



LEANDRO FARIA DA SILVA

**PREVISÃO DO RISCO DE INSOLVÊNCIA EMPRESARIAL
NO BRASIL: UMA ABORDAGEM ESTATÍSTICA E DE
APRENDIZADO DE MÁQUINA**

**LAVRAS - MG
2026**

LEANDRO FARIA DA SILVA

**PREVISÃO DO RISCO DE INSOLVÊNCIA EMPRESARIAL NO BRASIL: UMA
ABORDAGEM ESTATÍSTICA E DE APRENDIZADO DE MÁQUINA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Administração, área de concentração em Estratégia de Negócios Globais e Finanças Corporativas, para a obtenção do título de Mestre.

Prof. Dr. Francisval de Melo Carvalho
Orientador

Prof. Dr. José Willer do Prado
Coorientador

**LAVRAS - MG
2026**

**Ficha Catalográfica elaborada pelo Sistema de Geração
de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados
informados pelo(a) próprio(a) autor(a).**

Silva, Leandro Faria da.

Previsão do risco de insolvência empresarial no Brasil : uma abordagem estatística e de aprendizado de máquina / Leandro Faria da Silva. - 2026. 207 p. : il.

Orientador: Francisval de Melo Carvalho
Coorientador: José Willer do Prado

Dissertação (Mestrado Acadêmico) - Universidade Federal de Lavras, 2026.
Bibliografia.

1. Risco de insolvência empresarial. 2. Análise multivariada. 3. Modelos de aprendizado de máquina. 4. Modelo dinâmico de Fleuriet. 5. Indicadores financeiros tradicionais. I. Melo Carvalho, Francisval de. II. Prado, José Willer do. III. Universidade Federal de Lavras. IV. Título.

LEANDRO FARIA DA SILVA

**PREVISÃO DO RISCO DE INSOLVÊNCIA EMPRESARIAL NO BRASIL: UMA
ABORDAGEM ESTATÍSTICA E DE APRENDIZADO DE MÁQUINA**

**PREDICTION OF CORPORATE INSOLVENCY RISK IN BRAZIL: A STATISTICAL
AND MACHINE LEARNING APPROACH**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Administração, área de concentração em Estratégia de Negócios Globais e Finanças Corporativas, para a obtenção do título de Mestre.

APROVADO em 23 de fevereiro de 2026.

Prof. Dr. Paulo Henrique Sales Guimarães UFLA

Prof. Dr. Ednilson Sebastião de Ávila UFLA

Dr. Francisval de Melo Carvalho
Orientador

Dr. José Willer do Prado
Coorientador

**LAVRAS - MG
2026**

*A minha noiva, Laís,
Dedico.*

AGRADECIMENTOS

Agradeço à Universidade Federal de Lavras (UFLA), ao Departamento de Administração e Economia (DAE) e ao Programa de Pós-Graduação em Administração da UFLA (PPGA/UFLA) pela oportunidade de cursar o mestrado, pelo espaço oferecido e pelo cuidado em cada procedimento, seja de natureza administrativa ou relacionado à minha formação como mestre.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos que, por meio deste recurso, possibilitou minha organização financeira e tornou viável a concretização desta realização pessoal e profissional.

Este trabalho, ainda que seja de minha autoria, foi construído a partir da orientação e da paciência do Professor Dr. Francisval de Melo Carvalho e do Professor Dr. José Willer do Prado, que não mediram esforços para esclarecer minhas dúvidas e conduzir meu aprendizado ao longo do programa, contribuindo de forma decisiva para a construção do meu conhecimento. Agradeço também a todos os professores do PPGA que me apoiaram e me ensinaram ao longo dessa trajetória. Aos professores do Departamento de Estatística (DES) com os quais tive contato, deixo meus sinceros agradecimentos por me acolherem em disciplina cursada e contribuições valiosas neste trabalho.

Agradeço a todos os meus colegas pela proza, pelas dicas e pelos momentos vividos no programa. Em especial, agradeço a doutoranda Taís por todo o apoio dado desde o início da minha formação, pelos trabalhos submetidos juntos e por me ensinar como funciona o universo da pesquisa acadêmica. Finalizo o programa de mestrado com este presente, a nossa amizade.

Agradeço, em especial, à Laís, com quem estou atualmente noivo, pela paciência, pelas orientações, pelas motivações nos momentos em que não acreditei no meu potencial, bem como pelo afeto, carinho e amor recebidos diariamente. Como sempre, guiando-me e dando-me forças para trilhar o caminho profissional no qual acredito. Gratidão por existir em minha vida e pela possibilidade de evoluir ao lado de alguém tão incrível.

Agradeço aos meus gatos, Floquinho e Leona, que, nos momentos de estresse ao tentar conciliar a intensa rotina de trabalho na iniciativa privada com o curso de mestrado, ofereceram conforto por meio de carinhos e afeto.

Aos meus pais, Cleide e Juscelino, pela oportunidade da vida.

Há mais pessoas e instituições com quem tenho o dever de agradecer. Mas esses eu deixo para agradecer na futura tese de doutorado.

*“Educar é crescer. E crescer é viver.
Educação é, assim, vida no sentido mais
autêntico da palavra.”
Anísio Teixeira.*

RESUMO

Esta dissertação teve como objetivo geral analisar o desempenho preditivo de técnicas estatísticas e de aprendizado de máquina na previsão do risco de insolvência de empresas brasileiras não financeiras listadas na Brasil, Bolsa, Balcão (B3), no período de 2010 a 2024. A pesquisa foi estruturada em três estudos empíricos que aplicaram diferentes metodologias a um conjunto de indicadores financeiros tradicionais e dinâmicos, relacionados principalmente à liquidez, endividamento, rentabilidade e gestão do capital de giro, buscando identificar as variáveis de maior capacidade explicativa na classificação de empresas solventes e insolventes. O primeiro estudo empregou a Análise Discriminante, que apresentou desempenho robusto, com AUC de 0,955 e acurácia de 93,3% na amostra de teste, evidenciando a relevância combinada de indicadores tradicionais e do Modelo Dinâmico de Fleuriet na separação dos grupos solvente e insolvente. O segundo estudo utilizou Regressão Logística com dados em painel, estimada por Efeitos Aleatórios, permitindo captar a variação entre empresas, a evolução temporal do risco e os efeitos individuais não observados. O modelo alcançou AUC de 0,915 e acurácia de 90,2%, destacando a importância de indicadores associados à estrutura de capital, liquidez, capital de giro e disponibilidade financeira. O terceiro estudo, de caráter comparativo, avaliou seis algoritmos de aprendizado de máquina, sendo eles a *Decision Tree*, *Random Forest*, *Support Vector Machine*, *K-Nearest Neighbors*, *Multilayer Perceptrons* e *Extreme Gradient Boosting*, revelando a superioridade das técnicas *ensemble*. O modelo desenvolvido por meio do *Extreme Gradient Boosting* obteve o melhor desempenho global, com AUC de 0,981 e acurácia de 95,4%, seguido de perto pelo *Random Forest*, com AUC de 0,978 e acurácia de 95,2%, evidenciando que indicadores de rentabilidade, estrutura de capital e liquidez, em conjunto, explicam a situação de insolvência empresarial. De forma integrada, os resultados mostram que tanto abordagens estatísticas quanto modelos de aprendizado de máquina são eficazes para antecipar a insolvência empresarial no contexto brasileiro, enquanto a análise de importância das variáveis confirma que indicadores tradicionais e dinâmicos relacionados à liquidez, endividamento, rentabilidade e capital de giro são determinantes na previsão e antecipação da insolvência. Sob a perspectiva científica, o estudo contribui ao incorporar múltiplas técnicas, utilizar uma amostra ampla e recente e integrar indicadores tradicionais e dinâmicos ao campo de estudos sobre insolvência, preenchendo lacunas na literatura nacional. No âmbito social, gerencial e institucional, os modelos propostos oferecem subsídios para melhorar políticas de crédito, avaliação de investimentos, renegociação de dívidas e práticas de monitoramento preventivo da saúde financeira, apoiando gestores, credores, investidores e órgãos reguladores na mitigação do risco de insolvência e na promoção da estabilidade econômica.

Palavras-chave: Risco de insolvência empresarial. Análise multivariada. Modelos de aprendizado de máquina. Modelo Dinâmico de Fleuriet. Indicadores Financeiros Tradicionais.

ABSTRACT

This dissertation aims to analyze the predictive performance of statistical and machine learning techniques in forecasting the risk of insolvency of Brazilian non-financial firms listed on B3 (Brasil, Bolsa, Balcão) over the period from 2010 to 2024. The research is structured into three empirical studies that apply different methodologies to a set of traditional and dynamic financial indicators, mainly related to liquidity, leverage, profitability, and working capital management, in order to identify the variables with the greatest explanatory power in the classification of solvent and insolvent firms. The first study employs Discriminant Analysis, which presents robust performance, with an AUC of 0.955 and an accuracy of 93.3% in the test sample, highlighting the combined relevance of traditional indicators and the Fleuriet Dynamic Model in separating solvent and insolvent groups. The second study uses panel data Logistic Regression estimated by Random Effects, allowing the model to capture cross-firm variation, the temporal evolution of insolvency risk, and unobserved individual effects, achieving an AUC of 0.915 and an accuracy of 90.2% and emphasizing the importance of indicators related to capital structure, liquidity, working capital, and financial availability. The third study, of a comparative nature, evaluates six machine learning algorithms—Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, Multilayer Perceptrons, and Extreme Gradient Boosting—and reveals the superiority of ensemble techniques, with the Extreme Gradient Boosting model obtaining the best overall performance, with an AUC of 0.981 and an accuracy of 95.4%, followed closely by Random Forest, with an AUC of 0.978 and an accuracy of 95.2%, indicating that profitability, capital structure, and liquidity indicators jointly explain corporate insolvency. In an integrated perspective, the results show that both statistical approaches and machine learning models are effective in anticipating corporate insolvency in the Brazilian context, while the analysis of variable importance confirms that traditional and dynamic indicators related to liquidity, leverage, profitability, and working capital are decisive in predicting and anticipating insolvency. From a scientific perspective, the study contributes by incorporating multiple techniques, using a large and recent sample, and integrating traditional and dynamic indicators into the field of insolvency research, thus filling gaps in the national literature. From a social, managerial, and institutional standpoint, the proposed models provide support for improving credit policies, investment appraisal, debt renegotiation, and preventive financial monitoring practices, assisting managers, creditors, investors, and regulators in mitigating insolvency risk and promoting economic stability.

Keywords: Corporate Insolvency risk. Multivariate analysis. Machine learning models. Fleuriet's Dynamic Model. Traditional Financial Indicators.

INDICADORES DE IMPACTO

O presente estudo apresenta impactos sociais, econômicos, tecnológicos e acadêmicos ao desenvolver modelos estatísticos e de aprendizado de máquina na previsão do risco de insolvência aplicados a empresas brasileiras não financeiras listadas na Brasil, Bolsa, Balcão (B3), assim como pela identificação dos indicadores financeiros mais relevantes para a análise da insolvência empresarial, derivados tanto das métricas tradicionais quanto do Modelo Dinâmico de Fleuriet, representando contribuições essenciais para gestores, investidores, instituições financeiras e órgãos reguladores, ao fornecer evidências objetivas sobre os fatores que aumentam ou reduzem o risco de deterioração financeira. No campo econômico, os achados permitem aprimorar políticas de crédito, avaliação de investimentos, renegociação de dívidas e o monitoramento contínuo da saúde financeira das empresas, fortalecendo a tomada de decisão estratégica e reduzindo assimetrias informacionais. Sob a perspectiva social, a antecipação de riscos possibilita intervenções preventivas que preservam empregos, estabilizam cadeias produtivas e promovem segurança econômica para diferentes públicos impactados pelas operações empresariais. No âmbito tecnológico, o desenvolvimento de modelos estatísticos e de aprendizado de máquina por meio dos softwares SPSS, R e Python, aliado à identificação das variáveis mais determinantes, estimula o uso de ferramentas analíticas avançadas e consolida práticas quantitativas inovadoras, ampliando o potencial de automação e precisão na análise do risco empresarial. Em termos acadêmicos e institucionais, a pesquisa contribui para preencher lacunas na literatura nacional ao integrar diferentes abordagens metodológicas, avaliar sua efetividade em um amplo conjunto de empresas brasileiras e evidenciar quais indicadores financeiros são mais relevantes para a previsão de insolvência. Essa combinação fortalece o campo de finanças corporativas, incentiva a replicação dos métodos, oferece base para estudos futuros e amplia a capacidade de formulação de políticas corporativas mais eficazes. Os impactos do trabalho alinham-se especialmente ao ODS 8 (Trabalho Decente e Crescimento Econômico), ao ODS 9 (Indústria, Inovação e Infraestrutura) e ao ODS 16 (Paz, Justiça e Instituições Eficazes), evidenciando que o aprimoramento da capacidade de prever insolvência, aliado à identificação dos fatores financeiros críticos, promove estabilidade organizacional, inovação analítica e maior transparência nos processos decisórios, fortalecendo a governança e a eficiência institucional.

IMPACT INDICATORS

The present study generates social, economic, technological, and academic impacts by developing statistical and machine learning models to predict insolvency risk in Brazilian non-financial companies listed on the Brasil, Bolsa, Balcão (B3), as well as by identifying the most relevant financial indicators for insolvency analysis, derived from both traditional metrics and the Dynamic Model of Fleuriet, representing essential contributions for managers, investors, financial institutions, and regulatory bodies by providing objective evidence on the factors that increase or reduce the likelihood of financial deterioration. From an economic perspective, the findings enable improvements in credit policies, investment evaluation, debt renegotiation, and the continuous monitoring of firms' financial health, strengthening strategic decision-making and reducing informational asymmetry. From a social standpoint, early risk detection allows preventive interventions that preserve jobs, stabilize production chains, and promote economic security for the various groups affected by corporate operations. In the technological sphere, the development of statistical and machine learning models using SPSS, R, and Python, combined with the identification of the most influential variables, encourages the adoption of advanced analytical tools and consolidates innovative quantitative practices, expanding the potential for automation and precision in corporate risk assessment. Academically and institutionally, the research contributes to addressing gaps in the national literature by integrating different methodological approaches, assessing their effectiveness across a broad set of Brazilian companies, and highlighting which financial indicators are most relevant for predicting insolvency. This combination strengthens the field of corporate finance, encourages methodological replication, supports future research, and enhances the capacity for formulating more effective corporate policies. The impacts of this work align especially with SDG 8 (Decent Work and Economic Growth), SDG 9 (Industry, Innovation and Infrastructure), and SDG 16 (Peace, Justice and Strong Institutions), demonstrating that improving insolvency prediction capacity, together with the identification of critical financial factors, promotes organizational stability, analytical innovation, and greater transparency in decision-making processes, thereby strengthening governance and institutional efficiency.

LISTA DE FIGURAS

Figura 1 – Estrutura do trabalho de dissertação.	24
Figura 2 – Fluxograma metodológico para classificação desta pesquisa.	33
Figura 3 – Estrutura da metodologia para a dissertação.	37
Figura 4 – Empresas classificadas como insolventes ao longo do tempo.	42
Figura 5 – Evolução das empresas insolventes no tempo.	45
Figura 6 – Quantidade anual de pedidos de recuperação judicial e extrajudicial.	48
Figura 7 – O Balanço Patrimonial no Modelo Dinâmico.	60
Figura 8 – Representação gráfica da Análise Discriminante de dois grupos.	69
Figura 9 – Termômetro de Kanitz da função discriminante.	81
Figura 10 – Curva ROC do modelo desenvolvido pela AD dos casos não selecionados.	83
Figura 11 – Representação gráfica a Regressão Logística.	104
Figura 12 – Curva ROC da Regressão Logística com Efeitos Aleatórios.	122
Figura 13 – Representação gráfica do Modelo DT.	144
Figura 14 – Representação gráfica do modelo RF.	145
Figura 15 – Representação gráfica do modelo SVM.	146
Figura 16 – Representação gráfica do modelo KNN.	148
Figura 17 – Representação gráfica do modelo não linear.	149
Figura 18 – Gráfico comparativo das curvas ROC.	158

LISTA DE QUADROS

Quadro 1 – Seleção de trabalhos sobre previsão de insolvência (continua).....	30
Quadro 2 – Indicadores do modelo dinâmico.....	49
Quadro 3 – Modelos tradicionais (continua).	50

LISTA DE TABELAS

Tabela 1 – Empresas categorizadas como insolventes (continua).....	39
Tabela 2 – Data da recuperação judicial e extrajudicial das empresas (continua).	46
Tabela 3 – Primeira classificação do Tipo de Estrutura Financeira de Fleuriet em 1978.	61
Tabela 4 – Classificação atual do Tipo de Estrutura Financeira de Fleuriet.	62
Tabela 5 – Classificação da amostra no Tipo de Estrutura Financeira.....	66
Tabela 6 – Variáveis Financeiras Utilizadas Neste Estudo (continua).....	67
Tabela 7 – VIF das variáveis utilizadas na construção do modelo.....	73
Tabela 8 – Estatística descritiva das variáveis.....	74
Tabela 9 – Resumo de processamento de caso de análise.	74
Tabela 10 – Resultado do teste Box's M.	75
Tabela 11 – Resultado do teste Wilk's Lambda.....	76
Tabela 12 – Coeficientes da função discriminante canônica não padronizada.	76
Tabela 13 – Funções em centroides de grupo.....	79
Tabela 14 – Probabilidades a priori para grupos.	80
Tabela 15 – Resultados da classificação.....	82
Tabela 16 – Métricas de desempenho do modelo obtidas a partir da Curva ROC.....	84
Tabela 17 – Variáveis Utilizadas na Construção do Modelo (continua).	102
Tabela 18 – VIF das variáveis a Regressão Logística para Dados em Painel.	110
Tabela 19 – Estatísticas descritivas as variáveis selecionadas.	111
Tabela 20 – Comparação dos modelos <i>Pooled Logit</i> e Nulo.	112
Tabela 21 – Resultado o teste Breusch-Pagan.....	113
Tabela 22 – Resultado do teste de Hausman.	113
Tabela 23 – Comparativo de coeficientes dos modelos logísticos.	114
Tabela 24 – Resultado do modelo de Efeitos Aleatórios selecionado.....	117
Tabela 25 – Métricas de performance do modelo.	122
Tabela 26 – Variáveis Utilizadas na Construção dos Modelos (continua).....	139
Tabela 27 – VIF das variáveis financeiras (continua).	155
Tabela 28 – Estatísticas descritivas dos indicadores financeiros.	156
Tabela 29 – Comparativo do desempenho dos modelos de aprendizado de máquina.	157
Tabela 30 – Importância das cinco principais variáveis.....	160

LISTA DE SIGLAS

AC	Ativo Circulante
AD	Análise Discriminante
AF	Ativo Financeiro
ALP	Ativos de Longo Prazo
ANC	Ativo Não Circulante
AO	Ativo Operacional
AT	Ativo Total
AUC	<i>Area Under the Curve</i>
B3	Brasil, Bolsa, Balcão
CDG	Capital de Giro
CS	Capital Social
CVM	Comissão de Valores Mobiliários
DB	Dívida Bruta
DL	Dívida Líquida
DT	<i>Decision Tree</i>
EBIT	<i>Earnings Before Interests and Taxes</i>
ILD	Índice de Liquidez Dinâmico
KNN	<i>K-Nearest Neighbors</i>
LL	Lucro Líquido
LSTM	<i>Long Short-Term Memory</i>
MLP	<i>Multilayer Perceptrons</i>
NCG	Necessidade do Capital de Giro
PF	Passivo Financeiro
PL	Patrimônio Líquido
PLN	Patrimônio Líquido Negativo
PNC	Passivo Não Circulante
PNCF	Passivo Não Circulante Financeiro
PO	Passivo Operacional
PT	Passivo Total
RecL	Receita Líquida
REJ	Recuperação Extrajudicial
RF	<i>Random Forest</i>

RJD	Recuperação Judicial
RL	Regressão Logística
RLO	Receita Líquida Operacional
RLP	Realizável a Longo Prazo
ROA	<i>Return on Assets</i>
ROC	<i>Receiver Operating Characteristic</i>
ROE	<i>Return on Equity</i>
SPSS	<i>Statistical Package for the Social Sciences</i>
SVM	<i>Support Vector Machine</i>
T	Saldo de Tesouraria
TEF	Tipo de Estrutura Financeira
VIF	Variance Inflation Factor
XGBoost	<i>Extreme Gradient Boosting</i>

SUMÁRIO

	PRIMEIRA PARTE	18
1	INTRODUÇÃO	18
1.1	Contextualização do tema	18
1.2	Problema de pesquisa	20
1.3	Objetivos	21
1.3.1	Geral	21
1.3.2	Específicos	21
1.4	Justificativas da Pesquisa	22
1.5	Estrutura do Trabalho	23
2	REFERENCIAL TEÓRICO	25
2.1	Insolvência Empresarial	25
2.2	Risco de Insolvência	27
3	METODOLOGIA GERAL	33
3.1	Definição da Amostra Corte Transversal	39
3.2	Definição da Amostra Dados em Painel	43
3.3	Definição dos Indicadores	48
	SEGUNDA PARTE – ARTIGOS	54
	ARTIGO 1 – PREVISÃO DE INSOLVÊNCIA EM EMPRESAS NÃO FINANCEIRAS LISTADAS NA B3: UMA ABORDAGEM UTILIZANDO A ANÁLISE DISCRIMINANTE	55
1	INTRODUÇÃO	57
2	REFERENCIAL TEÓRICO	59
2.1	Modelos Tradicionais	59
2.2	O Modelo Dinâmico de Fleuriet	59
2.3	Estudos de Previsão de Insolvência com Análise Discriminante	62
3	METODOLOGIA	65
3.1	Amostra	65
3.2	Definição dos Indicadores	67
3.3	Análise Discriminante	69
3.3.1	Procedimentos e Teste do Modelo Discriminante	71
4	RESULTADOS E DISCUSSÃO	73
4.1	Teste de Multicolinearidade entre as variáveis independentes	73
4.2	Construção do modelo Análise Discriminante	74
5	CONSIDERAÇÕES FINAIS	85
	REFERÊNCIAS	87
	ARTIGO 2 – UMA ABORDAGEM DO RISCO DE INSOLVÊNCIA DE EMPRESAS BRASILEIRAS NÃO FINANCEIRAS UTILIZANDO A REGRESSÃO LOGÍSTICA	
1	INTRODUÇÃO	93
2	REFERENCIAL TEÓRICO	95
3	METODOLOGIA	101
3.1	Definição da Amostra	101

3.2	Definição dos Indicadores	101
3.3	Modelo Longitudinal Logístico para Dados em Painel	103
3.3.1	Procedimentos e Teste do Modelo Logístico para Dados em Painel	105
4	RESULTADOS E DISCUSSÃO	110
4.1	Verificação de colinearidade entre as variáveis independentes e estatísticas descritivas	110
4.2	Desenvolvimento da Regressão Logística para Dados em Painel.....	112
4.2.1	Testes de especificação e seleção do modelo	112
4.2.2	Análise do Modelo Final de Efeitos Aleatórios	116
4.2.3	Avaliação da Performance Preditiva do Modelo	121
5	CONSIDERAÇÕES FINAIS	124
	REFERÊNCIAS	126
	ARTIGO 3 – PREVISÃO DO RISCO DE INSOLVÊNCIA: UMA ANÁLISE COMPARATIVA ENTRE DIFERENTES MÉTODOS DE APRENDIZADO DE MÁQUINA	130
1	INTRODUÇÃO	132
2	REFERENCIAL TEÓRICO	134
3	METODOLOGIA	138
3.1	Definição da Amostra	138
3.2	Definição dos Indicadores	139
3.3	Técnicas de Aprendizado de Máquina.....	140
3.3.1	Decision Tree (DT).....	143
3.3.2	Random Forest (RF).....	144
3.3.3	Support Vector Machine (SVM)	146
3.3.4	K-Nearest Neighbors (KNN).....	147
3.3.5	Neural Networks (NN).....	148
3.3.6	Extreme Gradient Boosting (XGBoost)	150
3.4	Procedimentos de Estimação e Avaliação dos Modelos	150
4	RESULTADOS E DISCUSSÃO	155
4.1	Análise de multicolinearidade e das estatísticas descritivas	155
4.2	Análise da construção dos modelos	157
4.2.1	Interpretação dos modelos desenvolvidos.....	162
5	CONSIDERAÇÕES FINAIS	165
	REFERÊNCIAS	167
	TERCEIRA PARTE – CONSIDERAÇÕES FINAIS	173
	REFERÊNCIAS	177
	APÊNDICE A – Código em R para construção da Regressão Logística para Dados em Painel do Artigo 2	186
	APÊNDICE B – Código em Python para elaboração dos modelos de aprendizado de máquina do Artigo 3	192

PRIMEIRA PARTE

1 INTRODUÇÃO

Nesta seção será fornecida uma compreensão inicial da temática objeto do estudo. Para isso, a introdução será dividida em subtópicos, sendo apresentados a contextualização do tema, o problema de pesquisa, seguindo para o objetivo geral do estudo, os objetivos específicos que serão desenvolvidos por meio de artigos, as justificativas da pesquisa, demonstrando a importância da temática. Por fim, será mostrado a estrutura do trabalho de forma esquematizada.

1.1 Contextualização do tema

O termo risco possui diversas definições na literatura, contudo, temos uma percepção de senso comum que é a probabilidade de ocorrência de dois ou mais eventos, sendo desconhecido qual será o resultado a partir de uma ação ou escolha (Areosa, 2010). À luz da definição, é possível concluir que embora a noção de risco possa prognosticar aspectos positivos, em geral, o termo é pensado como uma consequência negativa devido a expectativa de um determinado evento esperado não ocorrer, trazendo consequências indesejadas.

Os riscos sempre fizeram parte da condição da vida humana (Amaro, 2005). A título de exemplo, ao sair de casa sem o guarda-chuva em um dia com previsão de chuva, na expectativa de obtenção do conforto de não carregar o objeto na trajetória, a pessoa assume o risco. Nesse caso, chover é um fator inesperado e indesejado no qual o indivíduo assume o risco diante de sua escolha.

No contexto empresarial, toda tomada de decisão, seja em nível operacional ou estratégico, envolve algum grau de risco. Para realizar a gestão do risco, é necessário compreendê-lo em suas diferentes dimensões e adotar práticas que possibilitem sua identificação, análise e controle. Segundo Dionne (2013), a gestão de riscos evoluiu como uma disciplina multidisciplinar, integrando conhecimentos de economia, finanças e ciências atuariais, com o objetivo de minimizar perdas potenciais e aumentar a eficiência organizacional. Dentre os diversos tipos de risco enfrentados pelas organizações, destaca-se o risco de insolvência, que representa a possibilidade de a empresa não conseguir honrar seus compromissos financeiros.

Sendo assim, uma definição de solvente é, segundo Almeida, Sales e Nunes (2023), quando uma empresa possui a capacidade de cumprir com suas obrigações correntes e que apresente uma situação patrimonial e resultados positivos que gerem lucros para garantia de sobrevivência futura. Já a insolvência é a incapacidade da companhia em liquidar as suas obrigações correntes a cada vencimento (Minatto; Borba, 2022). A compreensão desses conceitos é fundamental para a implementação de estratégias eficazes que assegurem a saúde financeira das organizações e previnam crises econômicas. Além disso, a análise contínua da solvência e insolvência permite que os gestores identifiquem sinais de alerta precoces, possibilitando intervenções oportunas para manter a estabilidade financeira e operacional.

A ausência da gestão do risco de insolvência pode trazer consequências para a empresa, tais como aumento no custo do capital à medida em que o risco de insolvência vem sendo percebido por credores e investidores, onde podem exigir taxa de juros ou retornos maiores sobre o investimento para compensar o risco (Closset *et al.*, 2021). Ainda, a empresa com dificuldade financeira pode levar a perda de credibilidade e confiança com fornecedores, clientes e parceiros (Jafaar *et al.*, 2018), fazendo com que a empresa sofra alguma interrupção operacional ou recusa na captação de recursos próprios ou de terceiros, levando à liquidação de suas atividades de forma compulsória (Agrawal, 2023).

Estudos demonstram uma crescente na temática risco de insolvência empresarial (Prado *et al.*; 2016; Shi; Li, 2019; Almeida, 2023; Dasilas; Rigani, 2024; Roy; Vasa, 2025), o que evidencia o interesse pela sociedade no âmbito social e da pesquisa em responder as causas e efeitos na temática, bem como desenvolver modelos que possibilitem sua previsão, principalmente em cenários de instabilidade econômica.

Desta forma, torna-se relevante investigar formas de identificação antecipada da insolvência, pois sua contribuição pode levar a credores decidirem para quem ou não emprestar o recurso econômico e a promover a resiliência organizacional das empresas por meio de mudanças estratégicas a fim de manter a empresa com saúde financeira.

Nesse cenário, a literatura tem avançado na formulação de modelos preditivos que relacionam indicadores contábeis e econômicos à probabilidade de insolvência, buscando identificar sinais de deterioração financeira e apoiar decisões empresariais. Contudo, as aplicações dessas técnicas ainda são pouco exploradas no contexto brasileiro, o que reforça a necessidade de estudos que desenvolvam modelos de previsão por meio de abordagens estatísticas multivariadas e de aprendizado de máquina.

Diante do exposto, evidencia-se a complexidade do risco de insolvência no cenário empresarial atual, exigindo modelos preditivos que antecipem sinais de deterioração financeira.

Este trabalho explora técnicas de previsão para identificar precocemente fatores críticos que podem levar à falência e fortalecer a resiliência organizacional. A construção de modelos com diferentes técnicas estatísticas e de aprendizado de máquina visa ampliar a compreensão dos determinantes da insolvência e oferecer ferramentas práticas para mitigar riscos e assegurar a sustentabilidade das empresas a longo prazo.

1.2 Problema de pesquisa

Ainda que se tenha uma crescente na literatura envolvendo a temática da previsão do risco de insolvência, ainda existem lacunas a serem preenchidas, principalmente no contexto brasileiro. Isso ocorre em razão de estudos nacionais adotarem recortes setoriais ou amostras específicas, o que dificulta a generalização dos modelos de previsão para diferentes contextos econômicos e empresariais (Barbosa; Nobre, 2023). Os autores ainda destacam que os modelos estatísticos tradicionais frequentemente utilizados no Brasil apresentam desempenho modesto em termos de acurácia e capacidade preditiva, comprometendo sua utilidade como ferramentas confiáveis de apoio à tomada de decisão.

Outro ponto é a escassez de estudos que promovam uma análise comparativa entre técnicas de inteligência computacional aplicadas ao contexto brasileiro. A ausência de abordagens integrativas restringe o avanço do conhecimento sobre quais métodos são mais adequados para diferentes cenários de previsão de insolvência (Aranha; Gondrigo, 2024).

De forma complementar, Fontalvo-Herrera e De la hoz-Granadillo (2020) afirmam que é necessário expandir a aplicação dessas técnicas para abranger uma gama mais ampla de variáveis e um maior volume de empresas em vários setores poderia aumentar a robustez das avaliações de eficiência. Essa abordagem facilita uma compreensão mais abrangente do desempenho dos negócios, beneficiando gestores, credores e investidores com modelos mais robustos e aplicáveis a realidade brasileira.

Desta forma, realizar estudos levando em consideração diferentes técnicas de previsão, assim como analisar diversas variáveis financeiras que impactam na classificação do risco de insolvência são necessários, levando em consideração empresas nos seus diferentes setores e contextos inseridos. À luz das considerações acima mencionadas, este estudo se concentra em responder ao seguinte **problema de pesquisa**: Quais técnicas estatísticas multivariadas e de aprendizado de máquina apresentam melhor desempenho preditivo na identificação do risco de insolvência de empresas não financeiras listadas na Brasil, Bolsa, Balcão (B3), levando em consideração os indicadores financeiros dos modelos tradicionais e dinâmicos?

1.3 Objetivos

1.3.1 Geral

Para responder ao problema de pesquisa, este estudo tem como **objetivo geral** analisar o desempenho preditivo de modelos estatísticos e de aprendizado de máquina na previsão do risco de insolvência de empresas brasileiras não financeiras listadas na Brasil, Bolsa, Balcão (B3), no período de 2010 a 2024, considerando as variáveis explicativas de indicadores financeiros tradicionais e do modelo dinâmico de Fleuriet extraídas das demonstrações financeiras, identificando as variáveis mais relevantes para classificação de empresas solventes e insolventes. Com isso, pretende-se contribuir com o avanço do conhecimento científico na área de finanças, ao testar diferentes técnicas de previsão sob o contexto brasileiro, além de oferecer aplicações práticas que auxiliem gestores e analistas na antecipação de sinais de deterioração financeira. O estudo também visa preencher lacunas existentes na literatura nacional no desenvolvimento de modelos com abordagens metodológicas distintas em um contexto amplo e atualizado do ambiente empresarial brasileiro.

1.3.2 Específicos

Para atingir o objetivo geral, os seguintes objetivos específicos são estabelecidos:

- (Artigo 1) Analisar o desempenho classificatório do risco de insolvência, desenvolvendo um modelo de previsão por meio da técnica estatística Análise Discriminante e quais são as variáveis que mais influenciam na classificação de empresas solventes e insolventes.
- (Artigo 2) Avaliar a construção do modelo com a Regressão Logística na previsão do risco de insolvência empresarial, bem como identificar quais foram as variáveis independentes que contribuíram na classificação do risco.
- (Artigo 3) Comparar o desempenho preditivo no desenvolvimento de diferentes modelos de aprendizado de máquina na previsão do risco de insolvência e identificar quais variáveis explicativas mais importantes foram determinantes na previsão de insolvência empresarial.

1.4 Justificativas da Pesquisa

A temática do risco de insolvência por meio do uso de técnicas estatísticas e de aprendizado de máquina para previsão tem ganhado relevância em diversos contextos organizacionais, estando presente em processos de identificação e quantificação dos riscos financeiros que impactam diretamente todas as partes interessadas. Segundo Wyrobek (2018), antecipar o risco de insolvência é um elemento indispensável para tomada de decisão de credores e investidores, assim como para gestores no que tange à reorganização financeira prévia para fortalecimento da saúde financeira das empresas, garantindo a continuidade do negócio em um mercado dinâmico e incerto.

Como **relevância gerencial**, a pesquisa se justifica pela necessidade das empresas desenvolverem mecanismos que possibilitem a identificação precoce de situações de vulnerabilidade financeira. A aplicação de modelos preditivos atribui a gestores maior capacidade de agir preventivamente antes do agravamento da situação e comprometimento da saúde financeira (Acosta-González *et al.*, 2019; Cunha; Penedo; Barboza, 2024; Vásquez-Serpa *et al.*, 2025). Devido a competitividade das companhias, a utilização de métodos estatísticos e de aprendizado de máquina colabora na sobrevivência das organizações (Wyrobek, 2018). Além disso, os resultados podem reestruturar as políticas de crédito, renegociação de dívidas, avaliação de investimentos e reestruturação corporativa, colaborando na tomada de decisão em níveis estratégicos (Valdes; Aleaga; Vidal, 2014).

Sob a ótica da **relevância social**, a insolvência empresarial gera impactos que podem transcender os limites das próprias empresas. Os modelos de previsão podem auxiliar na intervenção oportuna, garantindo estabilidade econômica e geração de empregos e de todos os agentes econômicos envolvidos na cadeia (Jafaar *et al.*, 2018; Lima; Paulino; Fávero, 2022; Barbosa; Nobre, 2023).

No que se refere a **relevância acadêmica**, o presente estudo busca preencher lacunas ainda existentes na literatura nacional, principalmente na construção de diferentes métodos estatísticos e de aprendizado de máquina aplicadas à previsão de falência. Embora existam estudos relevantes no Brasil, ainda há a carência de estudos com maior quantidade de empresas independente do setor (Barbosa; Nobre, 2023), bem como uma maior abrangência de técnicas de previsão e de variáveis predictoras (Fontalvo-Herrera; De la hoz-Granadillo, 2020). Com isso, este estudo visa contribuir para o avanço no conhecimento científico na área de finanças corporativas e estimular novos estudos que possam ampliar e validar os achados apresentados, favorecendo o amadurecimento teórico e metodológico do campo.

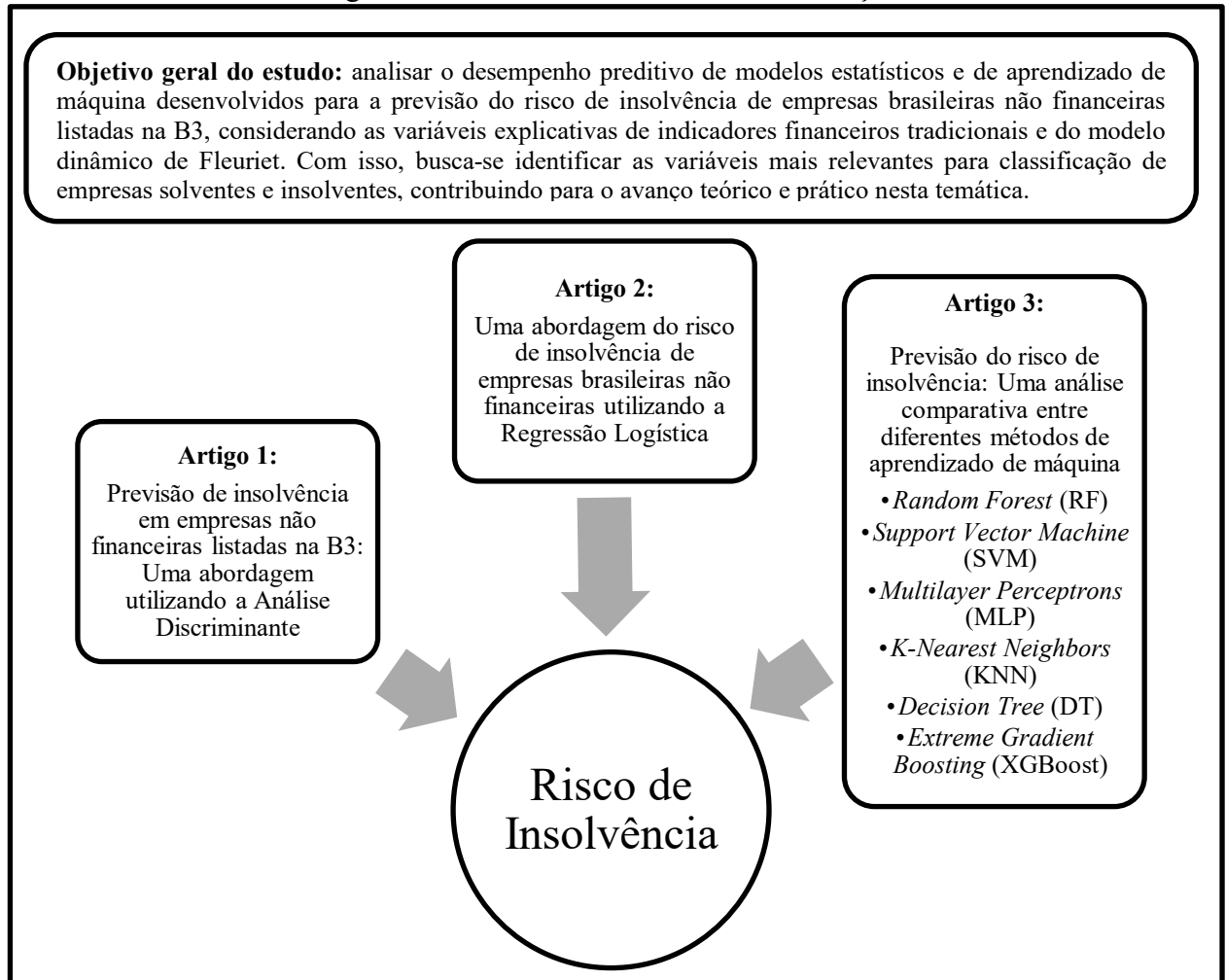
Na esfera **regulatória e institucional**, este estudo tem a sua relevância pela contribuição de métodos que podem ser utilizados em órgãos reguladores, instituições financeiras e investidores controladores ou não. Modelos de previsão mais precisos podem auxiliar na formulação de políticas públicas voltadas à mitigação de riscos sistêmicos e à promoção de um ambiente empresarial mais estável (Arora; Singh, 2020; Lee; Choi; Yoo, 2020).

