*” do XGBoost, que quantifica a contribuição de cada variável na construção das árvores de decisão ao longo do treinamento. A geração desse ranking permite identificar os principais fatores associados à dificuldade financeira, fornecendo informações para interpretação dos

resultados e explicabilidade do modelo. Essa análise tem sido amplamente empregada na literatura como ferramenta complementar à previsão, estando presente em estudos como os de Barboza e Altman (2024), Yang et al. (2021) e Geng et al. (2015), justamente para entender quais fatores se mostram mais discriminatórios a respeito da saúde financeira futura de uma empresa.

4 Resultados e Discussão

A avaliação das previsões para cada tipo de FD está exposta em matrizes de confusão e métricas de desempenho. Tais métricas serão comparadas com os resultados de Barboza e Altman (2024), que avaliou a performance dos modelos de LR e RF na predição de insolvência em empresas latino-americanas de todos os setores econômicos.

4.1 FD1 — Estrutura Contábil

Os resultados demonstrados pelas tabelas 4 e 5 mostram que o XGBoost conseguiu prever todos os 4 casos de FD1 presentes na base, sendo o Erro Tipo II igual a zero, o que também é observado no RF. Isso mostra uma alta sensibilidade de ambos os modelos de ML para esse tipo de problema financeiro em comparação com o LR, que teve Erro Tipo II de 19,88%. Por outro lado, o Erro Tipo I do XGBoost se mostrou abaixo do RF e do LR, revelando uma capacidade superior de separação das classes dentre os 3 modelos, conseguindo um AUC de 98,7%.

TABELA 4 - Matriz de Confusão de FD1, que se refere ao potencial de dificuldade financeira em relação a visão fundamentalista (contábil-financeira)

	Classe Predita: Positiva	Classe Predita: Negativa
Classe Real: Positiva	4 (TP)	0 (FN)
Classe Real: Negativa	21 (FP)	409 (TN)

Fonte: Dados da pesquisa.

TABELA 5 - Comparação de métricas para FD1 com previsão de 1 ano

Métrica	LR (Barboza; Altman, 2024)	RF (Barboza; Altman, 2024)	XGBoost (Este estudo)
AUC	–	–	98,66%
Acurácia (ACC)	90,77%	88,15%	95,16%
Brier Score (BS)	0,072	0,079	0,039
Erro Tipo I	8,63%	12,52%	4,88%
Erro Tipo II	19,88%	0%	0%

Fonte: Dados da pesquisa.

4.2 FD2 — (Des)valorização de Mercado

Os resultados da Tabela 6 e Tabela 7 mostram que o XGBoost é menos sensível ao problema financeiro de tipo FD2, conseguindo prever 152 das 172 empresas com FD2 e tendo um aumento considerável do Erro Tipo II, de 0% para 11,63%, quando comparado as previsões para FD1. A proporção de falsos positivos (Erro Tipo I) também foi maior, mostrando menor capacidade de diferenciar as classes, resultando em um menor AUC de 94,2%. De certa forma, isso mostra a dificuldade que existe em prever a tendência dos preços das ações.

TABELA 6 - Matriz de Confusão de FD2

	Classe Predita: Positiva	Classe Predita: Negativa
Classe Real: Positiva	152 (TP)	20 (FN)
Classe Real: Negativa	23 (FP)	164 (TN)

Fonte: Dados da pesquisa.

TABELA 7 - Comparação de métricas para FD2 com previsão de 1 ano

Métrica	LR (Barboza; Altman, 2024)	RF (Barboza; Altman, 2024)	XGBoost (Este estudo)
AUC	-	-	94,2%
Acurácia (ACC)	75,47%	86,75%	88%
Brier Score (BS)	0,187	0,099	0,1
Erro Tipo I	29,97%	14,56%	12,3%
Erro Tipo II	19,55%	12,06%	11,63%

Fonte: Dados da pesquisa.

