

RISCO DE INSOLVÊNCIA CORPORATIVA DE EMPRESAS BRASILEIRAS DE CAPITAL ABERTO: Uma análise preditiva setorial por meio de aprendizado de máquina

ALLISSON SILVA DOS SANTOS

UNIVERSIDADE FEDERAL DA PARAÍBA (UFPB)

THIÁGO SILVA DE MORAIS

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA (IFPB)

RENATA BRAGA BERENGUER DE VASCONCELOS

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA (IFPB)

CLAUDIANA DE SOUSA

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA (IFPB)

Agradecimento à órgão de fomento:

Agradecimento ao Conselho Nacional de Desenvolvimento Científico e Tecnológico

RISCO DE INSOLVÊNCIA CORPORATIVA DE EMPRESAS BRASILEIRAS DE CAPITAL ABERTO: Uma análise preditiva setorial por meio de aprendizado de máquina

1 INTRODUÇÃO

Em busca de melhor capacidade preditiva, estudos têm sido realizados para a identificação de dificuldades financeiras de empresas. Altman (1968) foi o grande precursor na definição de modelos preditivos para a investigação da saúde financeira organizacional. De acordo com Gimenes e Uribe-Opazo (2009), modelos de previsão de insolvência se estabelecem como instrumentos capazes de prever o fracasso empresarial, auxiliando seus usuários na tomada de decisões. Esses modelos utilizam de variáveis de diversos eixos, como: dados macroeconômicos, demonstrações contábeis, de mercado de atuação etc.

A crise financeira que se iniciou em 2008 trouxe à tona gargalos do ramo financeiro e da falta de concretude nas regulamentações financeiras, resultando na inatividade das empresas (GADGIL, 2021). Considerando a perspectiva dos diversos *stakeholders* que as companhias possuem, nasce a necessidade de identificar os riscos existentes na realidade organizacional. O tipo de risco que recebe foco nesta pesquisa é o de insolvência corporativa.

O risco de insolvência representa o indicativo de que as cláusulas contratuais determinadas com os credores, não são honradas ou são cumpridas, mas com grande dificuldade. Ressalta-se que o estágio de insolvência corporativa pode acarretar falência empresarial. O estágio de insolvência de uma empresa é compreendido pela insuficiência de recursos para o pagamento de dívidas determinadas, líquidas e vencidas (VODA *et al.*, 2021). A partir do momento que esse estado é percebido, um plano de ação pode ser definido para que não haja a piora de insolvência.

A administração financeira tem uma função crucial dentro das organizações, devido ao gerenciamento de dados relevantes para serem destrinchados na predição de riscos empresariais. Ao realizar previsões, o surgimento de dificuldades financeiras pode ser dificultado, devido a identificação dos gargalos insolventes e a resolução deles, antes da aparição de números negativos na demonstração dos dados contábeis (SUSS; TREITEL, 2019).

Santos, Besarria e Machado (2022) ao analisar a efetividade de modelos de aprendizado de máquina na identificação do risco de insolvência corporativa no contexto do mercado acionário brasileiro, evidenciou que o random forest é o melhor modelo preditivo. Porém, ao analisar apenas o setor financeiro, devido às suas especificidades, foi verificado que os algoritmos de SVM polinomial, o logit e o boosting apresentaram desempenhos superiores. Dessa forma, esses autores sugeriram que cada ramo empresarial fosse investigado individualmente, pois aspectos relacionados ao mercado de atuação podem definir o aparecimento de um alto risco de insolvência.

Nesse sentido, este estudo busca responder o seguinte problema: Como se comportam os modelos de aprendizado de máquina na identificação do risco de insolvência no mercado acionário brasileiro, conforme os setores de atuação? Os resultados desse estudo surgem com o apoio de dados de empresas brasileiras de capital aberto de treze anos (2010 – 2022), obtidos na plataforma Thomson Reuters Eikon.

Este trabalho possui relevância teórica, a partir do momento que traz uma ampliação de conteúdo sobre a acurácia preditiva de modelos de aprendizado de máquina, seguindo uma perspectiva empresarial setorial. Essa análise é interessante pois características relacionadas ao setor em que a empresa atua tendem a influenciar em estágios do risco de insolvência corporativa. Além disso, a pesquisa considera uma nacionalidade considerada emergente (brasileira), pelo qual não tem sido explorada com abrangência pela literatura nacional e internacional.

Os achados deste estudo são capazes de auxiliar na tomada de decisão de investidores, considerando que esses indivíduos necessitam de informações relevantes para decidir

racionalmente em qual ativo será mais vantajoso investir. A definição de empresas com menor ou maior risco de insolvência serve de apoio na atividade de avaliação de organizações, representando um critério decisivo para a delimitação do preço justo de compra do ativo.

Este estudo contribui para o cotidiano de empresários, ao se preocuparem com a notoriedade do valor da marca empresarial perante o mercado-alvo. No momento em que empresas são avaliadas positivamente (negativamente) pelo público, surge a expectativa de atrair resultados corporativos positivos (negativos). Organizações com uma má avaliação podem perder o interesse por parte de *stakeholders*. Dessa forma, estratégias podem ser tomadas para melhorar a visão das partes interessadas perante os atributos empresariais. Empresas com indicativo de menor risco de insolvência podem ter maior facilidade na manutenção da imagem corporativa.

Este estudo utiliza sete modelos de aprendizado de máquina, sendo seis supervisionados (naive bayes, random forest, logit, K-NN, bagging e boosting) e um não supervisionado (K-means). A partir das previsões efetuadas por esses modelos, os *stakeholders* tomarão melhores decisões, visto que conseguirão identificar a possibilidade de um alto ou baixo risco de insolvência antes de sua ocorrência.

2 REVISÃO DA LITERATURA

A insolvência corporativa é compreendida como o indicativo de que existem problemas para honrar os acordos feitos com os credores de uma empresa, experimentando problemas de liquidez, por insuficiência de recursos para executar ou continuar suas atividades (KALBUANA *et al.*, 2022). Algumas implicações podem ser geradas devido ao alto nível de insolvência corporativa, como: fornecedores só aceitarem pagamento à vista; diminuição considerável do valor de mercado e a chance de o consumidor cancelar suas compras na expectativa de não usufruir de bens e/ou serviços dentro das condições negociadas de atributos como: qualidade, prazo, quantidade, entre outros (BAE, 2012).