Em síntese, a presente pesquisa se justifica pela relevância multidimensional do tema abordado. Ao construir modelos com múltiplas abordagens estatísticas e aprendizado de máquina na previsão do risco de insolvência, o estudo visa não apenas contribuir para o avanço teórico e prático no campo das finanças, mas também oferecer subsídios concretos para a gestão empresarial, a formulação de políticas públicas e a promoção da estabilidade socioeconômica. O aprofundamento dessa temática revela-se, portanto, um passo essencial diante da crescente complexidade do ambiente corporativo e da demanda por decisões mais embasadas em dados e modelos analíticos de alta precisão.

1.5 Estrutura do Trabalho

Esta dissertação foi estruturada em duas partes complementares e integradas, com o propósito de oferecer uma visão abrangente e aprofundada do tema. A primeira parte contempla a introdução geral do tema objeto do estudo, destacando a sua relevância no contexto da área de finanças, além da apresentação do referencial teórico que fundamenta a pesquisa e a metodologia geral adotada. A segunda parte é composta por artigos independentes que integram este estudo, sendo apresentado, individualmente, introdução, referencial teórico, metodologia específica, resultados e discussão e considerações finais, de acordo com os objetivos e abordagens adotadas em cada estudo. Na Figura 1, a seguir, ilustra-se a estrutura conceitual da dissertação, destacando seus principais componentes da pesquisa e técnicas metodológicas utilizadas para alcançar os resultados e atingir o objetivo geral.

Figura 1 – Estrutura do trabalho de dissertação.



Fonte: Elaborado pelo autor (2025).

Evidencia-se por meio da Figura 1 que os três artigos concentram-se na temática do risco de insolvência. No primeiro artigo será realizado um estudo por meio da construção de um modelo utilizando a técnica estatística multivariada Análise Discriminante, a fim de validar o seu poder discriminante em empresas brasileiras, considerando diversos indicadores financeiros. No segundo artigo objetiva-se investigar a previsão de insolvência com o desenvolvimento de um modelo utilizando a Regressão Logística, com a finalidade de compreender o seu desempenho conforme amostra coletada e variáveis de controle selecionadas. Por fim, o terceiro artigo compreende em verificar a previsão de insolvência por meio do desenvolvimento de diferentes técnicas de aprendizado de máquina, com o objetivo de analisar quais os modelos possuem melhor capacidade de previsão e robustez. Com isso, espera-se avaliar o desempenho preditivo das diferentes abordagens estatísticas e de aprendizado de máquina, permitindo identificar os modelos mais eficazes e variáveis determinantes da previsão do risco de insolvência.

2 REFERENCIAL TEÓRICO

Neste tópico será apresentado o referencial teórico geral da dissertação. Buscou-se fundamentar conceitos essenciais para melhor compreensão da temática, sendo conceituados a insolvência empresarial e o risco de insolvência.

2.1 Insolvência Empresarial

Existem diversos fatores que podem levar uma empresa a ter dificuldades financeiras. De acordo com Assaf Neto e Silva (2012), a administração do capital de giro é um fator relevante para manter boa saúde financeira da empresa, onde a má gestão do ciclo financeiro da companhia poderá levá-la a falência. De acordo com Lisboa (2014), a análise adequada das demonstrações contábeis é essencial para sustentar as decisões gerenciais e garantir a estabilidade financeira das empresas.

No que se refere à alavancagem financeira, empresas muito endividadas ou com altos custos financeiros tendem a ter dificuldades financeiras, principalmente quando não conseguem administrar a sua dívida de forma eficaz e alavancar seu negócio com capital de terceiros (Teixeira; Brites, 2020; Adiputra; Ruslim, 2023).

Outro fator que corrobora com a insolvência empresarial é a ausência de governança corporativa. Segundo Crisóstomo e Melo Júnior (2015), empresas com boas práticas de governança corporativa tendem a apresentar indicadores financeiros mais saudáveis, sendo que a falta dessas práticas contribui para maior fragilidade financeira.

Segundo Jonnardi, Bangun e Natsir (2023), o fator que leva as empresas a terem dificuldades financeiras é a lucratividade econômica, onde quanto menor, maior é a probabilidade da empresa possuir dificuldade de liquidar com suas obrigações.

Estudos realizados com indicadores financeiros como os de Prado *et al.* (2020), Santos (2021) e Barbosa e Nobre (2023) demonstram que uma gestão adequada de indicadores financeiros atrelados a liquidez, endividamento e rentabilidade evitam o risco de crédito, enquanto a má administração do conjunto desses indicadores pode levar a empresa ao não cumprimento de suas obrigações.

Com relação a fatores externos, Huo, Chan e Miller (2024) destacam que pandemias, como a COVID-19 ocorrida entre 2019 e 2021, fatores políticos, mudança nos mercados e no comportamento do consumidor, somados a morosidade da adaptação das empresas, podem corroborar com as dificuldades financeiras e falência empresarial.

Diante desses desafios, as empresas buscam de forma legal se reorganizar financeiramente para a luta pela sobrevivência e quitação de suas dívidas. No Brasil, a lei vigente que regula a recuperação judicial, a extrajudicial e a falência do empresário e da sociedade empresária é a Lei nº 11.101, de 9 de fevereiro de 2005 (Brasil, 2005), revogando o Decreto-Lei nº 7.661, de 21 de junho de 1945 (Brasil, 1945), apresentando uma significativa necessidade de modernização dos instrumentos políticos aplicáveis à crise empresarial. Tal implementação visou a substituição de dispositivos ultrapassados e a criação de mecanismos que conciliem a preservação da atividade econômica com a justa satisfação dos credores.

Nesse sentido, a finalidade da legislação brasileira é de possibilitar a reestruturação financeira de empresa em dificuldades, promovendo o cumprimento das obrigações junto aos credores e a continuidade do negócio gerador de empregos e de renda. As recentes alterações introduzidas pela Lei nº 14.112 de 24 de dezembro de 2020 trouxeram inovações relevantes, muito em razão da crise econômica advinda pela pandemia do COVID-19, onde facilitou o acesso a financiamentos durante o período de recuperação, aprimorou os instrumentos de renegociação de dívidas e a aceleração na alienação de ativos, aumentando desta forma a eficácia dos mecanismos de recuperação judicial, extrajudicial e falência (Brasil, 2020). Com isso, observa-se que o atual arcabouço legal reflete com sua evolução normativa o acompanhamento e adaptação ao mercado dinâmico em que as organizações são expostas diariamente.

De acordo com Winter e Oliveira (2013), a recuperação judicial é um processo que envolve a apresentação de um plano de recuperação, que deve ser aprovado em assembleia geral de credores e homologado por juiz competente, que acompanhará toda a execução do plano, com a finalidade da empresa poder se reestruturar e manter as suas atividades, preservando os empregos e garantindo a continuidade do negócio.

Já a recuperação extrajudicial é uma alternativa mais ágil e menos onerosa que a recuperação judicial, sendo um acordo entre a empresa devedora e seus credores, havendo posterior homologação judicial apenas para validar de forma legal o plano. Segundo Moraes, Donadello e Andrade (2019), esse modelo permite maior autonomia na negociação, menor custo no processo. Embora seja menos utilizado do que a via judicial, a recuperação extrajudicial é fundamental para empresas que buscam resolver sua reestruturação financeira sem o envolvimento direto do Judiciário ao longo de todo o processo.

Por fim, a falência é decretada quando a empresa se mostra insolvente, ou seja, incapaz de cumprir suas obrigações financeiras, tornando inviável sua continuidade. Nesse caso, o processo visa a liquidação dos ativos da empresa para pagamento dos credores, seguindo uma

ordem legal de prioridade. Segundo Silva e Saito (2020), a falência, embora seja uma medida extrema, desempenha um papel importante no ambiente de negócios ao permitir a realização eficiente de recursos e preservar a confiança no sistema econômico. Contudo, os autores ainda destacam que a legislação brasileira busca evitar esta situação sempre que houver viabilidade econômica para reestruturação do negócio.

2.2 Risco de Insolvência

O risco de insolvência pode ser compreendido como a probabilidade de uma empresa não conseguir cumprir suas obrigações financeiras à medida que estas vencem, sendo um indicativo da fragilidade econômico-financeira de uma organização. Diferente da insolvência concretizada, o risco possui natureza preditiva e representa um alerta para a gestão financeira, investidores e credores, pois antecipa potenciais eventos de falência ou quebra empresarial (Chaglla, 2019).

À luz das possíveis ramificações da insolvência, pesquisadores formularam estruturas preditivas e examinaram seus impactos nas organizações. A temática surgiu no início do século XX, com destaque para o estudo seminal de Beaver (1966), que utilizou análise univariada para investigar a capacidade preditiva de indicadores financeiros isolados. Segundo Prado *et al.* (2016), o primeiro uso de técnica multivariada foi o seminal de Altman (1968), com a análise discriminante (AD) para classificar empresas solventes e insolventes. Já os primeiros estudos com aprendizado de máquina foram os de Ohlson (1980), Odom e Sharda (1990) e Tam (1991), ao introduzirem a Regressão Logística e Redes Neurais na previsão do risco de insolvência.

Estudos no exterior evidenciam a utilização de técnicas estatísticas multivariadas tradicionais para classificação e previsão, como Xie, Luo e Yu (2011), Affes e Hentati-Kaffel (2017), Lukason e Andresson (2019), Xhindi e Shestani (2020), Gajdosikova e Valaskova (2023) e Huo, Chan e Miller (2024). Esses estudos apresentam modelos como análise discriminante e regressão logística com boa capacidade de previsão e classificação.

Estudos recentes que utilizam técnicas de aprendizado de máquina para previsão de insolvência como Aljawazneh *et al.* (2021), Garcia (2022), Noh (2023) e Idhmad *et al.* (2024) evidenciam uma melhor acurácia nos modelos de previsão com relação aos modelos estatísticos tradicionais. De acordo com Vásquez-Serpa *et al.* (2025), as técnicas de aprendizado de máquina comumente utilizadas para os estudos na temática risco de insolvência são as *Multilayer Perceptrons* (MLP), *Random Forest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree* (DT) e *Extreme Gradient Boosting* (XGBoost).

No Brasil, estudos de previsão de insolvência foram realizados, podendo citar Kanitz (1974), Elizabetsky (1976), Altman, Baidya e Dias (1979), Silva (1982), Sanvicente e Minardi (1998), Akiama (2008), Stüpp (2015), Prado *et al.* (2018), Lima, Paulino e Fávero (2022), Shio e Sampaio (2024), nos quais utilizaram técnicas estatísticas como Regressão Múltipla, Análise Discriminante e Regressão Logística.

Estudos de previsão risco de insolvência que utilizam a aprendizado de máquina realizados no Brasil, mencionam-se os trabalhos de Prado *et al.* (2020), Pinto (2021), Santos (2021), Jesus e Besarria (2023), Cunha, Penedo e Barboza (2024) e Tavares *et al.* (2025) nos quais utilizaram técnicas como *Multilayer Perceptrons*, *Random Forest*, *Support Vector Machine* e *Decision Tree*.

Apesar das contribuições desses estudos para a previsão de insolvência no Brasil, algumas limitações devem ser destacadas. A análise se restringe a recortes específicos de dados e de variáveis independentes, o que limita a generalização dos resultados. A literatura nacional carece de modelos preditivos com alta acurácia e aplicabilidade, como apontam Barbosa e Nobre (2023). Aranha e Gondrige (2024) destacam a necessidade de aprimorar modelos existentes em razão das diferentes possibilidades de inserir nos modelos as variáveis predictoras e outras variáveis que podem afetar o desempenho das companhias. Além disso, há a carência nos estudos em integrar variáveis explicativas como indicadores financeiros tradicionais e dinâmicos na previsão de insolvência (Prado *et al.*, 2020).

A compreensão teórica do risco de insolvência requer o exame da evolução dos modelos de previsão que buscam identificar, a partir de dados financeiros, a probabilidade de falência ou deterioração patrimonial das empresas. Nesse sentido, diversos estudos clássicos e contemporâneos contribuíram para a consolidação desse campo de pesquisa, tanto no exterior quanto no Brasil.

Os estudos sobre risco de insolvência evoluíram a partir dos modelos estatísticos clássicos de previsão, que consolidaram a relação entre indicadores de liquidez, rentabilidade e endividamento e a solvência empresarial. Essas variáveis continuam sendo a base dos modelos preditivos contemporâneos, agora aprimorados por técnicas de aprendizado de máquina que ampliam a precisão e a capacidade de generalização.

Durante as décadas seguintes, os modelos estatísticos dominaram a literatura. No Brasil, técnicas como análise discriminante e regressão logística foram amplamente empregadas, a exemplo dos trabalhos de Kanitz (1974), Altman, Baidya e Dias (1979), Sanvicente e Minardi (1998), Akiama (2008) e mais recentemente Stüpp (2015), Lima, Paulino e Fávero (2022) e

Shio e Sampaio (2024). Esses estudos reforçaram a importância dos indicadores de liquidez, endividamento e rentabilidade na previsão da solvência empresarial.

O avanço das tecnologias e o aumento da disponibilidade de dados levaram à incorporação das técnicas de aprendizado de máquina na temática da previsão de falências. Conforme já mencionado, os primeiros esforços nesse sentido foram realizados por Odom e Sharda (1990) e Tam (1991), que utilizaram Redes Neurais Artificiais (*Neural Networks*) para melhorar a capacidade preditiva dos modelos. Posteriormente, técnicas como *Support Vector Machines* (SVM), *Decision Trees* (DT), *Random Forest* (RF), *K-Nearest Neighbors* (KNN) e *Extreme Gradient Boosting* (XGBoost) passaram a ser empregadas com resultados expressivos, como evidenciado nos estudos de Aljawazneh *et al.* (2021), Garcia (2022), Noh (2023) e Idhmad *et al.* (2024).

Pesquisas mais recentes têm comparado o desempenho de técnicas estatísticas com algoritmos de aprendizado de máquina. Por exemplo, Shetty, Musa e Brédart (2022) utilizaram SVM, XGBoost e Redes Neurais para prever falências de empresas belgas, alcançando acurácia superior a 82% com apenas três indicadores financeiros. De forma semelhante, Kothuru *et al.* (2022) compararam SVM, Random Forest e XGBoost, concluindo que os modelos de aprendizado de máquina apresentam desempenho superior aos tradicionais. Já Barboza, Basso e Kimura (2021) destacam que, apesar do avanço dos algoritmos, técnicas estatísticas ainda podem superar modelos de aprendizado de máquina quando bem calibradas com variáveis adequadas.

No Brasil, o uso dessas técnicas também tem avançado. Prado *et al.* (2020), Santos (2021), Jesus e Besarria (2023) e Cunha, Penedo e Barboza (2024) testaram diferentes modelos baseados em aprendizado de máquina, como *Random Forest*, *Decision Tree* e *Neural Networks*, aplicados a amostras de empresas nacionais, com resultados promissores. Esses estudos demonstram que técnicas como SVM e RF apresentam alta acurácia preditiva, reforçando seu potencial como ferramentas de apoio à gestão de risco financeiro.

Estudos de revisão sistemática e meta-análises também corroboram o potencial das técnicas de aprendizado de máquina. Por exemplo, Sulistiani *et al.* (2021) confirmam que modelos de inteligência computacional baseados em *machine learning* e *deep learning* superam estatísticas tradicionais em diversos contextos, enquanto Yuvaraj e Poongodi (2024) reforçam que variáveis como o tipo de algoritmo, o equilíbrio dos dados e a indústria analisada são determinantes para a acurácia dos modelos preditivos.

De forma complementar, Billios, Seretidou e Stavropoulos (2024) evidenciam que, apesar da eficácia dos indicadores financeiros tradicionais, sua estabilidade preditiva pode ser

impactada por mudanças nas condições econômicas, indicando que a robustez dos modelos estatísticos depende da adaptação ao contexto específico analisado.

De acordo com Vásquez-Serpa *et al.* (2025), a superioridade das técnicas de inteligência artificial como *Long Short-Term Memory* (LSTM), XGBoost e redes neurais profundas na previsão de risco de falência, especialmente em instituições financeiras. Já o estudo de Pereira e Martins (2015), com foco na produção científica nacional, mostra que, embora as técnicas estatísticas sejam predominantes nos trabalhos acadêmicos brasileiros, há uma tendência crescente na adoção de métodos baseados em aprendizado de máquina, ainda que de forma incipiente.

Nesse contexto, o Quadro 1 apresenta uma síntese diversos estudos sobre previsão do risco de insolvência, destacando as técnicas utilizadas, a acurácia alcançada, os indicadores financeiros empregados e, quando disponível, as fórmulas dos modelos aplicados. Essa consolidação permite uma visão comparativa do desempenho e das abordagens metodológicas adotadas por diferentes autores, reforçando a relevância de se avaliar empiricamente os modelos que serão aplicados nesta dissertação.

Quadro 1 – Seleção de trabalhos sobre previsão de insolvência (continua).

Autores	Técnica	Nível de precisão
Altman (1968)	Análise discriminante	95%.
Kanitz (1974)	Regressão múltipla e análise discriminante	Media o nível de cada empresa individualmente (com a aplicação menor rejeição de financiamentos e empréstimos).
Elisabetsky (1976)	Regressão multilinear	74,07% das empresas boas corretamente classificadas e 62,96% das empresas ruins.
Matias (1978)	Análise discriminante	Das 50 empresas solventes, 44 foram classificadas corretamente (88%). Das empresas insolventes, 45 foram classificadas de forma correta (90%).
Altman, Baidya e Dias (1979)	Análise discriminante	78% (três anos antes da situação-problema); 88% (um ano de antecedência).
Ohlson (1980)	Regressão logística	85,1%.
Silva (1982)	Análise discriminante	90% - Empresas solventes e 83% - empresas insolventes.

Quadro 1 – Seleção de trabalhos sobre previsão de insolvência (continuação).

Autores	Técnica	Nível de precisão
Odom e Sharda (1990)	Análise discriminante e neural networks	Média da acurácia da análise discriminante: 73,74% Média da acurácia da neural networks: 80,58%
Tam (1991)	Redes neurais Regressão logística Análise discriminante KNN Árvore de decisão	Redes neurais: 85,2% Regressão logística: 77,3% Análise discriminante: 75% KNN: 77,2% Árvore de decisão: 79,5%
Sanvicente e Minardi (1998)	Análise discriminante	81,8%
Akiama (2008)	Regressão logística	72,7%
Stüpp (2015)	Análise discriminante e regressão logística	Análise discriminante: acerto de 76,1%; Regressão logística: acerto de 76,2%.
Prado <i>et al.</i> (2018)	Análise discriminante	87,1%
Prado <i>et al.</i> (2020)	Redes Neurais artificiais	Modelo MT: 96,9% de acerto global; Modelo MT&D: 100% de acerto.
Pinto (2021)	Naive Bayes SVM Linear SVM Polynomial SVM Radial Regressão logística Redes neurais Floresta aleatória XGBoost	Naive Bayes: 94,96% SVM Linear: 95,28% SVM Polynomial: 95,09% SVM Radial: 96,51% Regressão logística: 96,55% Redes neurais: 96,38% Floresta aleatória: 97,8% XGBoost: 98,38%

Quadro 1 – Seleção de trabalhos sobre previsão de insolvência (conclusão).

Autores	Técnica	Nível de precisão
Santos (2021)	Floresta aleatória, <i>naive bayes</i> , <i>logit</i> , K-NN, SVM (linear, polinomial e de base radial), <i>bagging</i> e <i>boosting</i>	Floresta aleatória: 95,72% <i>naive bayes</i> : 91,01% <i>logit</i> : 92,51% K-NN: 94,01% SVM (linear, polinomial e de base radial): 92,8%, 93,72% e 95,22% <i>Bagging</i> : 94,44% <i>Boosting</i> : 94,86%
Lima, Paulino e Fávero (2022)	Regressão logística Análise discriminante	Regressão logística: 93,51% Análise discriminante: 93,80%
Jesus e Besarria (2023)	Naive bayes Random Forest SVM Regressão logística AdaBoost Árvore de decisão	Naive bayes: 86,25% Random Forest: 92,5% SVM: 85% Regressão logística: 85% AdaBoost: 91,25% Árvore de decisão: 93,75%
Cunha, Penedo e Barboza (2024)	Random Forest	98%
Shio e Sampaio (2024)	Cox Regressão logística	Cox: 84% Regressão logística: 94%
Tavares <i>et al.</i> (2025)	Redes neurais	95%

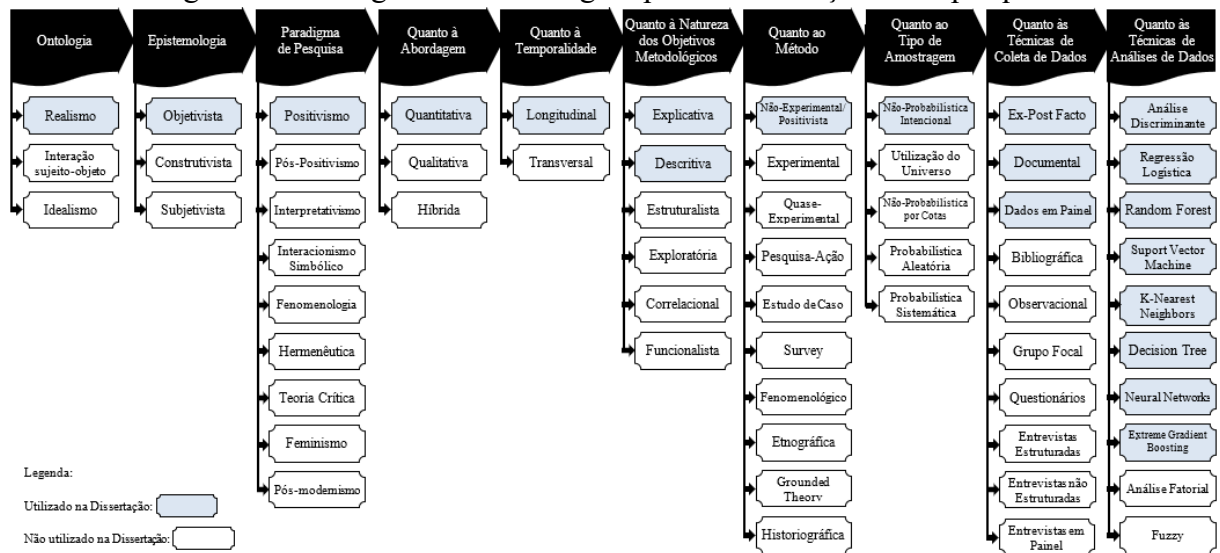
Fonte: Elaborado pelo autor (2025).

Esses achados sustentam a escolha metodológica deste estudo, que desenvolve modelos empregando técnicas estatísticas e de aprendizado de máquina na previsão do risco de insolvência de empresas brasileiras listadas na B3, uma vez que é observada a carência de estudos na comparação entre técnicas na temática do risco de insolvência, principalmente nos estudos com empresas brasileiras, buscando identificar quais modelos apresentam maior robustez e aplicabilidade prática no contexto nacional.

3 METODOLOGIA GERAL

Este tópico apresenta a metodologia geral adotada na realização do estudo. De acordo com Prado (2019), os estudos científicos podem ser classificados a partir de diferentes perspectivas, incluindo ontologia, epistemologia, paradigma de pesquisa, abordagem metodológica, temporalidade, natureza dos objetivos, método utilizado, tipo de amostragem, técnicas de coleta de dados e técnicas de análise dos dados. A Figura 2 ilustra as classificações atribuídas a esta pesquisa em conformidade com cada uma dessas categorias.

Figura 2 – Fluxograma metodológico para classificação desta pesquisa.



Fonte: Adaptado de Prado (2019, p. 56).

A ontologia pode ser definida como a concepção filosófica que orienta a forma como o pesquisador entende a realidade, ou seja, o que existe e como esse “ser” pode ser compreendido no contexto da investigação científica. Nesse sentido, ela antecede e fundamenta escolhas epistemológicas e metodológicas ao definir a natureza dos fenômenos a serem investigados (Brancher, 2018). Dentre as abordagens ontológicas, destacam-se o realismo e o idealismo. O realismo parte do pressuposto de que existe uma realidade objetiva, independente da consciência ou interpretação humana, e que pode ser investigada de forma empírica e racional (Carvalho; Varenne; Braga, 2014). Já o idealismo considera que a realidade é construída pela mente humana, sendo moldada por crenças, percepções e experiências individuais (Grishin, 2024). No presente estudo, adota-se a perspectiva realista, considerando que os fenômenos investigados existem de forma independente e objetiva, sendo possível conhecê-los por meio da análise sistemática e fundamentada.

Epistemologia diz respeito ao modo como o conhecimento é concebido e está diretamente ligada à ontologia, pois define como o pesquisador compreende a possibilidade de conhecer a realidade. Segundo Gomez (2015), pode ser classificada em objetivismo, subjetivismo e construtivismo. O objetivismo defende que o conhecimento é independente do sujeito e pode ser alcançado por meio de métodos rigorosos que minimizam interferências subjetivas. Essa visão sustenta que os significados e verdades existem fora da mente humana e podem ser revelados por meio da observação sistemática e lógica dedutiva (Rosa, 2015). Neste estudo, adota-se a perspectiva epistemológica objetivista, por acreditar na existência e significado dos fenômenos de forma independente da interpretação do observador.

O paradigma de pesquisa representa a visão de mundo que orienta todo o processo investigativo, articulando pressupostos ontológicos, epistemológicos e metodológicos. Segundo Alharahsheh e Pius (2020), o positivismo é um dos paradigmas mais tradicionais e valoriza a objetividade, a quantificação e a busca por leis gerais por meio da observação empírica e da mensuração. Associado à ontologia realista e à epistemologia objetivista, o positivismo entende que a realidade existe independentemente do observador e pode ser conhecida por meio de métodos científicos rigorosos. Neste estudo, adota-se o paradigma positivista, pois se acredita que os fenômenos investigados podem ser medidos e explicados com base em dados objetivos.

Quanto à abordagem metodológica, este estudo se classifica como quantitativo, pois busca mensurar variáveis e analisar relações causais entre elas de forma objetiva e sistemática. A abordagem quantitativa é caracterizada pelo uso de instrumentos estruturados, aplicação de técnicas estatísticas e foco na generalização dos resultados (Gonzales; Neves; Santos, 2018). Segundo Utsumi *et al.* (2007), pesquisas quantitativas frequentemente utilizam delineamentos experimentais, quase-experimentais ou descritivos, baseando-se na análise matemática para explicar fenômenos. Assim, a investigação foi conduzida por meio da coleta e análise de dados numéricos, visando alcançar resultados objetivos e replicáveis.

Quanto à temporalidade, este estudo é classificado como longitudinal, pois observa e analisa os fenômenos ao longo de um período contínuo (Prado *et al.*, 2019). Assim, considerando que este estudo analisa a previsão de insolvência de empresas no período de 2010 a 2024, justifica-se sua classificação como longitudinal, dado o acompanhamento temporal extenso e sistemático dos dados financeiros.

Em relação à natureza dos objetivos metodológicos, este estudo se classifica como descritivo e explicativo, uma vez que busca tanto apresentar as características do fenômeno estudado quanto compreender as variáveis que o influenciam. A pesquisa descritiva tem como

foco a observação e caracterização de aspectos de um fenômeno sem interferência do pesquisador, enquanto a explicativa visa identificar causas e relações entre variáveis (Utsumi *et al.*, 2007); (Barros *et al.*, 2016). Assim, ao analisar o risco de falência por meio de dados objetivos, o estudo não apenas descreve o fenômeno, mas também explica suas possíveis causas.

Quanto ao método, esta pesquisa é classificada como não-experimental, uma vez que as variáveis observadas não são manipuladas, mas analisadas em seu estado natural. Segundo Camayd e Freire (2020), esse método é comum em estudos quantitativos que buscam compreender relações entre variáveis a partir da observação de dados reais, sem intervenção direta do pesquisador. Essa abordagem é compatível com o paradigma positivista, ao assumir que os fenômenos podem ser analisados de forma objetiva e mensurável. Assim, os dados são coletados tal como se apresentam na realidade, permitindo a identificação de padrões e relações estatisticamente significativas.

Quanto ao tipo de amostragem, este estudo se classifica como não-probabilístico e intencional, pois os dados foram coletados com base em critérios específicos e definidos previamente, em vez de sorteio aleatório. Esse tipo de amostragem é adequado quando se deseja investigar um grupo com características particulares ou com relevância direta para os objetivos da pesquisa. Ferreira *et al.* (2021), em seu estudo sobre fluxos de caixa e endividamento em operadoras de planos de saúde, também utilizaram uma amostragem não-probabilística, destacando que a seleção foi limitada a empresas com dados disponíveis e relevantes para o tema.

De forma complementar, a amostra compreende as empresas brasileiras não financeiras de capital aberto listadas na Brasil, Bolsa e Balcão (B3) no período de 2010 a 2024. A variável dependente é de natureza dicotômica, sendo classificadas como solvente e insolvente. Os critérios de classificação para as empresas insolventes são aqueles que, durante o período analisado, ajuizaram o patrimônio líquido negativo, pedido de recuperação judicial ou de recuperação extrajudicial.

A delimitação da amostra a partir de 2010 justifica-se pela adoção das normas internacionais de contabilidade (IFRS) no Brasil, instituídas pelas Leis nº 11.638/2007 e nº 11.941/2009. Assim, o intervalo de 2010 a 2024 assegura a padronização nos processos de reconhecimento, mensuração e divulgação das demonstrações financeiras, permitindo que as variáveis independentes sejam calculadas sob premissas uniformes em todo o período analisado.

Para o artigo 1, no qual foi construído o modelo a partir da Análise Discriminante, a amostra foi selecionada por corte transversal, onde para cada empresa em que se apresentou pela primeira vez no período analisado como insolvente, foi selecionada uma empresa solvente do

mesmo setor e com ativo total informado no Balanço Patrimonial das Demonstrações Financeiras próximo ao da insolvente, indicando similaridade nos negócios.

Nos artigos 2 e 3, no qual foram desenvolvidos os modelos com a Regressão Logística para dados em painel e outros seis modelos de aprendizado de máquina, a amostra foi coletada no formato de dados em painel, em que considera as empresas empilhadas pelo período abrangido neste estudo, sendo que os modelos utilizados na previsão aceitam esta estrutura de dados.

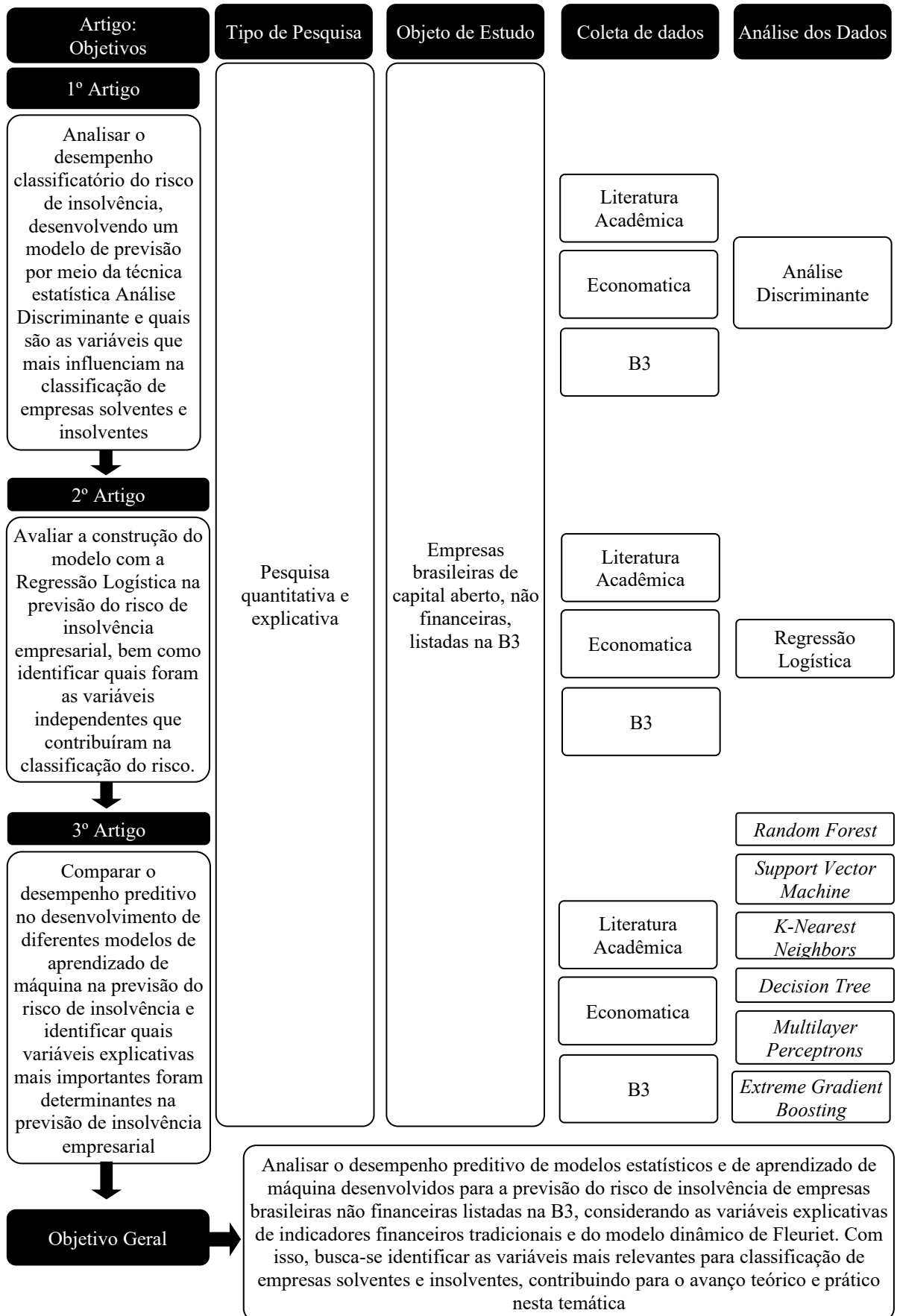
Quanto às técnicas de coleta de dados, este estudo utiliza as abordagens documental, *ex-post facto* e dados em painel. A pesquisa é documental porque os dados são extraídos diretamente das demonstrações financeiras publicadas pelas empresas listadas na B3, fontes oficiais e não manipuladas. Trata-se também de um estudo *ex-post facto*, pois os dados analisados se referem a eventos já ocorridos de 2010 a 2024, não sendo possível interferir em suas variáveis (Costa; Espejo; Almeida, 2011). Além disso, será utilizado painel de dados, combinando informações de diferentes empresas ao longo do tempo, o que permite explorar efeitos dinâmicos e relações entre variáveis de forma mais robusta (Steppan *et al.*, 2015).

Por fim, quanto as técnicas empregadas para o desenvolvimento dos modelos, o presente estudo utiliza a Análise Discriminante no primeiro artigo, a Regressão Logística para dados em painel no segundo artigo e no terceiro artigo utiliza-se seis diferentes técnicas de aprendizado de máquina, sendo elas a *Random Forest*, *Support Vector Machine*, *K-Nearest Neighbors*, *Decision Tree*, *Multilayer Perceptrons* e *Extreme Gradient Boosting*.

Adicionalmente, para o artigo 1, foi utilizada a técnica *hold-out*, onde foram realizadas subamostras para que 74,1% dos dados sejam destinados ao treino dos modelos e 25,9% à validação pelo subconjunto de teste. Já nos artigos 2 e 3, respeitando o aspecto temporal, foi realizado o método *hold-out* por empresa, onde 70% da amostra foi separada para o conjunto de dados para treino e 30% para o conjunto de teste, levando em consideração a validação temporal na ordem cronológica. Tais procedimentos visam a otimização da performance de cada método desenvolvido e maior rigor na avaliação da capacidade preditiva.

A Figura 3 sintetiza os procedimentos metodológicos adotados nesta dissertação, reunindo os principais elementos abordados em cada artigo, sendo eles o tipo de pesquisa, o objeto de estudo, a estratégia de coleta e as técnicas de análise de dados. Essa estrutura permite visualizar de forma integrada a abordagem metodológica utilizada para alcançar os objetivos específicos de cada estudo.

Figura 3 – Estrutura da metodologia para a dissertação.



Fonte: *Framework* adaptado de Antonialli (2018, p. 54).