O RF também teve aumento significativo no Erro Tipo II e mostrou um desempenho similar ao XGBoost quanto às demais métricas analisadas. Por outro lado, o Erro Tipo II do LR não teve grandes variações comparado com o FD1, porém, teve um aumento de mais de 20% no Erro Tipo I, mostrando que os 3 modelos, de alguma forma, tiveram maior dificuldade para separar as classes quando o tipo de dificuldade financeira prevista é o FD2.

Essa diferença na previsão de FD1 e FD2 pode ser explicada pelos critérios utilizados em suas definições. O tipo FD1 é baseado na relação entre o EBITDA e as despesas financeiras e reflete diretamente a estrutura contábil da empresa. Como esses dados são mais estáveis e derivam de demonstrativos financeiros padronizados, o modelo consegue identificar com maior precisão os padrões associados a esse tipo de dificuldade. Por outro lado, o tipo FD2 está relacionado à queda no valor de mercado da empresa, um indicador que é muito mais suscetível a fatores externos, como a percepção dos investidores, a volatilidade macroeconômica e eventos geopolíticos, refletindo as informações e as notícias conhecidas (Canton et al., 2021). Esses elementos introduzem ruído e incerteza nos dados, o que torna mais difícil para o modelo identificar com precisão empresas em risco.

4.3 Resultados de FD3 e FD4

Para FD3, o XGBoost teve acurácia de 88,24% e BS de 0,098, superiores ao LR (68,82%, 0,203) e desempenho próximo de RF (87,3%, 0,092). Já para FD4 o XGBoost alcançou acurácia de 94,93% e BS de 0,051, mantendo a superioridade em relação ao LR (88,56%, 0,089) e se mostrando ligeiramente melhor que o RF (91,61%, 0,067).

As outras métricas, para FD3 e FD4, não foram comparadas devido à falta de informações.

TABELA 8 - Matriz de Confusão de FD3

	Classe Predita: Positiva	Classe Predita: Negativa
Classe Real: Positiva	152 (TP)	21 (FN)
Classe Real: Negativa	30 (FP)	231 (TN)

Fonte: Dados da pesquisa.

TABELA 9 - Comparação de métricas para FD3 com previsão de 1 ano

Métrica	LR (Barboza; Altman, 2024)	RF (Barboza; Altman, 2024)	XGBoost (Este estudo)
AUC	-	-	94,14%
Acurácia (ACC)	68,82%	87,3%	88,24%
Brier Score (BS)	0,203	0,092	0,098
Erro Tipo I	-	-	11,49%
Erro Tipo II	-	-	12,14%

Fonte: Dados da pesquisa.

Comparando os três algoritmos, concluímos que os modelos baseados em XGBoost, aparentemente, superam os resultados de Barboza e Altman (2024). Em particular, é superior a LR nas previsões de dificuldade financeira, na maioria dos cenários. Além disso, os modelos

com XGBoost apresentaram desempenho similar ou superior ao RF. Isso pode ser explicado pela especialização setorial, já que o LR e o RF foram aplicados em uma base com empresas de todos os setores, enquanto o XGBoost utilizou uma base apenas de empresas do setor de utilidade pública, o que contribuiu para a maior homogeneidade e qualidade da predição ao evitar ruídos presentes em setores com estrutura financeira muito distinta. Um possível marcador que validaria essa explicação é o estudo das variáveis relevantes comparado com o estudo referência, algo que é discutido a seguir.

TABELA 10 - Matriz de Confusão de FD4

	Classe Predita: Positiva	Classe Predita: Negativa
Classe Real: Positiva	3 (TP)	0 (FN)
Classe Real: Negativa	22 (FP)	409 (TN)

Fonte: Dados da pesquisa.

TABELA 11 - Comparação de métricas para FD4 com previsão de 1 ano

Métrica	LR (Barboza; Altman, 2024)	RF (Barboza; Altman, 2024)	XGBoost (Este estudo)
AUC	-	-	97,25%
Acurácia (ACC)	88,56%	91,61%	94,93%
Brier Score (BS)	0,089	0,067	0,051
Erro Tipo I	-	-	5,1%
Erro Tipo II	-	-	0%

Fonte: Dados da pesquisa.