No cenário brasileiro, o quantitativo de organizações que tiveram recuperações judiciais deferidas e concedidas e decretaram falência, cresceu anualmente, no período de 2011 a 2019 (SERASA EXPERIAN, 2021). Esse crescimento de empresas com dificuldades financeiras poderia ser mitigado pela definição de prazos mais flexíveis, novas linhas de crédito, juros mais baixos etc.

A avaliação de desempenho financeiro tornou-se uma prática essencial para estimar os riscos de inadimplência de uma empresa. Dessa forma, pesquisadores e analistas desenvolveram variados modelos que auxiliam na prevenção das consequências procedentes da falência, o que permite determinar a tendência de uma empresa enfrentar problemas financeiros (GIMENES; URIBE-OPAZO, 2009).

A insolvência corporativa não implica automaticamente na falência de uma empresa, mas é um dos antecessores recorrentes (ZHOU *et al.*, 2022). As falências impõem grandes custos aos atores participantes do mercado, bem como à economia como um todo. Nesse sentido, a previsão do risco de insolvência tem sido um tópico crucial para *stakeholders*, como agências de *rating*, investidores, instituições financeiras, gestores corporativos e acionistas (ZHANG *et al.*, 2022). Com isso, as partes interessadas de uma empresa necessitam manter a atenção para indicadores que possam gerar indicativos de um desempenho financeiro desagradável e, assim, pensar em ações estratégicas para amparar seus interesses.

As empresas devem analisar de maneira constante informações quantitativas e qualitativas a fim de acompanhar seu desempenho e considerar se há tendência à insolvência (BATISTA; CRUZ; PICCOLI, 2020). Grande parte das informações quantitativas são de cunho financeiro, objetivas e relacionadas ao lucro da empresa, sendo possível desenvolver diversos tipos de análises a partir desses índices (BORATYŃSKA, 2016).

A capacidade de prever dificuldades financeiras pode fornecer *insights* valiosos para a recuperação econômica de uma organização que está enfrentando ameaças à sua continuidade operacional. O risco de insolvência de uma organização pode ter várias causas, tanto de natureza interna quanto externa. Realizar uma análise contábil-financeira para avaliar o estado de insolvência, por meio de indicadores de solvência financeira, desempenha um papel crucial na detecção de uma situação em que os recursos são insuficientes para atender às obrigações. Essa análise oferece uma visão clara do estado do patrimônio e auxilia na identificação de medidas necessárias para enfrentar as dificuldades financeiras (VODA *et al.*, 2021).

A justificativa para a necessidade de prever a insolvência corporativa reside na possibilidade de intervir em seu estado atual, uma vez que quando uma empresa se torna insolvente, a probabilidade de falência é elevada. Ao avaliar as organizações, é essencial considerar a interdependência entre os indicadores das demonstrações contábeis. Com informações preditivas em mãos, os profissionais de gestão podem se dedicar ao planejamento da recuperação financeira, visando restaurar a solvência da empresa (VODA *et al.*, 2021).

Com a evolução da tecnologia da informação e, conseqüentemente, da inteligência artificial, pesquisas têm utilizado de técnicas de aprendizado de máquina, que tem demonstrado grande precisão (QIAN; WANG; YUAN; GAO; SONG, 2022). A título de exemplo, Barboza, Kimura e Altman (2017) usaram dados de organizações norte-americanas do período de 1985 a 2013 e testaram os modelos SVM linear e radial, análise discriminante linear, bagging, random forest, boosting, redes neurais e logit. Os melhores resultados foram para os modelos random forest, bagging e boosting.

Tran *et al.* (2022) objetivaram comparar o poder preditivo de algoritmos de aprendizado de máquina, no conjunto de dados de organizações listadas no Vietnã de 2010 a 2021. Os resultados mostraram superioridade do random forest em comparação a alguns modelos, como: logit, SVM, redes neurais e decision tree. Santos, Besarria e Machado (2022) encontraram resultados similares para dados de empresas brasileiras de capital aberto. Malakauskas e Lakštutienė (2021) também encontraram resultados que corroboram o poder preditivo do random forest, ao investigar dados de pequenas e médias empresas de estados bálticos.

Nesse sentido, um modelo preciso e eficiente de previsão de risco de insolvência consegue gerar sinais de alerta antecipado para credores e investidores. A partir desses alertas, é possível visualizar o apoio por parte dos administradores na tomada de decisões operacionais e apoio das instituições financeiras para definir os indivíduos supervisores que estarão em busca de manter a estabilidade do mercado financeiro (ZHANG *et al.*, 2022).

No tocante às agências de *rating* e as instituições financeiras, além de obter as previsões, elas necessitam avaliar a acurácia desses modelos, para verificar quais são os mais justos, e se existe alguma diferenciação pelo tipo de indústria ou região. Além disso, os gestores querem identificar quais características que influenciam o surgimento de dificuldades financeiras e quais métricas precisam ser melhoradas para evitar a insolvência e a falência (ZHANG *et al.*, 2022).

As empresas que se encontram em estado de insolvência enfrentam problemas para conseguir capital no mercado devido às atitudes negativas advindas das agências de crédito, bancos e instituições financeiras. O motivo para que isso aconteça pode estar presente na desconfiança dessas empresas em relação às organizações insolventes, pois já existe uma batalha para quitar empréstimos anteriores (GUPTA; MAHAKUD, 2022).

A capacidade de prever a insolvência corporativa tem o potencial de evitar ou mitigar perdas financeiras para as organizações. Essas perdas resultam em custos que afetam diretamente os colaboradores, gestores, acionistas e fornecedores. Além disso, uma análise aprofundada revela que empresas com alto risco de insolvência podem impactar a sociedade como um todo, uma vez que os impostos gerados por essas empresas são utilizados para melhorar as condições de vida da população (STUPP; FLACH; MATTOS, 2018). Por conta

disso, as partes interessadas precisam ficar atentas na avaliação da situação financeira das oportunidades de investimento para que possam detectar o nível de saúde financeira esperada das organizações (WU; MA; OLSON, 2022).