Para garantir a confiabilidade e o bom ajuste dos modelos desenvolvidos para a previsão do risco de insolvência, foi utilizado o teste da Curva ROC (*Receiver Operating Characteristic*) e AUC (*Area Under the Curve*). A curva ROC é um gráfico da taxa de verdadeiros positivos (sensibilidade) em relação à taxa de verdadeiros negativos (especificidade). Ele é usado para avaliar o desempenho de um modelo de classificação binária, evidenciando o *trade-off* entre sensibilidade e especificidade, enquanto os valores de AUC resume o desempenho global dos modelos desenvolvidos, em que quanto mais próximo de 1, melhor é o desempenho do modelo. (Kamalov; Leung, 2015).

No contexto da previsão de risco de insolvência, a curva ROC ajuda a determinar o ponto de corte ideal que maximiza a capacidade do modelo de classificar corretamente entidades solventes e insolventes, conforme apresentado nos trabalhos de Stüpp (2015), Pinto (2021) e Santos (2021).

A AUC é uma medida escalar que quantifica a capacidade geral do modelo de discriminar entre as classes positivas e negativas. É a área sob a curva ROC e varia de 0 a 1, com 1 indicando discriminação perfeita e 0,5 sugerindo nenhuma discriminação melhor do que chance aleatória (Kamalov; Leung, 2015).

Na previsão de risco de insolvência, um valor de AUC mais alto indica um modelo de melhor desempenho na distinção entre entidades com probabilidade de inadimplência e aquelas que não o são (Kochański, 2024).

Além da curva ROC e do valor de AUC, outras métricas fundamentais na avaliação de modelos preditivos incluem a Sensibilidade (*Recall*), que mede a proporção de verdadeiros positivos corretamente identificados pelo modelo, e a Especificidade, que indica a capacidade do modelo de identificar corretamente os verdadeiros negativos. A Precisão representa a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas. De forma complementar, a Matriz de Confusão fornece uma visão geral dos resultados de classificação, mostrando as quantidades de verdadeiros positivos, falsos positivos (Erro Tipo 1), verdadeiros negativos e falsos negativos (Erro Tipo 2). Por fim, a Acurácia mede a proporção de previsões corretas (positivas e negativas) em relação ao total de casos analisados, sendo amplamente utilizada como métrica de desempenho (James *et al.*, 2023).

Já o *F1-Score*, conforme destacado por Firmansyah e Astuti (2024), representa a média harmônica entre precisão e *recall*, fornecendo uma medida equilibrada do desempenho do modelo, especialmente em conjuntos de dados desbalanceados. Essa métrica é particularmente útil quando se busca um compromisso entre a capacidade de identificar corretamente os

positivos e a de evitar falsos alarmes, complementando as demais medidas de avaliação mencionadas por James *et al.* (2023).

3.1 Definição da Amostra Corte Transversal

A amostra foi selecionada por meio de um corte transversal, identificando empresas que, durante o período analisado, solicitaram recuperação judicial (RJD), recuperação extrajudicial (REJ) ou apresentaram patrimônio líquido negativo (PLN). Essas empresas foram classificadas como insolventes. Para cada empresa insolvente, foram coletados os dados financeiros referentes a **um ano antes da ocorrência do evento**.

Para fins de equiparação da amostra, foram selecionadas empresas solventes que não apresentaram nenhum dos três critérios de insolvência ao longo do período de estudo, sendo a empresa solvente deveria pertencer ao mesmo setor de atuação e possuir um valor de ativo total próximo ao da empresa insolvente correspondente, garantindo assim a similaridade entre os negócios comparados.

A Tabela 1 apresenta as empresas, assim como mostra a data do evento, o valor do ativo em milhares de reais na data de 31 de dezembro do respectivo ano e o setor de atuação conforme classificação da plataforma Economatica.

Tabela 1 – Empresas categorizadas como insolventes (continua).

Empresa Insolvente	Critério	Data Evento	Ativo dezembro	Setor	Empresa Solvente	Ativo dezembro
Agrogalaxy	RDJ	2024	3.738.665	Agro e Pesca	Brasilagro	3.796.374
Alphaville	PLN	2018	3.687.925	Construção	Moura Dubeux	3.703.295
Altus S/A	PLN	2016	165.859	Siderur & Metalur	Tekno	303.926
Americanas	PLN	2022	35.262.271	Comércio	Dufry AG	34.902.654
Azevedo	PLN	2017	197.974	Construção	Priner	185.704
Azul	PLN	2019	26.324.690	Transporte Serviç	JSL	22.704.576
B Tech Eqi	PLN	2016	139.327	Alimentos e Beb	Excelsior	128.171
Bardella	RDJ	2019	998.578	Máquinas Indust	Stara	1.015.783
Bbmlogistica	PLN	2024	1.204.694	Transporte Serviç	Cvc Brasil	3.483.324
Biosev	PLN	2015	16.653.445	Outros	Localiza	16.598.979
BR Home	PLN	2016	473.397	Construção	Fica	471.419
BR Pharma	PLN	2017	215.074	Comércio	Track Field	208.166
Braskem	PLN	2020	112.941.225	Química	Valefert	20.401.625
Buettner	RDJ	2011	445.061	Textil	Ind Cataguas	437.562

Tabela 1 – Empresas categorizadas como insolventes (continuação).

Empresa Insolvente	Critério	Data Evento	Ativo dezembro	Setor	Empresa Solvente	Ativo dezembro
Cabinda Part	PLN	2012	36	Outros	Sudeste S/A	36
Cambuci	PLN	2016	342.212	Textil	Encorpar	329.407
Capitalpart	PLN	2014	174	Telecomunicações	Sul 116 Part	12.381
Ccx Carvao	PLN	2016	56.008	Mineração	Cent Açú	61.605
Ceee-D	PLN	2015	5.384.023	Energia Elétrica	Ceb	5.396.981
Ciabrasf	PLN	2024	29.880	Outros	Atom Educ	28.357
Cims	PLN	2013	26	Outros	Prompt Part	27
Cinesystem	PLN	2020	242.432	Outros	Allis Part	242.602
Clarion	RDJ	2013	2.242.759	Alimentos e Beb	Josapar	2.267.485
Compass Gas	PLN	2019	1.374	Petróleo e Gas	Brava	543.627
Contax	PLN	2017	2.433.617	Outros	Locamerica	2.400.641
Coteminas	PLN	2023	2.937.484	Textil	Vulcabras	2.989.103
Dexxos Par	RDJ	2013	1.330.746	Outros	Brisanet	1.335.934
Dtcom Direct	PLN	2022	13.036	Outros	Gama Part	14.827
Embpar S/A	PLN	2011	1.210.991	Comércio	Grupo Soma	1.208.206
Eneva	RDJ	2014	12.658.605	Energia Elétrica	Ampla Energ	12.694.690
Equatorial	RDJ	2012	19.104.023	Energia Elétrica	Cesp	19.091.897
Eternit	RDJ	2018	814.037	Minerais não Met	Nadir Figuei	932.181
Fer Heringer	PLN	2018	2.611.337	Química	Providencia	2.633.019
Fibam	RDJ	2014	155.576	Siderur & Metalur	Metisa	400.467
Flex S/A	PLN	2021	482.430	Outros	Ser Educa	481.733
Gol	PLN	2014	17.927.732	Transporte Serviç	All Norte	18.089.923
Iguacu Cafe	PLN	2011	1.261.754	Alimentos e Beb	Wow	1.255.870
Inepar	RDJ	2014	4.798.815	Outros	Zamp S.A.	4.806.406
Infracomm	PLN	2024	1.559.223	Comércio	D1000vfarma	1.502.408
Invepar	PLN	2020	27.888.120	Transporte Serviç	Ecorodovias	27.961.280
Joao Fortes	PLN	2019	1.854.651	Construção	Tecnisa	1.855.326
Karsten	PLN	2013	696.072	Textil	Pettenati	685.195
Laep	PLN	2012	934.636	Alimentos e Beb	Cacique	974.726
Lark Maqs	PLN	2011	50.626	Outros	Eletron	50.464
Light S/A	RDJ	2023	25.545.393	Energia Elétrica	Isa Energia	25.600.077
Log-In	PLN	2017	1.739.087	Transporte Serviç	Tegma	1.894.682
Longdis	PLN	2014	38	Outros	Harpia Part	245

Tabela 1 – Empresas categorizadas como insolventes (continuação).

Empresa Insolvente	Critério	Data Evento	Ativo dezembro	Setor	Empresa Solvente	Ativo dezembro
Lupatech	PLN	2011	3.221.959	Siderur & Metalur	Ferbasa	3.696.121
Mangels Indl	RDJ	2013	857.802	Siderur & Metalur	Aliperti	874.747
Mendes Jr	PLN	2016	1.465.855	Construção	Kepler Weber	1.435.381
Met Duque	RDJ	2014	336.567	Siderur & Metalur	Aco Altona	496.474
Metal Iguacu	PLN	2020	100.673	Siderur & Metalur	Panatlantica	631.774
Metalfrio	PLN	2015	1.957.131	Máquinas Indust	Romi	1.980.413
Minerva	PLN	2015	13.496.002	Alimentos e Beb	M.Diasbranco	13.292.421
Minupar	PLN	2012	567.102	Alimentos e Beb	Fornodeminas	377.581
MMX Miner	PLN	2014	5.818.322	Mineração	Litela	5.199.229
Mundial	PLN	2015	1.441.222	Siderur & Metalur	Confab	4.063.815
Nutriplant	PLN	2016	111.713	Química	Nortcquimica	122.045
OGX Petroleo	PLN	2013	10.305.596	Petróleo e Gas	Enauta Part	9.326.038
Oi	RDJ	2016	125.525.587	Telecomunicações	Telef Brasil	128.370.993
Opport Energ	PLN	2015	331.655	Outros	Eletromidia	332.724
OSX Brasil	PLN	2014	10.870.854	Veiculos e peças	Ioehp-Maxion	10.780.523
Padtec	PLN	2015	555.298	Software e Dados	Bemobi Tech	599.485
Parapanema	PLN	2020	5.289.568	Siderur & Metalur	Usiminas	54.778.934
PDG Realt	PLN	2016	7.104.938	Construção	Direcional	7.107.949
Plascar Part	PLN	2016	888.180	Veiculos e peças	Bic Monark	1.097.710
Pomifrutas	PLN	2016	141.663	Agro e Pesca	Rasip Agro	318.531
Portx	PLN	2011	3.287.538	Transporte Serviç	Santos Brp	3.277.293
Pq Hopi Hari	PLN	2013	583.066	Outros	Espacolaser	621.486
Rede Energia	RDJ	2012	5.717.243	Energia Elétrica	Bandeirante Energ	5.776.689
Renova	PLN	2018	3.759.389	Energia Elétrica	Statkraft	3.710.215
Rossi Resid	PLN	2018	3.636.184	Construção	Gafisa	3.613.348
Santanense	RDJ	2024	617.557	Textil	Azzas 2154	612.135
Santos Bras	PLN	2015	60	Transporte Serviç	Conc Rio Ter	43.287
Saraiva Livr	RDJ	2018	1.399.801	Outros	Cremer	1.425.773
Selectpart	PLN	2014	66	Telecomunicações	Daleth Part	19.900
Sequoia Log	REJ	2024	1.265.986	Transporte Serviç	Wilson Sons	3.570.639
Springs	PLN	2023	2.193.962	Textil	Cia Hering	2.282.438
SPTuris	PLN	2021	373.799	Outros	Pacific Rdsl	379.101

Tabela 1 – Empresas categorizadas como insolventes (conclusão).

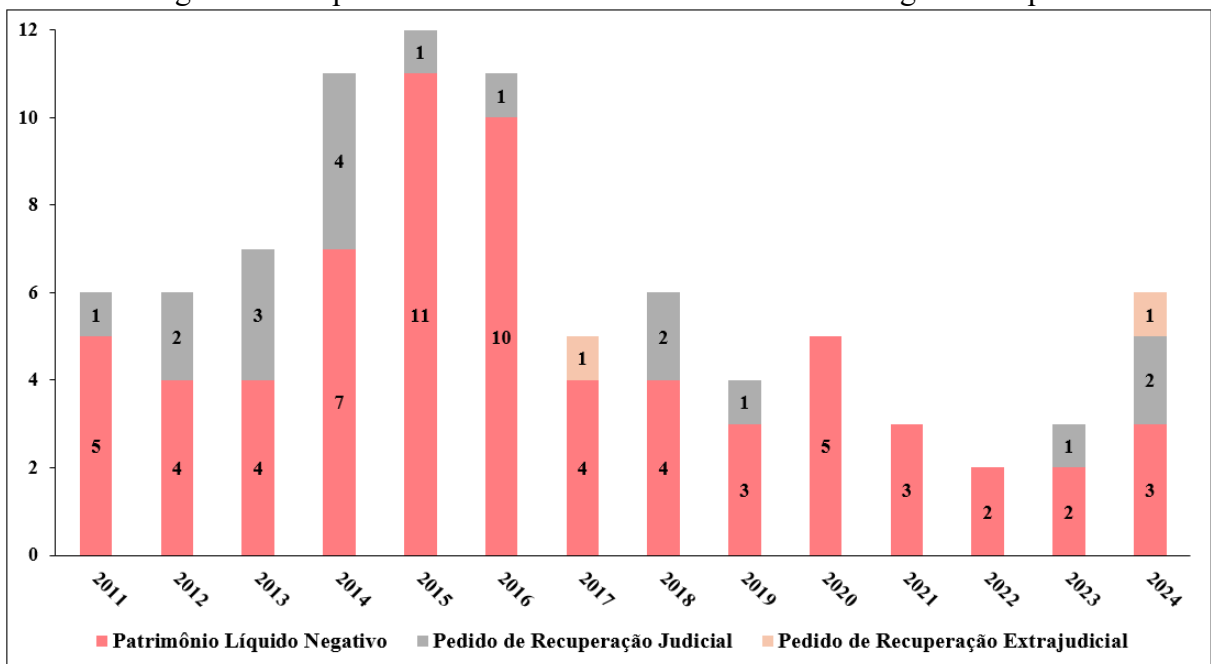
Empresa Insolvente	Critério	Data Evento	Ativo dezembro	Sector	Empresa Solvente	Ativo dezembro
Sultepa	RDJ	2015	1.920.995	Construção	Cury S/A	1.892.974
Taurus Armas	PLN	2015	1.659.940	Siderur & Metalur	Sid Nacional	67.085.968
Tectoy	PLN	2014	128.963	Outros	Sondotecnica	128.675
Telebras	PLN	2015	4.147.706	Telecomunicações	Desktopsigma	3.607.306
Triunfo Part	REJ	2017	6.809.558	Transporte Serviç	Hidrovias	6.824.760
Veste	PLN	2021	1.813.672	Textil	Cedro	1.377.396
Viver	PLN	2015	2.122.153	Construção	Kallas	2.144.383
Wetzel S/A	PLN	2012	403.905	Veiculos e peças	Fras-Le	1.589.332

Fonte: Adaptado de Costa (2023)

Nota: Ativo total em milhares de reais.

Com isso, a amostra compreende 87 empresas insolventes e 87 solventes não financeiras de capital aberto, totalizando 174 companhias cuja ações são negociadas na B3. Com relação as empresas insolventes, 18 empresas protocolaram o pedido de recuperação judicial, 2 empresas ajuizaram o pedido de recuperação extrajudicial e as demais 67 empresas apresentaram o patrimônio líquido negativo. A Figura 4 apresenta a quantidade de empresas que durante o período objeto deste estudo foram classificadas como insolventes.

Figura 4 – Empresas classificadas como insolventes ao longo do tempo.



Fonte: Elaborado pelo autor (2025).

Aprofundando a análise da amostra, a Figura 4 fragmenta os eventos de insolvência anuais de acordo com os três critérios utilizados. Verifica-se a predominância do Patrimônio Líquido Negativo como o principal indicador de dificuldade financeira na grande maioria dos anos, sugerindo que prejuízos acumulados que superam o valor do capital próprio investido é um sinalizador mais frequente na amostra do que as medidas legais formais, como os pedidos de Recuperação Judicial e Extrajudicial. Nesse sentido, essa distribuição evidencia a importância de se utilizar indicadores financeiros para a detecção precoce do risco de insolvência empresarial.

3.2 Definição da Amostra Dados em Painel

Para a coleta da amostra em dados em painel, este estudo utilizou os mesmos critérios de seleção das variáveis, considerando as empresas brasileiras listadas na Brasil, Bolsa, Balcão (B3), no período de 2010 a 2024, com exceção do setor financeiro. Como o modelo em painel considera simultaneamente as dimensões de empresa e tempo, foram incluídas na amostra todas as companhias que apresentaram informações financeiras em pelo menos um ano dentro deste intervalo. As observações sem dados disponíveis foram excluídas para garantir a consistência da base, uma vez que o valor zero é tratado como absoluto pelo modelo e interfere diretamente na estimação, podendo distorcer a representação real da estrutura financeira das empresas quando utilizado como substituto para dados ausentes.

Os critérios de classificação das empresas em solventes e insolventes seguiram os mesmos parâmetros adotados na amostra em corte transversal, contudo, na estrutura em dados em painel, a classificação foi realizada no próprio ano de ocorrência do evento de insolvência, e não de forma defasada, de modo a capturar todos os períodos em que a empresa apresentou a condição de insolvência ao longo do tempo. Os dados foram organizados em periodicidade anual, permitindo apreender a dinâmica do risco de insolvência, bem como identificar padrões de deterioração financeira associados a diferentes contextos econômicos.

A utilização de dados em painel possibilita observar o comportamento individual das empresas e as variações entre elas no tempo, aumentando a robustez da análise e a capacidade de captar efeitos específicos de cada firma e de cada período. Além disso, essa abordagem favorece uma avaliação mais precisa do desempenho dos modelos preditivos, uma vez que considera tanto as características estruturais das empresas quanto as flutuações conjunturais que podem afetar sua solvência. Essa perspectiva longitudinal, portanto, contribui para uma compreensão mais abrangente dos determinantes do risco de insolvência no contexto

corporativo brasileiro, permitindo ainda analisar a persistência do estado de solvência ou insolvência ao longo do tempo e a recorrência de episódios de deterioração financeira, aspectos relevantes em estudos que investigam fenômenos de natureza dinâmica.

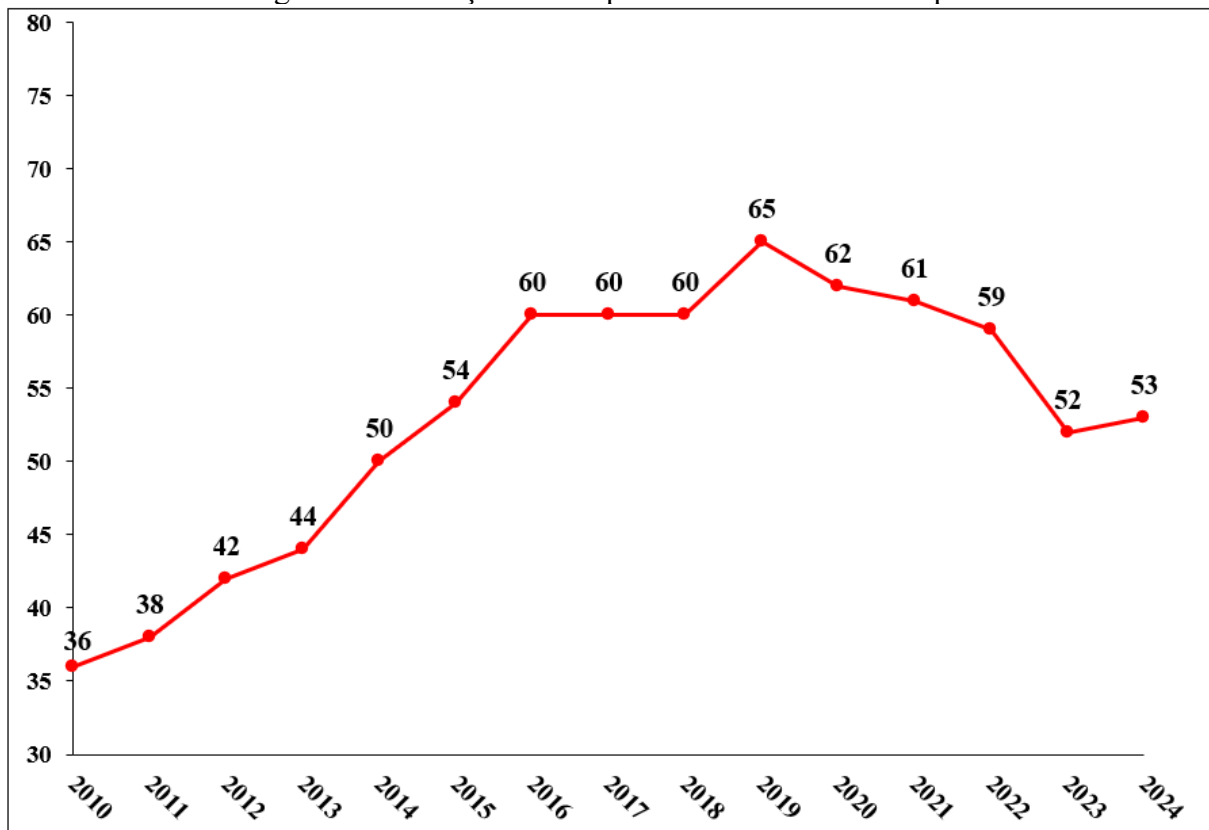
Com base nos critérios de seleção adotados, a amostra final é composta por 501 empresas, denominados como indivíduos para este estudo, onde 132 empresas apresentaram situação de insolvência em pelo menos um dos anos analisados, enquanto 369 registraram apenas situações de solvência no período. Como nem todas as empresas possuem informações para todos os anos, o que é comum nesta temática de estudo, o banco de dados foi estruturado como um painel não balanceado.

Cada linha da amostra representa uma empresa em um ano específico com dados válidos. Essa linha inclui o nome da empresa, o setor econômico, o ano da observação, a variável dependente indicando solvência ou insolvência e as variáveis independentes, sendo elas os indicadores financeiros selecionados neste estudo. Essa estrutura resultou em 5.247 linhas, tratadas como observações neste estudo. Entre elas, 796 correspondem a situações de insolvência, enquanto 4.451 representam situações de solvência, caracterizando um desbalanceamento de classes, sendo comum em pesquisas sobre previsão de insolvência.

Por fim, considerando o número total de observações e o número de variáveis analisadas, obteve-se uma matriz com 120.681 valores efetivamente utilizados nas análises, já descontadas as ausências de informação financeira, o que assegura a consistência da base de dados e a confiabilidade dos procedimentos estatísticos empregados ao longo do estudo.

A Figura 5 ilustra a distribuição das observações classificadas como insolventes ao longo do período de estudo, evidenciando a evolução temporal da ocorrência dos eventos de insolvência na amostra analisada e permitindo identificar períodos de maior concentração desses eventos, contribuindo para a compreensão da dinâmica do risco de insolvência ao longo do tempo.

Figura 5 – Evolução das empresas insolventes no tempo.



Fonte: Elaborado pelo autor (2025).

A Figura 5 detalha a evolução anual das situações de insolvência no Brasil, onde em 2010 havia 36 empresas na situação de risco de crédito, o que em 2024 já apresentavam um total de 53 empresas com sinal de deterioração financeira, um aumento de 47,2%, reforçando a necessidade de estudos desenvolvendo modelos preditivos e identificação de variáveis que sirvam como ferramentas estratégicas para os *stakeholders* para mitigação do risco de crédito.

Para classificação das empresas que homologaram o pedido de recuperação judicial ou extrajudicial, foram coletadas a data do pedido por meio da plataforma de consulta de documentos da Comissão de Valores Mobiliários (CVM) (CVM, 2025), e no site de relacionamento com investidores de cada empresa, onde é divulgado nas Demonstrações Financeiras o Fato Relevante informando a data do pedido e do encerramento, se ocorreu. A Tabela 2 apresenta as empresas que ajuizaram o pedido, a data do pedido e do encerramento da ação.

Tabela 2 – Data da recuperação judicial e extrajudicial das empresas (continua).

Empresa	Setor Economica	Tipo	Data do Pedido	Data do Encerramento
Agrogalaxy	Agro e Pesca	RJD	18/09/2024	-
Americanas	Comércio	RJD	19/01/2023	-
Bardella	Máquinas Indust	RJD	26/07/2019	-
Buettner	Textil	RJD	05/05/2011	-
Schlosser	Textil	RJD	06/04/2011	13/08/2018
Coteminas	Textil	RJD	06/05/2024	-
Santanense	Textil	RJD	06/05/2024	-
Sultepa	Construção	RJD	03/07/2015	-
Dommo	Petróleo e Gás	RJD	30/10/2013	-
Fibam	Siderur & Metalur	RJD	14/10/2014	-
Pq Hopi Hari	Outros	RJD	24/08/2016	-
Inepar	Outros	RJD	29/08/2014	-
Joao Fortes	Construção	RJD	27/04/2020	-
Lupatech	Siderur & Metalur	RJD	22/05/2015	14/03/2023
BR Pharma	Comércio	RJD	09/01/2018	-
MMX Miner	Mineração	RJD	25/11/2016	-
Met Duque	Siderur & Metalur	RJD	03/02/2014	-
Oi	Telecomunicações	RJD	20/06/2016	15/12/2022
Oi	Telecomunicações	RJD	02/03/2023	-
OSX Brasil	Veiculos e peças	RJD	19/01/2024	-
Paranapanema	Siderur & Metalur	RJD	30/11/2022	-
PDG Realt	Construção	RJD	22/02/2017	14/10/2021
Pomifrutas	Agro e Pesca	RJD	24/01/2018	-
Pet Manguinh	Petróleo e Gas	RJD	17/01/2013	-
Renova	Energia Elétrica	RJD	16/10/2019	12/02/2025
Sansuy	Outros	RJD	15/05/2007	04/12/2020
Springs	Textil	RJD	06/05/2024	-
Tecnosolo	Outros	RJD	03/08/2012	-
Teka	Textil	RJD	26/10/2012	-
Tex Renaux	Textil	RJD	05/05/2010	-
Viver	Construção	RJD	16/09/2016	17/12/2021
Clarion	Alimentos e Beb	RJD	05/06/2013	-
Eneva	Energia Elétrica	RJD	09/12/2014	30/06/2016

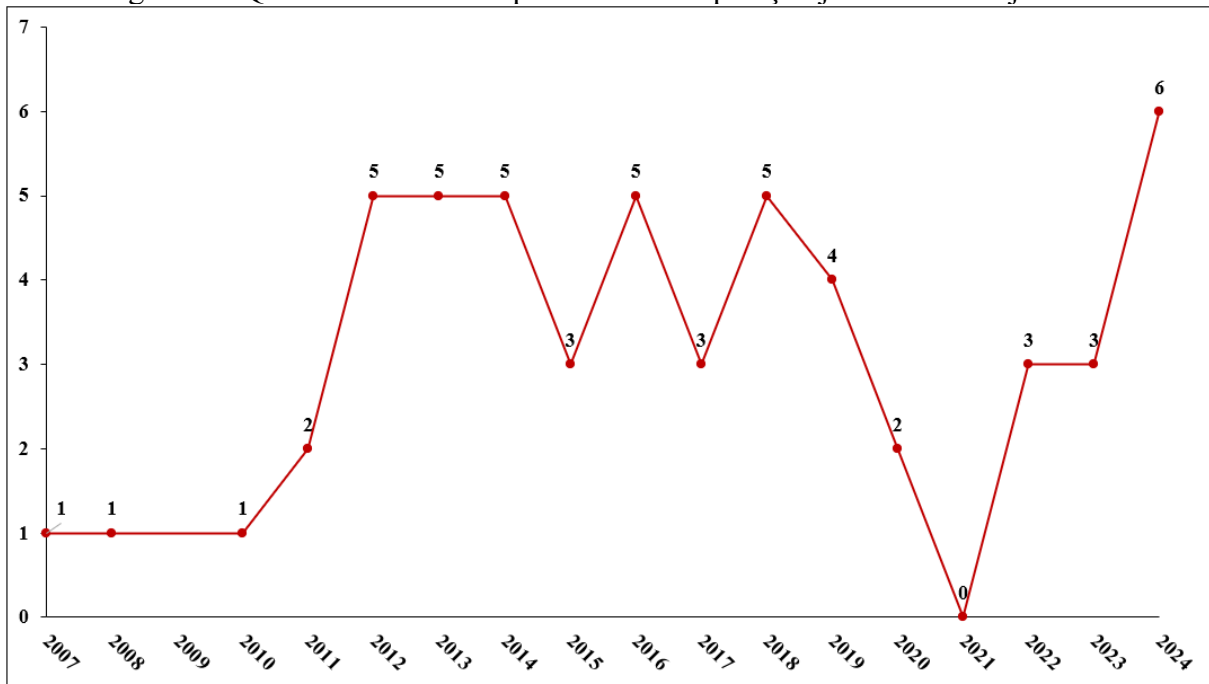
Tabela 2 – Data da recuperação judicial e extrajudicial das empresas (conclusão).

Empresa	Setor Economica	Tipo	Data do Pedido	Data do Encerramento
Equatorial	Energia Elétrica	RJD	28/02/2012	01/12/2014
Eternit	Minerais não Met	RJD	19/03/2018	09/08/2024
Light S/A	Energia Elétrica	RJD	12/05/2023	-
Mangels Indl	Siderur & Metalur	RJD	01/11/2013	15/03/2017
Met Duque	Siderur & Metalur	RJD	03/02/2014	-
Rede Energia	Energia Elétrica	RJD	23/11/2012	01/09/2016
Saraiva Livr	Outros	RJD	23/11/2018	-
Veste	Textil	RJD	08/06/2020	28/09/2020
Dexxos Par	Outros	RJD	15/04/2013	12/11/2020
DHB	Veiculos e peças	RJD	13/03/2015	-
Fer Heringer	Química	RJD	04/02/2019	16/02/2022
Hoteis Othon	Outros	RJD	27/11/2018	-
Wetzel S/A	Veiculos e peças	RJD	03/02/2016	-
Const Beter	Construção	RJD	12/09/2008	-
Rossi Resid	Construção	RJD	19/09/2022	-
Lark Maqs	Outros	RJD	11/05/2012	-
Contax	Outros	REJ	30/12/2019	07/06/2022
Contax	Outros	REJ	07/06/2022	-
Nutriplant	Química	REJ	02/10/2017	-

Fonte: Elaborado pelo autor (2025).

A Tabela 2 detalha as empresas da amostra que recorreram ao processo de Recuperação Judicial (RJD) ou Extrajudicial (REJ), especificando o setor de atuação e as datas de início e encerramento de cada processo. Observa-se uma diversidade de setores impactados, como Têxtil, Energia Elétrica, Construção e Siderurgia e Metalurgia, o que demonstra a amplitude do fenômeno da dificuldade financeira no cenário corporativo brasileiro. Para visualizar a dinâmica temporal e a frequência desses eventos de forma consolidada, os dados foram agregados anualmente. A Figura 6, a seguir, ilustra a quantidade de pedidos de recuperação registrados a cada ano.

Figura 6 – Quantidade anual de pedidos de recuperação judicial e extrajudicial.



Fonte: Elaborado pelo autor (2025).

A análise da Figura 6 revela uma nítida flutuação no número de pedidos de recuperação ao longo do tempo. Após um período de baixa incidência entre 2007 e 2010, nota-se um salto expressivo em 2011, inaugurando uma fase de maior intensidade que se estende até 2018. Um ponto de destaque é a acentuada queda subsequente, que culmina na ausência de pedidos em 2021, um comportamento que pode estar associado a medidas de suporte econômico durante a pandemia ou da liquidação imediata das empresas sem utilizar medidas legais para o contorno do cenário econômico da companhia. Já no período de 2022 a 2024 demonstra uma reversão abrupta dessa tendência, com uma escalada que atinge o maior pico da série histórica em 2024, sinalizando uma deterioração recente e acentuada nas condições de crédito das empresas.

3.3 Definição dos Indicadores

Os indicadores foram definidos com base na tabela Notação das Fórmulas de Cálculo dos Indicadores de Prado *et al.* (2018, p. 349), que são os indicadores dinâmicos do modelo Fleuret indicados em demais estudos na temática conforme apresentados no Quadro 2.

Quadro 2 – Indicadores do modelo dinâmico.

Cód.	Indicadores	Fórmula	Autores	Sinal Esperado
X1	CDG sobre Ativo	CDG / AT	Prado <i>et al.</i> (2018)	-
X2	CDG sobre Receita Líquida	$CDG / RecL$		-
X3	NCG sobre Ativo	NCG / AT	Brito, Assaf Neto e Corrar (2009), Carvalho (2004), Minussi, Damacena e Ness Junior (2002) e Sanvicente e Minardi (1998)	+
X4	NCG sobre Receita Líquida	$NCG / RecL$	Brito, Assaf Neto e Corrar (2009), Carvalho (2004), Eifert (2003), Horta (2010), Minussi, Damacena e Ness Junior (2002) e Sanvicente e Minardi (1998)	+
X5	Saldo de Tesouraria sobre Ativo	T / AT	Eifert (2003)	-
X6	Saldo de Tesouraria sobre Receita Líquida	$T / RecL$	Melo e Coutinho (2007)	-
X7	Passivo Financeiro sobre Ativo Circulante	PF / AC	Horta (2010) e Vieira (2008)	+
X8	Tipo de Estrutura Financeira	TEF	Melo e Coutinho (2007)	-
X9	Termômetro de Liquidez – TL	$T / ([NCG])$	Horta (2010) e Vieira (2008)	-
X10	Saldo de Tesouraria	$T = AF - PF$	Melo e Coutinho (2007)	-
X11	Necessidade de Capital de Giro	$NCG = AO - PO$	Melo e Coutinho (2007)	+
X12	Capital de Giro	$CDG = PNC - ANC$	Melo e Coutinho (2007)	-
X13	Endividamento Financeiro	$(PF + PNCF) / AT$	Brito, Assaf Neto e Corrar (2009)	+

Fonte: Prado *et al.* (2018, p. 349).

Além das variáveis no modelo dinâmico, foram utilizadas neste estudo os indicadores tradicionais selecionados por Costa (2023, p. 106). Tais indicadores econômico-financeiros foram utilizados em estudos anteriores relacionados a previsão de insolvência empresarial, sendo os principais Altman (1968), Kanitz (1974), Elizabetsky (1976), Matias (1978), Altman, Baydia e Dias (1979), Silva (1982), Sanvicenti e Minardi (1998), Akiama (2008), Stüpp (2015) e Prado *et al.* (2020), que corresponderam num total de 23 variáveis nos modelos tradicionais,

aqui renomeadas a partir da *X14* levando em consideração a ordem do Quadro 2, até *X35*, demonstrada no Quadro 3.

Quadro 3 – Modelos tradicionais (continua).

Nome	Indicadores	Fórmula	Autores	Sinal Esperado
X14	Ativo Circulante sobre Passivo Circulante	AC / PC	Kanitz (1974), Matias (1978), Silva (1982), Stüpp (2015).	-
X15	<i>Receita líquida operacional</i> sobre <i>Ativo Total</i>	RLO / AT	Altman (1968), Altman, Baidya e Dias (1979), Stüpp (2015).	-
X16	<i>Lucro Antes de Juros e Impostos</i> sobre <i>Ativo Total</i>	$EBIT / AT$	Altman (1968), Altman, Baidya e Dias (1979).	-
X17	<i>Passivo Circulante + Passivo não Circulante</i> sobre <i>Ativo Total</i>	$PC + PNC / AT$	Silva (1982), Stüpp (2015).	+
X18	<i>Ativo Circulante + Realizável a Longo Prazo</i> sobre <i>Passivo Circulante + Passivo não Circulante</i>	$AC + RLP / PC + PNC$	Kanitz (1974), Stüpp (2015).	-
X19	<i>Caixa e equivalente de caixa</i> sobre <i>Ativo não circulante – Realizável a Longo Prazo</i>	$Disponibilidade / AC + RLP$	Elisabetsky (1976), Silva (1982).	-
X20	Estoque sobre Ativo Total	$Estoque / AT$	Elisabetsky (1976), Silva (1982).	+
X21	Fornecedores sobre Ativo Total	$Fornecedores / AT$	Matias (1978), Silva (1982).	+
X22	Patrimônio Líquido sobre Ativo Total	PL / AT	Akiama(2008), Altman, Baidya e Dias (1979), Matias (1978), Sanvicente e Minardi (1998).	-
X23	<i>Ativo Circulante – Estoques</i> sobre <i>Passivo Circulante</i>	$AC - Estoques / PC$	Kanitz (1974).	-
X24	Patrimônio Líquido sobre Passivo Total	PL / PT	Altman, Baidya e Dias (1979), Sanvicente e Minardi (1998).	-

Quadro 3 – Modelos tradicionais (continuação).

Nome	Indicadores	Fórmula	Autores	Sinal Esperado
X25	Ativo Não Circulante sobre Patrimônio Líquido	ANC / PL	Stüpp (2015), Akiama (2008).	+
X26	<i>Ativo Circulante – Passivo Total</i> sobre <i>Ativo Total</i>	AC – PT / AT	Sanvicente e Minardi (1998).	-
X27	Financiamentos e Empréstimos Bancários sobre Ativo Circulante	DB / AC	Matias (1978).	+
X28	<i>Lucro Antes de Juros e Impostos – Despesas Financeiras + Receitas Financeiras</i> sobre Ativo Total	EBIT – DF + RF / AT	Sanvicente e Minardi (1998).	-
X29	Passivo Circulante sobre Ativo Total	PC / AT	Elisabetsky (1976).	+
X30	Patrimônio Líquido – Capital Social sobre Ativo Total	PL – CS / AT	Sanvicente e Minardi (1998).	-
X31	<i>Ativo Total – Obrigações sociais e trabalhistas, Impostos a pagar</i> sobre <i>Patrimônio Líquido</i>	AT – Obrigações sociais e trabalhistas e impostos a pagar / PL	Silva (1982).	-
X32	<i>Ativo não circulante – Realizável a Longo Prazo</i> sobre <i>Patrimônio Líquido</i>	ANC – RLP / PL	Akiama (2008).	+
X33	<i>Lucro Antes de Juros e Impostos</i> sobre Despesas Financeiras	EBIT / Impostos sobre Despesas Financeiras	Sanvicente e Minardi (1998).	-
X34	<i>Lucro Antes de Juros e Impostos</i> sobre Dívida Líquida	EBIT / DL	Stüpp (2015).	-

Quadro 3 – Modelos tradicionais (conclusão).