4.4 Análise da Importância das Variáveis

A análise da importância das variáveis revela como o modelo XGBoost identificou diferentes padrões de risco financeiro, variando conforme a definição de FD. Cada tipo de FD apresenta um conjunto distinto de variáveis-chave, o que reforça a capacidade do modelo de ajustar seu foco preditivo.

Em FD1, as variáveis mais relevantes (Figura 1) concentram-se em indicadores associados à rentabilidade. Segundo Chen (2011), esse tipo de variável pertence à categoria de *earning ability*, ou seja, refletem a capacidade da empresa de gerar lucros.

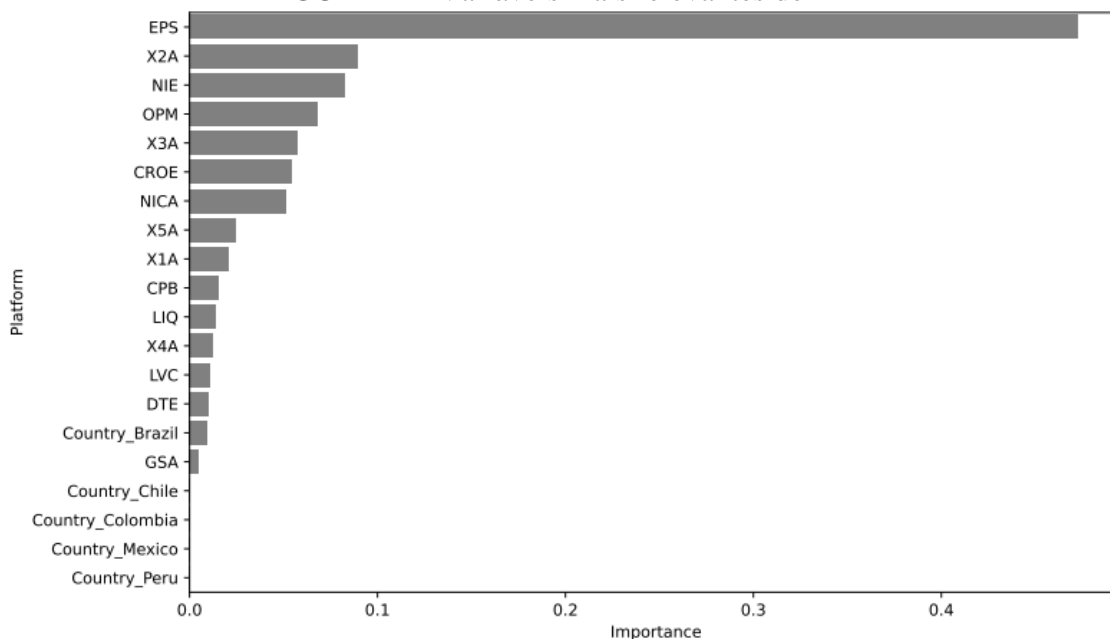
Dentre os atributos apontados como mais importantes pelo modelo, destacam-se: EPS (lucro por ação), NIE (lucro líquido sobre patrimônio líquido), OPM (margem operacional) e NICA (lucro líquido sobre ativo circulante), todos associados à lucratividade. O CROE também surgiu como variável relevante, ele representa a variação do ROE entre 2 anos consecutivos, considerando, assim, a evolução de uma medida de *earning ability*. Levando em conta que o FD1 é baseado em EBITDA comparado às despesas financeiras, é coerente que o modelo tenha priorizado indicadores contábeis clássicos de rentabilidade e fluxo operacional.

Neste estudo, o EPS foi classificado como a mais importante variável na previsão de FD1. Geng et al. (2015) ressaltam que ela desempenha um papel de destaque na previsão da deterioração da lucratividade, sendo um indicador da capacidade de expansão de capital. Além disso, os autores classificam a NICA como um importante indicador da capacidade de desenvolvimento de negócios e, portanto, uma variável relevante para a predição de dificuldade financeira.