3 METODOLOGIA

Neste estudo, emprega-se o uso de aprendizado de máquina não supervisionado para a detecção de risco de insolvência corporativa. O modelo escolhido para o agrupamento de empresas com alto ou baixo risco de insolvência foi o K-means. A variável utilizada no K-means foi o *Z-Score* sugerido por Citterio e King (2023), sendo definido como o número de desvios padrão abaixo da média que a rentabilidade de uma organização deve cair antes do patrimônio deixar de ser positivo, conforme Equação 1.

$$Z - Score = \frac{Média(ROA) + Média(Patrimônio Líquido / Ativo Total)}{\sigma(ROA)} \quad (1)$$

Os valores médios e desvios padrão da Equação 1 foram calculados considerando um período de cinco trimestres. Para a análise empírica deste estudo, foram utilizados modelos supervisionados de aprendizado de máquina, como naive bayes, random forest, logit, K-NN, bagging e boosting. Esses modelos foram selecionados por serem algoritmos relevantes no contexto de aprendizado de máquina supervisionado. Tais métodos de aprendizado supervisionado têm sido aplicados em estudos que abordam questões relacionadas às dificuldades financeiras de empresas (BARBOZA; KIMURA; ALTMAN, 2017).

Os dados foram divididos em conjuntos de treinamento e teste para o aprendizado de máquina. O conjunto de treinamento foi composto por 75% das observações, enquanto o conjunto de teste foi formado pelos dados restantes, selecionados aleatoriamente. A padronização foi aplicada para melhor adequação dos dados aos modelos. A efetividade dos modelos foi avaliada pelas seguintes métricas: acurácia, sensibilidade, especificidade, *Mean Absolut Error* (MAE), *Root Mean Squared Error* (RMSE), *Area Under the ROC Curve* (AUC) e a curva *Receiver Operating Characteristic* (ROC).

Para a execução desta pesquisa foram utilizadas cinco variáveis preditivas: tamanho da empresa, *Return On Assets* (ROA), liquidez corrente, alavancagem financeira e a razão do capital de giro pelo ativo total. Essas variáveis foram utilizadas em pesquisas anteriores com objetivo similar (BARBOZA; KIMURA; ALTMAN, 2017). O tamanho da organização pode representar um determinante considerável para a definição da insolvência corporativa, pelo qual grandes organizações podem possuir menor risco, por conta da capacidade de diversificação. No quesito ROA, as organizações com maior rentabilidade precisam de menos dívidas para executar suas atividades. Dessa forma, quanto menor (maior) a dívida, possivelmente menor (maior) será o risco de insolvência corporativa.

A inclusão da variável de liquidez corrente como variável independente se baseia na capacidade de cumprir obrigações de curto prazo, uma vez que uma maior liquidez corrente indica uma maior capacidade de pagamento de dívidas a terceiros. No que diz respeito à alavancagem financeira, em países em desenvolvimento, a estrutura de capital das empresas costuma depender de financiamento, e, portanto, espera-se que o valor desse indicador seja menor para organizações com baixo risco de insolvência (SANTANA FILHO *et al.*, 2019).

Por fim, no que diz respeito à proporção entre o capital de giro e o total de ativos, o capital de giro é considerado um indicador da habilidade da empresa em cumprir suas obrigações de curto prazo. Essa habilidade é avaliada em relação ao tamanho da organização, representado pelo total de ativos. Utilizando as cinco variáveis independentes mencionadas, foi estimada a Equação 2 para analisar a previsão do risco de insolvência corporativa das empresas da amostra.

$$Z - Score_{it} = \beta_0 + \beta_1 TAM_{it} + \beta_2 ROA_{it} + \beta_3 LIQC_{it} + \beta_4 ALAV_{it} + \beta_5 CGAT_{it} + \varepsilon_{it} \quad (2)$$

Em que: Z-Score representa o risco de insolvência da organização *i* no momento *t*; TAM representa o tamanho da organização *i* no momento *t*; ROA é o retorno sobre os ativos da organização *i* no momento *t*; LIQC é a liquidez corrente da empresa *i* no momento *t*; ALAV representa a alavancagem financeira da empresa *i* no momento *t*; CGAT representa o capital de giro dividido pelo ativo total da empresa *i* no momento *t*; β_0 é uma constante; β_n representa os coeficientes das variáveis independentes; ε representa o tamanho do erro.

Quadro 1 – Operacionalização das variáveis independentes

Variáveis independentes	Operacionalização	Autores
Tamanho (TAM)	TAM = ln(Ativo Total)	Shahwan e Habib (2020)
Retorno sobre os ativos (ROA)	ROA = Lucro Líquido / Ativo Total	Chen, Chen e Lien (2020)
Alavancagem financeira (ALAV)	ALAV = Passivo Oneroso / Ativo total	Shahwan e Habib (2020)
Liquidez Corrente (LIQC)	LIQC = Ativo Circulante / Passivo Circulante	Wang, Chen, Chu (2018)
Capital de Giro / Ativo total (CGAT)	CGAT = Capital de Giro / Ativo total	Barboza, Kimura e Altman (2017)

Os dados foram obtidos a partir da plataforma Thomson Reuters Eikon, que compila as informações de empresas brasileiras com ações negociadas na Brasil, Bolsa, Balcão [B]³. Os dados analisados são trimestrais e obedecem ao período de 2010 a 2022 e foram divididos por setores de atuação. Observações foram excluídas da amostra por não apresentarem todos os dados necessários à execução dos algoritmos, ou pelo setor de atuação não apresentar a quantidade mínima de 300 observações para a realização dos testes.