Nome	Indicadores	Fórmula	Autores	Sinal Esperado
X35	<i>Passivo Circulante</i> sobre <i>Passivo Circulante + Passivo Não Circulante</i>	$PC / PC + PNC$	Stüpp (2015).	+

Fonte: Adaptado de Costa (2023).

Com isso, o total de variáveis coletadas para análise neste estudo compreende em 35 indicadores considerando os modelos tradicionais e o modelo dinâmico.

Como as variáveis são calculadas com base nos dados de empresas de diversos tamanhos e setores diferentes, as variáveis que são indicadores financeiros possuem uma distribuição assimétrica e caudas pesadas.

Para reduzir valores discrepantes das variáveis que podem comprometer as estimativas dos modelos desenvolvidos, foi necessário realizar o processo de winsorização dos dados, considerando o valor máximo de 5%, reduzindo os efeitos dos *outliers*, gerando resultados mais confiáveis e garantindo que não se tenha conclusões equivocadas em cada modelo construído (Nguyen; Viviane; Jabeur, 2023). Como se trata de indicadores financeiros provenientes de empresas de diferentes tamanhos e setores, é esperado que haja certa assimetria na distribuição das variáveis. Ainda assim, com o procedimento adotado, busca-se que a assimetria se aproxime de três e a curtose de zero, características de uma distribuição normal (Fávero e Belfiore, 2017).

De forma complementar, para os artigos 2 e 3, foi realizado previamente o procedimento de padronização das variáveis via *Z-score*, em que transforma os valores de cada variável para que a média seja igual a zero e o desvio padrão igual a 1, enquanto no artigo 1 a padronização foi realizada automaticamente pelo próprio *software* estatístico, que transforma as variáveis internamente antes da estimação do modelo. Assim, assegurou-se a homogeneidade das escalas em todos os modelos analisados, evitando distorções provocadas por diferentes unidades de medida.

Segundo Firmansyah e Astuti (2024), a padronização é realizada para reduzir os valores extremos e garantir que cada variável seja comparável uma com a outra, padronizando a amplitude, uma vez que variáveis com amplitudes diferentes podem trazer um viés de interpretação nos modelos construídos. O valor resultante da padronização *Z-score* indica o valor da variável menos a sua média sobre o desvio padrão.

Para assegurar a robustez dos modelos e evitar distorções estatísticas, procedeu-se à verificação da multicolinearidade entre as variáveis independentes utilizadas nos três artigos desta dissertação. Inicialmente, foi analisada a matriz de correlação, considerando como indicativo de colinearidade valores superiores a 0,8, conforme orienta Gujarati e Porter (2011). Em complemento, aplicou-se o *Variance Inflation Factor* (VIF), no qual valores acima de 10 sugerem elevada colinearidade e comprometimento na precisão das estimativas dos parâmetros. Esse procedimento de avaliação e eventual exclusão de variáveis altamente correlacionadas foi adotado de forma sistemática em todos os artigos, garantindo consistência metodológica e maior confiabilidade nos resultados obtidos.

Conforme ressaltam James *et al.* (2023), a multicolinearidade é uma das principais fontes de instabilidade em modelos lineares e preditivos, pois eleva a variância dos estimadores e reduz sua capacidade de generalização para novos dados. A adoção de critérios objetivos de correlação e VIF, portanto, alinha-se ao princípio da regularização, amplamente discutido por Kuhn e Johnson (2013), que consiste em limitar a influência de variáveis excessivamente correlacionadas para reduzir o erro de previsão fora da amostra. Essa prática, embora simples, cumpre função análoga às técnicas de penalização como o Ridge e o LASSO, descritas pelos autores, atuando como um controle prévio de complexidade do modelo.

Além da remoção de variáveis altamente correlacionadas, a consistência dos modelos também foi assegurada mediante o uso de procedimentos de reamostragem, como a validação cruzada (*k-fold cross-validation*), que permite estimar o desempenho preditivo médio em diferentes partições da amostra. Essa técnica, conforme argumentam Kuhn e Johnson (2013), é eficaz para reduzir o risco de *overfitting*, garantindo que os resultados obtidos não sejam específicos do conjunto de treinamento, mas reflitam o comportamento do modelo em observações não utilizadas na estimação.

Desse modo, a estratégia adotada nesta pesquisa, composta pela verificação do VIF, correlação e validação cruzada, está em conformidade com as recomendações teóricas contemporâneas sobre seleção e validação de variáveis preditoras, garantindo equilíbrio entre parcimoniosidade, estabilidade estatística e capacidade preditiva dos modelos testados.

SEGUNDA PARTE – ARTIGOS

Para atingir os objetivos específicos da dissertação, neste tópico foram desenvolvidos os seguintes artigos:

- i. Artigo 1: Previsão de insolvência em empresas não financeiras listadas na b3: uma abordagem utilizando a análise discriminante.
- ii. Artigo 2: Uma abordagem do risco de insolvência de empresas brasileiras não financeiras utilizando a Regressão Logística.
- iii. Artigo 3: Previsão do risco de insolvência: Uma análise comparativa entre diferentes métodos de aprendizado de máquina.

ARTIGO 1 – PREVISÃO DE INSOLVÊNCIA EM EMPRESAS NÃO FINANCEIRAS LISTADAS NA B3: UMA ABORDAGEM UTILIZANDO A ANÁLISE DISCRIMINANTE

RESUMO

Este estudo teve como objetivo analisar o desempenho classificatório do risco de insolvência, desenvolvendo um modelo de previsão por meio da técnica estatística Análise Discriminante e quais são as variáveis que mais influenciam na classificação de empresas solventes e insolventes brasileiras não financeiras listadas na B3. As variáveis independentes foram indicadores financeiros tradicionais e do Modelo Dinâmico de Fleuriet. A amostra foi composta por 174 empresas, distribuídas em 87 pares formados por uma insolvente, identificada por Patrimônio Líquido Negativo, Pedido de Recuperação Judicial ou Extrajudicial, e uma solvente do mesmo setor e porte, no período de 2010 a 2024. Utilizaram-se 35 indicadores financeiros extraídos da literatura e calculados a partir das demonstrações financeiras anuais. Os resultados apontaram o Saldo de Tesouraria sobre a Receita Líquida e o Passivo Circulante Financeiro somado ao Passivo Financeiro Não Circulante sobre o Ativo Total como os indicadores do Modelo Dinâmico de Fleuriet que contribuíram para a classificação. Os indicadores tradicionais mais discriminantes foram a Liquidez Corrente, o Lucro Antes dos Juros e Impostos sobre o Ativo Total e Lucro Antes dos Juros e Impostos sobre Impostos sobre Despesas Financeiras. O modelo apresentou acurácia de 93,3% na amostra de teste e AUC de 0,955, confirmando elevada capacidade discriminatória. Como limitações, destacam-se a dificuldade de atender integralmente à premissa de normalidade multivariada e a restrição ao contexto brasileiro de empresas abertas não financeiras. Para pesquisas futuras, recomenda-se ampliar o escopo amostral, incorporar variáveis qualitativas e macroeconômicas e aplicar a metodologia em diferentes países e setores.

Palavras-chave: Insolvência. Modelo Dinâmico. Análise Discriminante.

ABSTRACT

This study aimed to analyze the classification performance of insolvency risk by developing a predictive model based on Discriminant Analysis and identifying the variables that most influence the classification of Brazilian non-financial listed companies as solvent or insolvent on B3. The independent variables comprised traditional financial indicators and measures derived from the Fleuriet Dynamic Model. The sample consisted of 174 firms, organized into 87 matched pairs formed by one insolvent firm—identified by negative equity, filing for judicial or extrajudicial reorganization—and one solvent firm from the same industry and similar size, covering the period from 2010 to 2024. A total of 35 financial indicators were extracted from the literature and calculated using annual financial statements. The results indicate that Treasury Balance over Net Revenue and the ratio of Financial Current Liabilities plus Non-Current Financial Liabilities to Total Assets were the Fleuriet Dynamic Model indicators that contributed most to classification. Among traditional indicators, Current Liquidity, Earnings Before Interest and Taxes over Total Assets, and Earnings Before Interest and Taxes over Financial Expenses were the most discriminant. The model achieved an accuracy of 93.3% on the test sample and an AUC of 0.955, confirming its high discriminatory power. As limitations, the study highlights the difficulty of fully meeting the assumption of multivariate normality and the restriction to the Brazilian context of non-financial publicly traded firms. Future research is encouraged to expand the sample scope, incorporate qualitative and macroeconomic variables, and apply the methodology to different countries and sectors.

Keywords: Insolvency. Dynamic Model. Discriminant Analysis.

1 INTRODUÇÃO

A insolvência representa uma condição financeira precária caracterizada pela incapacidade de uma entidade, seja um indivíduo ou uma organização corporativa, de cumprir suas obrigações de dívida quando elas se tornarem vencidas e pagáveis, sinalizando assim um risco significativo à sua viabilidade econômica.

As iniciativas de pesquisa pioneiras relacionadas à previsão de falências, notadamente aquelas conduzidas por Beaver (1966) e Altman (1968), avançaram significativamente no campo ao introduzir técnicas analíticas multivariadas sofisticadas, junto com indicadores econômicos pertinentes, que aprimoraram coletivamente a precisão e a confiabilidade da previsão da insolvência corporativa.

Os estudos de Beaver (1966) e Altman (1968) não apenas estabeleceram uma base sólida para o desenvolvimento subsequente de diversas metodologias destinadas a avaliar rigorosamente a viabilidade financeira e a saúde geral de várias empresas, mas também desempenharam um papel fundamental na evolução de estratégias abrangentes de gerenciamento de risco que agora são parte integrante das práticas financeiras contemporâneas. Esses trabalhos pioneiros impactaram profundamente a avaliação de dificuldades financeiras e a adoção de medidas preventivas contra a insolvência, indo além das contribuições teóricas.

O capital de giro representa um componente fundamental das finanças corporativas, pois engloba o gerenciamento complexo e a supervisão estratégica dos ativos e passivos circulantes de uma empresa, a fim de maximizar o valor gerado pelos itens circulantes (Machado; Machado; Callado, 2006). De acordo com Assaf Neto e Silva (2012), a visão mais correta para administração do capital é de baixo para cima, por meio da gestão de recursos de longo prazo que excede as aplicações de mesma titularidade.

Dentre os avanços do conceito de capital de giro, destaca-se o Modelo Dinâmico de Fleuriet, desenvolvido na década de 1970, que apresenta uma análise estruturada baseada na necessidade de capital de giro, capital circulante líquido e saldo de tesouraria, permitindo identificar a estrutura financeira das empresas de maneira dinâmica.

O Modelo Dinâmico de Fleuriet redefine a forma de leitura do Balanço Patrimonial das empresas, classificando as contas patrimoniais dos ativos e passivos com base nas atividades operacionais e não operacionais, introduzindo um conceito diferente do tradicional capital de giro ao analisar estruturas de recursos permanentes e de longo prazo com as necessidades operacionais do curto prazo, permitindo identificar se o financiamento operacional é realizado de maneira sustentável e gerando saldos de tesouraria para o negócio ou se há risco financeiro.

Prado *et al.* (2018) investigaram o uso desse modelo para a avaliação de risco de crédito, reforçando sua relevância para o contexto brasileiro de previsão de insolvência. De acordo com Marion (2019), os indicadores tradicionais combinam diversas análises a partir da estrutura das Demonstrações Financeiras apresentadas e divulgadas pelas empresas, sendo o tripé de análises fundamentais os índices de liquidez relacionadas a situação financeira, endividamento que compõem a estrutura de capital e por fim a rentabilidade que demonstra a situação econômica.

Com isso, Carneiro e Bastos (2022) enfatizam que a análise de indicadores contábeis não deve se restringir aos indicadores tradicionais de medidas de risco, desempenho e lucratividade, mas integrar aos indicadores do Modelo Dinâmico de Fleuriet que podem contribuir com a gestão do capital de giro das companhias.

Diante desse contexto, o presente estudo tem como **objetivo** analisar o desempenho classificatório do risco de insolvência, desenvolvendo um modelo de previsão por meio da técnica estatística Análise Discriminante e quais são as variáveis que mais influenciam na classificação de empresas solventes e insolventes de empresas brasileiras não financeiras listadas na Brasil, Bolsa, Balcão (B3), no período de 2010 a 2024. A amostra é composta por empresas solventes e insolventes, balanceadas por setor econômico e porte, a fim de assegurar comparabilidade entre os grupos analisados.

A pesquisa contempla, de forma integrada, indicadores financeiros tradicionais e variáveis derivadas do Modelo Dinâmico de Fleuriet, buscando compreender quais dessas medidas contribuem para a identificação antecipada de sinais do risco de crédito. Essa abordagem permite avaliar se o modelo discriminante é capaz de distinguir com precisão empresas em diferentes estágios de solvência, ampliando a robustez empírica das evidências no contexto brasileiro.

2 REFERENCIAL TEÓRICO

2.1 Modelos Tradicionais

Os indicadores tradicionais são inicialmente extraídos a partir das Demonstrações Financeiras das empresas e então aplicadas para identificação do grau de liquidez, endividamento e rentabilidade em uma análise inicial (Marion, 2019). Segundo Fleuriet e Zeidan (2015), para realizar as análises relacionadas a solvência empresarial eram utilizados índices de liquidez corrente, seca e geral, extraídas do Balanço Patrimonial. Embora relevantes, esses índices não captam o ciclo de conversão de caixa nem a dinâmica da empresa na formação de um saldo de tesouraria saudável para liquidar suas obrigações.

Prado *et al.* (2020), mencionam que embora os indicadores tradicionais possuam a sua importância, tais indicadores não refletem a dinâmica do capital de giro e da necessidade de capital de giro, nem como estas duas estruturas financeiras alavancam o saldo de tesouraria, lacuna que motivou a criação do modelo dinâmico.

2.2 O Modelo Dinâmico de Fleuriet

Na Figura 7 apresenta-se o Balanço Patrimonial com as contas classificadas de acordo com o modelo dinâmico.

Figura 7 – O Balanço Patrimonial no Modelo Dinâmico.

ATIVO		PASSIVO	
ATIVO CIRCULANTE CONTAS ERRÁTICAS	<i>Circulante</i> Financeiro Numerário em caixa, Bancos conta movimentos, Títulos e valores mobiliários.	PASSIVO CIRCULANTE CONTAS ERRÁTICAS	<i>Circulante</i> Financeiro Duplicatas descontadas, Empréstimos bancários em curto prazo, Dividendos a pagar, Imposto de renda a recolher etc.
	Operacional Duplicatas a receber, Estoques de produtos acabados, Estoques de produção em andamento, Estoques de matérias-primas, Adiantamento a fornecedores etc.		Operacional Fornecedores de matérias-primas Salários e encargos sociais, Impostos e taxas.
ATIVO NÃO CIRCULANTE CONTAS NÃO CÍCLICAS	<i>Ativo Não Circulante</i> Permanente Empréstimos a terceiros, Títulos a receber etc. Investimentos, Imobilizados.	PASSIVO NÃO CIRCULANTE CONTAS NÃO CÍCLICAS	<i>Passivo Não Circulante</i> Permanente Exigível em longo prazo, Empréstimos bancários, Debêntures, Financiamentos etc. Patrimônio Líquido Capital social, Reservas, Lucros ou prejuízos acumulados

Fonte: Adaptado de Fleuriet (2003).

Fleuriet, Kehye e Blanc (1978) desenvolveram o Modelo Dinâmico de Gestão Financeira, adaptado para diferentes realidades econômicas. O modelo classifica as contas do Balanço Patrimonial em cíclicas (de rápida rotatividade), não cíclicas (movimentação lenta) e erráticas (movimento descontínuo). Além disso, introduz os conceitos de Necessidade de Capital de Giro (NCG), Saldo de Tesouraria e Efeito Tesoura.

No ciclo financeiro, as saídas de caixa precedem as entradas, gerando uma aplicação permanente de fundos, com a NCG representando a diferença entre ativos e passivos cíclicos, conforme representada na Equação 1.

$$NCG = \textit{ativo cíclico} - \textit{passivo cíclico} \quad (1)$$

A NCG, quando positiva, evidencia a aplicação permanente de fundos. Quando a NCG é financiada por recursos de curto prazo, como empréstimos bancários, por exemplo, o risco de crédito aumenta (Fleuriet, 2003).

O modelo dinâmico conta também com o conceito de Capital de Giro (CDG) demonstrado na Equação 2, sendo este caracterizado pela diferença entre passivos e ativos não cíclicos, também denominados de permanentes (Fleuriet, 2003).

$$CDG = \textit{passivo permanente} - \textit{ativo permanente} \quad (2)$$

Denomina-se Saldo de Tesouraria (T), segundo Fleuriet (2003) a diferença entre as contas erráticas do ativo conforme apresenta a Equação 3, como as contas que compõem a disponibilidade imediata da empresa (caixas, bancos e aplicações financeiras) e do passivo financeiro (empréstimos e financiamentos).

$$T = \textit{ativo errático} - \textit{passivo errático} \quad (3)$$

Ainda, de acordo com Fleuriet e Zeidan (2015), temos o Índice de Liquidez Dinâmico (IDL) como mostra a Equação 4 que mede qual percentual de ALP (Ativos de Longo Prazo) e de NCG (Necessidade de Capital de Giro) está coberta por recursos de longo prazo. O índice negativo significa que a empresa está exposta ao risco de insolvência no curto prazo.

$$IDL = T / (ALP + NCG) \quad (4)$$

Por meio da análise do ciclo econômico das empresas, Fleuriet (1978) sugere inicialmente uma classificação em quatro tipos de estrutura financeira, conforme apresentado no Tabela 3.

Tabela 3 – Primeira classificação do Tipo de Estrutura Financeira de Fleuriet em 1978.

	CDG	NCG	T	Classificação
Tipo 1	+	+	-	Insatisfatória
Tipo 2	+	+	+	Sólida
Tipo 3	-	+	-	Muito ruim
Tipo 4	+	-	+	Excelente

Fonte: Adaptado de Fleuriet (1978).

Atualmente são seis tipos de classificação de estrutura financeira conforme demonstrado por Braga (1991), alterando também a ordem dos tipos com relação ao original, conforme demonstrado no Tabela 4.

Tabela 4 – Classificação atual do Tipo de Estrutura Financeira de Fleuriet.

	CDG	NCG	T	Classificação
Tipo 1	+	-	+	Excelente
Tipo 2	+	+	+	Sólida
Tipo 3	+	+	-	Insatisfatória
Tipo 4	-	+	-	Péssima
Tipo 5	-	-	-	Muito ruim
Tipo 6	-	-	+	Alto risco

Fonte: Adaptado de Braga (1991); Marques e Braga (1995).

O Modelo Fleuriet e Zeidan (2015) classifica as empresas em seis tipos com base na gestão do capital de giro e liquidez. O tipo 1 tem excelente liquidez, com grande reserva de tesouraria. O tipo 2 mantém saldo positivo de tesouraria, mesmo com necessidade de capital de giro positiva. O tipo 3 enfrenta risco de liquidez, com saldo de tesouraria negativo. O tipo 4 financia ativos não circulantes com dívidas de curto prazo, expondo-se a riscos. O tipo 5 combina dívidas de curto prazo e necessidade de capital de giro negativa, aumentando o risco de insolvência. O tipo 6 mantém saldo de caixa positivo, mas pode ter liquidez corrente inferior a 1. Estudos como os de Prado *et al.* (2018) e Silva *et al.* (2022) reforçam a importância desse modelo na gestão de risco de crédito e insolvência, destacando a análise setorial e de mercado.

2.3 Estudos de Previsão de Insolvência com Análise Discriminante

O estudo seminal de Altman (1968) foi um dos primeiros a utilizar a técnica de Análise Discriminante para previsão de insolvência empresarial. A amostra compreendeu 66 empresas industriais norte-americanas, sendo 33 classificadas como insolventes 1 ano antes do patrimônio líquido negativo e 33 solventes entre os anos de 1946 e 1965, pareadas conforme tamanho e setor. Foram utilizadas cinco variáveis financeiras, sendo elas o capital de giro sobre total do ativo, lucro retido sobre total do ativo, lucro antes dos juros e impostos sobre total do ativo, valor de mercado do patrimônio líquido sobre valor contábil da dívida e vendas sobre total do ativo. O modelo apresentou acurácia de 95% na amostra, demonstrando a robustez inicial da técnica mesmo com número limitado de variáveis.

Prado *et al.* (2018) realizaram um estudo de previsão do risco de crédito baseada no Modelo Dinâmico de Fleuriet, aplicada a empresas brasileiras com dados do período entre 2015 e 2019. A amostra foi composta por 121 empresas, sendo 70 classificadas como solventes e 51 insolventes. O modelo de Análise Discriminante foi desenvolvido a partir de 13 indicadores financeiros do modelo dinâmico, obtendo uma acurácia de 87,1% na amostra de teste, evidenciando a capacidade da função discriminante em diferenciar grupos com base em variáveis do modelo de Fleuriet. As variáveis mais relevantes na classificação foram o capital de giro sobre o ativo total, a necessidade de capital de giro sobre a receita líquida, o tipo de estrutura financeira, o termômetro de liquidez dinâmico e passivo financeiro de curto e longo prazo sobre o ativo total.

Prado (2019) analisou a aplicação de técnicas estatísticas para previsão de insolvência de empresas brasileiras. A amostra foi composta por 121 empresas, sendo 70 solventes e 51 insolventes. Foram utilizados 35 indicadores financeiros explorados na literatura, incluindo O modelo de Análise Discriminante apontou como variáveis mais relevantes a margem líquida e o grau de endividamento. A acurácia obtida na amostra de teste foi de 90,9%, o que demonstra a efetividade da técnica para o contexto brasileiro. As variáveis mais relevantes para classificação de empresas insolventes foram o capital de giro sobre o ativo total e o tipo de estrutura financeira.

No trabalho de Lima, Paulino e Fávero (2022) buscou-se investigar o emprego da Análise Discriminante na previsão do risco de insolvência em empresas brasileiras no período de 2010 a 2020. A amostra foi composta de 2.776 observações e 72 empresas. Foram utilizadas 7 variáveis financeiras 3 relacionadas a fatores ambientais, sociais e de governança. O modelo apresentou uma acurácia de 93,8% na amostra de teste, sendo as variáveis ambientais as mais importantes no modelo, bem como imobilizado sobre o patrimônio líquido, governança, capital de terceiros, endividamento, retorno sobre o ativo, giro do ativo e composição do endividamento.

No trabalho de Costa (2023), foi aplicado o modelo de Análise Discriminante para previsão do risco de insolvência em empresas brasileiras de capital aberto entre 2010 e 2021. A autora utilizou 27 indicadores financeiros iniciais distribuídos entre liquidez, endividamento, rentabilidade e atividade, além de uma variável qualitativa relativa à estrutura de capital. A amostra foi balanceada, composta por 86 empresas insolventes e 86 empresas solventes, para equiparação. O modelo apresentou acurácia de 70,83%.

Huo, Chan e Miller (2024) desenvolveram um modelo preditivo para falências no setor de restaurantes nos Estados Unidos com base na técnica de Análise Discriminante. A amostra

foi composta por 72 empresas, sendo 36 falidas e 36 não falidas entre os anos de 2000 e 2019. As variáveis independentes foram 20 indicadores financeiros relacionados a liquidez, rentabilidade, alavancagem, eficiência operacional e estrutura de capital. O modelo apresentou acurácia de 80% na amostra de teste, mostrando-se altamente eficaz na identificação de empresas insolventes.

Apesar das evidências apresentadas na literatura referente a modelos construídos a partir da técnica Análise Discriminante na previsão do risco de insolvência, observa-se que os estudos existentes ainda apresentam limitações relevantes principalmente no contexto brasileiro. Em geral, os trabalhos utilizam amostras restritas com relação ao período de análise, de setores e do porte das empresas, além de empregar conjuntos de variáveis pouco integradas, analisando de forma isolada indicadores financeiros tradicionais ou variáveis derivadas do Modelo Dinâmico de Fleuriet.

De forma complementar, há a carência de estudos que analisem de forma integrativa variáveis financeiras tradicionais e do modelo dinâmico no contexto das empresas brasileiras não financeiras, de forma a construir um modelo com amostras recentes e que integrem este conjunto de variáveis para identificação de empresas solventes e insolventes. Nesse sentido, permanece aberta a necessidade de investigar se a Análise Discriminante, aplicada a um conjunto integrado de indicadores financeiros, é capaz de oferecer elevada capacidade classificatória e evidenciar os determinantes do risco de insolvência no cenário corporativo brasileiro.

3 METODOLOGIA

Com relação ao *locus* da pesquisa, tem-se por objeto as demonstrações financeiras que foram coletadas na plataforma *online* Economatica no período de 2010 a 2024. Foram utilizados dados anuais, para adequação e cálculo dos indicadores tradicionais e do modelo dinâmico. Os filtros realizados referem-se às empresas brasileiras ativas na B3 de capital aberto, com exceção do setor financeiro, que possui particularidades operacionais, bem como no reconhecimento, mensuração, apresentação e divulgação dos fatos contábeis.

A amostra compreende empresas que, nesse período, apresentaram patrimônio líquido negativo ou protocolaram pedido de recuperação judicial. Para realizar a previsão e classificação das empresas solventes e insolventes, foram utilizados os dados das demonstrações financeiras referentes ao exercício imediatamente anterior ao pedido de recuperação judicial ou à ocorrência de patrimônio líquido negativo. Dessa forma, cada empresa foi considerada apenas uma vez no corte transversal, o que assegura a independência entre as observações da amostra.

Para fins de equiparação, a cada empresa insolvente foi associada uma empresa solvente do mesmo setor de atuação e com valor do ativo total mais próximo possível, de modo a garantir maior comparabilidade entre os pares. Esse procedimento resultou na formação de 87 pares, totalizando 174 empresas, cujos dados foram coletados na plataforma Economatica, em consonância com a metodologia empregada em estudos anteriores (Brito; Assaf Neto; Corrar, 2009; Costa, 2023; Huo *et al.*, 2024; Prado, 2016; Sanvicente; Minardi, 1998).

Os resultados foram gerados pelo programa estatístico *Statistical Package for the Social Sciences* (SPSS), versão 26, ano 2019.

3.1 Amostra

A distribuição das empresas por setor econômico evidencia a diversidade da amostra analisada. Ao todo, foram identificadas 17 categorias, incluindo uma denominada Outros (17 empresas insolventes), que reúne setores com baixa representatividade individual. As demais categorias são: Construção (9), Siderurgia e Metalurgia (9), Transporte e Serviços (9), Têxtil (7), Alimentos e Bebidas (6), Energia Elétrica (6), Comércio (4), Telecomunicações (4), Química (3), Veículos e Peças (3), Agro e Pesca (2), Máquinas Industriais (2), Mineração (2), Petróleo e Gás (2), Minerais Não Metálicos (1) e Software e Dados (1).

Ao classificar a amostra no Tipo de Estrutura Financeira (TEF) do Modelo Dinâmico de Fleuriet, verifica-se que a classificação dos tipos está de acordo com as informações na literatura (Prado *et al.* 2019). O Tabela 5 apresenta a amostra classificada no TEF.

Tabela 5 – Classificação da amostra no Tipo de Estrutura Financeira.

	CDG	NCG	T	Classificação	Solventes	Insolventes	Total
Tipo 1	+	-	+	Excelente	6	4	10
Tipo 2	+	+	+	Sólida	25	7	32
Tipo3	+	+	-	Insatisfatória	33	25	58
Tipo 4	-	+	-	Péssima	19	41	60
Tipo 5	-	-	-	Muito ruim	3	9	12
Tipo 6	-	-	+	Alto risco	1	1	2
				Total	87	87	174

Adaptado de Prado *et al.* (2019, p. 355)

Com relação as empresas solventes da amostra, nota-se que a maior parte se encontra nas três primeiras classificações, sendo elas Excelente (6 indivíduos), Sólida (25) e Insatisfatória (33), representando 74% da amostra de solventes. Já para a amostra das empresas insolventes, a maior parte encontra-se nas três últimas classificações, sendo as piores no TEF, apresentando 59% da amostra de insolventes. Assim, os resultados evidenciam que o Tipo de Estrutura Financeira introduzido pelo Modelo Dinâmico está em conformidade com a amostra coletada pelos critérios estabelecidos neste estudo, sendo eles o Patrimônio Líquido Negativo, o Pedido de Recuperação Judicial e o Pedido de Recuperação Extrajudicial.

Uma situação interessante ocorre com as duas empresas classificadas como Tipo 6 (Alto risco) no qual evidencia diferenças objetivas que explicam seus desfechos distintos quanto à solvência.

A empresa solvente Bemobi Tech, setor Software e Dados, embora igualmente enquadrada no tipo de maior risco do Modelo Dinâmico de Fleuriet, mantém níveis menos extremos de alavancagem, estrutura patrimonial menos desgastada e indicadores operacionais com menor volatilidade do que a empresa insolvente SPTuris, do setor Outros. Essa combinação sustenta sua capacidade de honrar obrigações financeiras e evita o avanço para um quadro de insolvência, mesmo diante de fragilidades estruturais. Assim, os resultados demonstram que, dentro de um mesmo tipo de estrutura financeira, a intensidade dos desequilíbrios é o elemento decisivo que diferencia empresas que permanecem solventes daquelas que evoluem para a insolvência.

3.2 Definição dos Indicadores

Para este estudo, utilizou-se as variáveis de Prado *et al.* (2018) e Costa (2023), agrupando indicadores financeiros de modelos comumente utilizados, denominados indicadores tradicionais, e do modelo dinâmico de Fleuriet, correspondendo inicialmente em um total de 35 variáveis independentes a serem analisadas. Para o desenvolvimento do modelo, os dados foram winsorizados a 5% (2,5% em cada cauda da distribuição), a fim de reduzir o viés causado por *outliers*, limitando valores extremos e garantindo maior robustez às estimativas. Com relação a padronização das variáveis, foi realizada durante a construção da função discriminante, dispensando a necessidade de transformação prévia.

Para avaliação de ausência de multicolinearidade entre as variáveis, sendo um pressuposto para o modelo a ser desenvolvido a partir do método estatístico Análise Discriminante, (Corrar et al., 2007; Endri, 2018), foi utilizado o VIF (Variance Inflation Factor), onde valores acima de 10 indicam que a variável possui alta colinearidade (Gujarati; Porter, 2011). Além disso, foi utilizado a matriz de correlação para verificar a correlação entre as variáveis. Segundo Gujarati e Porter (2011), correlações maiores que 0,8 é indicativo de colinearidade entre as variáveis. A verificação prévia do VIF e da Matriz de Correlação cumpre papel análogo aos métodos de regularização, limitando o número efetivo de variáveis ativas e controlando a variância dos estimadores (Kuhn; Johnson, 2013; James *et al.*, 2023). A Tabela 6 apresenta as variáveis financeiras a serem analisadas estatisticamente neste estudo.

Tabela 6 – Variáveis Financeiras Utilizadas Neste Estudo (continua).

Nome	Indicador Financeiro
X2	$\frac{\text{Capital de Giro}}{\text{Receita Líquida}}$
X4	$\frac{\text{Necessidade de Capital de Giro}}{\text{Receita Líquida}}$
X6	$\frac{\text{Saldo de Tesouraria}}{\text{Receita Líquida}}$
X7	$\frac{\text{Passivo Financeiro}}{\text{Ativo Circulante}}$
X9	$\frac{\text{Saldo de Tesouraria}}{\text{Necessidade de Capital de Giro}}$

Tabela 6 – Variáveis Financeiras Utilizadas Neste Estudo (conclusão).

Nome	Indicador Financeiro
	<i>Passivo Circulante Financeiro</i>
	+
X13	$\frac{\text{Passivo Não Circulante Financeiro}}{\text{Ativo Total}}$
X14	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$
X15	$\frac{\text{Receita Líquida Operacional}}{\text{Ativo Total}}$
X16	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Ativo Total}}$
	<i>Disponibilidade</i>
X19	$\frac{\text{Ativo Circulante}}{\text{Realizável a Longo Prazo}}$
	+
X20	$\frac{\text{Estoque}}{\text{Ativo Total}}$
X21	$\frac{\text{Fornecedores}}{\text{Ativo Total}}$
X26	$\frac{\text{Ativo Circulante} - \text{Passivo Total}}{\text{Ativo Total}}$
	<i>Patrimônio Líquido</i>
	-
X30	$\frac{\text{Capital Social}}{\text{Ativo Total}}$
X33	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Impostos sobre Despesas Financeiras}}$
X34	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Dívida Líquida}}$

Fonte: Elaborado pelo autor (2025).

Dessa forma, o modelo mantém as variáveis com maior contribuição econômica e estatística à discriminação entre empresas solventes e insolventes, evitando distorções causadas por relações espúrias ou redundantes entre indicadores contábeis.

Após realizados os testes e análises, as 16 variáveis que sobraram e que foram utilizadas no modelo são X2, X4, X6, X7, X9, X13, X14, X15, X16, X19, X20, X21, X27, X30, X33, X34.

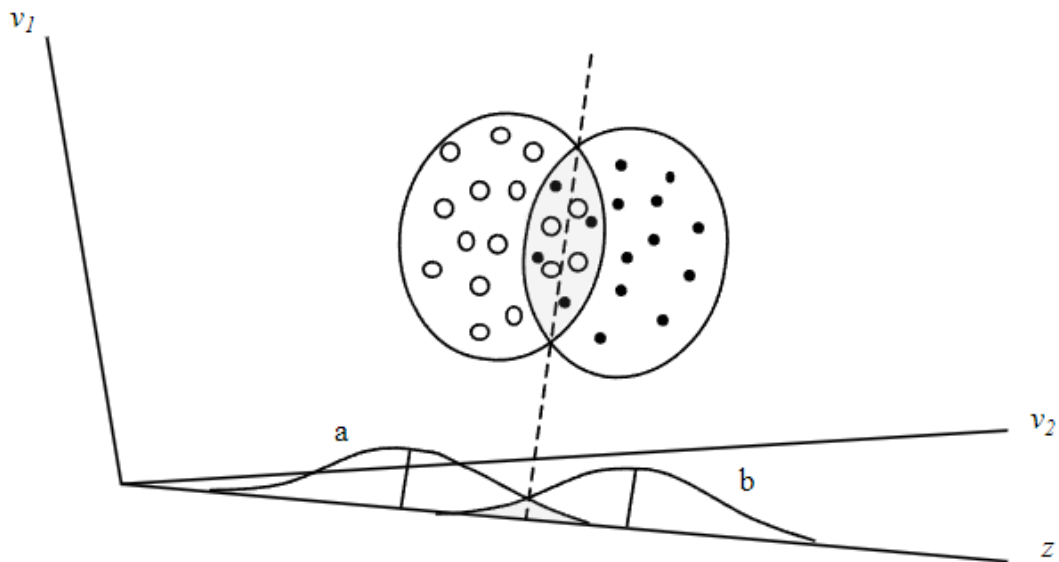
3.3 Análise Discriminante

A Análise Discriminante (AD) é um método estatístico amplamente utilizado na previsão de insolvência empresarial devido à sua simplicidade conceitual e baixo custo computacional. O método estima os centros de classes e a dispersão dos dados com base na média aritmética e na matriz de covariância da amostra, sendo eficaz quando as suposições sobre distribuição dos dados são atendidas (Hubert *et al.*, 2024).

Segundo Loza e Loza (2023), a AD permite uma avaliação rápida da saúde financeira de uma empresa utilizando um conjunto restrito de indicadores, facilitando a identificação precoce de riscos de falência. Altman (1968) foi pioneiro na aplicação desse método, desenvolvendo um modelo de previsão de falência com alta precisão e aplicabilidade em diferentes contextos econômicos (Antonowicz e Antonowicz, 2023).

Virgillito e Famá (2008) afirmam que, quando se trata de duas medidas, V_1 e V_2 , divididas em solventes e insolventes representados pelas elipses A e B, distinguidas por pontos pequenos e grandes conforme apresentado na Figura 8, bem como seus respectivos universos e o eixo Z determinado pela função discriminante baseada em índices, ao desenhar uma linha reta que atravessa a área de sobreposição das duas elipses e projetá-la em um novo eixo Z. De acordo com Hair *et al.* (2009), observa-se que a área de interseção entre as distribuições univariadas A e B é a menor possível quando comparada a todas as outras linhas que poderiam ser desenhadas através da área de sobreposição das elipses.

Figura 8 – Representação gráfica da Análise Discriminante de dois grupos.



Fonte: Prado *et al.* (2018, p. 350)

A interpretação dos resultados da AD também será fundamental para ratificar a classificação das empresas consideradas com alto grau de solvência ou insolvência por meio da análise dos indicadores do modelo dinâmico e tradicional (Virgillito e Famá, 2008).

A Equação 5 apresenta o modelo AD, que combina diferentes variáveis explicativas, ponderadas pelo impacto que causam na função, de forma a classificar os grupos analisados.

$$Z = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (5)$$

Em que:

Z , representa a variável dependente dicotômica, que indica uma pontuação ou score discriminante da observação em relação ao grupo;

a , é a constante da função quando todo $X_i = 0$;

b_n , é o coeficiente discriminante que cada variável independente tem, refletindo o peso relativo de cada indicador financeiro na separação entre os grupos;

X_n , são os valores das variáveis independentes.

A Análise Discriminante é um método estatístico multivariado relevante e amplamente aplicada na previsão de insolvência empresarial devido à sua robustez e interpretabilidade. Estudos recentes, como Huo *et al.* (2024), demonstram que, apesar da crescente adoção de modelos logísticos e de aprendizado de máquina, a Análise Discriminante apresenta bom desempenho na previsão de classificação. Além disso, Wieprow e Gawlik (2021) reforçam sua eficácia ao utilizá-la para avaliar o risco de insolvência no setor de turismo, destacando sua capacidade de classificação precisa mesmo em cenários de crise econômica.

Embora modelos de aprendizado de máquina tenham sido incorporados nos estudos de previsão do risco de crédito nos últimos anos, como nos trabalhos de Feng (2022), Xia *et al.* (2023) e Emmanuel *et al.* (2024), a Análise Discriminante (AD) apresenta vantagens específicas.