Yang et al. geraram *rankings* das variáveis mais importantes tanto com o próprio XGBoost, quanto com o *framework* SHAP. Em ambos os casos, o EPS foi colocado como um indicador de alta relevância e, segundo os autores, a média de EPS das empresas saudáveis foi muito acima das empresas com FD. No presente estudo, a variável EPS também apresentou diferenças estatisticamente significativas entre os grupos de empresas com e sem dificuldade

financeira em todos os tipos de FD. No FD1 a média de EPS das empresas insolventes foi de -3,58, enquanto a média entre as empresas solventes foi de 0,57.

FIGURA 1 – Variáveis mais relevantes de FD1



Fonte: Dados da pesquisa.

Além das variáveis clássicas de rentabilidade, o modelo classificou como importantes as medidas de Altman (1968), sendo a mais importante dentre elas o X2A (Medida de Rentabilidade), que pode ser definida como o valor total dos lucros reinvestidos e/ou perdas de uma empresa ao longo de sua vida, o que, implicitamente, considera a idade da empresa. Empresas mais jovens, por exemplo, tendem a ter um X2A baixo por não terem tido tempo de acumular lucros, o que as coloca em maior risco de falência, refletindo a maior incidência de falhas em empresas jovens na realidade. A segunda medida de Altman mais relevante, segundo o modelo, é o X3A (Medida de Eficiência Operacional), que é calculada dividindo os lucros da empresa antes de juros e impostos pelos seus ativos totais; ela é definida por Altman (1968) como a "verdadeira produtividade dos ativos da empresa" sendo um indicador de alta relevância.

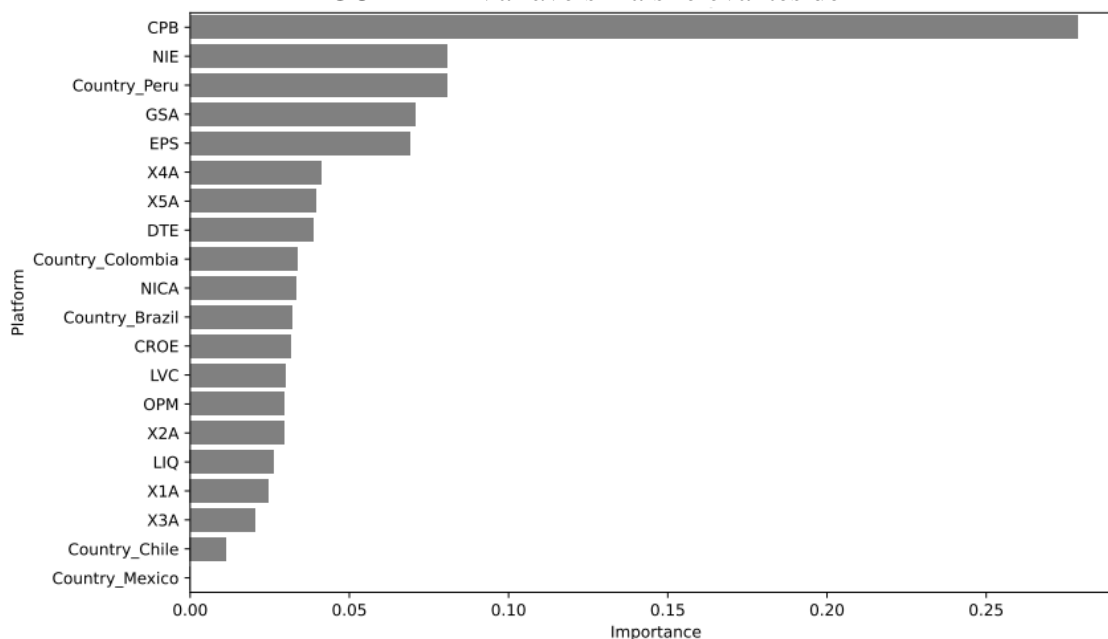
O modelo também revela o CPB como uma variável relevante e, mesmo que não esteja dentre as mais importantes, é, segundo Barboza e Altman (2024), um indicador da eficiência do mercado em ajustar o preço das ações de empresas com risco de dificuldade financeira.