As nomenclaturas para os setores de atuação que foram considerados por esta pesquisa são as definidas pela North American Industry Classification System (NAICS), sendo elas: *Real Estate and Rental and Leasing* (Bens Imóveis e Aluguel e Leasing), *Utilities* (Serviços de Utilidade Pública), *Manufacturing* (Manufatura), *Retail Trade* (Comércio de Varejo), *Information* (Informação), *Transportation and Warehousing* (Transporte e Armazenagem) e *Construction* (Construção). A análise estatística utilizou dos resultados evidenciados com o apoio dos softwares R e RStudio, utilizando de suas bibliotecas.

4 ANÁLISE DOS RESULTADOS

Inicialmente, os dados foram utilizados para alcançar medidas estatísticas básicas: mediana, média e desvio padrão (DP). Os resultados dessas medidas podem ser observados a partir da Tabela 1, para cada setor estudado na pesquisa. Em relação ao Z-Score, compreende-se que, quanto maior o valor, melhor o nível de solvência da organização. Em relação aos valores de média, o setor de atuação que apresenta melhor (pior) resultado é o de Informação (Transporte e Armazenagem) com o valor de 114,23 (55,32). A situação do pior resultado se altera quando analisada a mediana, pelo qual o setor de Comércio de Varejo ocupa esse espaço com o valor de 40,31. Ressalta-se que entre as variáveis estudadas, o Z-Score apresenta o maior

desvio padrão, indicando uma alta volatilidade no quesito insolvência das empresas brasileiras de capital aberto.

Considerando os valores de média, as empresas de maior (menor) tamanho são do setor de Transporte e Armazenagem (Construção). Já considerando os valores de mediana, as empresas do setor de Serviços de Utilidade Pública possuem maior tamanho. Dessa forma, as empresas de Transporte e Armazenagem ou de Serviços de Utilidade Pública podem possuir um menor risco em suas operações, por conta da possibilidade de diversificação do negócio. Em consonância a esse resultado, as empresas de Serviços de Utilidade Pública possuem maior ROA, considerando os valores centrais, o que sugere que essas organizações podem precisar de menos dívidas para executar suas atividades, devido ao alto rendimento em suas operações.

Em relação ao grau de alavancagem, o setor de Transporte e Armazenagem apresentou um maior valor de mediana e média, pelo qual é esperado que o valor desse indicador seja menor para organizações com baixo risco de insolvência (SANTANA FILHO *et al.*, 2019). Sobre a razão entre o capital de giro e o total de ativos, o setor de Construção apresentou maiores resultados para a média e mediana, sendo uma indicação de melhor capacidade empresarial de quitar suas obrigações, comparado a outros setores.

Tabela 1: Estatística descritiva

Setor de Atuação	Métricas	Z-Score	TAM	ROA	ALAV	LIQ	CGAT
Real Estate and Rental and Leasing N = 379	Mediana	62,708	21,950	0,007	0,350	1,990	0,196
	Média	92,987	21,700	0,005	0,370	2,081	0,213
	DP	96,476	1,130	0,023	0,149	0,964	0,203
Utilities N = 1611	Mediana	50,740	22,880	0,013	0,328	1,203	0,040
	Média	69,738	22,850	0,013	0,328	1,724	0,046
	DP	71,305	1,306	0,022	0,140	2,805	0,135
Manufacturing N = 3611	Mediana	43,299	21,490	0,009	0,313	1,897	0,212
	Média	65,798	21,770	0,011	0,311	2,266	0,213
	DP	72,630	1,932	0,034	0,176	2,228	0,179
Retail Trade N = 661	Mediana	40,312	22,490	0,008	0,306	1,666	0,220
	Média	61,371	22,470	0,009	0,288	1,761	0,212
	DP	64,805	1,370	0,021	0,136	0,661	0,140
Information N = 372	Mediana	75,375	21,970	0,008	0,141	1,505	0,083
	Média	114,231	22,340	0,006	0,185	2,230	0,137
	DP	125,230	2,196	0,033	0,150	2,251	0,203
Transportation and Warehousing N = 510	Mediana	45,987	22,100	0,007	0,422	1,022	0,003
	Média	55,319	23,080	0,007	0,388	1,197	0,011
	DP	44,860	1,451	0,023	0,195	0,750	0,130
Construction N = 701	Mediana	59,933	21,810	0,006	0,280	2,348	0,355
	Média	88,609	21,610	0,004	0,274	2,587	0,319
	DP	107,880	1,284	0,031	0,141	1,376	0,217

Fonte: Dados da pesquisa

Com o intuito de avaliar os modelos de aprendizado de máquina, foram utilizados os valores de sensibilidade, especificidade, acurácia, MAE, RMSE, AUC e a visualização gráfica da curva ROC. Os valores referentes à sensibilidade e especificidade podem ser vistos na Tabela 2.

Ao analisar os valores de sensibilidade, que verifica a capacidade dos modelos de identificarem corretamente os exemplos positivos de alto risco de insolvência, o K-NN apresentou melhor resultado, isoladamente, para os setores de Bens Imóveis e Aluguel e Leasing, Serviços de Utilidade Pública e Manufatura. O naive bayes apresentou melhor resultado para os setores de Informação e Construção. Conjuntamente, o K-NN e o naive bayes apresentaram melhor resultado para o setor de Comércio de Varejo. Além disso, para o setor de

Transporte e Armazenagem, houve um empate entre quatro modelos: K-NN, naive bayes, bagging e boosting.