Entre elas, destaca-se o fato de oferecer um modelo parcimonioso, que minimiza o risco de *overfitting*, fenômeno em que o modelo se ajusta excessivamente aos dados de treinamento e falha em generalizar para novos dados, resultando em previsões não confiáveis. Além disso, a AD não depende de grandes volumes de dados para fornecer resultados consistentes. Esse cenário evidencia o *trade-off* entre interpretabilidade e precisão, frequentemente observado na comparação entre métodos estatísticos tradicionais e técnicas avançadas de aprendizado de máquina.

Assim, sua aplicação neste estudo justifica-se tanto pela validação empírica recente quanto pela necessidade de um modelo estatisticamente sólido e de fácil aplicabilidade para a previsão de insolvência.

3.3.1 Procedimentos e Teste do Modelo Discriminante

No modelo discriminante, a variável dependente foi dividida em dois grupos: solventes e insolventes. A amostra foi separada em dois subgrupos, um para desenvolver a função discriminante (70% da amostra) e outro para testar o modelo (30%), utilizando a técnica de *cross-validation*. O método *Stepwise* foi empregado para selecionar as variáveis independentes, com base no poder discriminante de cada uma.

Para assegurar a validade estatística do modelo construído a partir da Análise Discriminante, foram aplicados o teste de igualdade de médias entre os grupos, o teste de *Box' M* e o teste de *Wilk's Lambda*.

O teste de igualdade de médias, complementado pelo teste F-ANOVA, tem como objetivo verificar se as médias das variáveis independentes diferem entre os grupos de empresas, permitindo identificar aquelas com maior poder discriminante (Corrar *et al.*, 2007).

As hipóteses formuladas são:

H₀: As médias das variáveis independentes individuais são iguais entre os grupos de empresas e, portanto, não há poder discriminante.

H₁: Pelo menos uma das médias das variáveis independentes individuais difere entre os grupos, indicando capacidade discriminante significativa.

Para validação do teste, foi utilizado o grau de significância de 5%.

Já o teste de *Box's M* foi aplicado para verificar a igualdade das matrizes de covariância entre os grupos. Nesse teste, são formuladas as seguintes hipóteses:

H₀: As matrizes de covariância populacionais são iguais entre os grupos.

H₁: Pelo menos uma das matrizes de covariância difere entre os grupos.

A decisão é baseada no nível de significância adotado de 5%. Segundo Corrar *et al.* (2007), a rejeição da hipótese nula indica violação da premissa de normalidade multivariada,

podendo afetar a precisão da função discriminante, especialmente em amostras pequenas ou com distribuição não normal.

Por fim, o teste U de *Wilks' Lambda* foi aplicado para verificar a capacidade discriminante do modelo, isto é, se as variáveis selecionadas realmente permitem separar as empresas em grupos distintos.

As hipóteses consideradas foram:

H₀: As variáveis, em conjunto, possuem médias iguais entre os grupos, indicando ausência da capacidade discriminante.

H₁: As variáveis, em conjunto, apresentam médias diferentes entre os grupos, indicando capacidade discriminante significativa.

A discriminação entre os grupos foi determinada pelo menor valor de Lambda e nível de significância inferior a 5%. A atribuição dos grupos foi definida com base na probabilidade a priori, ponderando o tamanho de cada grupo e seu peso relativo na amostra.

A atribuição dos grupos foi definida com base na probabilidade a priori, ponderando o tamanho de cada grupo e seu efeito.

4 RESULTADOS E DISCUSSÃO

4.1 Teste de Multicolinearidade entre as variáveis independentes

Um dos pressupostos da Análise Discriminante é a ausência de multicolinearidade entre as variáveis (Corrar *et al.*, 2007; Endri, 2018). A Tabela 7 apresenta os valores de VIF das variáveis utilizadas na construção do modelo discriminante.

Tabela 7 – VIF das variáveis utilizadas na construção do modelo.

Nome	VIF
X2	2,66
X4	3,77
X6	1,55
X7	2,32
X9	1,38
X13	3,66
X14	2,08
X15	4,54
X16	5,83
X19	2,25
X20	2,55
X21	2,8
X26	4,23
X30	3,33
X33	1,88
X34	1,21

Fonte: Elaborado pelo autor (2025).

Por meio da Tabela 7, observamos que o fator de inflação da variância está abaixo de 10, indicando que as variáveis independentes não possuem colinearidade severa, permitindo maior robustez na previsão do modelo.

As estatísticas descritivas antes da transformação dos dados são apresentadas na Tabela 8, informando os valores de média, mínimo, máximo, desvio padrão, variância, assimetria e curtose.

Tabela 8 – Estatística descritiva das variáveis.

Variável	Média	Mínimo	Máximo	D. Padrão	Variância	Assimetria	Curtose
X2	0,175	-3,525	7,164	1,662	2,762	2,234	8,623
X4	0,748	-0,700	8,125	1,711	2,926	3,293	10,585
X6	-4,94e-07	-8,00e-06	1,00e-06	1,53e-06	2,36e-12	-3,600	14,215
X7	0,423	1,00e-06	3,22	0,661	0,436	3,017	9,51
X9	-0,622	-17,392	20,016	6,022	36,261	0,624	4,849
X13	0,239	1,00e-06	0,697	0,2	0,04	0,821	-0,389
X14	6,265	1,00e-06	156,829	26,155	684,105	5,53	28,929
X15	0,579	1,00e-06	1,835	0,49	0,24	0,746	-0,191
X16	-0,09	-1,68	0,247	0,396	0,157	-3,091	9,02
X19	0,202	1,00e-06	0,994	0,267	0,071	1,876	2,626
X20	0,082	1,00e-06	0,334	0,095	0,009	0,987	-0,028
X21	0,071	1,00e-06	0,345	0,074	0,005	1,917	4,297
X27	-0,543	-0,989	1,00e-06	0,272	0,074	0,561	-0,537
X30	-0,693	-16,865	0,672	3,052	9,317	-4,709	21,448
X33	1,019	-9,649	24,155	5,757	33,14	2,3	7,859
X34	0,242	-11,042	13,093	3,389	11,487	0,752	7,826

Fonte: Elaborado pelo autor (2025).

As estatísticas descritivas das variáveis revelam a presença de heterogeneidade, onde o valor das médias possui uma grande variação, além de diferentes assimetrias e curtoses acentuadas, o que indica a presença de caudas pesadas. Tais constatações não inviabilizam a previsão pela Análise Discriminante e reforçam a diferença de comportamentos financeiros das empresas da amostra devido as suas particularidades de negócios e de setor.

4.2 Construção do modelo Análise Discriminante

A Tabela 9 evidencia a amostra geral dos indivíduos utilizadas para gerar a função discriminante, bem como os indivíduos que foram separados em treino e teste para realizar a previsão de separação dos grupos.

Tabela 9 – Resumo de processamento de caso de análise.

Amostra	Qtd.	%
Selecionado - treinamento	129	74,1%
Não selecionado - validação	45	25,9%
Total	174	100%

Fonte: Elaborado pelo autor (2025).

Dos 174 indivíduos que compreendem o total da amostra (100%), 129 (74,1%) formam a amostra de treinamento do modelo, e o restante, 45 indivíduos, é a amostra de validação do modelo para previsão da classificação dos grupos.

O primeiro foi o de igualdade de médias, cujo resultado indicou que as variáveis X6 (Saldo de Tesouraria sobre a Receita Líquida), X7 (Passivo Financeiro sobre o Ativo Circulante), X13 (Passivo Circulante Financeiro e Passivo Não Circulante Financeiro sobre o Ativo Total), X14 (Ativo Circulante sobre o Passivo Circulante), X19 (Disponibilidade sobre Ativo Circulante e Realizável a Longo Prazo), X21 (Fornecedores sobre o Ativo Total) e X33 (Lucro Antes dos Juros e Impostos sobre Impostos Sobre Despesas Financeiras), apresentaram valores de significância inferiores a 5% ($p\text{-valor} < 0,05$). Desta forma, rejeita-se H_0 , indicando que estas são variáveis possuem forte poder de separação entre empresas solventes e insolventes.

Na sequência a Tabela 10 apresenta o resultado do teste *Box's M*, aplicado para verificar a hipótese nula de igualdade das matrizes de covariância entre os grupos de empresas solventes e insolventes.

Tabela 10 – Resultado do teste Box's M.

Z				
Box's M	Aprox.	df1	df2	Sig.
703,656	44,901	15	59507,614	0,001

Fonte: Elaborado pelo autor (2025).

O resultado do teste mostra que a significância é menor que 0,001, inferior ao nível de significância de 5%, o que significa violação da premissa. Dessa forma, rejeita-se a hipótese nula (H_0) de igualdade das matrizes de covariância, indicando violação da premissa de homogeneidade entre os grupos analisados. Essa violação pode ser fruto tanto do tamanho da amostra quanto pela assimetria na composição dos grupos, indicando ausência de normalidade multivariada.

Apesar disso, diversos estudos têm demonstrado que essa premissa raramente se sustenta em contextos empresariais com indicadores financeiros assimétricos e amostras de setores heterogêneos. Voda *et al.* (2021) e Prado *et al.* (2019) reforçam que, mesmo em cenários sem normalidade multivariada comprovada, modelos de AD mantêm elevado poder de previsão, desde que submetidos a critérios rigorosos de seleção de variáveis e validação estatística. Isso respalda a escolha da AD nesta pesquisa, considerando sua interpretabilidade e acurácia em contextos reais de previsão de insolvência.

Na sequência realizou-se o teste U (*Wilk's Lambda*) apresentado na Tabela 11, que avalia a capacidade discriminante do modelo.

Tabela 11 – Resultado do teste Wilk's Lambda.

Variáveis Inseridas	Lambda	df1	df2	df3	Estatística	df1	df2	Sig.
X33 (EBIT/ImpDespFin)	0,860	1	1	127	20,681	1	127	0,001
X13 ((PF+PNCF)/AT)	0,702	2	1	127	26,689	2	126	0,001
X16 (EBIT/AT)	0,681	3	1	127	19,520	3	125	0,001
X14 (AC/PC)	0,648	4	1	127	16,815	4	124	0,001
X6 (T/RecL)	0,627	5	1	127	14,660	5	123	0,001

Fonte: Elaborado pelo autor (2025).

Observa-se que na análise discriminante tem-se 7 melhores variáveis para a função discriminante, e todas apresentaram significância inferior a 0,05 (p-valor de 0,05), rejeitando H_0 , confirmando que o conjunto de variáveis possui capacidade discriminante estatisticamente significativa. Nesse sentido, as variáveis que melhor explicam o modelo são X6 (Saldo de Tesouraria sobre a Receita Líquida), X13 (Passivo Circulante Financeiro e Passivo Não Circulante Financeiro sobre o Ativo Total), X14 (Ativo Circulante sobre o Passivo Circulante), X16 (Lucro Antes dos Juros e Impostos sobre o Ativo Total) e X33 (Lucro Antes dos Juros e Impostos sobre Impostos Sobre Despesas Financeiras). As demais não foram selecionadas.

Na Tabela 12, apresentam-se os coeficientes estimados da função discriminante, que sintetizam o peso relativo de cada indicador na separação dos grupos.

Tabela 12 – Coeficientes da função discriminante canônica não padronizada.

Variáveis	Função
X6 (T/RecL)	-206225,484
X13 ((PF+PNCF)/AT)	3,809
X14 (AC/PC)	-0,018
X16 (EBIT/AT)	-1,575
X33 (EBIT/ImpDespFin)	-0,090
(Constante)	-1,019

Fonte: Elaborado pelo autor (2025).

A partir dos coeficientes apresentados na Tabela 12, estima-se a função discriminante canônica não padronizada, que representa a combinação linear dos indicadores selecionados pelo modelo. A Equação 6 apresenta a forma final da função discriminante obtida neste estudo.

$$\begin{aligned}
 Z = & -1,019 - 206225,484 \cdot \left(\frac{T}{RecL} \right) + 3,809 \cdot \left(\frac{(PF + PNCF)}{AT} \right) - 0,018 \cdot \left(\frac{AC}{PC} \right) \\
 & - 1,575 \cdot \left(\frac{EBIT}{AT} \right) - 0,090 \cdot \left(\frac{EBIT}{ImpDespFin} \right)
 \end{aligned}
 \tag{6}$$

A seguir, apresentam-se as análises individuais das variáveis selecionadas pelo modelo, considerando evidências da literatura.

O coeficiente negativo associado a variável X6 (Saldo de Tesouraria sobre a Receita Líquida) indica que maiores níveis de folga financeira reduzem a probabilidade de insolvência, exatamente o sinal esperado para esse indicador, cuja lógica teórica pressupõe que pressões provocadas no ciclo de conversão de caixa levam a empresa ao risco de crédito. O trabalho de Prado *et al.* (2018) testou indicadores financeiros relacionados ao modelo dinâmico de Fleuriet, no qual não encontrou significância estatística para este indicador financeiro, levando em consideração a amostra e variáveis testados. Neste estudo, este indicador financeiro apresentou significância no modelo discriminante e está em conformidade com o que foi proposto por Fleuriet. Assim, a significância e o sinal esperado de T/RecL confirmam que a deterioração das contas erráticas eleva o risco de insolvência.

A variável X13 (Passivo Financeiro Circulante e Passivo Financeiro não Circulante sobre o Ativo Total) apresentou coeficiente positivo, único entre as selecionadas, confirmando exatamente o sinal esperado para um indicador de alavancagem financeira, onde quanto maior o peso da dívida financeira na estrutura de capital, maior o risco na tentativa de alavancagem financeira por meio do capital de recurso, e nesse sentido, maior a probabilidade de insolvência. Isso converge diretamente com Prado *et al.* (2018), que identificaram o endividamento financeiro como um dos principais discriminantes de risco. Essa coerência também é observada em Lima, Paulino e Fávero (2022), que demonstram que a composição da dívida exerce influência significativa sobre a probabilidade de falência.

Já os achados de Belo, Cescon e Cescon (2022) evidenciam que indicadores de endividamento, como o Exigível Total sobre o Ativo e ao Patrimônio Líquido, são relevantes na caracterização da insolvência, uma vez que o excesso de capital de terceiros intensifica o risco de deterioração financeira das empresas. Portanto, o sinal positivo e sua significância estatística reforçam que empresas mais alavancadas têm maior propensão a enfrentar restrições de liquidez e deterioração financeira.

Já o coeficiente negativo da variável X14 (Ativo Circulante sobre o Passivo Circulante) confirma a expectativa teórica de que maiores níveis de liquidez corrente reduzem o risco de

insolvência, uma vez que refletem maior capacidade de honrar obrigações de curto prazo. Estudos como Sanvicente e Minardi (1998) e Prado *et al.* (2019) destacam que a liquidez operacional e a gestão do capital de giro são determinantes para a solvência. Esse resultado é compatível com Belo, Cescon e Cescon (2022), que utilizam a liquidez corrente como um dos principais indicadores nos modelos de fator de insolvência. O sinal encontrado está, portanto, plenamente alinhado à teoria, em que empresas com o índice de liquidez corrente mais elevado tendem a permanecer solventes, reduzindo a pressão financeira de curto prazo.

O coeficiente negativo observado para X16 (Lucro Antes dos Juros e Impostos sobre o Ativo Total) confirma o sinal esperado para indicadores de rentabilidade, onde quanto maior a capacidade de geração de resultado operacional, menor a probabilidade de insolvência. A literatura reforça essa lógica, mesmo quando não utiliza diretamente este indicador financeiro. Prado *et al.* (2019) demonstram que indicadores de desempenho como ROE e margem líquida são discriminantes importantes, e Lima, Paulino e Fávero (2022) apontam o retorno sobre o ativo como variável central na previsão de falência.

Os achados corroboram com o estudo de Belo, Cescon e Cescon (2022), que demonstram que o EBIT sobre o Ativo Total é variável significativa de rentabilidade operacional, evidenciando que a incapacidade de geração de resultados operacionais está diretamente associada ao risco de insolvência. Portanto, o sinal negativo encontrado reforça que maior eficiência operacional reduz substancialmente a chance de a empresa se tornar insolvente.

A variável X33 (Lucro Antes dos Juros em Impostos sobre os Impostos sobre as Despesas Financeiras) apresentou coeficiente negativo, indicando que maior capacidade de cobertura das despesas financeiras diminui o risco de insolvência, exatamente o comportamento esperado para indicadores de cobertura. Embora Sanvicente e Minardi (1998) não tenham identificado significância desse índice, os autores destacam que métricas de cobertura são sensíveis à alavancagem e ao custo da dívida, relacionando-se diretamente à resiliência financeira da empresa.

De forma complementar, o estudo de Valaskova *et al.* (2023) evidencia o índice de cobertura de juros (Lucro Antes dos Juros e Impostos sobre Despesas Financeiras) é uma variável financeira significativa para a avaliação do risco financeiro ao analisar o indicador em ambientes de estresse econômico como o ocorrido na pandemia da COVID-19. Assim, o sinal negativo e sua significância reforçam que empresas capazes de gerar EBIT suficiente para cobrir o serviço da dívida possuem menor probabilidade de falência, principalmente em ambientes de juros elevados.

Em conjunto, os indicadores selecionados pela função discriminante revelam um padrão consistente de distinção entre empresas solventes e insolventes, combinando métricas de liquidez, endividamento e desempenho operacional que, em sua totalidade, refletem a solidez financeira da firma.

A presença de X13 $((PF+PNCF)/AT)$ como único coeficiente positivo reforça o papel central da alavancagem financeira no risco de insolvência, enquanto os coeficientes negativos de X6 $(T/RecL)$, X14 (AC/PC) , X16 $(EBIT/AT)$ e X33 $(EBIT/ImpDespFin)$ confirmam que maiores níveis de folga financeira, liquidez corrente, rentabilidade e capacidade de cobertura das despesas financeiras reduzem significativamente o risco de insolvência.

Esses sinais, todos alinhados às expectativas teóricas, evidenciam que o modelo discrimina adequadamente as empresas a partir de elementos estruturais do ciclo financeiro e operacional, demonstrando convergência com a literatura que enfatiza a interação entre liquidez, desempenho e estrutura de capital como determinantes fundamentais da continuidade empresarial.

Finalizando o processo de classificação da amostra, foram considerados os centróides dos grupos, que representam as médias das funções discriminantes para cada categoria de empresas. Conforme apresentado na Tabela 13, o grupo das empresas solventes apresenta centroide igual a -0,847, enquanto o grupo das empresas insolventes apresenta valor de 0,692. Esses centroides funcionam como pontos de referência para a função discriminante canônica, permitindo identificar o *cut-off point* que define a fronteira de separação entre os dois grupos. Assim, observações com escores próximos a valores positivos tendem a ser classificadas como solventes, ao passo que escores mais negativos indicam insolvência, evidenciando a capacidade discriminante do modelo.

Tabela 13 – Funções em centroides de grupo.

Y	Função
Solvente	-0,847
Insolvente	0,692

Fonte: Elaborado pelo autor (2025).

Já a Tabela 14 apresenta as probabilidades a priori utilizadas na análise discriminante, definidas com base na proporção de casos observados em cada grupo da amostra. Verifica-se que 45% das empresas pertencem ao grupo das solventes e 55% ao grupo das insolventes, proporções que refletem o comportamento empírico da base de dados e servem como ponderadores iniciais no processo de classificação. Essas probabilidades são fundamentais para

o cálculo das probabilidades posteriores (posteriori), uma vez que ajustam a decisão de classificação conforme a distribuição real das observações entre os grupos analisados.

Tabela 14 – Probabilidades a priori para grupos.

Y	A priori	Casos utilizados na análise
Solvente	0,450	58
Insolvente	0,550	71
Total	1,000	129

Fonte: Elaborado pelo autor (2025).

Para definição do ponto de corte que servirá para classificação dos casos, deve-se calcular os custos iguais de erros de classificação, ponderando a relação centróides *versus* probabilidades (Corrar *et al.*, 2007). Sendo assim, a Equação 7 apresenta qual foi o ponto de corte para a função discriminante.

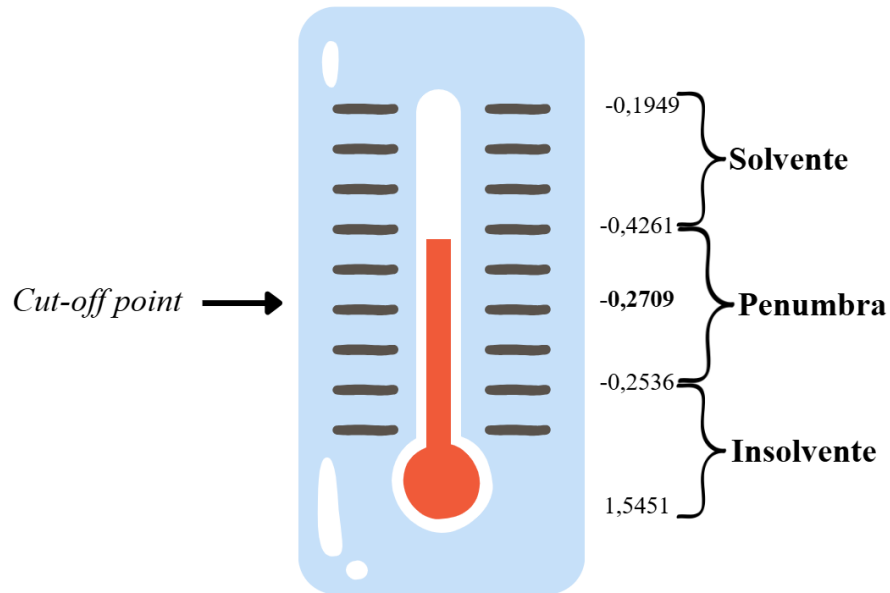
$$\text{Ponto de corte} = \frac{N_2 C_1 + N_1 C_2}{N_1 + N_2} = \frac{29(-0,847) + 16(0,692)}{29 + 16} = -0,2998 \quad (7)$$

Assim, o quando o valor calculado pela função discriminante canônica for maior que o ponto de corte (-0,2998), será considerado como uma empresa insolvente e, caso contrário, solvente.

Para complementar a interpretação da função discriminante desenvolvida neste estudo, foi desenvolvida uma nova ótica a partir do Termômetro de Insolvência de Kanitz, utilizando os scores discriminantes dos 45 casos não selecionados utilizados para a construção do modelo preditivo, extraíndo os centroides e o desvio padrão de cada grupo.

A Figura 9 demonstra que as empresas que possuem situação de solvência possuem um *score* inferior a -0,4261 foram classificadas como solvente. O ponto de corte de -0,2709 obtido por meio do Termômetro de Kanitz demonstra que empresas com *score* inferior a este valor foram classificadas como insolvente. Já a camada chamada de penumbra, segundo Kanitz (1974), a empresa que estiver no intervalo é classificada como situação indefinida.

Figura 9 – Termômetro de Kanitz da função discriminante.



Fonte: Elaborado pelo autor (2025).

Observa-se que scores compreendidos entre $-0,4261$ e $-0,2536$ situam-se na zona de penumbra, na qual o modelo matemático não apresenta poder discriminatório suficiente para classificar a empresa de forma segura como solvente ou insolvente. Por fim, valores inferiores a $-0,2536$ são classificados como insolventes, sinalizando maior fragilidade financeira e elevado risco de insolvência, conforme captado pela combinação dos indicadores econômico-financeiros utilizados na função discriminante.

Por fim, quanto aos resultados da classificação, a Tabela 15 evidencia que 82,2% dos casos originais selecionados foram classificados corretamente. Esse resultado demonstra a consistência inicial do modelo ao distinguir entre empresas solventes e insolventes.

Para verificar a robustez da função, foi realizado o teste *cross-validated* de Lachenbruch, também denominado *leave-one-out cross-validation* (LOOCV). Nesse procedimento, cada caso da amostra de desenvolvimento é retirado, o modelo é reestimado sem essa observação e, em seguida, aplicado sobre ela. O processo se repete para todas as observações, permitindo avaliar a capacidade de classificação do modelo sem a participação prévia da empresa analisada. Nesse cenário, obteve-se novamente um percentual de acertos de 82,2%, confirmando a estabilidade da função discriminante.

Pelos resultados auferidos, observa-se que, dentre as 16 empresas previamente classificadas como insolventes, apenas 1 foi incorretamente classificada como solvente. Tal situação evidencia um ponto de atenção, pois poderia induzir credores a investir ou conceder crédito a uma empresa que, na prática, já apresentava elevado risco de insolvência. Ainda assim,

a baixa incidência desse tipo de erro reforça a utilidade do modelo como ferramenta de apoio à decisão no contexto da previsão de insolvência.

Tabela 15 – Resultados da classificação.

		Y	Associação ao grupo prevista			
			Solvente	Insolvente	Total	
Casos selecionados	Original	Contagem	Solvente	49	9	58
			Insolvente	14	57	71
		%	Solvente	84,5	15,5	100,0
			Insolvente	19,7	80,3	100,0
	Com validação cruzada	Contagem	Solvente	49	9	58
			Insolvente	14	57	71
%		Solvente	84,5	15,5	100,0	
		Insolvente	19,7	80,3	100,0	
Casos não selecionados	Original	Contagem	Solvente	27	2	29
			Insolvente	1	15	16
		%	Solvente	89,7	10,3	100,0
			Insolvente	6,3	93,8	100,0

Fonte: Elaborado pelo autor (2025).

O principal resultado é o que se aplica a amostra teste, pois indica se a função consegue classificar bem empresas que não participaram a construção. Nesse sentido, a função conseguiu classificar 93,3% dos casos não selecionados de forma aleatória, apresentando excelente capacidade de previsão na classificação de grupos.

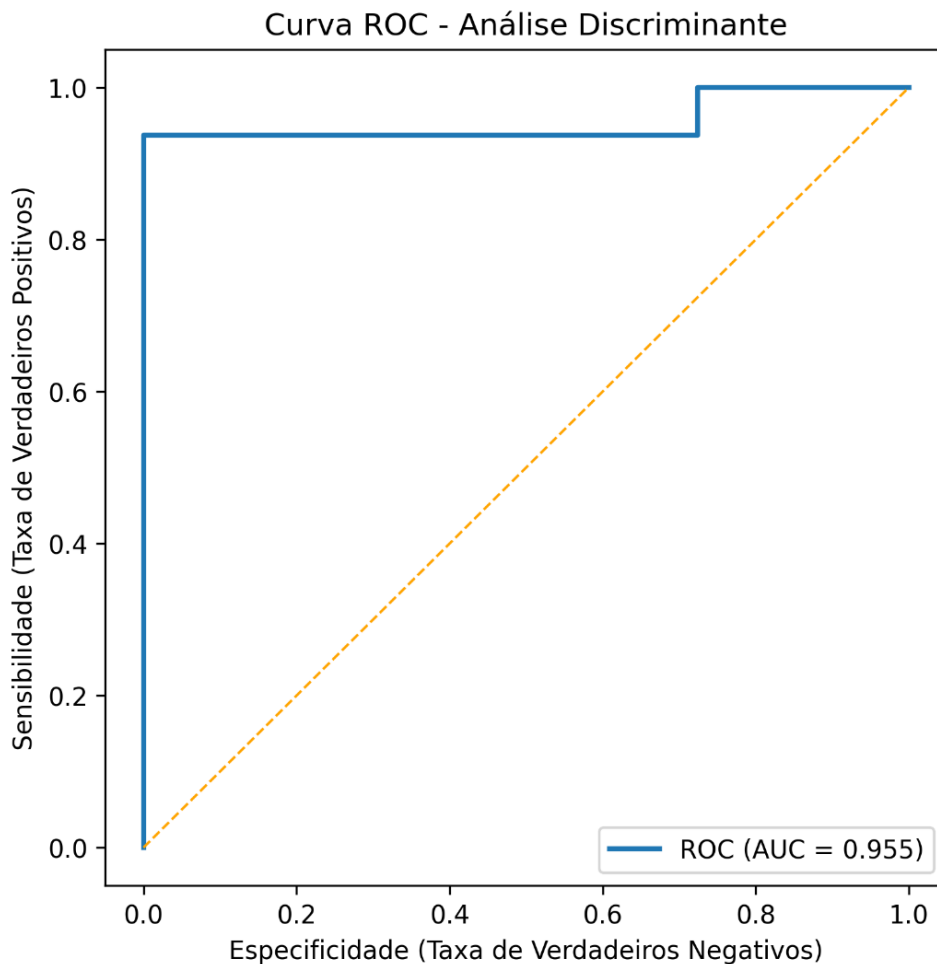
Esse desempenho obtido na amostra de teste é comparável aos estudos clássicos e contemporâneos que também aplicaram a Análise Discriminante para previsão de insolvência. Altman (1968), em seu modelo pioneiro, atingiu 95% de acurácia na amostra, enquanto Prado *et al.* (2018) e Prado (2019) alcançaram 87,1% e 90,9%, respectivamente, em empresas brasileiras. Resultados semelhantes foram encontrados por Lima, Paulino e Fávero (2022), com 93,8%, e por Costa (2023), ainda que com desempenho mais modesto de 70,8%. Já em estudos internacionais, como o de Huo *et al.* (2024), a técnica apresentou 80% de acurácia no setor de restaurantes norte-americano.

Esses achados reforçam que a taxa de 93,3% obtida neste trabalho se encontra dentro do intervalo de desempenho relatado na literatura, confirmando a robustez da técnica quando aplicada a diferentes contextos e conjuntos de variáveis.

De forma complementar, as principais variáveis discriminantes identificadas no presente estudo, sendo eles os indicadores de liquidez, rentabilidade, estrutura financeira e o passivo financeiro de curto e longo prazo sobre o ativo total, variável típica do Modelo Dinâmico de Fleuriet, também se mostraram consistentes com os resultados reportados por Prado *et al.* (2018), Prado (2019), Costa (2023) e Lima, Paulino e Fávero (2022). Esses elementos reforçam que, mesmo diante de limitações metodológicas conhecidas da Análise Discriminante, a técnica mantém elevada capacidade preditiva, sobretudo em amostras balanceadas e com variáveis selecionadas de forma criteriosa.

Já quanto a robustez do modelo, foi verificado por meio da Curva ROC (*Receiver Operating Characteristic*) para a amostra de teste (casos não selecionados), apresentada na Figura 10. O gráfico permite observar a frequência relativa entre verdadeiros positivos (sensibilidade) e falsos positivos (especificidade).

Figura 10 – Curva ROC do modelo desenvolvido pela AD dos casos não selecionados.



Conforme evidencia a Figura 10, a AUC (*Area Under the Curve*) obtida foi de 0,955, valor considerado elevado próximo ao limite de 1, demonstrando a capacidade do modelo em avaliar corretamente as classes insolventes (verdadeiros positivos) e solventes (verdadeiros negativos). Nesse sentido, Kochański (2024) reforça que valores de AUC mais altos ou próximos de 1 refletem maior capacidade de discriminação do modelo analisado, tornando essa métrica essencial para validação de modelos de previsão. Com isso, o valor obtido nesse estudo confirma a robustez da classificação da Análise Discriminante e seu desempenho superior ao acaso, ou seja, AUC maior que 0,5.

Na sequência, a Tabela 16 complementa essa análise ao informar os principais indicadores de desempenho relacionados a Curva ROC, sendo eles a AUC, acurácia, precisão, sensibilidade, especificidade, *F1-score*, VP (Verdadeiros Positivos), FP (Falsos Positivos), VN (Verdadeiros Negativos), FN (Falsos Negativos).

Tabela 16 – Métricas de desempenho do modelo obtidas a partir da Curva ROC.

AUC	Acurácia	Precisão	Sensibilidade	Especificidade	F1 Score	VP	FP	VN	FN
0,955	0,933	0,882	0,938	0,931	0,909	15	1	27	2

Fonte: Elaborado pelo autor (2025).

Observa-se que a acurácia é de 93,3%, mostrando que a maior parte das observações foram classificadas corretamente, assegurando a confiabilidade no modelo. A precisão, de 88,2%, evidencia que a maior parte das empresas classificadas como insolventes estavam nessa condição, reduzindo os falsos positivos.

A sensibilidade, de 93,8%, demonstra a elevada capacidade do modelo em identificar corretamente as empresas insolventes, enquanto a especificidade, de 93,1%, confirma sua eficiência na correta classificação das solventes. Esses resultados atendem ao que se espera em estudos de previsão de insolvência, nos quais é preferível priorizar a correta identificação de empresas insolventes, ainda que isso implique eventualmente classificar uma empresa solvente como insolvente. Dessa forma, os achados obtidos alinham-se ao esperado para a temática, reforçando a utilidade do modelo.

Por fim, o *f1 score*, de 90,9%, demonstra o equilíbrio entre precisão e sensibilidade, validando a consistência do modelo. Os valores absolutos de VP (15), FP (3), VN (26), FN (1) confirmam a baixa incidência de erros na classificação. Em conjuntos, os resultados estão de acordo com o alto valor de AUC, resultando que a Análise Discriminante é uma ferramenta robusta para previsão de insolvência de empresas brasileiras.

5 CONSIDERAÇÕES FINAIS

Este estudo investigou a aplicabilidade dos indicadores financeiros do Modelo Dinâmico de Fleuriet e dos Modelos Tradicionais na previsão de insolvência de empresas brasileiras listadas na B3 por meio do desenvolvimento de um modelo de classificação utilizando a técnica estatística Análise Discriminante.

Os resultados indicaram o Saldo de Tesouraria sobre a Receita Líquida (T/RecL) e o Passivo Circulante Financeiro e Passivo Financeiro Não Circulante Financeiro sobre o Ativo Total (PF + PNCF/AT) como os indicadores no Modelo Dinâmico de Fleuriet que contribuíram no modelo de classificação de empresas solventes e insolventes.

Já no Modelo Tradicional, os indicadores que melhor discriminam os grupos foram o Ativo Circulante sobre o Passivo Circulante, o Lucro Antes dos Juros e Impostos sobre o Ativo Total e o Lucro Antes dos Juros e Impostos sobre Impostos Sobre Despesas Financeiras (EBIT/Impostos Sobre Despesas Financeiras).

A técnica multivariada Análise Discriminante foi capaz de classificar 93,3% dos casos não selecionados, e a Curva ROC apresentou uma AUC de 0,955, possuindo, portanto, ótima capacidade de previsão do modelo. Trabalhos como os de Huo *et al.* (2024), Prado *et al.* (2018), Prado *et al.* (2019) e Antonowics (2023) também confirmam o bom poder de previsão da insolvência empresarial utilizando o modelo para risco de falência empresarial, corroborando com o estudo inicial de Altman (1968) ao utilizar a mesma técnica na equação Z-Score.

Do ponto de vista prático, os resultados destacam variáveis-chave que podem orientar credores, investidores e gestores na avaliação de risco de crédito e na tomada de decisão empresarial. Assim, o estudo avança o debate acadêmico e, ao mesmo tempo, entrega evidências aplicáveis ao monitoramento preventivo da saúde financeira de empresas listadas na B3.

Apesar da utilidade prática dos modelos desenvolvidos e do excelente poder discriminante, o estudo apresenta limitações. A principal delas é que, a Análise Discriminante parte da premissa da normalidade multivariada, a qual dificilmente é atendida em estudos que utilizam indicadores financeiros, dada sua natureza assimétrica e a heterogeneidade entre setores e empresas. Tal condição pode, em algum grau, comprometer a capacidade preditiva do modelo, ainda que, conforme os resultados demonstraram, o desempenho obtido tenha sido satisfatório. Outra limitação refere-se ao escopo deste estudo, que se concentrou em determinados indicadores financeiros e no contexto de empresas brasileiras de capital aberto, excetuando-se aquelas do setor financeiro, o que restringe a generalização dos achados.

Como perspectivas para pesquisas futuras, recomenda-se a incorporação de outros indicadores financeiros que possam ampliar a capacidade de discriminação entre empresas solventes e insolventes. Além disso, sugere-se a aplicação do modelo a amostras de empresas do setor financeiro, a organizações de diferentes países e a períodos adicionais de observação, de modo a verificar a robustez dos resultados em diferentes contextos e ampliar a comparabilidade internacional.

Por fim, conclui-se que a Análise Discriminante, aplicada a um conjunto integrado de indicadores financeiros tradicionais e dinâmicos, é capaz de prever insolvência com elevado grau de acurácia. Os achados oferecem uma contribuição teórica ao demonstrar que a combinação entre os dois conjuntos de indicadores amplia a capacidade discriminatória do modelo, reforçando a relevância do Modelo Dinâmico de Fleuriet em estudos contemporâneos sobre insolvência. Do ponto de vista prático, os resultados fornecem subsídios para que gestores, credores e investidores identifiquem precocemente situações de fragilidade financeira, apoiando decisões de crédito, alocação de capital e políticas de gestão do capital de giro.

REFERÊNCIAS

- ASSAF NETO, A.; SILVA, C. A. T. **Administração do Capital de Giro**. 4. ed. Atlas, 2012.
- BEAVER, W. H. **Financial Ratios as Predictors of Failure-Reply**. *Journal of Accounting Research*, v. 4, p. 123-127, 1966.
- ALTMAN, E. I. **Financial Ratios, Discriminant Analysis And The Prediction Of Corporate Bankruptcy**. *The Journal of Finance*, v. 23, n. 4, p. 589–609, 1968. DOI: <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>.
- ANTONOWICZ, P.; ANTONOWICZ, A. **The Capacity to Forecast Enterprise Insolvency on the Polish Market Using the Precursory Altman Z-Score Model (1968)**. *Annales Universitatis Mariae Curie-Skłodowska*, v. 56, n. 5, p. 7–26, 2023. DOI: <https://doi.org/10.17951/h.2022.56.5.7-26>.
- BELO, D. P. C.; CESCÓN, J. A.; CESCÓN, J. A. P. **Análise do fator de insolvência das empresas com patrimônio líquido negativo listadas na Brasil, Bolsa e Balcão (B3)**. *Contabilometria – Brazilian Journal of Quantitative Methods Applied to Accounting*, Monte Carmelo, v. 9, n. 1, p. 111–129, jan./jun. 2022.
- BRAGA, R. **Análise avançada do capital de giro**. *Caderno de Estudos*, 01-20, 1991. DOI: <https://doi.org/10.1590/S1413-92511991000100003>.
- BRITO, G. A. S., ASSAF NETO, A.; CORRAR, L. J. **Sistema de classificação de risco de crédito: uma aplicação a companhias abertas no Brasil**. *Revista Contabilidade & Finanças*, v. 20, n. 5, p. 1, 2009. DOI: 10.1590/S1519-70772009000300003.
- CARNEIRO, W. F.; BASTOS, E. V. P. **Modelo contábil fleuriet para a tomada de decisão nos subsetores agropecuários e de alimentos processados brasileiros**. *Revista Produção e Desenvolvimento*, 8(1), v8, p. 612, 2022.
- CORRAR, L. J., PAULO, E.; DIAS FILHO, J. M. **Análise Multivariada Para os Cursos de Administração, Ciências Contábeis e Economia**. Editora Atlas Ltda., 2007.
- COSTA, T. R. D. **Risco De Insolvência: Uma Análise de Diferentes Métodos de Previsão e do Efeito da Estrutura de Capital**. Dissertação de Mestrado. Universidade Federal de Lavras., 2023. Disponível em: <http://repositorio.ufla.br/jspui/handle/1/57037>. Acesso em: 30 de abril de 2025.
- EMMANUEL, I., SUN, Y.; WANG, Z. **A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method**. *Journal of Big Data*, v. 11, n. 1, p. 23, 2024. DOI: <https://doi.org/ez26.periodicos.capes.gov.br/10.1186/s40537-024-00882-0>.
- ENDRI. **Prediction of bankruptcy potential company of drinking water region**. *International Journal of Business and Management Invention*, v. 7, n. 5, p. 75–85, 2018. DOI: <https://dx.doi.org/10.2139/ssrn.3808630>.