No caso do FD2 (Figura 2), relacionado à desvalorização de mercado, destaca-se o CPB como principal variável preditiva. Este indicador reflete a variação na relação entre o valor de mercado da empresa e seu valor patrimonial contábil de um ano para o outro, expondo a mudança na percepção dos investidores quanto ao valor da empresa. A média do CPB foi significativamente diferente entre os dois grupos: empresas saudáveis apresentaram valorização média de 0,26, enquanto empresas com FD registraram desvalorização média de -0,50. Essa diferença reforça o poder discriminativo do CPB, sugerindo que a deterioração financeira tende a ser precedida por uma queda expressiva na avaliação de mercado da empresa em relação ao seu patrimônio.

O GSA, classificado como a quarta variável mais relevante, representa o crescimento das vendas ao longo de dois anos consecutivos e se enquadra na categoria de Desempenho da Gestão (*Management Performance*), conforme definido por Chen (2011). Sua importância pode ser explicada por ser um sinal precoce de desempenho futuro, sendo que a desaceleração ou

retração nas receitas frequentemente precede deteriorações em lucratividade, geração de caixa e percepção de valor por parte do mercado.

FIGURA 2 – Variáveis mais relevantes de FD2



Fonte: Dados da pesquisa.

Assim como observado no FD1, as variáveis propostas por Altman (1968) também se destacam como relevantes neste contexto. O X4A (Medida de Mercado) é descrito pelo autor como a margem de segurança que a empresa possui antes que seus passivos ultrapassem o valor de seus ativos, levando à insolvência. Essa variável expressa a relação entre a valorização de mercado da empresa e o seu grau de endividamento, sendo, assim, relevante para FD2 por incorporar diretamente a percepção do mercado.

Considerando que o valor de mercado pode ser fortemente influenciado por fatores macroeconômicos e geopolíticos, o modelo FD2 atribui importância também à localização geográfica (país-sede) das empresas analisadas, o que não ocorre com a mesma intensidade em FD1. Esse destaque é coerente com a natureza do FD2, que capta variações na avaliação de mercado, sensíveis a contextos regionais específicos. Vale ressaltar que os dados utilizados neste estudo se estendem até o ano de 2019 e, portanto, não contemplam os efeitos da pandemia de COVID-19, que notadamente impactou os mercados globais a partir de 2020.