Tabela 2: Matriz de Confusão

Setor de atuação	Modelo	VP	FP	FN	VN	Sensib.	Especif.
Real Estate and Rental and Leasing	K-NN	15	4	4	72	0,7895	0,9474
	Naive Bayes	13	6	8	68	0,6842	0,8947
	Logit	0	19	1	75	0,0000	0,9868
	Random Forest	10	9	4	72	0,5263	0,9474
	Bagging	9	10	3	73	0,4737	0,9605
	Boosting	8	11	5	71	0,4211	0,9342
Utilities	K-NN	13	29	14	347	0,3095	0,9612
	Naive Bayes	1	41	3	358	0,0238	0,9917
	Logit	0	42	0	361	0,0000	1,0000
	Random Forest	8	34	3	358	0,1905	0,9917
	Bagging	10	32	7	354	0,2381	0,9806
	Boosting	6	36	7	354	0,1429	0,9806
Manufacturing	K-NN	54	50	38	725	0,5192	0,9502
	Naive Bayes	19	85	26	737	0,1827	0,9659
	Logit	4	100	5	758	0,0385	0,9934
	Random Forest	42	62	13	750	0,4038	0,9830
	Bagging	43	61	20	743	0,4135	0,9738
	Boosting	26	78	16	747	0,2500	0,9790
Retail Trade	K-NN	6	5	6	148	0,5455	0,9610
	Naive Bayes	6	5	11	143	0,5455	0,9286
	Logit	0	11	0	154	0,0000	1,0000
	Random Forest	3	8	4	150	0,2727	0,9740
	Bagging	4	7	3	151	0,3636	0,9805
	Boosting	3	8	4	150	0,2727	0,9740
Information	K-NN	11	15	6	61	0,4231	0,9104
	Naive Bayes	15	11	16	51	0,5769	0,7612
	Logit	7	19	3	64	0,2692	0,9552
	Random Forest	14	12	3	64	0,5385	0,9552
	Bagging	12	14	3	64	0,4615	0,9552
	Boosting	11	15	2	65	0,4231	0,9701
Transportation and Warehousing	K-NN	1	5	7	115	0,1667	0,9426
	Naive Bayes	1	5	9	113	0,1667	0,9262
	Logit	0	6	0	122	0,0000	1,0000
	Random Forest	0	6	2	120	0,0000	0,9836
	Bagging	1	5	3	119	0,1667	0,9754
	Boosting	1	5	1	121	0,1667	0,9918
Construction	K-NN	16	19	18	122	0,4571	0,8714
	Naive Bayes	20	15	38	102	0,5714	0,7286
	Logit	0	35	3	137	0,0000	0,9786
	Random Forest	15	20	14	126	0,4286	0,9000
	Bagging	17	18	14	126	0,4857	0,9000
	Boosting	14	21	17	123	0,4000	0,8786

Fonte: Dados da pesquisa

Em relação à especificidade, que mensura a capacidade dos modelos em identificarem corretamente os exemplos negativos de alto risco de insolvência, o boosting apresentou o melhor resultado para o setor de Informação. Para os demais setores, os melhores valores foram

denotados pelo modelo logit. Apesar desse resultado, o modelo logit mostrou uma forte tendência de definir as observações como baixo risco de insolvência, o que pode representar um problema para a identificação correta de empresas com alto risco de insolvência. Em contradição, esse cenário com o logit não ocorre na pesquisa de Santos, Besarria e Machado (2022) ao analisar as empresas financeiras vinculadas a [B]³.

Na Tabela 3, é possível denotar os valores de Acurácia, MAE e RMSE para cada teste efetuado. A acurácia indica a proporção de observações corretamente classificadas em comparação ao total de observações. O MAE mensura a média das diferenças absolutas entre as previsões do modelo e os seus valores verdadeiros. O RMSE mede a raiz quadrada da média dos erros quadrados entre as previsões e os seus valores verdadeiros.

Tabela 3: Acurácia, MAE e RMSE

Setor de atuação	Modelo	Acurácia	MAE	RMSE
Real Estate and Rental and Leasing	K-NN	0,9158	0,0842	0,2901
	Naive Bayes	0,8526	0,1474	0,3839
	Logit	0,7895	0,2105	0,4588
	Random Forest	0,8632	0,1368	0,3699
	Bagging	0,8632	0,1368	0,3699
	Boosting	0,8316	0,1684	0,4104
	K-NN	0,8933	0,1067	0,3266
Utilities	Naive Bayes	0,8908	0,1092	0,3304
	Logit	0,8958	0,1042	0,3228
	Random Forest	0,9082	0,0918	0,3030
	Bagging	0,9032	0,0968	0,3111
	Boosting	0,8933	0,1067	0,3267
	K-NN	0,8985	0,1015	0,3186
	Naive Bayes	0,8720	0,1280	0,3578
Manufacturing	Logit	0,8789	0,1211	0,3480
	Random Forest	0,9135	0,0865	0,2941
	Bagging	0,9066	0,0934	0,3057
	Boosting	0,8916	0,1084	0,3293
	K-NN	0,9333	0,0667	0,2582
	Naive Bayes	0,9030	0,0970	0,3114
	Logit	0,9333	0,0667	0,2582
Retail Trade	Random Forest	0,9273	0,7272	0,2697
	Bagging	0,9394	0,0606	0,2462
	Boosting	0,9273	0,0727	0,2697
	K-NN	0,7742	0,2258	0,4752
	Naive Bayes	0,7097	0,2903	0,5388
	Logit	0,7634	0,2366	0,4864
	Random Forest	0,8387	0,1613	0,4016
Information	Bagging	0,8172	0,1828	0,4275
	Boosting	0,8172	0,1828	0,4275
	K-NN	0,9062	0,0938	0,3062
	Naive Bayes	0,8906	0,1094	0,3307
	Logit	0,9531	0,0469	0,2165
	Random Forest	0,9375	0,0625	0,2500
	Bagging	0,9375	0,0625	0,2500
Transportation and Warehousing	Boosting	0,9531	0,0469	0,2165
	K-NN	0,7886	0,2114	0,4598
	Naive Bayes	0,6971	0,5503	0,3029
	Logit	0,7829	0,2171	0,4660
	Random Forest	0,8057	0,1943	0,4408
	Bagging	0,8171	0,1829	0,4276
	Boosting	0,7829	0,2172	0,4660
Construction				

Fonte: Dados da pesquisa

Em relação às métricas da Tabela 3, o modelo K-NN apresentou melhor resultado para o setor de Bens Imóveis e Aluguel e Leasing; o modelo random forest apresentou melhor resultado para os setores de Serviços de Utilidade Pública, Manufatura e Informação; o modelo bagging apresentou melhor resultado para os setores de Comércio de Varejo e Construção; e os modelos logit e boosting apresentaram melhores resultados, conjuntamente, para o setor de Transporte e Armazenagem.