FENG, Y. **Bank Green Credit Risk Assessment and Management by Mobile Computing and Machine Learning Neural Network under the Efficient Wireless Communication**. *Wireless Communications and Mobile Computing*, v. 2022, p. 1–11, 2022. DOI: <https://dx.doi.org/10.1155/2022/3444317>.

FLEURIET, M. **A Dinâmica Financeira das Empresas Brasileiras**. 1. ed. 2003.

FLEURIET, M; KEHDYE, R; BLANC, G. **A Dinâmica Financeira das Empresas Brasileiras: Um Novo Método de Análise, Orçamento e Planejamento Financeiro**. 1. ed. 1978.

FLEURIET, M; ZEIDAN, R. **O Modelo Dinâmico de Gestão Financeira**. 1. ed. 2015. Alta Books.

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria básica**. 5. ed. Porto Alegre: AMGH, 2011.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009.

HUBERT, M., RAYMAEKERS, J.,; ROUSSEEUW, P. J. **Robust discriminant analysis**. 2024. DOI: <https://doi-org.ez26.periodicos.capes.gov.br/10.1002/wics.70003>.

HUO, Y., CHAN, L. H.; MILLER, D. **Bankruptcy prediction for restaurant firms: A comparative analysis of multiple discriminant analysis and logistic regression**. *Journal of Risk and Financial Management*, v. 17, n. 39, p. 9, 2024. DOI: <https://doi.org/10.3390/jrfm17090399>.

IBM SPSS Statistics for Windows. Version 26.0. Armonk: IBM Corp., 2019.

KANITZ, S. **Como Prever Falências de Empresas**. *Revista Negócios em Exame*, p. 95–102, 1974.

KOCHAŃSKI, Błażej. **The shape of an ROC curve in the evaluation of credit scoring models**. *Statistics in Transition new series*, v. 25, n. 2, p. 205–218, jun. 2024. Disponível em: <https://doi.org/10.59170/stattrans-2024-022>.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. New York: Springer, 2013. DOI: <https://doi.org/10.1007/978-1-4614-6849-3>.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An introduction to statistical learning: with applications in Python**. Nova York: Springer, 2023. DOI: <https://doi.org/10.1007/978-3-031-38747-0>.

LIMA, F. G.; PAULINO, C. T.; FÁVERO, L. P. L. **ESG e Machine Learning: O Impacto na Previsão de Insolvência de Empresas Brasileiras**. F. G., *Contabilometria*, v. 11, p. 1, 2022.

LOZA, S.; LOZA, D. **Application of discriminant models for evaluating the probability of bankruptcy on the example of jsc “ukrzaliznytsya”**. *Zbirnik Naukovih Prac’*

Deržavnogo Univerzitetu Infrastrukturi Ta Tehnologij. 2023. DOI: <https://doi.org/10.32703/2664-2964-2023-53-6-16>.

MACHADO, M. A. V., MACHADO, M. R.; CALLADO, A. L. C. **Análise dinâmica e o financiamento das necessidades de capital de giro das pequenas e médias empresas localizadas na cidade de João Pessoa, PB: um estudo exploratório**. Base Revista de Administração e Contabilidade da UNISINOS, v. 3, n. 2, p. 139-149, 2006.

MARION, J. C. **Análise das demonstrações contábeis**. São Paulo: Atlas, 2019.

PRADO, J. W. D., CARVALHO, F. D. M., BENEDICTO, G. C. D., ALCÂNTARA, V. D. C.; SANTOS, A. C. D. **Uma abordagem para análise do risco de crédito utilizando o modelo Fleuriet**. Revista de Educação e Pesquisa em Contabilidade (REPeC), v. 12, n. 3, p. 341–363, 2018. DOI: <https://doi.org/10.17524/repec.v12i3.1816>.

PRADO, José Willer do. **Risco de crédito: uma abordagem utilizando análise discriminante, regressão logística e redes neurais artificiais**. 2016. 228 f. Dissertação (Mestrado em Administração) – Universidade Federal de Lavras, Lavras, 2016.

PRADO, J. W., CARVALHO, F. D. M., BENEDICTO, G. C. D.; LIMA, A. L. R. **Analysis of credit risk faced by public companies in Brazil: An approach based on discriminant analysis, logistic regression and artificial neural networks**. Estudios Gerenciales, v. 347, p. –360, 2019. DOI: <https://doi.org/10.18046/j.estger.2019.153.3151>.

PRADO, J. W. D., VILAMAIOR, A. G., CAMPOS, A. C.; NASCIMENTO, T. B. P. D. **Previsão da Insolvência Empresarial Utilizando Redes Neurais Artificiais**. Revista Gestão e Desenvolvimento, v. 17, n. 2, p. 136, 2020. DOI: <https://doi.org/10.17524/repec.v12i3.1816>.

SANVICENTE, A. Z.; MINARDI, A. M. A. F. **Identificação de indicadores contábeis significativos para a previsão de concordata de empresas**. Instituto Brasileiro de Mercado de Capitais, Working Paper, p. 1-12, 1998.

SILVA, J. L., SILVA, M. L. Q.; NOBRE, F. C. **Determinantes da insolvência: avaliação da relação entre o fator de insolvência, rentabilidade, tamanho e modelo dinâmico de capital de giro**. Caderno de Administração, v. 30, n. 2, p. 41, 2022. DOI: <https://doi.org/10.4025/cadadm.v30i2.61284>.

VALASKOVA, K.; DURANA, P.; KLIESTIK, T.; VOJTEKOVA, S. **Level of pandemic consequences for the indebtedness of the Slovak hospitality sector**. Investment Management and Financial Innovations, Sumy, v. 20, n. 4, p. 335–349, 2023. DOI: [https://doi.org/10.21511/imfi.20\(4\).2023.27](https://doi.org/10.21511/imfi.20(4).2023.27).

VIRGILLITO, S. B.; FAMÁ, R. **Estatística multivariada na construção de modelos para análise do risco de crédito e previsão de insolvência de empresas**. Revista Integração, v. 53, n. 13, p. 105-118, 2008. DOI: <https://doi.org/10.1016/j.ejco.2024.100084>.

VODA, A. D. *et al.* **Corporate bankruptcy and insolvency prediction model**. Technological and Economic Development of Economy, v. 27, n. 5, p. 1039–1056, 2021. DOI: <https://doi.org/10.3846/tede.2021.15106>.

WIEPROW, J.; GAWLIK, A. **The use of discriminant analysis to assess the risk of bankruptcy of enterprises in crisis conditions using the example of the tourism sector in Poland.** *Risks*, v. 9, n. 7, p. 8, 2021. DOI: <https://doi.org/10.3390/risks9040078>.

XIA, Y., XU, T., WEI, M.-X., WEI, Z.-K.; TANG, L.-J. **Predicting Chain's Manufacturing SME Credit Risk in Supply Chain Finance Based on Machine Learning Methods.** *Sustainability*. v. 15, n. 2, p. 1087, 2023. DOI: <https://doi.org/10.3390/su15021087>.

ARTIGO 2 – UMA ABORDAGEM DO RISCO DE INSOLVÊNCIA DE EMPRESAS BRASILEIRAS NÃO FINANCEIRAS UTILIZANDO A REGRESSÃO LOGÍSTICA

RESUMO

O estudo teve como objetivo avaliar a construção do modelo com a Regressão Logística na previsão do risco de insolvência empresarial, bem como identificar quais foram as variáveis independentes que contribuíram na classificação do risco. A técnica foi aplicada a empresas brasileiras não financeiras listadas na B3 entre 2010 e 2024. A amostra em painel foi composta por 5.247 observações anuais de 501 empresas. Inicialmente foram considerados 17 indicadores financeiros tradicionais e dinâmicos como variáveis independentes. A estimação por Efeitos Aleatórios permitiu capturar simultaneamente a variação entre empresas, a evolução temporal do risco e os efeitos individuais não observados, preservando a totalidade da amostra. O modelo apresentou excelente desempenho, com AUC de 0,915 e acurácia de 90,2%. As variáveis mais relevantes para explicar a probabilidade de insolvência foram: Patrimônio Líquido menos Capital Social sobre o Ativo Total, Disponibilidades sobre Ativo Circulante e Realizável a Longo Prazo, Ativo Circulante sobre Passivo Circulante, Necessidade de Capital de Giro, Capital de Giro sobre Receita Líquida, Financiamentos e Empréstimos Bancários sobre Ativo Total, Saldo de Tesouraria sobre Receita Líquida e Fornecedores sobre Ativo Total. Esses resultados reforçam que a adequada gestão da liquidez, da estrutura de capital e do capital de giro é determinante para prevenir a deterioração financeira. O estudo é limitado ao contexto brasileiro e à utilização de um único método estatístico. Pesquisas futuras podem ampliar a análise para empresas de outros países ou do setor financeiro, incorporar novas variáveis, incluindo qualitativas, e comparar o desempenho com métodos de aprendizado de máquina.

Palavras-chave: Risco de insolvência. Indicadores Financeiros. Regressão Logística. Dados em Painel.

ABSTRACT

This study aimed to evaluate the construction of a Logistic Regression model for predicting corporate insolvency risk, as well as to identify the independent variables that contributed to firm classification. The technique was applied to Brazilian non-financial companies listed on Brasil, Bolsa, Balcão (B3) over the period from 2010 to 2024. The panel data sample comprised 5,247 annual observations from 501 firms. Initially, 17 traditional and dynamic financial indicators were considered as independent variables. Estimation using Random Effects allowed the model to simultaneously capture cross-firm heterogeneity, the temporal evolution of insolvency risk, and unobserved individual effects, while preserving the full sample. The model exhibited strong predictive performance, achieving an AUC of 0.915 and an accuracy of 90.2%. The most relevant variables explaining the probability of insolvency were equity minus paid-in capital over total assets, cash and cash equivalents over current and long-term receivables, current assets over current liabilities, working capital requirement, working capital over net revenue, bank loans and financing over total assets, treasury balance over net revenue, and trade payables over total assets. These findings reinforce that effective management of liquidity, capital structure, and working capital is critical to preventing financial distress. The study is limited to the Brazilian context and to the use of a single statistical method. Future research may extend the analysis to firms from other countries or the financial sector, incorporate additional variables, including qualitative measures, and compare the results with machine learning approaches.

Keywords: Insolvency risk. Financial indicators. Logistic regression. Panel data.

1 INTRODUÇÃO

A insolvência empresarial pode ser entendida como o não cumprimento de suas obrigações devido a problemas de liquidez (Stupp, 2015). Com isso, faz-se necessário o estudo de previsão de falências, a fim de evitar a ocorrência de tal situação. Segundo Aranha e Gondrige (2024), o tema de previsão de insolvência tem sido bastante discutido, devido a existência de trabalhos que utilizam variáveis preditoras diferentes, sendo estes diversos indicadores econômico-financeiros que podem afetar os modelos de previsão.

De acordo com Shi e Li (2019), passadas mais de cinco décadas desde o estudo seminal de Altman (1968), que introduziu o modelo *Z-Score* a partir da combinação de indicadores econômico-financeiros e da Análise Discriminante para classificar empresas solventes e insolventes, a literatura passou a avançar significativamente na compreensão do risco de falência corporativa. Nesse contexto, a insolvência passou a ser interpretada como um processo antecipação do risco de crédito, associado à incapacidade da empresa em honrar seus compromissos financeiros no curto e no longo prazo. Assim, diferentes técnicas estatísticas e conjuntos de variáveis passaram a ser explorados com o objetivo de aprimorar a capacidade preditiva dos modelos e antecipar situações de inadimplência e falência empresarial.

No contexto brasileiro, o interesse pela temática surgiu, conforme evidenciado por Altman, Baidya e Dias (1979), em um ambiente marcado por instabilidade macroeconômica, elevadas taxas de inflação, restrições de crédito e maior exposição das empresas ao risco de liquidez. Esses fatores reforçaram a necessidade de instrumentos quantitativos capazes de mensurar o risco de crédito corporativo e identificar, de forma antecipada, empresas com maior probabilidade de falência. Nos anos subsequentes, a literatura nacional consolidou esse campo de pesquisa por meio do desenvolvimento e da aplicação de modelos estatísticos e de aprendizado de máquina voltados à previsão da insolvência empresarial, ampliando o debate sobre os determinantes financeiros do risco de crédito e a eficiência dos modelos de classificação.

A regressão logística, nesse contexto, se apresenta como uma ferramenta estatística poderosa para a previsão de eventos binários, como a ocorrência ou não da insolvência empresarial. Conforme evidenciado no estudo de Schio e Sampaio (2022), essa técnica mostrou-se altamente eficaz ao modelar o risco de recuperação judicial em empresas brasileiras, com base em dados contábeis e indicadores econômico-financeiros. Essa evidência reforça a aplicabilidade da regressão logística como metodologia preditiva confiável para fins de

avaliação da saúde financeira das empresas, subsidiando decisões estratégicas por parte de investidores, gestores e agentes reguladores.

Estudo de revisão de técnicas de previsão do risco de insolvência como os de Pereira e Martins (2015), Prado *et al.* (2016), Shi e Li (2019), Billios, Seretidou e Stavropoulos (2024) e Vásquez-Serpa *et al.* (2025) evidenciam uma crescente na temática e uso de diferentes técnicas e variáveis para previsão. Ainda, identifica-se uma ampliação do uso da Regressão Logística como técnica estatística multivariada aplicada à previsão do risco de insolvência, destacando-se sua adequação à modelagem de eventos binários e seu desempenho preditivo consistente reportado na literatura. Apesar da recorrência dessa abordagem nos estudos de revisão, observa-se uma lacuna relevante no contexto brasileiro, dando ênfase a utilização de indivíduos como empresas brasileiras não financeiras de capital aberto, em um período longitudinal mais recente, utilizando indicadores financeiros que abrangem os grupos de liquidez, endividamento e rentabilidade, observando tanto o aspecto econômico quanto a geração de caixa.

Diante do exposto, observa-se a relevância crescente da utilização de modelos estatísticos robustos na identificação antecipada de situações de risco financeiro nas organizações. Neste contexto, o presente artigo tem como **objetivo** avaliar a construção do modelo com a Regressão Logística na previsão do risco de insolvência empresarial, bem como identificar quais foram as variáveis independentes que contribuíram na classificação do risco, a partir da aplicação dessa técnica em empresas brasileiras listadas na B3. Busca-se, com isso, analisar o desempenho preditivo do modelo e identificar as principais variáveis que influenciam a probabilidade de uma empresa entrar em estado de insolvência, contribuindo para a literatura nacional e para a prática da gestão financeira.

2 REFERENCIAL TEÓRICO

O primeiro estudo de previsão do risco de insolvência realizado utilizando a técnica estatística Regressão Logística foi de Ohlson (1980). A amostra utilizada pelo autor compreende em 105 empresas não financeiras que entraram com pedido legal de falência entre 1970 e 1976, e 2058 empresas não financeiras, ambas de capital aberto, onde as variáveis utilizadas foram 9 indicadores financeiros. Com isso, o modelo ajustado com Ohlson obteve uma acurácia de 96,12%, superando na época as expectativas com relação a análise discriminante utilizada por outros autores anteriores a este estudo. Os indicadores financeiros mais relevantes no estudo foram o capital de terceiros sobre o ativo total, retorno sobre o ativo (ROA) e variação do lucro líquido, indicando que endividamento e rentabilidade do negócio impactam diretamente o risco de insolvência empresarial.

Após a divulgação do estudo de Ohlson (1980), diversos outros estudos foram realizados utilizando a técnica regressão logística com amostras e variáveis diversas a fim de se verificar a acurácia do modelo sob diferentes perspectivas de empresas globais e outros indicadores financeiros que pudessem vir a ser relevantes para classificação e previsão de insolvência de empresas.

Observa-se no mundo trabalhos como o Brédart (2014) que conseguiu no seu modelo uma acurácia de 83,82%, onde as amostras coletadas foram na Amex, Nasdaq e na NYSE no período de 2000 a 2012. A amostra compreendeu em 435 empresas classificadas como insolventes, mediante empresas que entraram com pedido de falências conforme o Código de Falências dos Estados Unidos. Além disso, o autor equiparou cada indivíduo da amostra acrescentando mais 435 empresas consideradas saudáveis, totalizando 870 empresas na amostra. A combinação levou em consideração o setor e o tamanho da empresa medido pelo total do ativo constante no Balanço Patrimonial da empresa. Já a variável dependente foi dicotômica, assumindo 1 para insolvente e 0 para solvente. As variáveis dependentes foram indicadores financeiros utilizados no seminal de Ohlson (1980) e outros autores posteriores a este trabalho que utilizaram a regressão logística como técnica para previsão de insolvência empresarial. As variáveis independentes significantes no estudo foram o a lucratividade mensurada pelo retorno sobre o ativo (ROA), liquidez corrente e patrimônio líquido sobre o ativo total, concluindo que a rentabilidade das companhias, a liquidez e a lucratividade somada com investimentos de capital próprio contribuem para as empresas honrarem seus compromissos financeiros.

O trabalho de Fawzi, Kamaluddin e Sanusi (2015) atingiu uma acurácia de 82,1% ao utilizar indicadores financeiros relacionados ao fluxo de caixa das empresas. A amostra compreendeu 53 empresas classificadas como insolventes e outras 53 empresas caracterizadas como solventes levando em consideração o setor, o tamanho da empresa medido pelo ativo total e o ano fiscal mais próximo. O número total da amostra correspondeu, portanto, a 106 empresas. As variáveis significantes na construção do modelo logit para previsão do risco de crédito foram o fluxo de caixa operacional e despesas de juros sobre as despesas de juros, fluxo de caixa operacional sobre o capital de terceiros e o fluxo de caixa operacional sobre a dívida de longo prazo, evidenciando que a capacidade de conversão do resultado operacional em caixa com relação ao endividamento empresarial está diretamente ligada ao risco de falência das companhias.

Ao analisar a previsão de insolvência de empresas brasileiras no período de 2004 a 2013, Stüpp (2015) utilizou 29 indicadores financeiros e 44 empresas, sendo 22 classificadas como insolventes e outras 22 classificadas como solventes, fazendo a combinação da amostra entre solvente e insolvente considerando o setor e tamanho da empresa avaliando o tamanho do ativo. Como resultado, a regressão logística obteve uma acurácia de 76,2%, indicando boa capacidade de previsão. As variáveis significantes no estudo foram o exigível sobre o ativo, retorno sobre o patrimônio líquido, liquidez corrente, EBIT sobre a dívida líquida, grau de imobilização, grau de endividamento, ciclo financeiro, liquidez geral e giro do ativo.

No estudo de Abdullah *et al.* (2019), obteve-se uma acurácia máxima de 90% levando em consideração empresas pequenas e médias no qual foram testadas 1 ano antes da dificuldade financeira, sendo testadas nesse período 26 empresas insolventes e 26 empresas solventes, totalizando 52 indivíduos no total, utilizando indicadores financeiros nas variáveis independentes e a variável dicotômica 1 para insolvente e 0 para solvente na variável dependente. As variáveis independentes que apresentaram significância no modelo desenvolvido foram o capital de terceiros sobre o ativo total, passivo circulante sobre o ativo total, liquidez corrente, vendas sobre o ativo total, EBIT sobre o ativo total e lucro líquido sobre o capital social, evidenciando que endividamento, dívidas de curto prazo sobre a dívida total, a liquidez de curto prazo, a receita operacional e a lucratividade são fatores determinantes para a análise de insolvência.

Olaoye, Ogbemor e Okusami (2020) examinaram o efeito do risco financeiro sobre desempenho bancário em um painel de bancos comerciais nigerianos. A pesquisa avaliou variáveis como diversificação de risco, monitoramento de crédito, compliance e métricas tradicionais de desempenho ROE (Retorno sobre o Patrimônio Líquido) e ROA (Retorno sobre

o Ativo). Para decidir entre efeitos fixos e aleatórios, os autores aplicaram o teste de Hausman, encontrando que o modelo de efeitos aleatórios era adequado para o Modelo I, enquanto o Modelo II deveria ser estimado com efeitos fixos. A escolha metodológica evidencia que, mesmo quando há indícios de correlação entre efeitos individuais e regressoras em alguns modelos, há contextos financeiros em que efeitos aleatórios são preferíveis por manterem a eficiência e aproveitarem a variação entre unidades. Os resultados do estudo mostraram que os indicadores financeiros retorno sobre o ativo (ROA) e retorno sobre o patrimônio líquido (ROE) possuem forte capacidade explicativa dos modelos utilizados.

O trabalho de Pinto (2021) selecionou a amostra a partir das empresas do segmento atacado não financeiras registradas na Brasil, Bolsa e Balcão (B3), resultando em 473 empresas no período de 2011 a 2019, sendo um total de 8.766 observações colhidas trimestralmente, sendo, portanto, a amostra compreendida como dados em painel. Como variáveis independentes, foram testados 27 indicadores econômicos e financeiros. No modelo de regressão logística, o estudo obteve acurácia de 96,55% e curva ROC de 94% nos dados de teste, indicando um ótimo poder preditivo. As variáveis significantes no trabalho foram a composição do endividamento, lucros retidos sobre o ativo total, liquidez seca, estoque sobre o ativo total e a necessidade de capital de giro.

O estudo de Santos (2021) concentrou-se na previsão de insolvência de empresas brasileiras de capital aberto. A amostra foi coletada no tipo dados em painel, totalizando 5.610 observações sem dados faltantes, referentes ao período de 2010 a 2020. As variáveis independentes englobaram indicadores financeiros e macroeconômicas. Quando empregada a regressão logística, o modelo alcançou acurácia de 92,51%, demonstrando boa capacidade de classificação.

Lima, Paulino e Fávero (2022) avaliaram o impacto de métricas ESG na previsão de insolvência e de pedidos de recuperação judicial de empresas brasileiras, empregando a regressão logística e a análise discriminante com dados de 2010 a 2020. A amostra foi composta por empresas que recorreram à recuperação judicial e aquelas classificadas como insolventes, estruturando-se uma variável dependente *dummy* e utilizando como independentes indicadores financeiros tradicionais e métricas ambientais, sociais e de governança (ESG). No modelo de regressão logística alcançou-se acurácia de 93,8% na previsão de insolvência. As variáveis significantes no estudo foram a participação do capital de terceiros, passivo financeiro sobre o ativo total, composição do endividamento no curto e longo prazo, ativo imobilizado sobre o patrimônio líquido, receita líquida sobre o ativo total e retorno sobre o ativo.

Na sequência, o trabalho de Schio e Sampaio (2022) investigaram o fenômeno da recuperação judicial no Brasil entre 2013 e 2018. A amostra contemplou 142 instituições financeiras, de capital aberto ou fechado, públicas e privadas, acompanhadas mensalmente ao longo do período. As variáveis utilizadas foram indicadores de liquidez, rentabilidade, endividamento e rotatividade extraídos dos relatórios contábeis. Por meio da técnica multivariada Regressão Logística, tratado como variável binária (1 para pedido de recuperação, 0 caso contrário), obteve-se acurácia de 94% na previsão com um ano de antecedência, indicando ótima previsão do modelo. As variáveis financeiras determinantes na construção do modelo logit para previsão da insolvência empresarial foram a liquidez corrente, ROA, grau de endividamento e rotação do ativo.

O estudo de Silva e Pessanha (2022) analisou a previsão de insolvência de distribuidoras de energia elétrica brasileiras entre 2006 e 2019 utilizando dados em painel. A amostra compreendeu empresas não financeiras reguladas pelo setor elétrico, reunindo informações anuais ao longo de mais de uma década. Os autores empregaram indicadores econômico-financeiros amplamente utilizados na literatura, como liquidez corrente, endividamento, margem operacional e rentabilidade, para compor o conjunto de variáveis explicativas. O modelo logístico misto estimado apresentou bom desempenho preditivo, destacando-se o papel do endividamento e de lucros antes dos impostos como principais determinantes do risco de insolvência. As variáveis que compuseram o modelo logit para dados em painel foram 14 indicadores financeiros relacionados ao capital próprio, liquidez, endividamento e rentabilidade.

Um ponto metodológico relevante desse estudo refere-se à escolha dos autores pelo modelo de efeitos aleatórios. Como a variável de insolvência apresentava pouca variação temporal dentro de parte das distribuidoras, a estimação por efeitos fixos implicaria a exclusão automática de unidades sem mudança de status, reduzindo significativamente o tamanho útil da amostra. Para evitar perda informacional e viés de seleção, Silva e Pessanha (2022) optaram pelo modelo de efeitos aleatórios, que preserva todas as observações do painel e captura a heterogeneidade entre empresas, mantendo a consistência e a estabilidade preditiva do modelo logístico. Esse argumento é especialmente relevante no contexto do presente estudo, no qual grande parte das empresas também não apresenta variação temporal recorrente na situação de insolvência, reforçando a adequação da abordagem por efeitos aleatórios para a previsão do risco corporativo.

Com relação ao estudo de Xuezhou *et al.* (2022), investigou-se a relação entre crescimento empresarial e risco de insolvência em um painel composto por 284 empresas não

financeiras listadas no Paquistão entre 2013 e 2017. As variáveis independentes incluíram medidas de crescimento interno e externo, além da maturidade do endividamento, potenciais colaterais fixos e indicadores de controle. O risco de insolvência foi utilizado como variável dependente. Os autores aplicaram modelos com PCSE (*Panel-Corrected Standard Errors*) e FGLS (*Feasible Generalized Least Squares*), ambos métodos que assumem variação entre empresas, aproximando-se conceitualmente do modelo de efeitos aleatórios, uma vez que não eliminam a heterogeneidade individual como ocorre no modelo de efeitos fixos. O estudo demonstrou que a maturidade da dívida exerce papel mediador significativo entre crescimento e risco de insolvência, enquanto os colaterais fixos atuam como moderadores.

Por fim, Jesus e Besarria (2023) concentrou-se na previsão de risco de insolvência bancária, combinando indicadores contábeis extraídas das Demonstrações Financeiras de instituições financeiras listados na B3. A amostra incluiu 12 instituições públicas e privadas, totalizando 396 observações trimestrais entre o primeiro trimestre de 2012 e o primeiro trimestre de 2021. No modelo logístico, com variável dependente dicotômica (1 para alto risco, 0 para baixo risco), obteve-se acurácia de 85%. As variáveis independentes relacionadas aos indicadores financeiros que foram significantes no modelo desenvolvido foram o retorno sobre o ativo e patrimônio líquido sobre o ativo total.

Apesar dos avanços significativos na aplicação da Regressão Logística para previsão de insolvência empresarial, algumas limitações ainda são observadas nos estudos analisados. Uma das principais limitações refere-se ao número relativamente restrito de estudos que utilizam dados em painel. O uso de dados em painel permite considerar a variação temporal e individual das empresas, o que pode aumentar a robustez e a acurácia dos modelos preditivos. Além disso, períodos curtos e poucos setores podem enviesar os modelos, uma vez que mesmos setores ou períodos curtos causar correlação entre as observações.

Outra limitação importante é a abrangência temporal relativamente curta de muitos estudos, que analisam janelas históricas restritas, limitando a capacidade de capturar mudanças estruturais ou ciclos econômicos distintos. A presente pesquisa busca superar essas limitações ao aplicar a Regressão Logística com uma base de dados composta por Demonstrações Financeiras de empresas brasileiras no período de 2010 a 2024. Ao utilizar uma amostra nacional ampla e atualizada, bem como um conjunto expandido de variáveis explicativas, este estudo pretende oferecer um modelo mais robusto e alinhado à realidade das empresas brasileiras, contribuindo tanto para o desenvolvimento teórico quanto para práticas gerenciais e decisões estratégicas mais eficazes no contexto corporativo e regulatório nacional.

Por fim, observa-se que muitos estudos empregam amostras reduzidas, como nos trabalhos de Fawzi, Kamaluddin e Sanusi (2015) e Abdullah *et al.* (2019), o que pode limitar a generalização dos resultados. Além disso, grande parte dessas pesquisas ainda se apoia em conjuntos de indicadores relativamente padronizados, herdados de trabalhos seminais como o de Ohlson (1980).

Uma oportunidade de avanço está na inclusão de variáveis derivadas do modelo dinâmico de Fleuriet, que analisa o ciclo financeiro e o saldo de disponibilidade das empresas. Esses indicadores permitem capturar aspectos estruturais e dinâmicos da saúde financeira, enriquecendo as técnicas de previsão e tornando-as mais sensíveis às múltiplas dimensões do risco empresarial.

Diante desse cenário, o presente estudo apresenta uma abordagem mais ampla e atualizada, que combina métricas tradicionais e indicadores dinâmicos, aplicados a um conjunto de dados mais robusto e representativo. Ao fazer isso, busca-se não apenas contribuir para o avanço da literatura sobre previsão de insolvência no contexto brasileiro, mas também oferecer subsídios práticos para gestores, investidores e formuladores de políticas, fortalecendo a tomada de decisão no âmbito corporativo e regulatório.

3 METODOLOGIA

O método estatístico multivariado aplicado neste estudo foi o Modelo Longitudinal Logístico para Dados em Painel.

O *locus* da pesquisa é composto pela base de dados formada a partir das Demonstrações Financeiras apresentadas e divulgadas por empresas brasileiras não financeiras de capital aberto, com ações negociadas na Brasil, Bolsa e Balcão (B3), obtidas por meio da plataforma Economatica, abrangendo o período de 2010 a 2024.

Para a geração dos testes e resultados obtidos, foi utilizado neste estudo a linguagem de programação *R*, versão 4.4.2, de 31 de outubro de 2024, e o código para geração do modelo foi produzido na plataforma *R Studio*, versão 2025.09.01, de 01 de setembro de 2025.

3.1 Definição da Amostra

Para a definição da variável dependente, a classificação das empresas foi realizada por observação anual, considerando a situação econômico-financeira de cada empresa em cada período analisado. Assim, uma empresa foi classificada como insolvente em determinado ano quando apresentou patrimônio líquido negativo, pedido de recuperação judicial ou pedido de recuperação extrajudicial naquele período específico. Nessas situações, a variável *dummy* assumiu o valor 1, enquanto nos demais anos em que tais condições não foram observadas, a empresa foi classificada como solvente, recebendo valor 0.

Os dados utilizados possuem estrutura de dados em painel, combinando as dimensões transversal (empresas) e temporal (anos), o que permite que uma mesma empresa seja classificada como solvente em determinados períodos e insolvente em outros, de acordo com a evolução de sua condição financeira ao longo do tempo. A amostra final é composta por 5.247 observações anuais, referentes a 501 empresas, das quais 796 observações foram classificadas como insolventes e 4.451 como solventes. Esta forma de coleta de dados está em conformidade com os trabalhos de Pinto (2021), Santos (2021) e Jesus e Besarria (2023), que utilizaram a técnica regressão logística com dados em painel para previsão do modelo.

3.2 Definição dos Indicadores

Para o propósito desta investigação, foram empregadas as variáveis delineadas por Prado *et al.* (2018) e Costa (2023), categorizando indicadores financeiros derivados de modelos

frequentemente utilizados, denominados indicadores tradicionais, juntamente com o modelo dinâmico de Fleuriet, culminando em um total abrangente de 35 variáveis independentes.

A Tabela 17 apresenta as variáveis que compuseram o Modelo Longitudinal Logístico para Dados em Painel. Essas variáveis resultam da etapa de pré-processamento descrita na metodologia, que incluiu a winsorização a 5%, a exclusão de indicadores com correlação superior a 0,8 e VIF maior que 10, bem como a padronização via Z-score das variáveis independentes, garantindo comparabilidade entre diferentes escalas financeiras.

Tabela 17 – Variáveis Utilizadas na Construção do Modelo (continua).

Nome	Indicador Financeiro
X2	$\frac{\text{Capital de Giro}}{\text{Receita Líquida}}$
X4	$\frac{\text{Necessidade de Capital de Giro}}{\text{Receita Líquida}}$
X6	$\frac{\text{Saldo de Tesouraria}}{\text{Receita Líquida}}$
X7	$\frac{\text{Passivo Financeiro}}{\text{Ativo Circulante}}$
X9	$\frac{\text{Saldo de Tesouraria}}{\text{Necessidade de Capital de Giro}}$
X11	$\text{Necessidade de Capital de Giro}$
X12	Capital de Giro
X14	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$
X15	$\frac{\text{Receita Líquida Operacional}}{\text{Ativo Total}}$
X16	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Ativo Total}}$
X19	$\frac{\text{Disponibilidade}}{\text{Ativo Circulante}}$ + $\text{Realizável a Longo Prazo}$
X20	$\frac{\text{Estoque}}{\text{Ativo Total}}$
X21	$\frac{\text{Fornecedores}}{\text{Ativo Total}}$
X27	$\frac{\text{Financiamentos e Empréstimos Bancários}}{\text{Ativo Circulante}}$
X30	$\frac{\text{Patrimônio Líquido}}{\text{Capital Social}}$
X33	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Impostos sobre Despesas Financeiras}}$

Tabela 17 – Variáveis Utilizadas na Construção do Modelo (conclusão).

Nome	Indicador Financeiro
X34	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Dívida Líquida}}$

Fonte: Elaborado pelo autor (2025).

Podemos observar que, de 35 variáveis independentes, 17 foram mantidas para análise no modelo construído. Essa filtragem possibilita maior robustez na estimação do modelo e reduz o risco de distorções associadas à multicolinearidade entre os indicadores financeiros.

3.3 Modelo Longitudinal Logístico para Dados em Painel

De acordo com Corrar *et al.* (2007), a Regressão Logística (RL) foi desenvolvida na década de 1960 com a finalidade de desenvolver um modelo preditivo ou explicativo quando a variável dependente fosse de natureza dicotômica, como é o caso de problemas de classificação de risco de insolvência de empresas. Além disso, esta técnica tem como objetivo estimar a probabilidade de ocorrência de determinado evento ou prever o enquadramento em uma categoria. Sendo assim, segundo os autores, os resultados obtidos da variável por meio da RL permite não apenas a classificação das categorias, como também da probabilidade de o evento ocorrer.

A Equação 8 apresenta a expressão geral da RL de natureza binária. Segundo Fávero e Belfiore (2017), o modelo estima a probabilidade de ocorrência do evento em estudo para cada observação, e não os valores previstos da variável dependente.

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}} \quad (8)$$

Em que:

P_i representa a probabilidade do evento de interesse ocorrer para a observação i ;

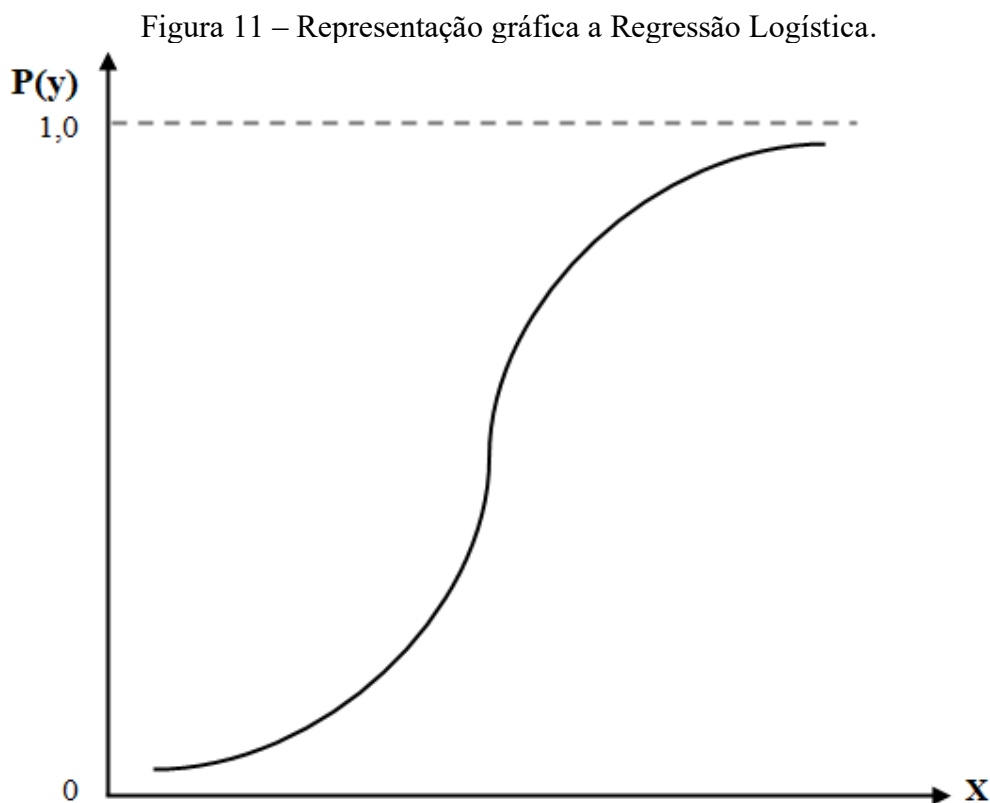
e , é a base do logaritmo natural, servindo para criar a curvatura da função logística, controlando como a probabilidade cresce ou decresce;

β_0 , intercepto ou constante, sendo o valor inicial quando as variáveis independentes X são iguais a zero;

$\beta_1, \beta_2, \dots, \beta_k$, representa os coeficientes da regressão, mostrando o efeito de uma variável independente X_j sobre a *log-odds* (logaritmo da chance de ocorrência) do evento;

$X_{1i}, X_{2i}, \dots, X_{ki}$, indica as variáveis independentes (preditoras), ou seja, são as informações que explicam, ou não, a ocorrência do evento para cada observação i .