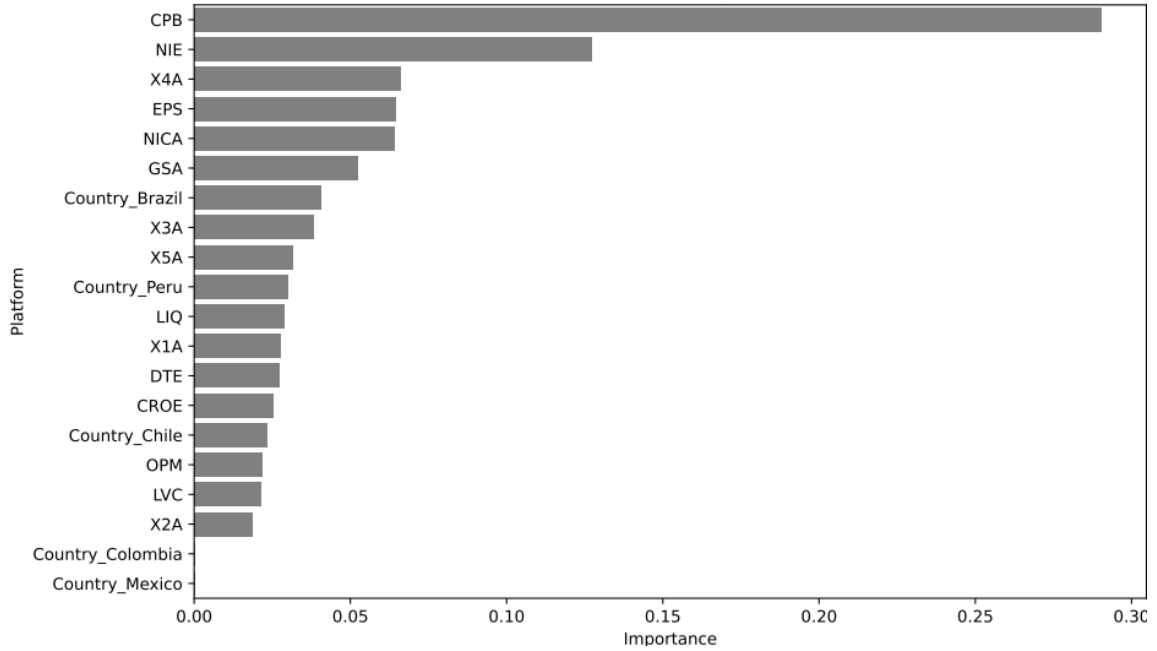
A importância estatística da variável Peru pode estar relacionada a características econômicas específicas desse país no período analisado. As empresas peruanas apresentaram maior NIE (0,15) e CPB (0,038) em relação à média (0,09 e -0,078, respectivamente), além de uma menor frequência de FD.

Por fim, o ranking de importância inclui ainda variáveis associadas à rentabilidade, como o NIE e o EPS, que refletem a capacidade da empresa de gerar retorno sobre o capital investido.

A análise das variáveis mais relevantes nos modelos FD3 e FD4 – ilustrado nas Figuras 3 e 4, respectivamente – revela um padrão coerente com a forma como essas definições foram construídas. Como o FD2 apresenta uma proporção de casos significativamente maior do que o FD1, sendo 516 e 44 ocorrências, respectivamente, o FD3 (FD1 ou FD2) tende a ter um ranking de variáveis mais importantes similar ao do FD2, por ser o mais frequente, o que explica

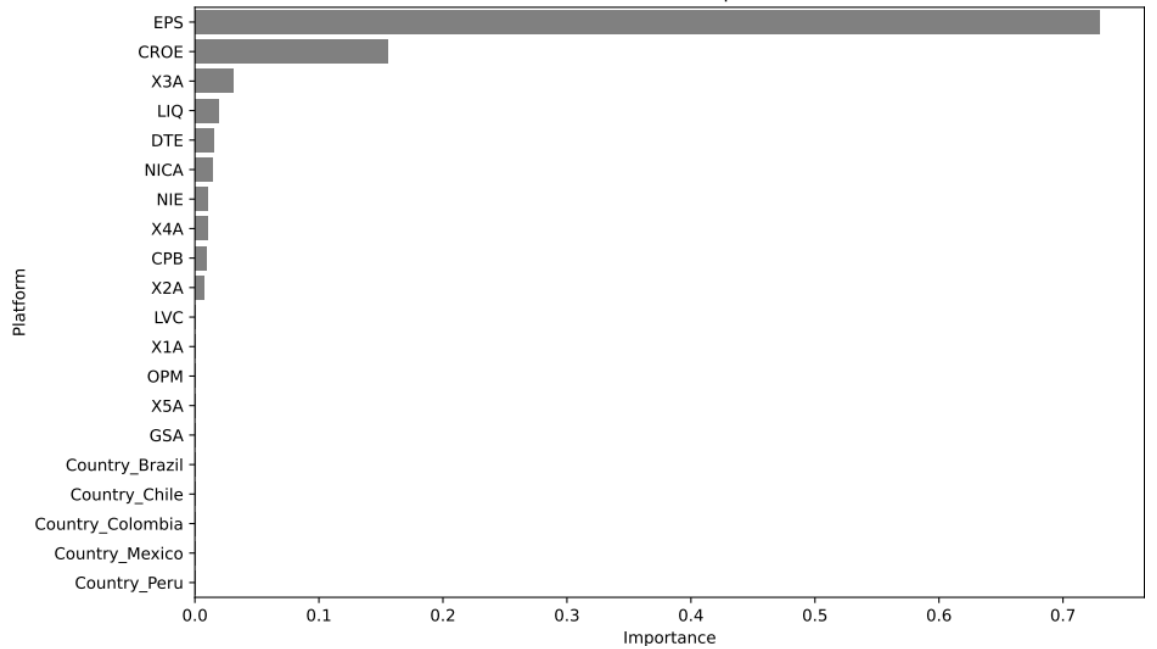
o CPB e NIE no topo. Por outro lado, o FD4 (FD1 e FD2), tende a ter um *ranking* similar ao FD1, ou seja, o menos frequente, resultando no EPS como variável mais relevante.