No que concerne à Tabela 4, encontram-se disponíveis os valores de AUC para cada teste realizado, pelo qual mensura a capacidade de distinguir as observações entre as classes positiva e negativa, sendo identificado que quanto maior o valor, melhor o modelo. Nesse sentido, o modelo K-NN apresentou melhor resultado para os setores de Bens Imóveis e Aluguel e Leasing e de Serviços de Utilidade Pública; o bagging apresentou melhor resultado para o setor de Manufatura; o random forest apresentou melhor resultado para os setores de Comércio de Varejo, de Informação e de Construção; e o naive bayes apresentou melhor resultado para o setor de Transporte e Armazenagem.

Dessa forma, o mais destacado pela métrica AUC foi o random forest. Em consonância com os resultados desta pesquisa, os achados de Santos, Besarria e Machado (2022) e Barboza, Kimura e Altman (2017) demonstram que o random forest tem sido destaque em desempenho preditivo para detecção do risco de insolvência. A Figura 1 reforça os resultados de AUC, a partir da curva ROC, que quanto mais próximo estiver do canto superior esquerdo do gráfico, o desempenho é melhor.

Tabela 4: AUC

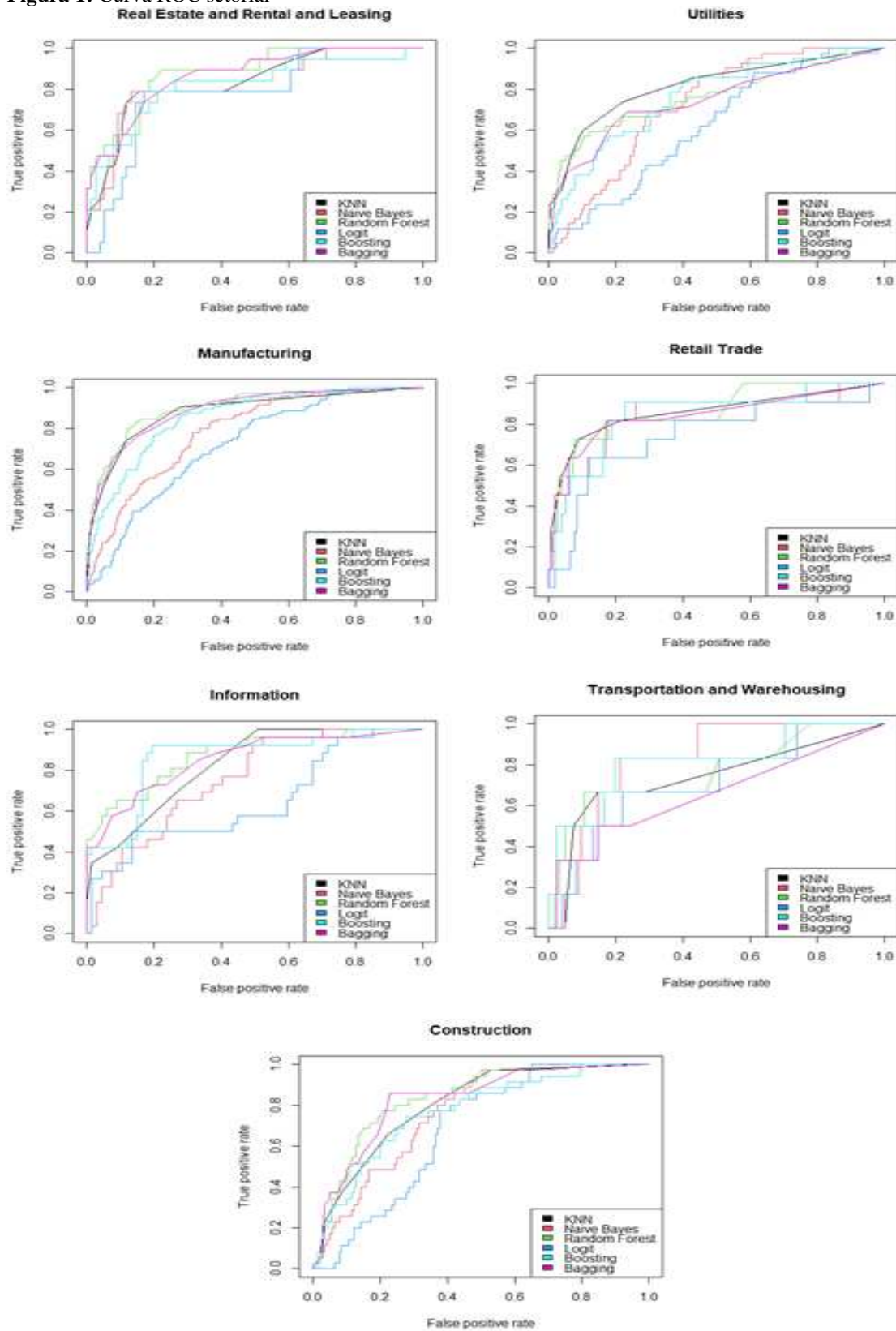
Modelo	Real Estate and Rental and Leasing	Utilities	Manufacturing	Retail Trade	Information	Transportation and Warehousing	Construction
K-NN	0,8923	0,8136	0,8802	0,8560	0,8269	0,7363	0,8036
Naive Bayes	0,8165	0,7254	0,7857	0,8459	0,7572	0,8265	0,7631
Logit	0,7832	0,6216	0,7178	0,7444	0,6487	0,7131	0,6784
Random Forest	0,8802	0,7657	0,9004	0,8625	0,8737	0,7650	0,8432
Bagging	0,8657	0,7469	0,8886	0,8400	0,8533	0,6557	0,8259
Boosting	0,8199	0,7569	0,8539	0,8447	0,8628	0,8142	0,7696

Fonte: Dados da pesquisa

Diante das análises apresentadas, a depender da métrica utilizada para avaliar os acertos de aprendizado de máquina, há escolhas de modelos diferentes para cada setor. Diante das métricas de sensibilidade, acurácia, MAE, RMSE, AUC o modelo K-NN apresentou melhores resultados para o setor de Bens Imóveis e Aluguel e Leasing. Além disso, esse modelo apresentou melhores resultados para o setor de Serviços de Utilidade Pública, ao considerar os valores de AUC e sensibilidade. Dessa forma, o K-NN representa um forte candidato para as empresas destes dois ramos, visando a sua adoção nas práticas preditivas de risco de insolvência.