Segundo Corrar *et al.* (2007) e Fávero (2017), a fórmula da Regressão Logística forma em uma curva S, também conhecida como curva sigmoide, conforme apresentado na Figura 11. No eixo X estão os valores das variáveis independentes e no eixo Y está o valor da probabilidade P . Quando o valor das variáveis X estão muito baixos, a probabilidade de sucesso do evento ocorrer é próxima de 0. Portanto, na medida em que os valores de X aumentam, a probabilidade de sucesso se aproxima de 1.



Fonte: Corrar *et al.* (2007, p. 290).

Segundo Corrar *et al.* (2007), a curva sigmoide na RL ocorre em razão da necessidade de mudanças mais expressivas na função logística na medida em que se aproxima das extremidades 0 e 1. Com isso, para que, por exemplo, a probabilidade avance de 97% para 98%, o incremento a ser realizado nas variáveis independentes é maior do que obter uma evolução de 48% para 49% (aproximadamente no meio da curva).

Ao utilizar o modelo logit para dados em painel, ou seja, um modelo longitudinal logístico, a chance de ocorrência de um evento se dá a cada instante no tempo, definida pela Equação 8 como a expressão geral (Fávero e Belfiore, 2017).

$$\ln(\text{chance}_{Y_{it}=1}) = a_i + \beta_1 \cdot X_{1it} + \beta_2 \cdot X_{2it} + \dots + \beta_k \cdot X_{kit} \quad (9)$$

Em que:

$\ln = (\text{Chance}_{Y_{it}=1})$, é o logaritmo da razão de chances de o evento 1 ocorrer;

Y_{it} , variável dependente binária, 0 ou 1;

i , indivíduo (empresa);

t , instante no tempo;

a_i , intercepto específico no tempo, podendo ser efeito fixo ou aleatório, que controla as características invariantes no tempo;

X_{kit} , variável explicativa k , medida de uma empresa i no instante t ;

β_k , coeficiente de regressão associado a variável X_{kit} .

Na Equação 10 observa-se a transformação da expressão geral para que a probabilidade de ocorrência fique entre 0 e 1 (Fávero e Belfiore, 2017), conforme demonstrado na Equação 9, agora, com o instante t inserido na equação apresentada.

$$P_{it} = \frac{e^{(a_i + \beta_1 \cdot X_{1it} + \beta_2 \cdot X_{2it} + \dots + \beta_k \cdot X_{kit})}}{1 + e^{(a_i + \beta_1 \cdot X_{1it} + \beta_2 \cdot X_{2it} + \dots + \beta_k \cdot X_{kit})}} \quad (10)$$

Com isso, é possível estimar os parâmetros do logaritmo da função de máxima verossimilhança, estimando o modelo adequação da regressão logística para dados em painel.

3.3.1 Procedimentos e Teste do Modelo Logístico para Dados em Painel

Para garantir a robustez do modelo logístico para dados em painel na previsão do risco de insolvência de empresas, foram realizados diagnósticos de testes e procedimentos executados. Segundo Fávero e Belfiore (2017), para garantir a ausência de multicolinearidade foram retiradas as variáveis que possuem correlação maior que 0,8 e VIF maior que 10. Este estudo segue os procedimentos verificados nos trabalhos de (Abdullah *et al.*, 2019; Fávero; Belfiore, 2017; Jongh *et al.*, 2015; Stüpp, 2015; Brédart, 2014), e alinhado aos princípios de parcimônia e regularização discutidos por autores de modelagem preditiva, a eliminação de redundâncias cumpre papel análogo a penalizações ao controlar a complexidade efetiva do conjunto de regressoras e, assim, o erro fora da amostra (Kuhn; Johnson, 2013; James *et al.*, 2023).

Na sequência, os dados foram winsorizados a 5% a fim de reduzir valores extremos das variáveis independentes. Para a validação preditiva, foi adotada a estratégia de *hold-out* agrupado por empresa, na qual aproximadamente 70% das firmas foram utilizadas para o treinamento e 30% para o teste do modelo. Essa divisão assegura que uma mesma empresa não apareça em ambos os subconjuntos, prevenindo vazamento de informação (*data leakage*) e possibilitando uma avaliação realista do desempenho do modelo em empresas não observadas durante o processo de estimação.

Além disso, foi realizada a padronização das variáveis via *Z-score*, em que a transformação do valor das variáveis faz com que o desvio padrão seja igual a 1 e média 0. O emprego desta técnicas faz-se necessário nos indicadores financeiros do conjunto de dados desta amostra, uma vez que a amplitude de diferentes variáveis pode gerar uma má interpretação pelos modelos com relação a importância das variáveis, viesando os resultados e selecionando equivocadamente variáveis atribuídas como mais ou menos importantes do que outras.

O primeiro teste realizado foi a estimação de um modelo nulo, contendo apenas a constante comparado com um modelo logístico (*Pooled Logit*) que inclui todas as variáveis independentes propostas. Esse procedimento verifica se as variáveis independentes conseguem explicar melhor o modelo em relação ao nulo.

As hipóteses são formuladas da seguinte forma:

H₀: Todos os coeficientes das variáveis independentes são simultaneamente iguais a zero e, portanto, o modelo Pooled Logit não melhora o ajuste em relação ao modelo nulo.

H₁: Pelo menos um coeficiente é diferente de zero e, com isso, o modelo Pooled Logit apresenta ajuste significativamente superior ao modelo nulo.

Segundo Fávero e Belfiore (2017), o modelo *Pooled Logit* funciona como se a base de dados seja uma grande *cross-section*, considerando a existência de erros-padrão robustos com agrupamento por indivíduo, para controle da autocorrelação dos resíduos do indivíduo (neste estudo, empresa) ao longo do tempo.

Para critério de decisão, rejeita-se a hipótese nula se o p-valor da Razão de Verossimilhança (LLR) for menor que o nível de significância de 5%.

O passo seguinte na especificação do modelo de painel foi verificar se a heterogeneidade não observada entre as empresas era estatisticamente relevante. Para isso, foi aplicado o Teste de Breusch-Pagan, que, Segundo Fávero e Belfiore (2017), identifica a presença efeitos individuais específicos.

As hipóteses formuladas foram:

H₀: A variância dos efeitos individuais é igual a zero, não havendo heterogeneidade entre as empresas. O modelo Pooled é o mais adequado.

H₁: A variância dos efeitos individuais é diferente de zero, existindo heterogeneidade entre as empresas, sendo necessário um modelo de painel.

Se o nível de significância for inferior ao adotado de 5%, rejeita-se H_0 , indicando presença de efeitos individuais significativos e necessidade de estimar um modelo de efeitos fixos ou aleatórios.

Uma vez confirmada a presença de efeitos individuais, o teste de Hausman é empregado para identificar, segundo Fávero e Belfiore (2017), se o modelo de Efeitos Fixos ou de Efeitos Aleatórios é mais robusto para o desenvolvimento do modelo, avaliando se os efeitos individuais não observados estão correlacionados com as variáveis explicativas.

As hipóteses são:

H₀: Não há correlação entre os efeitos individuais e as variáveis explicativas, sendo o modelo de Efeitos Aleatórios é consistente e eficiente.

H₁: Existe correlação entre os efeitos individuais e as variáveis explicativas e, portanto, o modelo de Efeitos Fixos é consistente e o modelo de Efeitos Aleatórios é inconsistente.

A decisão baseia-se no p-valor do teste, em que se rejeita H_0 quando a significância for menor que 0,05, optando-se pelo modelo de Efeitos Fixos.

Outro teste realizado é o da Razão de Verossimilhança, que avalia a significância conjunta das variáveis explicativas se o modelo final estimado por meio do *Log-Likelihood*.

As hipóteses do teste são:

H₀: Todos os coeficientes $\beta_{1t} = \beta_{2t} = \dots = \beta_{kt} = 0$, portanto as variáveis independentes não explicam a probabilidade de insolvência.

H₁: Pelo menos um coeficiente $\beta_{it} \neq 0$, e as variáveis independentes explicam significativamente a probabilidade de insolvência.

Rejeita-se H_0 se o p-valor do teste for menor que 0,05, indicando que o modelo é estatisticamente significativo.

Na sequência foi estimada a medida *McFadden's R²*, também conhecida como Pseudo R². Essa métrica avalia a qualidade de ajuste do modelo, indicando o quanto a inclusão das variáveis independentes melhora a capacidade preditiva em comparação a um modelo nulo (sem variáveis explicativas). A Equação 11 apresenta a expressão do teste a ser realizado.

$$pseudo R^2 = \frac{-2LL_0 - (-2LL_{m\acute{a}x})}{-2LL_0} \quad (11)$$

Em que:

-2LL₀, representando o valor Log-Likelihood do Modelo Nulo, pressupondo que nenhuma variável explicativa influencia na variável dependente, servindo como um modelo de comparação, onde este não tem nenhuma informação sobre as variáveis explicativas; e

-2LL_{máx}, em que indica o quão bem o modelo explica os dados, onde quanto menor esse valor, melhor é o ajuste do modelo.

Por fim, o teste Wald que verifica a significância estatística de cada coeficiente do modelo (Fávero e Belfiore, 2017).

As hipóteses são:

H₀: O coeficiente da variável $\beta_i = 0$, sendo a variável independente não contribui para o modelo).

H₁: O coeficiente da variável $\beta_i \neq 0$ e, portanto, a variável independente é estatisticamente significativa no modelo.

A rejeição da hipótese nula indica que o coeficiente é estatisticamente significativo a 5% de probabilidade.

De acordo com Corrar *et al.* (2007), diferentemente dos outros testes que buscaram avaliar o modelo logístico com um todo, este tem a finalidade de aferir o grau de significância de cada coeficiente da equação logística, inclusive a constante. A Equação 12 apresenta a expressão do teste Wald.

$$Wald = \left(\frac{b}{S.E} \right)^2 \quad (12)$$

Em que:

b = coeficiente de uma variável independente incluída no modelo; e

S. E. = erro-padrão.

Por fim, outro fator a ser observado e inserido na RL é o *cut-off*, que, segundo Fávero (2017), refere-se ao ponto de corte que o pesquisador escolhe para a classificação dos indivíduos em um grupo ou em outro em decorrência das probabilidades de ocorrência ou não do evento, neste estudo, a insolvência e a solvência. Para este estudo, o valor atribuído ao *cut-off* é de 0,0039.

Cabe destacar que, em modelos logísticos para dados em painel, a análise de resíduos clássica, como normalidade, homocedasticidade ou independência serial, não é aplicável, uma vez que a variável dependente é dicotômica e, portanto, os resíduos não seguem uma distribuição normal.

Em substituição, a avaliação do modelo se dá por meio de testes de especificação global, como a Razão de Verossimilhança e da medida Pseudo R², da significância individual dos coeficientes, aferida pelo teste Wald, e da escolha entre efeitos fixos ou aleatórios por meio dos testes de Breusch-Pagan e Hausman.

Além disso, a robustez do modelo foi examinada por métricas preditivas usuais em modelos logísticos, como a Curva ROC, a AUC e a matriz de confusão. Dessa forma, a avaliação do desempenho e da adequação do modelo priorizou os critérios consagrados na literatura de regressão logística para dados em painel, em conformidade com Fávero e Belfiore (2017).

4 RESULTADOS E DISCUSSÃO

4.1 Verificação de colinearidade entre as variáveis independentes e estatísticas descritivas

A Tabela 18 apresenta os valores VIF das variáveis que compuseram o Modelo Longitudinal Logístico para Dados em Painel.

Tabela 18 – VIF das variáveis a Regressão Logística para Dados em Painel.

Nome	VIF
X2	3,28
X4	2,24
X6	2,38
X7	3,16
X9	1,11
X11	1,52
X12	1,45
X14	2,55
X15	4,18
X16	2,50
X19	2,09
X20	2,41
X21	2,85
X27	4,59
X30	2,08
X33	1,83
X34	1,12

Fonte: Elaborado pelo autor (2025).

Observa-se que todas as variáveis apresentaram valores de VIF inferiores ao limiar crítico adotado, indicando ausência de colinearidade severa entre os indicadores financeiros. Esses resultados confirmam a adequação do conjunto de variáveis selecionado para a estimação do modelo longitudinal logístico, assegurando maior estabilidade das estimativas e confiabilidade na interpretação dos efeitos associados a cada indicador.

A Tabela 19 apresenta as estatísticas descritivas das variáveis independentes antes da transformação dos dados via winsorização e padronização, indicando os valores de média, mínimo, máximo, desvio padrão, variância, assimetria e curtose.

Tabela 19 – Estatísticas descritivas as variáveis selecionadas.

Variável	Média	Mínimo	Máximo	D. Padrão	Variância	Assimetria	Curtose
X2	-0,081	-6,175	2,918	1,297	1,682	-2,569	11,658
X4	0,368	-1,222	4,460	0,916	0,840	2,867	9,809
X6	-3,983E-07	-9,714E-06	8,716E-07	1,686E-06	2,843E-12	-4,810	22,873
X7	0,217	0,000	2,035	0,396	0,157	3,166	10,735
X9	-0,251	-10,852	12,230	3,319	11,013	0,718	6,680
X11	1,766E+06	-5,896E+05	1,940E+07	4,126E+06	1,702E+13	3,063	8,995
X12	-9,244E+04	-1,078E+07	7,619E+06	2,827E+06	7,995E+12	-1,224	5,944
X14	1,340	1,00e-06	9,427	1,829	3,346	2,696	8,621
X15	0,411	1,00e-06	1,727	0,463	0,214	1,073	0,429
X16	0,018	-0,644	0,253	0,144	0,021	-2,750	10,536
X19	0,118	1,00e-06	0,771	0,178	0,032	2,039	3,996
X20	0,054	1,00e-06	0,310	0,086	0,007	1,537	1,227
X21	0,051	1,00e-06	0,326	0,074	0,005	2,093	4,358
X27	0,836	1,00e-06	7,265	1,432	2,050	2,982	9,680
X30	-0,539	-14,183	0,467	2,515	6,326	-4,709	21,644
X33	1,335	-7,111	18,102	3,897	15,188	2,334	8,367
X34	0,186	-5,913	7,447	1,808	3,271	0,823	8,516

Fonte: Elaborado pelo autor (2025).

As estatísticas descritivas apresentadas na Tabela 19 evidenciam a heterogeneidade das variáveis explicativas utilizadas no modelo de regressão logística para dados em painel. Observa-se que algumas variáveis possuem médias próximas de zero e baixa dispersão, como X6 (Saldo de Tesouraria sobre Receita Líquida) e X20 (Estoque sobre Ativo Total), enquanto outras apresentam valores extremos e alta variabilidade, como X11 (Necessidade de Capital de Giro), X12 (Capital de Giro) e X33 (Lucros Antes dos Juros e Impostos sobre Impostos sobre Operações Financeiras).

Esses resultados refletem a diversidade de escalas e magnitudes financeiras típicas em estudos de insolvência. Além disso, os coeficientes de assimetria e curtose indicam distribuições não normais para várias variáveis, com presença de caudas longas e *outliers*, sendo um comportamento comum em séries econômico-financeiras de empresas de diferentes setores e tamanhos.

Dessa forma, reforça-se a importância do uso de modelos logísticos longitudinais robustos, capazes de lidar com heterogeneidade setorial, variabilidade temporal e não normalidade dos dados, proporcionando maior estabilidade estatística e confiabilidade preditiva.

4.2 Desenvolvimento da Regressão Logística para Dados em Painel

Nesta seção, são apresentados os resultados do modelo preditivo de insolvência empresarial. Realizou-se os testes de especificação e robustez que levaram ao modelo final, analisando os coeficientes estimados, as variáveis que mais influenciaram no modelo e a avaliação da performance em um conjunto de dados de teste, não observado anteriormente pelo modelo.

4.2.1 Testes de especificação e seleção do modelo

A primeira etapa realizada foi o Teste de Razão de Verossimilhança. O teste constitui na comparação do modelo *Pooled Logit* com todas as variáveis independentes com o modelo nulo, composto apenas pela constante. A Tabela 20 apresenta a comparação dos modelos, onde o objetivo foi verificar se a inclusão das variáveis explicativas melhora significativamente o ajuste do modelo.

Tabela 20 – Comparação dos modelos *Pooled Logit* e Nulo.

Modelo	Log-Likelihood	AIC	Pseudo R ² (McFadden)	Estatística	p-valor
Nulo	-1549,11	3100,22	-	-	-
<i>Pooled Logit</i>	-706,52	1449,05	0,5439	1685,2	0,001

Fonte: Elaborado pelo autor (2025).

Os resultados demonstram que o modelo melhora ao incluir as variáveis, passando de um Log Likelihood do modelo Nulo de -1549,11 para um Log-Likelihood de -706,52, onde mais próximo de zero, melhor é o modelo. Além disso, o Teste de Razão de Verossimilhança (LLR) confirma que essa melhora é estatisticamente significativa, evidenciando um p-valor menor que 0,001, rejeitando a hipótese nula, confirmando que o conjunto de variáveis explicativas contribui significativamente para explicar a probabilidade de insolvência.

Além disso, o Pseudo R² de 0,5439 indica que o modelo com as variáveis consegue explicar aproximadamente 54,39% da variação da probabilidade de insolvência em comparação com o modelo nulo. Portanto, os resultados validam a robustez do modelo conforme seu aumento significativo do poder explicativo por meio das variáveis e justificam o prosseguimento da análise de modelos em painel mais robustos.

Em seguida, foi conduzido o Teste de Breusch–Pagan, cujo objetivo é avaliar a presença de efeitos individuais específicos das empresas (heterogeneidade não observada), a fim de determinar se um modelo de painel é mais adequado que o modelo *Pooled*. A Tabela 21 apresenta o resultado do teste.

Tabela 21 – Resultado do teste Breusch-Pagan.

Estatística	p-valor
4450,1	0,001

Fonte: Elaborado pelo autor (2025).

O resultado do teste de Breusch-Pagan é altamente significativo, levando à rejeição da hipótese nula, evidenciando a presença de efeitos individuais específicos das empresas, reforçando a necessidade da construção de um modelo de painel para capturar a variabilidade específica de cada firma ao longo do tempo.

Após confirmar a necessidade de um modelo de painel, aplicou-se o Teste de Hausman, utilizado para decidir entre o modelo de Efeitos Fixos e o de Efeitos Aleatórios, com base na correlação entre os efeitos individuais e as variáveis explicativas. A Tabela 22 apresenta o resultado do teste de Hausman.

Tabela 22 – Resultado do teste de Hausman.

Estatística	p-valor
4442,0	0,001

Fonte: Elaborado pelo autor (2025).

O resultado apresentado na Tabela 22 demonstra que o Teste de Hausman foi altamente significativo ($p\text{-valor} < 0,001$), levando à rejeição de H_0 de ausência de correlação entre os efeitos individuais e as variáveis explicativas. Esse resultado indica que, teoricamente, o modelo de Efeitos Fixos seria o mais consistente, uma vez que controla a possível endogeneidade proveniente de fatores não observados correlacionados com as variáveis independentes.

Ainda que o resultado do Teste de Hausman aponte, em termos teóricos, para a preferência pelo modelo de Efeitos Fixos, é importante reconhecer suas limitações práticas e conceituais no contexto desta pesquisa. De acordo com Hill *et al.* (2020), o teste baseia-se na comparação entre estimadores de efeitos fixos e aleatórios sob a suposição de que ambos são consistentes, o que nem sempre se verifica em modelos logísticos para dados em painel,

especialmente quando o evento de interesse é raro e há significativa perda de observações no modelo de efeitos fixos.

Segundo Nourin *et al.* (2025), em amostras desbalanceadas ou com baixa variação temporal, como ocorre na previsão de insolvência, na qual poucas empresas alteram seu status ao longo do período, o teste tende a rejeitar o modelo de efeitos aleatórios mesmo quando este é mais parcimonioso e adequado à finalidade preditiva. Desta forma, o resultado do teste de Hausman deve ser interpretado com cautela e complementado por critérios empíricos de ajuste e estabilidade, uma vez que, em contextos com dados binários, painéis desbalanceados e eventos raros, o teste pode não constituir um critério decisório confiável (King; Zeng, 2001).

Após a aplicação dos testes de especificação e identificação da presença de heterogeneidade entre as empresas, tornou-se necessário comparar empiricamente o desempenho dos diferentes modelos de regressão logística, sendo eles o *Pooled Logit*, Efeitos Fixos e Efeitos Aleatórios, a fim de avaliar qual estrutura melhor representa os dados e fornece estimativas mais consistentes. Essa etapa complementa os testes de Breusch-Pagan e Hausman, permitindo observar como a inclusão ou o tratamento dos efeitos individuais impacta a qualidade do ajuste e a parcimônia do modelo.

A Tabela 23 apresenta o resumo comparativo dos três modelos estimados, considerando as medidas de ajuste Log-Likelihood e AIC, a capacidade explicativa por meio do Pseudo R² de McFadden e o número de observações e empresas incluídas na amostra de treino.

Tabela 23 – Comparativo de coeficientes dos modelos logísticos.

Modelo	Log-Likelihood	AIC	Pseudo R²	Amostra (Treino)	Empresas (Treino)
Pooled Logit	-706,52	1449,05	0,5439	3651	350
Efeitos Aleatórios	-482,11	1002,22	0,3460	3651	350
Efeitos Fixos	-199,74	433,47	0,4498	1331	101

Fonte: Elaborado pelo autor (2025).

A partir dos resultados, observa-se que o modelo *Pooled* apresenta o maior Pseudo R² (0,5439), porém tal métrica não deve ser utilizada como critério de seleção entre modelos de painel, pois ignora completamente a heterogeneidade não observada e pode inflar artificialmente a capacidade explicativa ao não controlar diferenças intrínsecas entre as empresas. Dessa forma, embora informativo, o Pseudo R² não substitui critérios de parcimônia e consistência econométrica como o AIC e a adequação teórica do modelo.

A especificação de Efeitos Fixos apresentou melhor ajuste dentro da amostra, com o menor Log-Likelihood (-199,74), porém sua estimação mostrou-se limitada pela natureza da variável dependente. No modelo logístico, apenas empresas que apresentam variação temporal no status de insolvência (uma mesma companhia que alternam entre 0 e 1 ao longo do período) contribuem para a estimação dos parâmetros. Esse mecanismo de identificação é amplamente conhecido na literatura. Conforme destaca Baltagi (2005), o *logit* de Efeitos Fixos utiliza apenas unidades com mudanças na classificação ao longo do tempo, eliminando automaticamente todos os grupos com resultados invariantes. Tal característica reduz substancialmente a amostra útil, restringe o poder explicativo e gera estimativas instáveis, especialmente quando o evento de interesse é raro ou apresenta baixa variabilidade temporal, o que é típico em estudos de previsão de insolvência. Além disso, o modelo logístico de efeitos fixos está sujeito ao problema dos parâmetros incidentais, que compromete a consistência dos estimadores quando o número de períodos é limitado, como também discutido por Baltagi (2005).

Por outro lado, o modelo de Efeitos Aleatórios demonstrou equilíbrio entre qualidade de ajuste e estabilidade dos coeficientes estimados. Essa especificação permite capturar simultaneamente a variação entre empresas (*between*) e a variação temporal dentro de cada empresa (*within*), sem descartar observações e mantendo a eficiência da estimação. Baltagi (2005) ressalta que modelos de efeitos aleatórios são adequados quando as empresas observadas são tratadas como uma amostra extraída de uma população mais ampla, sendo este cenário perfeitamente compatível com o presente estudo, que analisa empresas brasileiras de capital aberto ao longo de vários anos. Ademais, ao contrário do modelo condicional de Efeitos Fixos, a abordagem de Efeitos Aleatórios preserva os efeitos específicos não observados das empresas, permitindo recuperar probabilidades individuais e tornando o modelo aplicável para fins preditivos.

Do ponto de vista teórico, a escolha pelo modelo de Efeitos Aleatórios é amplamente sustentada pela literatura econométrica. Baltagi (2005) enfatiza que essa especificação é adequada quando as unidades observadas fazem parte de uma população maior, sendo uma condição típica de estudos empresariais com dados em painel, como o presente. O autor também aponta que modelos de efeitos fixos sofrem perda significativa de graus de liberdade e podem apresentar estimadores instáveis quando existe baixa variação temporal na variável dependente, cenário que corresponde exatamente ao encontrado nesta pesquisa, em que apenas uma parcela reduzida das empresas apresentou mudança no status de insolvência. Além disso, quando não há evidências robustas de correlação entre os efeitos inobserváveis e as variáveis explicativas, a especificação de efeitos aleatórios permanece consistente e mais parcimoniosa.

A literatura aplicada reforça esses argumentos ao lidar com situações semelhantes. Silva e Pessanha (2022), ao analisar a previsão de insolvência em distribuidoras de energia elétrica, também constataram baixa variação temporal na variável dependente, o que inviabilizou a estimação por efeitos fixos. Assim, os autores adotaram um modelo logístico com efeitos aleatórios para preservar a amostra completa e obter estimativas estáveis, solução metodológica no qual está completamente alinhada à estratégia deste estudo.

Esse padrão também é observado em outros trabalhos empíricos que lidam com risco financeiro e dados em painel. Olaoye, Ogbemor e Okusami (2020), ao estudar risco e desempenho bancário na Nigéria, e Xuezhou *et al.* (2022), ao investigar o impacto do crescimento das firmas sobre o risco de insolvência, optaram por especificações com efeitos aleatórios para capturar adequadamente a heterogeneidade entre unidades e evitar perda substancial de informações decorrentes da exclusão de firmas sem variação temporal. Assim, a adoção do modelo de Efeitos Aleatórios neste estudo é tecnicamente fundamentada e consistente tanto com a teoria econométrica quanto com as práticas metodológicas observadas na literatura recente de insolvência e risco financeiro.

Diante desses aspectos teóricos e empíricos, a escolha pelo modelo *Logit* de Efeitos Aleatórios mostra-se a mais apropriada para representar o comportamento longitudinal da insolvência empresarial, assegurando consistência econométrica, preservação da amostra completa e capacidade preditiva, elementos essenciais para os objetivos desta pesquisa.

4.2.2 Análise do Modelo Final de Efeitos Aleatórios

Conforme apontaram os testes de especificação apresentados anteriormente, o modelo de Efeitos Aleatórios foi o mais adequado para representar o comportamento longitudinal da insolvência empresarial. Essa abordagem permitiu capturar tanto a variação entre as empresas (*between*) quanto a variação temporal dentro das empresas (*within*), garantindo maior eficiência estatística e estabilidade na estimação dos parâmetros.

A Tabela 24 apresenta os resultados do modelo estimado com base na amostra de treino, abrangendo 3.651 observações de 350 empresas pertencentes a diversos setores econômicos. O modelo foi estimado utilizando padronização prévia das variáveis (*Z-score*), conforme as premissas metodológicas adotadas.

Tabela 24 – Resultado do modelo de Efeitos Aleatórios selecionado.

Variável	Coefficiente	Erro Padrão	Teste de Wald z	p-valor
(Constante)	-7,340	0,611	-12,021	0,001 ***
X30 ((PL-CS)/AT)	-2,793	0,375	-7,446	0,001 ***
X19 (DISP/(AC+RLP))	-0,816	0,205	-3,984	0,001 ***
X14 (AC/PC)	-0,861	0,272	-3,170	0,002 **
X11 (NCG)	-1,679	0,590	-2,845	0,004 **
X2 (CDG/RecL)	-0,835	0,336	-2,485	0,013 *
X27 (DB/AC)	0,501	0,220	2,284	0,022 *
X21 (FORNECEDORES/AT)	0,380	0,177	2,145	0,032 *
X6 (T/RecL)	-0,499	0,230	-2,170	0,030 *
X16 (EBIT/AT)	-0,369	0,203	-1,821	0,069
X33 (EBIT/IDF)	-0,435	0,324	-1,343	0,179
X7 (PF/AC)	0,175	0,161	1,084	0,278
X4 (NCG/RecL)	0,260	0,247	1,052	0,293
X12 (CDG)	0,184	0,230	0,803	0,422
X34 (EBIT/DL)	-0,107	0,138	-0,771	0,441
X9 (T/NCG)	0,0898	0,131	0,683	0,494
X15 (RLO/AT)	0,050	0,228	0,220	0,826
X20 (ESTOQUES/AT)	-0,037	0,264	-0,141	0,888
<i>Variância do efeito de grupo</i>	31,270	-	-	-
<i>Desvio-Padrão do efeito de grupo</i>	5,592	-	-	-
<i>Número de observações</i>	3651	-	-	-
<i>Número de empresas</i>	350	-	-	-
<i>Log-Likelihood</i>	-482,111	-	-	-
<i>AIC</i>	1002,22	-	-	-

Nota: *** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$. A variável *Group Var* representa a variância dos efeitos aleatórios entre os grupos (empresas).

Fonte: Elaborado pelo autor (2025).

A análise dos resultados do modelo Logit com Efeitos Aleatórios revela que oito variáveis apresentaram significância estatística Z no Teste de Wald, indicando contribuição relevante para explicar a probabilidade de insolvência das empresas. As variáveis X30, X19, X14, X11, X2 e X6 tiveram coeficientes negativos, sugerindo que o aumento desses indicadores está associado a uma redução na probabilidade de insolvência, enquanto as variáveis X27 e X21

apresentaram coeficientes positivos. Isso evidencia o papel da otimização da estrutura de capital, da liquidez e da gestão eficiente do capital de giro na sustentabilidade financeira das firmas.

Por meio da estimação do modelo apresentada na Tabela 24, informando os coeficientes e as variáveis significantes, é possível elaborar a função da probabilidade de ocorrência do evento ocorrer, no caso, da solvência ou insolvência, a partir do modelo construído pelo método da Regressão Logística com dados em painel, representada pela Equação 13.

$$P_{it} = \frac{e^{(-7,340-2,793 \cdot \left(\frac{PL-CS}{AT}\right) - 0,816 \cdot \left(\frac{DISP}{(AC+RLP)}\right) - 0,861 \cdot \left(\frac{AC}{PC}\right) - 1,679 \cdot (NCG) - 0,835 \cdot \left(\frac{CDG}{ReCL}\right) + 0,501 \cdot \left(\frac{DB}{AC}\right) + 0,380 \cdot \left(\frac{FORNECEDORES}{AT}\right) - 0,499 \cdot \left(\frac{T}{ReCL}\right))}}{1 + e^{(-7,340-2,793 \cdot \left(\frac{PL-CS}{AT}\right) - 0,816 \cdot \left(\frac{DISP}{(AC+RLP)}\right) - 0,861 \cdot \left(\frac{AC}{PC}\right) - 1,679 \cdot (NCG) - 0,835 \cdot \left(\frac{CDG}{ReCL}\right) + 0,501 \cdot \left(\frac{DB}{AC}\right) + 0,380 \cdot \left(\frac{FORNECEDORES}{AT}\right) - 0,499 \cdot \left(\frac{T}{ReCL}\right)}} \quad (13)$$

Após a construção da função logística do modelo, foi feita a análise individual das variáveis significativas, discutindo os sinais estimados e esperados, a coerência teórica e a convergência com os achados da literatura aplicada à previsão de insolvência.

A variável X30 (Patrimônio Líquido subtraído do Capital Social sobre o Ativo Total) apresentou o maior impacto do modelo, indicando que empresas com maior proporção de lucros retidos e reservas acumuladas sobre o ativo total exibem menor propensão à insolvência. O sinal negativo confirma a expectativa teórica de que a rentabilidade acumulada representa a base mais saudável e recorrente de capitalização, reduzindo a dependência de capital de terceiros. Evidências semelhantes são encontradas nos estudos de Ohlson (1980), Brédart (2014), Stüpp (2015), Lima, Paulino e Fávero (2022), Fawzi, Kamaluddin e Sanusi (2015), Abdullah *et al.* (2019), Prado *et al.* (2019), Pinto (2021), Lima Paulino e Fávero (2022), Schio e Sampaio (2022) e Silva e Pessanha (2022) que identificam a rentabilidade e o fortalecimento patrimonial por meio da lucratividade para financiar os recursos da empresa como um dos pilares de proteção contra o risco de insolvência, sobretudo em ambientes instáveis, uma vez que o fortalecimento do patrimônio líquido reduz a probabilidade de risco de crédito nas empresas. Dessa forma, o resultado reforça que a acumulação de lucros, e não apenas a integralização de capital, constitui importante indicador para solvência da empresa.

Já a variável X19 (Disponibilidade sobre o Ativo Circulante e Realizável a Longo Prazo) apresentou coeficiente negativo, demonstrando que maior disponibilidade financeira em relação ao conjunto do ativo circulante e realizável de longo prazo reduz o risco de insolvência.

Essa interpretação é coerente com Brédart (2014), que destaca os indicadores de liquidez como preditores altamente eficazes da sobrevivência empresarial no longo prazo, sendo determinante na previsão de insolvência, especialmente em empresas com ciclos operacionais mais voláteis. Logo, o resultado evidencia que reservas financeiras são fundamentais para absorver choques, assegurar pagamentos e evitar rupturas no ciclo operacional.

Com coeficiente negativo e alta significância, a variável X14 (Ativo Circulante sobre o Passivo Circulante) confirma o papel da liquidez corrente como amortecedor natural de risco. Empresas cujo ativo circulante cobre confortavelmente o passivo circulante possuem maior capacidade de honrar compromissos imediatos, reduzindo a probabilidade de insolvência. Ohlson (1980), Abdullah *et al.* (2019), Lima, Paulino e Fávero (2022) e Schio e Sampaio (2022) identificaram que a liquidez corrente é um dos indicadores mais consistentes na previsão insolvência organizacional, o que reforça sua importância no contexto brasileiro. Assim, o sinal negativo confirma a expectativa teórica, mostrando que liquidez operacional robusta protege a empresa de pressões súbitas sobre fluxo de caixa.

Agora, com relação a variável X11 (Necessidade de Capital de Giro), apresentou coeficiente negativo, contrariando a expectativa teórica inicial de que NCG elevada aumentaria o risco de insolvência. No entanto, essa inversão encontra justificativa no próprio Modelo Dinâmico de Fleuriet, que estabelece que a NCG não deve ser analisada isoladamente. Uma NCG positiva é típica da maioria das empresas brasileiras e sinaliza apenas que a operação demanda financiamento temporário, o risco real depende da relação entre Necessidade de Capital de Giro, Capital de Giro e Saldo de Tesouraria. Assim, quando o CDG e o T são suficientes para financiar a NCG, esta deixa de representar fragilidade e passa a compor uma estrutura equilibrada. Em outras palavras, uma NCG elevada somente é prejudicial quando não há CDG e T que a sustentem. Os resultados confirmam o achado de Pinto (2021), onde a NCG é relevante na previsão do risco de insolvência no contexto brasileiro. Esse comportamento é compatível com a lógica descrita por Fleuriet e Zeidan (2015) ao informar que a saúde financeira depende da interação dinâmica entre os três indicadores financeiros.

Ao analisar variável X2 (Capital de Giro sobre a Receita Líquida) significativa no modelo, verifica-se que o coeficiente é negativo, revelando que empresas com maior volume de capital de giro próprio em relação às receitas enfrentam menor probabilidade de insolvência. Esse resultado converge com a teoria de Fleuriet, que considera o CDG o principal amortecedor do ciclo operativo. Estudos como Ohlson (1980) reforçam que estruturas de capital baseadas em recursos permanentes proporcionam maior resiliência ao ciclo financeiro e reduzem a pressão

sobre liquidez. Assim, o modelo confirma que o CDG exerce papel estratégico na solvência empresarial, mitigando descasamentos temporais entre entradas e saídas de caixa.

O coeficiente positivo associado à variável X27 (Dívida Bruta composta pelos Empréstimos e Financiamentos de Curto e Longo Prazo sobre o Ativo Circulante) evidencia que empresas com maior dependência de empréstimos bancários enfrentam maior risco de insolvência. Esse achado é amplamente documentado na literatura. Segundo Fawzi, Kamaluddin e Sanusi (2015), que analisaram o fluxo de caixa como premissa para identificar as causas da deterioração do caixa, demonstram que o capital de terceiros associados com as dívidas de curto prazo tem forte poder preditivo de falência e redução na disponibilidade financeira, especialmente pela pressão imediata que exercem sobre o caixa. Portanto, o sinal positivo esperado e a relevância em outros estudos confirmam que a concentração de dívidas financeiras no curto prazo amplifica a exposição ao risco.

A variável X21 (Fornecedores sobre o Ativo Total) apresentou coeficiente positivo, indicando que maior dependência de fornecedores como fonte de financiamento operacional está associada ao aumento da probabilidade de insolvência. Fawzi, Kamaluddin e Sanusi (2015) demonstraram que firmas em deterioração ampliam o uso do crédito comercial como forma de sustentar operações em meio a restrições de caixa. Resultado semelhante é encontrado em Schio e Sampaio (2022), que relatam que elevações no passivo operacional estão entre os principais sinais precoces de insolvência no Brasil. Assim, o sinal positivo reforça a lógica financeira de que a postergação de pagamentos, embora alivie o caixa no curto prazo, aumenta a vulnerabilidade da empresa.

Por fim, o coeficiente negativo associado à variável X6 (Saldo de Tesouraria sobre a Receita Líquida) demonstra que maior saldo de tesouraria proporcional à receita reduz a probabilidade de insolvência. Considerado um dos pilares do Modelo de Fleuriet, o T capta a folga financeira residual após o financiamento das operações. Prado *et al.* (2019) evidenciam que o Saldo de Tesouraria é indicador sensível de deterioração financeira, principalmente quando o CDG e a NCG se desalinham. Estudos como o de Lima, Paulino e Fávero (2022) e Schio e Sampaio (2022) reforçam que métricas de liquidez operacional e flexibilidade financeira possuem elevado poder preditivo. Nesse sentido, o sinal negativo confirma a interpretação de que empresas com maior folga de tesouraria mantêm maior resiliência frente a choques de liquidez.

Os resultados confirmam que a liquidez, a alavancagem e a estrutura de capital são dimensões centrais na explicação do risco de insolvência em empresas brasileiras, reforçando

a importância da eficiência operacional e do controle da estrutura de capital como variáveis determinantes para a solvência empresarial.