FIGURA 3 – Variáveis mais relevantes de FD3



Fonte: Dados da pesquisa.

FIGURA 4 – Variáveis mais relevantes de FD4



Fonte: Dados da pesquisa.

Os resultados obtidos superaram os modelos de LR e RF propostos por Barboza e Altman (2024), com destaque para FD1 e FD4, cujas acurácias superaram 94%, com erro tipo II igual a zero. Além disso, observou-se que o modelo para FD1 prioriza variáveis contábeis ligadas à rentabilidade, enquanto o modelo para FD2 dá ênfase a variáveis de mercado e apresenta sensibilidade a fatores geográficos, como o país de origem da empresa.

A interseção entre o desempenho preditivo e a ordenação das variáveis mais relevantes indica que a robustez do XGBoost não se deve apenas à sua elevada acurácia, mas também à capacidade de identificar de forma consistente atributos economicamente significativos. Com isso, o modelo não apenas realiza previsões com precisão superior aos métodos tradicionais, como também oferece explicabilidade e fundamentação analítica para a avaliação de risco financeiro corporativo, se provando uma ferramenta útil tanto para profissionais de mercado, que buscam previsibilidade, quanto para gestores internos, que buscam compreender os fatores que levam a empresa a entrar em dificuldade financeira.

5 Conclusões

Este artigo avaliou a capacidade preditiva do modelo XGBoost para antecipar eventos de insolvência em empresas latino-americanas do setor de utilidade pública com base em dados contábeis e de mercado entre 2000 e 2019. A partir de quatro definições distintas de FD, analisamos não apenas a performance preditiva do modelo, mas também a relevância das variáveis envolvidas, buscando aliar acurácia à capacidade explicativa.

Os resultados mostraram que o XGBoost apresentou desempenho superior aos modelos de LR e RF utilizados por Barboza e Altman (2024), especialmente nos modelos FD1 e FD4, com acurácias superiores a 94% e erro tipo II igual a zero — um resultado expressivo em termos de predição de risco financeiro. Mesmo em definições mais complexas como FD2 e FD3, o modelo manteve níveis elevados de precisão.

Além da performance, o modelo permitiu identificar as variáveis mais relevantes em cada definição de FD. No FD1 (baseado em dados contábeis), destacaram-se indicadores de rentabilidade como EPS, NIE e X2A. Já no FD2 (queda de valor de mercado), o CPB e o GSA assumiram maior importância, evidenciando a influência da percepção do mercado e do desempenho de vendas.

A especialização setorial do modelo pode ter contribuído para a qualidade dos resultados, sugerindo que abordagens segmentadas são promissoras para melhorar a acurácia e a interpretabilidade das previsões. Dessa forma, este trabalho reforça a utilidade do XGBoost como ferramenta prática e explicável na gestão de riscos financeiros corporativos, e abre caminho para pesquisas futuras que explorem outros setores e horizontes de previsão mais longos.