Para o setor de Informação, as métricas que apresentaram resultados compatíveis foi o de acurácia e AUC, apontando o random forest como um bom modelo. Santos, Besarria e Machado (2022), Barboza, Kimura e Altman (2017), Malakauskas e Lakštutienė (2021) e Tran *et al.* (2022) corroboram esse resultado, ao evidenciar o random forest, como um dos melhores modelos para prever o risco de insolvência, mas sem uma perspectiva setorial.

Figura 1: Curva ROC setorial



Fonte: Dados da pesquisa

Os demais setores não encontraram consonância entre as métricas para definir qual seria o modelo mais indicado para previsão do risco de insolvência. Diferentemente, a partir de dados de empresas do setor de Manufatura, Bae (2012) buscou melhorar a precisão da previsão do

risco de insolvência e utilizou de modelos como o logit, análise discriminante múltipla, *support vector machine* (SVM) de base radial, *multi-layer perceptron* (MLP), algoritmos de árvore de classificação (C5.0) e redes bayesianas. Os resultados demonstraram que o *support vector machine* (SVM) de base radial demonstrou maior precisão na predição.

Por fim, esses resultados podem contribuir para uma melhor predição da insolvência corporativa, evitando ou reduzindo perdas financeiras por parte das organizações. Essas perdas acarretam despesas que têm um impacto direto nos funcionários, líderes, acionistas e fornecedores (STUPP; FLACH; MATTOS, 2018). Ao examinar os modelos preditivos alinhados aos setores, as empresas com um alto risco de insolvência terão maior acurácia na definição do seu status e poderão tomar decisões estratégicas com mais assertividade.

5 CONSIDERAÇÕES FINAIS

Este trabalho, com o objetivo de analisar o comportamento de modelos de aprendizado de máquina na identificação do risco de insolvência no mercado acionário brasileiro, considerando o foco setorial, foi alcançado com êxito para os setores de Bens Imóveis e Aluguel e Leasing, Serviços de Utilidade Pública, Manufatura, Comércio de Varejo, Informação, Transporte e Armazenagem e Construção.

De acordo com algumas métricas, o modelo K-NN apresentou melhores resultados para os setores de Bens Imóveis e Aluguel e Leasing de Serviços de Utilidade Pública. Para o setor de Informação, algumas métricas demonstraram resultados compatíveis para o modelo random forest, o apontando o como um bom modelo. Dessa forma, o K-NN e o random forest representam fortes candidatos para utilização das organizações destes setores, visando a melhoria das predições efetuadas com o intuito de capturar o risco de insolvência.

A eficácia dos modelos é considerada de suma importância para as empresas, investidores e outras partes interessadas, uma vez que permite a adoção de medidas para mitigar o risco de insolvência e reavaliar a adequação dos ativos para inclusão em carteiras de investimento. Nesse sentido, os atores ligados aos negócios saberão com maior precisão a situação financeira dessas organizações.

Apesar de encontrar esses resultados, esse estudo apresentou algumas limitações. A primeira limitação evidencia-se na exclusão dos demais setores da pesquisa, sendo estes a maioria, não conseguindo atendê-los a partir dos resultados desta pesquisa. Outra limitação encontra-se no foco em dados quantitativos, deixando de lado os dados qualitativos advindos de relatórios organizacionais, de governança, de redes sociais, ou de outros meios, que poderiam robustecer os resultados da pesquisa.

As limitações apresentadas sugerem ideias oportunas de pesquisas futuras. Primeiramente, poderia ser feita uma análise setorial que comporte os dados de todos os setores, podendo utilizar dados de empresas listadas em bolsas de valores estrangeiras, para suprir a pequena quantidade de observações verificadas na [B]³ para alguns setores. Além disso, poderiam ser incorporados indicadores qualitativos, na perspectiva de ampliar o corpo de variáveis preditivas pretendidas por este estudo.

Por fim, há o desejo que esta pesquisa estimule a colaboração entre especialistas em aprendizado de máquina e finanças, a fim de aprimorar os sistemas preventivos para o risco de insolvência. Esses sistemas possuem a função fundamental de alertar os gestores, investidores e partes interessadas sobre a situação financeira das organizações, sendo um insumo informacional estratégico para a tomada de decisão.

REFERÊNCIAS

ALTMAN, Edward L. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. **The Journal of Finance**, v. 23, n. 4, p. 589-609, set. 1968.

BAE, Jae Kwon. Predicting financial distress of the South Korean manufacturing industries. **Expert Systems with Applications**, v. 39, n. 10, p. 9159-9165, ago. 2012. Elsevier BV. <http://dx.doi.org/10.1016/j.eswa.2012.02.058>.

BARBOZA, Flavio; KIMURA, Herbert; ALTMAN, Edward. Machine learning models and bankruptcy prediction. **Expert Systems with Applications**, v. 83, p. 405-417, out. 2017. Elsevier BV. <http://dx.doi.org/10.1016/j.eswa.2017.04.006>.

BORATYŃSKA, Katarzyna. FsQCA in corporate bankruptcy research. An innovative approach in food industry. **Journal of Business Research**, v. 69, n. 11, p. 5529-5533, nov. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.jbusres.2016.04.166>.

CITTERIO, Alberto; KING, Timothy. The role of Environmental, Social, and Governance (ESG) in predicting bank financial distress. **Finance Research Letters**, v. 51, p. 103411, jan. 2023. Elsevier BV. <http://dx.doi.org/10.1016/j.frl.2022.103411>.

CHEN, Chih-Chun; CHEN, Chun-Da; LIEN, Donald. Financial distress prediction model: the effects of corporate governance indicators. **Journal of Forecasting**, v. 39, n. 8, p. 1238-1252, 22 abr. 2020. Wiley. <http://dx.doi.org/10.1002/for.2684>.