Do ponto de vista teórico, este trabalho contribui para a literatura ao validar o uso de modelos de aprendizado de máquina, como o XGBoost, em contextos emergentes e setoriais, ampliando o escopo de aplicação desses métodos para além dos mercados desenvolvidos e propondo a análise de setores econômicos específicos. Já do ponto de vista prático, os resultados demonstram o potencial do XGBoost como uma ferramenta robusta e interpretável para a detecção precoce de deterioração financeira, oferecendo suporte a analistas, reguladores e investidores na formulação de estratégias de mitigação de risco, alocação de capital e decisões de crédito.

Por fim, este trabalho apresenta algumas limitações que devem ser consideradas. O modelo foi testado com um horizonte temporal restrito a apenas um ano, aplicando-se a empresas de um único setor (utilidade pública) e com recorte geográfico limitado à América Latina. Essas escolhas, embora justificadas pela proposta do estudo, restringem a generalização dos resultados. Pesquisas futuras podem ampliar a análise para múltiplos setores econômicos, o que permitiria comparar diferenças na importância das variáveis e avaliar se há ganho de desempenho preditivo com modelos específicos por setor. Além disso, novos estudos podem explorar diferentes horizontes temporais e aplicar outras técnicas de ML em contextos geográficos diversos, a fim de investigar a robustez, aplicabilidade e interpretabilidade dos algoritmos em cenários distintos de previsão de dificuldade financeira.

Referências

ALTMAN, Edward I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. **The journal of finance**, v. 23, n. 4, p. 589-609, 1968.

ALTMAN, E. I. Applications of Distress Prediction Models: What Have We Learned After 50 Years from the Z-Score Models? **International Journal of Financial Studies**, v. 6, n. 3, p. 70, 2 ago. 2018.

BARBOZA, Flavio; ALTMAN, Edward. Predicting financial distress in Latin American companies: A comparative analysis of logistic regression and random forest models. **The North American Journal of Economics and Finance**, v. 72, p. 102158, 2024.

BARBOZA, Flavio. Data and Code (in R) for predicting Financial Distress in Latin America [Conjunto de Dados]. V1. 2023. Harvard Dataverse. <https://doi.org/10.7910/DVN/ODLGNJ>.

BEAVER, William H. Financial ratios as predictors of failure. **Journal of accounting research**, p. 71-111, 1966.

CANTON, Cristiane et al. Dificuldade financeira e desempenho de mercado de organizações latino-americanas. **Revista Facultad de Ciencias Económicas: Investigación y Reflexión**, v. 29, n. 1, p. 11-26, 2021.

CARDONA, Laura; GUTIÉRREZ, Marcela; AGUDELO, Diego A. Volatility transmission between US and Latin American stock markets: Testing the decoupling hypothesis. **Research in International Business and Finance**, v. 39, p. 115-127, 2017.

CHEN, Mu-Yen. Predicting corporate financial distress based on integration of decision tree classification and logistic regression. **Expert systems with applications**, v. 38, n. 9, p. 11261-11272, 2011.

GENG, Ruibin; BOSE, Indranil; CHEN, Xi. Prediction of financial distress: An empirical study of listed Chinese companies using data mining. **European Journal of Operational Research**, v. 241, n. 1, p. 236-247, 2015.

GEPP, Adrian; KUMAR, Kuldeep. Predicting financial distress: A comparison of survival analysis and decision tree techniques. **Procedia Computer Science**, v. 54, p. 396-404, 2015.

HUANG, Yu-Pei; YEN, Meng-Feng. A new perspective of performance comparison among machine learning algorithms for financial distress prediction. **Applied Soft Computing**, v. 83, p. 105663, 2019.

OHLSON, James A. Financial ratios and the probabilistic prediction of bankruptcy. **Journal of accounting research**, p. 109-131, 1980.

PINDADO, Julio; RODRIGUES, Luis; DE LA TORRE, Chabela. Estimating financial distress likelihood. **Journal of Business Research**, v. 61, n. 9, p. 995-1003, 2008.

YANG, He et al. The extraction of early warning features for predicting financial distress based on XGBoost model and shap framework. **International Journal of Financial Engineering**, v. 8, n. 03, p. 21410