CRUZ, J. A. E.; BATISTA, A. M.; RIBEIRO PICCOLI, P. G. Modelo de insolvência aplicado a organizações cooperativas médicas: análise do mercado brasileiro. **Revista de Gestão e Organizações Cooperativas**, v. 7, n. 13, 2020. DOI: 10.5902/2359043240561.

GADGIL, Aashish A. Machine Learning based Intelligent Financial Crisis Prediction Models. **International Journal of Engineering Trends and Applications: A Review**, v. 8, n. 3, p. 30-35, 2021.

GIMENES, R. M. T.; URIBE-OPAZO, M. A. Modelos Multivariantes para a Previsão de Insolvência em Cooperativas Agropecuárias: Uma Comparação entre a Análise Discriminante e de Probabilidade Condicional - Logit. **Contabilidade Vista & Revista**, v. 14, n. 3, p. 45-64, 2009.

GUPTA, Gaurav; MAHAKUD, Jitendra. Impact of financial distress on investment-cash flow sensitivity: evidence from emerging economy. **International Journal of Managerial Finance**, v. 19, n. 4, p. 713-743, 18 jul. 2022. Emerald. <http://dx.doi.org/10.1108/ijmf-03-2022-0102>.

KALBUANA, Nawang; TAQI, Muhamad; UZLIAWATI, Lia; RAMDHANI, Dadan. The Effect of Profitability, Board Size, Woman on Boards, and Political Connection on Financial Distress Conditions. **Cogent Business & Management**, v. 9, n. 1, p. 1-22, 20 nov. 2022. Informa UK Limited. <http://dx.doi.org/10.1080/23311975.2022.2142997>.

MALAKAUSKAS, Aidas; LAKLTUTIENĖ, Aušrinė. Financial Distress Prediction for Small and Medium Enterprises Using Machine Learning Techniques. **Engineering Economics**, v. 32,

n. 1, p. 4-14, 26 fev. 2021. Kaunas University of Technology (KTU). <http://dx.doi.org/10.5755/j01.ee.32.1.27382>.

QIAN, Hongyi; WANG, Baohui; YUAN, Minghe; GAO, Songfeng; SONG, You. Financial distress prediction using a corrected feature selection measure and gradient boosted decision tree. **Expert Systems with Applications**, v. 190, p. 116202, mar. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.eswa.2021.116202>.

SANTANA FILHO, Júlio César; OLIVEIRA, Elis Regina; SANTOS, Geovane Camilo; OLIVEIRA, ElcioDihl. Análise dos índices de desempenho econômico-financeiro dos clubes de futebol do campeonato brasileiro de 2014 a 2018: antes e após o profut. **Brazilian Journal of Development**, v. 5, n. 7, p. 9733-9764, 2019. <http://dx.doi.org/10.34117/bjdv5n7-149>.

SANTOS, Allisson Silva dos; BESARRIA, Cassio da Nobrega; MACHADO, Márcio André Veras. Previsão de insolvência corporativa por meio de aprendizado de máquina. In: ENCONTRO DA ANPAD - ENANPAD, 46., 2022, On-Line. **Anais [...]**. On-Line: Enanpad, 2022. p. 1-22.

SERASA EXPERIAN. **Indicadores econômicos**. 2021. Disponível em: <https://www.serasaexperian.com.br/conteudos/indicadores-economicos/>. Acesso em: 06 jun. 2022.

SHAHWAN, Tamer Mohamed; HABIB, Ahmed Mohamed. Does the efficiency of corporate governance and intellectual capital affect a firm's financial distress? Evidence from Egypt. **Journal of Intellectual Capital**, v. 21, n. 3, p. 403-430, 5 mar. 2020. Emerald. <http://dx.doi.org/10.1108/jic-06-2019-0143>.

STUPP, Diego Rafael; FLACH, Leonardo; MATTOS, Luísa Karam de. Analysis of the impact of adopting international accounting standards in predicting the insolvency of businesses listed on the BM&FBovespa brazilian stock exchange. **Race - Revista de Administração, Contabilidade e Economia**, v. 17, n. 2, p. 397-422, 28 ago. 2018. Universidade do Oeste de Santa Catarina. <http://dx.doi.org/10.18593/race.v17i2.16094>.

SUSS, Joel; TREITEL, Henry. Predicting bank distress in the UK with machine learning. **Staff Working Paper: Bank of England**, p. 1-45, out. 2019

TRAN, Kim Long; LE, Hoang Anh; NGUYEN, Thanh Hien; NGUYEN, Duc Trung. Explainable Machine Learning for Financial Distress Prediction: evidence from vietnam. **Data**, v. 7, n. 11, p. 160, 14 nov. 2022. MDPI AG. <http://dx.doi.org/10.3390/data7110160>.

VODA, Alina Daniela; DOBROTĂ, Gabriela; ȚÎRCĂ, Diana Mihaela; DUMITRAȘCU, Dănuț Dumitru; DOBROTĂ, Dan. Corporate bankruptcy and insolvency prediction model. **Technological and Economic Development of Economy**, v. 27, n. 5, p. 1039-1056, 19 ago.2021. Vilnius Gediminas Technical University. <http://dx.doi.org/10.3846/tede.2021.15106>.

WANG, Gang; CHEN, Gang; CHU, Yan. A new random subspace method incorporating sentiment and textual information for financial distress prediction. **Electronic Commerce Research and Applications**, v. 29, p. 30-49, maio 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.elerap.2018.03.004>.

WU, Desheng; MA, Xiyuan; OLSON, David L. Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. **Decision Support Systems**, v. 159, p. 113814, ago. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.dss.2022.113814>.

ZHANG, Zijiao; WU, Chong; QU, Shiyong; CHEN, Xiaofang. An explainable artificial intelligence approach for financial distress prediction. **Information Processing & Management**, v. 59, n. 4, p. 102988, jul. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.ipm.2022.102988>.

ZHOU, Fanyin; FU, Lijun; LI, Zhiyong; XU, Jiawei. The recurrence of financial distress: a survival analysis. **International Journal of Forecasting**, v. 38, n. 3, p. 1100-1115, jul. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.ijforecast.2021.12.005>.