De forma complementar, observa-se que a combinação entre indicadores financeiros tradicionais e aqueles derivados do Modelo Dinâmico de Fleuriet potencializa a capacidade explicativa do modelo, ao integrar medidas de liquidez e capital de giro dinâmico com variáveis clássicas de estrutura de capital e desempenho. Essa complementaridade evidencia que a análise conjunta dos dois grupos de indicadores permite capturar tanto o aspecto estático quanto o dinâmico da saúde financeira das empresas, proporcionando uma visão mais abrangente e precisa sobre os determinantes da insolvência.

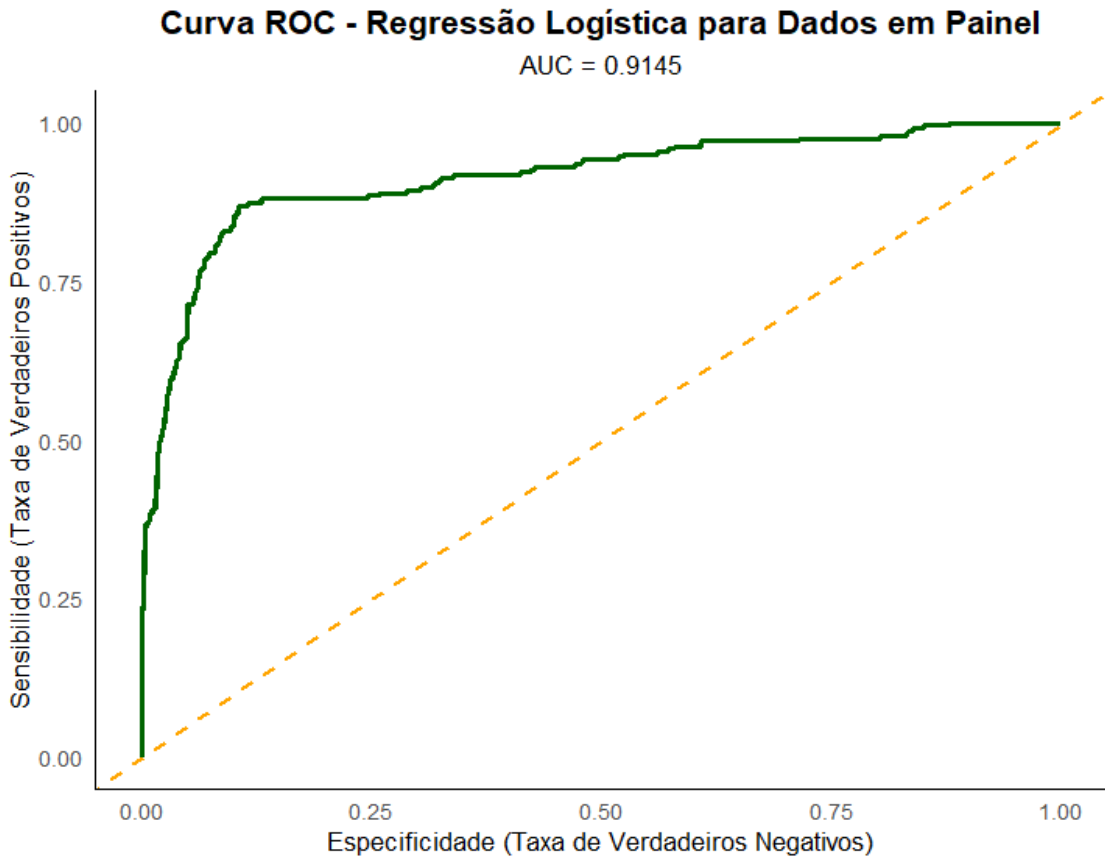
Neste estudo, incorporou-se de forma integrativa os indicadores financeiros tradicionais e dinâmicos. Enquanto os trabalhos citados enfatizaram a importância da rentabilidade, da liquidez corrente e da alavancagem, o presente estudo amplia o escopo da análise ao considerar a interação entre a estrutura financeira estática e a gestão dinâmica dos recursos operacionais, conforme introduzido pelo modelo de Fleuriet. Essa integração permite compreender com maior profundidade como as variações na estrutura de capital e na folga de caixa influenciam a probabilidade de insolvência, oferecendo uma contribuição original à literatura empírica brasileira sobre previsão de risco de crédito.

Dessa forma, o modelo de Regressão Logística de Efeitos Aleatórios consolidou-se como a abordagem mais adequada para representar o comportamento longitudinal da insolvência empresarial, capturando tanto a variação entre empresas quanto a variação temporal dentro de cada firma. Além de apresentar coeficientes estatisticamente significativos e coerentes com a teoria financeira de forma integrativa com indicadores financeiros tradicionais e dinâmicos, o modelo demonstrou elevada capacidade de ajuste e estabilidade nas estimativas.

4.2.3 Avaliação da Performance Preditiva do Modelo

A capacidade preditiva do modelo foi avaliada no conjunto de dados de teste, correspondente a 30% da amostra total. O modelo demonstrou elevado poder discriminante entre empresas solventes e insolventes, alcançando uma Área Sob a Curva ROC (AUC) de 0,9145, valor considerado excelente segundo os critérios de avaliação de modelos binários. A Figura 12 apresenta a Curva ROC da Regressão Logística para Dados em Painel, evidenciando a expressiva separação entre as classes e o desempenho consistente do modelo.

Figura 12 – Curva ROC da Regressão Logística com Efeitos Aleatórios.



Fonte: Elaborado pelo autor (2025).

A Tabela 25, a seguir, apresenta as principais métricas de desempenho obtidas com base no conjunto de teste.

Tabela 25 – Métricas de performance do modelo.

AUC	Acurácia	Precisão	Sensibilidade	Especificidade	F1 Score	VP	FP	VN	FN
0,915	0,902	0,6469	0,800	0,921	0,715	196	107	1244	49

Fonte: Elaborado pelo autor (2025).

Os resultados apresentados evidenciam que o modelo *Logit* com Efeitos Aleatórios atingiu uma AUC de 0,915, indicando um excelente poder discriminante entre empresas solventes e insolventes. Esse valor revela que o modelo é capaz de distinguir corretamente, em 91,5% das vezes, as observações de empresas em diferentes condições financeiras, demonstrando elevada consistência estatística e robustez preditiva.

A Acurácia de 90,2% reforça o desempenho global do modelo, representando a proporção total de classificações corretas. A Sensibilidade de 80% mostra que o modelo identificou corretamente oito em cada dez empresas insolventes, o que é essencial em estudos de previsão de risco, onde a identificação antecipada de insolvências tem maior valor gerencial

do que falsos alarmes. Já a Especificidade de 92,1% demonstra a capacidade do modelo em reconhecer com alta precisão as empresas solventes, evitando classificações incorretas que poderiam comprometer decisões de crédito ou investimento.

A Precisão de 64,7% e o F1-Score de 0,7153 indicam um equilíbrio satisfatório entre a capacidade de prever corretamente casos positivos e a proporção de previsões positivas que realmente se confirmaram. Esses resultados evidenciam que, embora o modelo seja ligeiramente mais conservador ao sinalizar insolvência, ele mantém um equilíbrio saudável entre a minimização de falsos positivos e a detecção de casos reais de insolvência devido à base de dados deste estudo ser desbalanceada, onde a maioria das observações da amostra apresentam classificações de solvência.

Em conjunto, os indicadores confirmam que o modelo *Logit* para dados em painel é estatisticamente sólido, eficiente e aplicável a contextos reais de avaliação de risco financeiro. O desempenho obtido reflete a eficácia da integração entre indicadores tradicionais e variáveis derivadas do Modelo Dinâmico de Fleuriet, cuja combinação permitiu capturar simultaneamente aspectos de liquidez, estrutura de capital e gestão operacional, dimensões que se mostraram determinantes para a previsão do risco de insolvência em empresas brasileiras.

5 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo avaliar a efetividade da Regressão Logística para Dados em Painel na previsão do risco de insolvência empresarial, a partir da aplicação dessa técnica em empresas brasileiras listadas na B3, com exceção do setor financeiro. A amostra, composta por dados em painel, permitiu capturar simultaneamente a variação entre empresas e ao longo do tempo, conferindo maior precisão às estimativas. Após realizados os testes estatísticos específicos para modelos longitudinais logísticos, foi identificada a existência de efeitos entre as diferentes empresas da amostra, o que justificou a adoção do modelo de Efeitos Aleatórios como abordagem mais adequada.

O modelo apresentou AUC de 0,915 e Acurácia de 90,2%, demonstrando excelente desempenho global na classificação entre empresas solventes e insolventes. Além disso, a Sensibilidade de 80,0% e a Especificidade de 92,1% confirmam a capacidade do modelo em identificar com alta precisão as empresas em risco de insolvência e, ao mesmo tempo, evitar classificações incorretas de companhias solventes. O F1-Score de 0,715 reforça o equilíbrio entre precisão e abrangência na detecção de casos reais, validando o desempenho robusto do modelo em um conjunto de dados desbalanceado.

Por meio da estatística z , representando o Teste de Wald individual para cada variável, foi possível observar que as principais variáveis que contribuíram para a classificação no modelo foram X30 (Patrimônio Líquido menos Capital Social sobre o Ativo Total), X19 (Disponibilidade sobre Ativo Circulante mais Realizável a Longo Prazo), X14 (Ativo Circulante sobre Passivo Circulante), X11 (Necessidade de Capital de Giro), X2 (Capital de Giro sobre Receita Líquida) e X6 (Saldo de Tesouraria sobre Receita Líquida), que apresentaram coeficientes negativos, indicando que o fortalecimento da estrutura patrimonial e da liquidez reduz a probabilidade de insolvência.

Por outro lado, X27 (Financiamentos e Empréstimos Bancários sobre Ativo Circulante) e X21 (Fornecedores sobre Ativo Total) apresentaram coeficientes positivos, sugerindo que o aumento do endividamento e da dependência de capital de terceiros ampliam o risco financeiro das empresas.

Esses achados reforçam a importância da gestão da estrutura de capital e da liquidez operacional como determinantes da solvência empresarial. O estudo corrobora evidências da literatura internacional, como Ohlson (1980), Brédart (2014) e Abdullah *et al.* (2019), ao demonstrar que indicadores de rentabilidade, alavancagem e capital de giro são essenciais na previsão de insolvência. Ao mesmo tempo, amplia a discussão ao integrar variáveis do Modelo

Dinâmico de Fleuriot, evidenciando que medidas de folga financeira e estrutura operacional dinâmica contribuem para maior acurácia na identificação antecipada do risco de crédito.

Este estudo teve como limitação a amostra ser realizada somente com empresas brasileiras de capital aberto listadas na B3, analisando indicadores financeiros coletados a partir da literatura científica selecionada, por meio da aplicação de um único método estatístico.

Como sugestão para estudos futuros, recomenda-se a análise amostral de empresas de capital aberto de outros países, ou exclusivamente do setor financeiro, nacional ou não, bem como a coleta de outras possíveis variáveis independentes, como outros indicadores financeiros ou variáveis qualitativas como a presença de governança corporativa ou métricas de sustentabilidade do negócio na temática da previsão de insolvência empresarial, podendo ainda ser aplicado em diferentes técnicas estatísticas ou outras de aprendizado de máquina.

Conclui-se, portanto, que este estudo contribui ao demonstrar empiricamente que a Regressão Logística para Dados em Painel, aliada a indicadores financeiros tradicionais e dinâmicos, é capaz de antecipar o risco de insolvência com elevada precisão, oferecendo uma abordagem metodológica útil tanto para pesquisadores quanto para gestores e analistas do mercado. Ao integrar indicadores financeiros tradicionais e dinâmicos o modelo longitudinal logístico para dados em painel construído neste estudo, a pesquisa amplia o escopo da literatura e fornece uma ferramenta preditiva mais sensível às mudanças estruturais das empresas ao longo do tempo.

REFERÊNCIAS

ABDULLAH, N. A. H., AHMAD, A. H., ZAINUDIN, N.; RUS, R. M. **Predicting Financially Distressed Small- and Medium-sized Enterprises in Malaysia**. *Global Business Review*, v. 20, n. 3, p. 627–639, 2019. DOI: <https://doi.org/10.1177/0972150919837053>.

ALTMAN, E. I. **Financial Ratios, Discriminant Analysis And The Prediction Of Corporate Bankruptcy**. *The Journal of Finance*, v. 23, n. 4, p. 589–609, 1968. DOI: <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>.

ALTMAN, E. I., BAIDYA, T. K.; DIAS, L. M. R. **Previsão de problemas financeiros em empresas**. *Revista de administração de empresas*, v. 19, p. 17-28, 1979. DOI: <https://doi.org/10.1590/S0034-75901979000100002>.

ARANHA, J. A. M.; GONDRIGE, E. D. O. **Estatística multivariada aplicada: Construção do Modelo de Previsão de Insolvência Aranha & Gondrige**. *CONTRIBUCIONES A LAS CIENCIAS SOCIALES*, v. 17, p. 8, 2024. DOI: <https://doi.org/10.55905/revconv.17n.8-279>.

BALTAGI, B. H. **Econometric analysis of panel data**. 3. ed. Chichester: John Wiley & Sons, 2005.

BILLIOS, D., SERETIDOU, D.; STAVROPOULOS, A. **The Power of Numerical Indicators in Predicting Bankruptcy: A Systematic Review**. *Journal of Risk and Financial Management*, v. 17, n. 10, p. 433, 2024. DOI: <https://doi.org/10.3390/jrfm17100433>.

BRÉDART, X. **Bankruptcy Prediction Model: The Case of the United States**. *International Journal of Economics and Finance*, v. 6, p. 3, 2014. DOI: <https://doi.org/10.5539/ijef.v6n3p1>.

CORRAR, L. J., PAULO, E.; DIAS FILHO, J. M. **Análise Multivariada Para os Cursos de Administração, Ciências Contábeis e Economia**. Editora Atlas Ltda., 2007.

COSTA, T. R. D. **Risco De Insolvência: Uma Análise de Diferentes Métodos de Previsão e do Efeito da Estrutura de Capital**. Dissertação de Mestrado. Universidade Federal de Lavras., 2023. Disponível em: <http://repositorio.ufla.br/jspui/handle/1/57037>. Acesso em: 30 de abril de 2025.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata**. Rio de Janeiro: Elsevier, 2017.

FAWZI, N. S., KAMALUDDIN, A.; SANUSI, Z. M. **Monitoring Distressed Companies through Cash Flow Analysis**. *Procedia Economics and Finance*, v. 28, p. 136–144, 2015. DOI: [https://doi.org.ez26.periodicos.capes.gov.br/10.1016/S2212-5671\(15\)01092-8](https://doi.org.ez26.periodicos.capes.gov.br/10.1016/S2212-5671(15)01092-8).

FLEURIET, M; ZEIDAN, R. **O Modelo Dinâmico de Gestão Financeira**. 1. ed. 2015. Alta Books.

HILL, E.; DAVIS, K. **Limitations of fixed-effects models for panel data**. Statistical Modeling Journal, Londres, v. 18, n. 4, p. 455-472, 2020. DOI: <https://doi.org/10.1177/0731121419863785>.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An introduction to statistical learning: with applications in Python**. Nova York: Springer, 2023. DOI: <https://doi.org/10.1007/978-3-031-38747-0>.

JESUS, D. P.; BESARRIA, C. D. N. **Machine learning and sentiment analysis: Projecting bank insolvency risk**. Research in Economics, v. 77, n. 2, p. 226–238, 2023. DOI: <https://doi.org/10.1016/j.rie.2023.03.001>.

JONGH, P. DE; JONGH, E. DE; PIENAAR, M.; GORDON-GRANT, H.; OBERHOLZER, M.; SANTANA, L. **The impact of pre-selected variance inflation factor thresholds on the stability and predictive power of logistic regression models in credit scoring**. ORION, v. 31, n. 1, p. 17-37, 2015. DOI: <https://doi.org/10.5784/31-1-162>.

KING, G.; ZENG, L. **Logistic regression in rare events data**. Political Analysis, Cambridge, v. 9, n. 2, p. 137-163, 2001. DOI: <https://doi.org/10.1093/oxfordjournals.pan.a004868>.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. New York: Springer, 2013. DOI: <https://doi.org/10.1007/978-1-4614-6849-3>.

LIMA, F. G.; PAULINO, C. T.; FÁVERO, L. P. L. **ESG e Machine Learning: O Impacto na Previsão de Insolvência de Empresas Brasileiras**. F. G., Contabilometria, v. 11, p. 1, 2022.

NOURIN, S.; NASIM, F.; ALI, T. **Exploring the nexus of Hausman test application in tourism economics**. International Journal of Econometrics and Statistics, Dhaka, v. 12, n. 1, p. 35-49, 2025. DOI: <https://doi.org/10.1057/s41599-025-05247-3>.

OHLSON, J. A. **Financial ratios and the probabilistic prediction of bankruptcy**. Journal of accounting research, 109-131, 1980. DOI: <https://doi.org/10.2307/2490395>.

OLAOYE, S. A.; OGBEBOR, P. I.; OKUSAMI, Y. A. **Financial risk and financial performance of deposit money banks listed in Nigeria**. International Journal of Business and Management Review, v. 8, n. 4, p. 78-100, 2020. DOI: <https://doi.org/10.37745/IJBMR.VOL8.NO4.78-100.2020>

PEREIRA, V. S.; MARTINS, V. F. **Estudos de previsão de falências—uma revisão das publicações internacionais e brasileiras de 1930 a 2015**. Revista Contemporânea de Contabilidade, v. 12, n. 26, p. 163-196, 2015. DOI: <https://doi.org/10.5007/2175-8069.2015v12n26p163>.

PINTO, A. C. **O poder preditivo dos modelos com aprendizado de máquina é superior aos modelos tradicionais para análise do risco de crédito?** Revista Debates em Economia Aplicada—REDEA, v. 1, p. 1, 2021.

PRADO, J. W.; DE CASTRO ALCÂNTARA, V.; DE MELO CARVALHO, F.; VIEIRA, K. C.; MACHADO, L. K. C.; TONELLI, D. F. **Multivariate analysis of credit risk and bankruptcy research data: A bibliometric study involving different knowledge fields (1968–2014)**. *Scientometrics*, v. 106, n. 3, p. 1007–1029, 2016. DOI: <https://doi.org/10.1007/s11192-015-1829-6>.

PRADO, J. W. D., CARVALHO, F. D. M., BENEDICTO, G. C. D., ALCÂNTARA, V. D. C.; SANTOS, A. C. D. **Uma abordagem para análise do risco de crédito utilizando o modelo Fleuriet**. *Revista de Educação e Pesquisa em Contabilidade (REPeC)*, v. 12, n. 3, p. 341–363, 2018. DOI: <https://doi.org/10.17524/repec.v12i3.1816>.

PRADO, J. W., CARVALHO, F. D. M., BENEDICTO, G. C. D.; LIMA, A. L. R. **Analysis of credit risk faced by public companies in Brazil: An approach based on discriminant analysis, logistic regression and artificial neural networks**. *Estudios Gerenciales*, v. 347, p. –360, 2019. DOI: <https://doi.org/10.18046/j.estger.2019.153.3151>.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2024. Versão 4.4.2. Disponível em: <https://www.R-project.org/>.

SANTOS, ALLISSON SILVA DOS. **Previsão de Insolvência Corporativa: Uma Análise de Empresas Brasileiras de Capital Aberto Por Meio de Aprendizado de Máquina**. 2021. 73 f. Dissertação de Mestrado (Mestrado em Administração). Universidade Federal da Paraíba, João Pessoa, 2021.

SCHIO, T. A.; SAMPAIO, A. V. **Previsão da Recuperação Judicial de Empresas no Brasil: Uma Investigação Empírica**. *Pensar Contábil*, v. 24, n. 8, p. 5, 2022.

SHI, Y.; LI, X. **A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms**. *Heliyon*, v. 5, n. 1, p. 2, 2019. DOI: <https://doi.org/10.1016/j.heliyon.2019.e02997>.

SILVA, S. F. P. da; PESSANHA, J. F. M. **Identificação de indicadores para previsão de insolvência das distribuidoras de energia elétrica por meio de regressão logística para dados em painel**. *Contabilometria – Brazilian Journal of Quantitative Methods Applied to Accounting*, Monte Carmelo, v. 9, n. 1, p. 73–91, jan./jun. 2022.

STÜPP, Diego Rafael. **Previsão de insolvência a partir de indicadores contábeis: evidências de empresas listadas na BM&FBOVESPA nos anos 2004-2013**. 2015. 137 f. Dissertação (Mestrado em Contabilidade) – Universidade Federal de Santa Catarina, Florianópolis, 2015.

VÁSQUEZ-SERPA, L.-J., RODRÍGUEZ, C., PÉREZ-NÚÑEZ, J.-R.; NAVARRO, C. **Challenges of Artificial Intelligence for the Prevention and Identification of Bankruptcy Risk in Financial Institutions: A Systematic Review**. *Journal of Risk and Financial Management*, v. 18, n. 1, p. 26, 2025. DOI: <https://doi.org/10.3390/jrfm18010026>.

XUEZHOU, H.; REHMAN, R. U.; JIANG, L.; KHAN, M. A.; LI, X.; JIANG, Y. **Does firm growth impede or expedite insolvency risk? A Mediated Moderation Model of**

Leverage Maturity and Potential Fixed Collaterals. *Frontiers in Environmental Science*, Lausanne, v. 10, p. 1–16, 2022. DOI: <https://doi.org/10.3389/fenvs.2022.841380>.

ARTIGO 3 – PREVISÃO DO RISCO DE INSOLVÊNCIA: UMA ANÁLISE COMPARATIVA ENTRE DIFERENTES MÉTODOS DE APRENDIZADO DE MÁQUINA

RESUMO

O estudo teve como objetivo comparar o desempenho preditivo no desenvolvimento de diferentes modelos de aprendizado de máquina na previsão do risco de insolvência e identificar quais variáveis explicativas mais importantes foram determinantes na previsão de insolvência empresarial, analisando as empresas brasileiras não financeiras listadas na B3 no período de 2010 a 2024, totalizando 5.247 observações e 501 empresas. Foram considerados 35 indicadores financeiros tradicionais e do modelo dinâmico. Os modelos desenvolvidos incluíram *Random Forest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree* (DT), *Multilayer Perceptrons* (MLP) e *Extreme Gradient Boosting* (XGBoost). Os resultados evidenciaram o modelo XGBoost com maior desempenho preditivo, com acurácia de 95,4% e AUC de 0,981, confirmando sua robustez e superioridade preditiva frente às demais técnicas. Os modelos RF, SVM e MLP também apresentaram resultados consistentes, enquanto DT e KNN mostraram limitações em capturar relações não lineares. A análise de importância das variáveis revelou que os principais determinantes da insolvência estão associados à estrutura patrimonial, liquidez e gestão do capital de giro. Destacam-se os indicadores Patrimônio Líquido menos Capital Social sobre Ativo Total, Ativo Circulante sobre Passivo Circulante, Disponibilidade sobre o Ativo Circulante e Realizável a Longo Prazo, Capital de Giro e Saldo de Tesouraria sobre Receita Líquida. Como limitações, ressalta-se o uso exclusivo de empresas brasileiras listadas e apenas variáveis financeiras. Sugere-se, para estudos futuros, incluir variáveis macroeconômicas, insolvência associada ao fluxo de caixa, variáveis como de governança corporativa e ESG, e construção de modelos baseados em técnicas avançadas de *deep learning*.

Palavras-chave: Risco de insolvência. Indicadores Financeiros. Aprendizado de Máquina.

ABSTRACT

This study aimed to compare the predictive performance of different machine learning models in forecasting corporate insolvency risk and to identify the most relevant explanatory variables determining business failure. The analysis considered Brazilian non-financial firms listed on Brasil, Bolsa, Balcão (B3) over the period from 2010 to 2024, totaling 5,247 firm-year observations from 501 companies. A set of 35 traditional financial indicators and dynamic measures was employed as explanatory variables. The models developed included Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree (DT), Multilayer Perceptrons (MLP), and Extreme Gradient Boosting (XGBoost). The results highlighted XGBoost as the best-performing model, achieving an accuracy of 95.4% and an AUC of 0.981, confirming its robustness and superior predictive capability compared to the other techniques. RF, SVM, and MLP also delivered consistent results, whereas DT and KNN showed limitations in capturing nonlinear relationships. Variable importance analysis revealed that the main determinants of insolvency are related to capital structure, liquidity, and working capital management. Key indicators include equity minus paid-in capital over total assets, current assets over current liabilities, cash and cash equivalents over current assets and long-term receivables, working capital, and treasury balance over net revenue. As limitations, the study is restricted to Brazilian listed firms and to the exclusive use of financial variables. Future research is encouraged to incorporate macroeconomic factors, cash flow-based insolvency measures, corporate governance and ESG variables, and to develop models based on advanced deep learning techniques to enhance predictive robustness and generalizability.

Keywords: Insolvency Risk. Financial Indicators. Machine Learning.

1 INTRODUÇÃO

O risco de insolvência empresarial pode ser entendido como a possibilidade de uma organização não conseguir honrar suas obrigações financeiras no prazo acordado, colocando em xeque sua continuidade operacional. Esse risco pode ser ocasionado por diversos fatores internos e externos, como má gestão, crises econômicas ou decisões financeiras inadequadas, afetando diretamente a liquidez da empresa e sua capacidade de geração de valor. A correta gestão de riscos, portanto, torna-se essencial para garantir a sustentabilidade organizacional, conforme destacado por Peres (2010), ao apontar que a mitigação dos riscos empresariais é uma resposta necessária às exigências do ambiente competitivo e globalizado.

Diante da importância desse tema, os estudos sobre previsão do risco de insolvência empresarial têm buscado aplicar técnicas estatísticas multivariadas aliadas a indicadores contábeis e financeiros como variáveis explicativas. Desde o trabalho pioneiro de Altman (1968), que utilizou análise discriminante e indicadores como liquidez, rentabilidade e alavancagem para compor o modelo *Z-score*, diversos pesquisadores passaram a desenvolver modelos preditivos com o objetivo de antecipar crises de liquidez. Mais recentemente, técnicas de aprendizado de máquina como redes neurais e modelos dinâmicos de análise financeira têm demonstrado potencial para aprimorar a acurácia dos modelos de previsão de insolvência, como apresentado por Prado *et al.* (2020), que alcançaram taxas de acerto superiores a 90% na previsão e classificação de empresas solventes e insolventes.

A literatura sobre previsão da insolvência empresarial está diretamente relacionada aos estudos de risco de crédito e falência, uma vez que a incapacidade de honrar obrigações financeiras representa a materialização mais severa do risco assumido por credores e investidores. O risco de crédito refere-se à probabilidade de inadimplência frente aos compromissos contratuais, enquanto a falência constitui a concretização desse risco em sua forma extrema, caracterizada pela deterioração financeira persistente e pela inviabilidade da continuidade operacional. Nesse contexto, os modelos de previsão de insolvência configuram instrumentos relevantes para a identificação antecipada de fragilidades financeiras e para o suporte à tomada de decisão em finanças corporativas (Góis *et al.*, 2020).

No contexto brasileiro, o interesse pelo tema da previsão de insolvência ganhou impulso com a aplicação de métodos estatísticos multivariados em pesquisas nacionais. Um exemplo significativo é o trabalho de Altman, Baidya e Dias (1979), que adaptaram a metodologia do *Z-Score* para empresas brasileiras, considerando particularidades econômicas e contábeis locais. Essa iniciativa pioneira não apenas confirmou a aplicabilidade do modelo no Brasil, como

também estimulou uma série de estudos posteriores voltados à melhoria da previsão da insolvência empresarial no cenário nacional, contribuindo para a formação de uma base sólida de conhecimento sobre o tema.

A utilização de técnicas de aprendizado de máquina tem crescido ao longo do tempo conforme evidenciado nos trabalhos devido a seus modelos preditivos apresentarem uma acurácia por vezes superiores aos modelos estatísticos (Prado *et al.*, 2016; Pereira; Martins, 2015; Billios; Seretidou; Stavropoulos, 2024; Vásquez-Serpa *et al.*, 2025), citando os modelos como *Random Forest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree* (DT), *Multilayer Perceptrons* (MLP) e *Extreme Gradient Boosting* (XGBoost) nos quais tem sido utilizadas para medir o poder de previsão.

De acordo com Costa (2023), existe uma carência nos estudos de previsão de insolvência nos quais utilizam técnicas de aprendizado de máquina, no que tange a comparação de modelos e principalmente em estudos cujo objeto de estudo são as empresas brasileiras.

Evidenciando a relevância gerencial, política, social e acadêmica que deste tema recente e pouco explorado no cenário brasileiro, este estudo tem como **objetivo** comparar o desempenho preditivo no desenvolvimento de diferentes modelos de aprendizado de máquina na previsão do risco de insolvência e identificar quais variáveis explicativas mais importantes foram determinantes na previsão de insolvência empresarial. Os modelos desenvolvidos neste estudo foram RF, SVM, KNN, DT, MLP e XGBoost e as empresas que compuseram a amostra foram as listadas Brasil, Bolsa e Balcão (B3), com exceção do setor financeiro, no período de 2010 a 2024.

2 REFERENCIAL TEÓRICO

O primeiro trabalho a utilizar alguma técnica de aprendizado de máquina na temática da previsão do risco de insolvência, segundo Prado *et al.* (2016), foi o de Odom e Sharda (1990), que empregou o modelo *Neural Networks*. O objetivo deste estudo foi comparar a técnica multivariada Análise Discriminante com a *Neural Networks* para verificar qual técnica possui melhor acurácia. O trabalho utilizou os mesmos indicadores financeiros de Altman (1968) no modelo discriminante. Com uma amostra total de 129 empresas, nos quais 65 classificadas como solventes e 64 não insolventes, foram realizados testes com diferentes proporções entre empresas falidas e não falidas, sendo que modelo *Neural Networks* apresentou na maioria dos testes melhor desempenho na previsão de falências.

Na sequência, o trabalho de Tam (1991) foi relevante na pesquisa ao evidenciar que o modelo *Neural Networks* apresentou melhor capacidade preditiva que demais modelos estatísticos e de aprendizado de máquina, sendo eles a Análise Discriminante, Regressão Logística, *K-Nearest Neighbors* e *Decision Tree*, utilizando como amostra 118 bancos e 19 variáveis financeiras. O estudo evidencia a importância de se testar diferentes modelos em um mesmo estudo, a fim de comparar para determinada situação ou amostra qual modelo apresenta melhor desempenho preditivo na temática.

Após a divulgação dos estudos iniciais de Odom e Sharda (1990) e Tam (1991), houve um crescente interesse na temática de previsão de falência de empresas e risco de crédito utilizando modelos de inteligência computacional, conforme apresentado nos estudos de Prado *et al.* (2016), Shi e Li (2019), Almeida (2023), Dasilas e Rigani (2024) e Roy e Vasa (2025), devido ao avanço da tecnologia que permite criar modelos preditivos mais robustos.

O trabalho de Barboza, Kimura e Altman (2017) apresentou um estudo comparativo com 8 diferentes modelos desenvolvidos a partir técnicas de aprendizado de máquina, sendo elas o *Support Vector Machine* (SVM), *Bagging*, *Boosting*, *Random Forest* (RF), *Neural Networks* (NN), e técnicas tradicionais como a Análise Discriminante (AD) e a Regressão Logística (RL), com amostra de 449 empresas norte americanas no período de 1985 a 2013, totalizando mais de 10.000 observações empresas-ano. Os resultados apontaram que RF obteve uma acurácia de 100%, seguido por *Bagging* (99,67%) e *Boosting* (96,21%), evidenciando que modelos de aprendizado apresentam melhor poder preditivo em relação as técnicas tradicionais.

No estudo realizado por Aljawazneh *et al.* (2021), utilizou-se técnicas comparativas de aprendizado de máquina, tais como *Random Forest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Multilayer Perceptron* (MLP) e *Extreme Gradient Boosting*

(XGBoost), compreendendo uma amostra total de 19.678 observações, sendo divididos por três conjuntos de dados, e 192 variáveis no total. O estudo ressalta que foi necessário realizar diversas técnicas de balanceamento de dados para equalizar as variáveis da amostra, sendo os indivíduos coletados em três países diferentes, em que ultrapassaram 97% de acurácia, indicando excelente poder preditivo. Contudo, o estudo reforça da necessidade de se testar outros métodos em outros contextos para verificar a autenticidade da robustez obtida por meio de técnicas de balanceamento dos dados.

De forma complementar, no trabalho de Garcia (2022) foram utilizados modelos como *Neural Networks* (NN), *Random Forest* (RF), *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM) e *Extreme Gradient Boosting* (XGBoost), com uma amostra composta por 41.933 observações de informações financeiras trimestrais de 1.824 empresas, no período de 2010 a 2018, excluindo-se empresas do setor financeiro. Com base em cinco variáveis financeiras, o autor obteve acurácia de 100% ao aplicar a técnica Random Forest. O autor destaca o problema de desbalanceamento de classes, onde a amostra compreendeu maior quantidade de observações solventes do que insolventes, o que levou a utilizar métodos de balanceamento de classes.

Noh (2023) utilizou modelos estatísticos como a Regressão Logística juntamente com modelos LSTM, RF, KNN e DT para verificação de quais modelos apresentam melhor desempenho. Utilizando uma amostra com 1020 empresas listadas na bolsa sul-coreana KOSPI (*Composite Stock Price Index*) e KOSDAQ (com dados coletados de 2012 a 2021, ao analisar 13 indicadores financeiros como variáveis independentes, demonstrando que o RF dentre as técnicas de aprendizado de máquina apresentou melhor acurácia, assumindo que o modelo LSTM não foi adequado para a análise realizada neste estudo.

Idhmad *et al.* (2024) evidenciaram que por meio de um estudo realizado com empresas de Taiwan, utilizando os modelos de aprendizado de máquina SVM, KNN, RF, *Naive Bayes* (NB), DT, *AdaBoost* e *Gradient Boosting Classifier* (GBC), o modelo que apresentou melhor desempenho foi o SVM com 98% de acurácia.

No Brasil, a aplicação de técnicas de aprendizado de máquina para a previsão de insolvência empresarial tem ganhado relevância nos últimos anos, embora ainda seja menos explorada do que no cenário internacional. Esses estudos buscam adaptar modelos amplamente utilizados no exterior às características do mercado e das empresas brasileiras, considerando fatores como disponibilidade de dados, particularidades contábeis e especificidades setoriais. Entre as abordagens investigadas, destacam-se Redes Neurais Artificiais, Máquinas de Vetores

de Suporte, Florestas Aleatórias e algoritmos de *Boosting*, geralmente avaliados com base em indicadores financeiros extraídos das demonstrações contábeis.

No trabalho de Prado *et. al.* (2020) evidenciou-se a utilização de NN na previsão de falência de empresas, utilizando indicadores financeiros tradicionais e do modelo dinâmico de Fleuriet (1978; 2003), totalizando 8 variáveis que foram aplicadas modelo. A amostra utilizada foi composta por empresas do setor comercial, sendo classificadas 50 empresas como insolventes e 50 como solventes, para fins de equiparação da amostra. Os autores concluíram que o modelo NN conseguiu atingir 100% de acurácia.

No estudo de Pinto (2021), utilizou-se as técnicas Naive Bayes, SVM, Regressão Logística, NN, RF e XGBoost, sendo a amostra do tipo dados em painel com 7.734 observações, sendo elas demonstrações financeiras de empresas listadas na B3, compreendendo um total de 423 empresas, utilizando 25 indicadores financeiros e 16 indicadores macroeconômicos. Os modelos apresentaram ótimo desempenho, sendo o RF com maior poder de previsão, com 98,38% de acurácia.

No estudo de Santos (2021), foram utilizados modelos RF, NB, Regressão Logística, KNN, SVM, *Bagging* e *Boosting*. A amostra média anual contemplou 299 empresas no período de 2010 a 2020, e os dados foram coletados em corte transversal com foco em setores diversos, com destaque também para a análise separada de instituições financeiras, sendo utilizadas 19 variáveis financeiras. O modelo RF foi o que apresentou melhor desempenho preditivo de forma geral, atingindo acurácia máxima de 95,72%.

O artigo de Jesus e Besarria (2023) inovou ao integrar análise de sentimentos de gestores bancários com técnicas de aprendizado de máquina na projeção do risco de insolvência de instituições financeiras brasileiras listadas na B3. A amostra foi composta por 12 bancos, com dados trimestrais de 2012 a 2021, configurando-se como painel de dados. Os modelos preditivos utilizados no estudo foram a Regressão Logística, NB, RF, AdaBoost, SVM e DT, sendo este último o que apresentou melhor desempenho na amostra de teste, com 93,75%.

Cunha, Penedo e Barboza (2024) realizaram uma análise focada exclusivamente na técnica RF para prever a insolvência de empresas brasileiras de capital aberto. A amostra foi composta por empresas que tiveram seu registro deslistado na B3 por insolvência entre os anos de 2005 e 2018. Foram utilizados dados extraídos da CVM e da plataforma Economatica, caracterizando a base como de corte transversal. O estudo utilizou 10 variáveis financeiras. O modelo RF obteve acurácia de 98%, mostrando-se altamente eficiente na predição de falências em empresas brasileiras.

Por fim, Tavares *et al.* (2025) implementaram uma abordagem inovadora para previsão de falência de pequenas e médias empresas (PMEs) com a utilização de *Convolutional Neural Network* (CNN), a partir da geração de imagens que representam os dados contábeis das empresas. A amostra abrangeu 14.456 empresas em corte transversal. O diferencial metodológico está na conversão de dados contábeis em imagens processáveis pela CNN, possibilitando a aplicação de técnicas avançadas normalmente utilizadas em reconhecimento visual. O modelo CNN alcançou 97,8% de acurácia.

Com base na literatura apresentada, observa-se que a aplicação de técnicas avançadas de aprendizado de máquina na previsão do risco de insolvência empresarial tem se mostrado altamente eficaz, especialmente em termos de acurácia e robustez preditiva. No entanto, a maioria dos estudos concentra-se em contextos internacionais, com foco em mercados desenvolvidos, o que pode limitar a aplicabilidade dos resultados a outras realidades econômicas, como a brasileira. Além disso, ainda são escassos os estudos que comparam de forma sistemática o desempenho de diferentes modelos de aprendizado de máquina utilizando uma base de dados brasileira ampla, atual e com características específicas, o que reforça a lacuna existente na literatura nacional sobre o tema.

Diante desse cenário, justifica-se a realização deste estudo, que visa avaliar comparativamente a efetividade de diferentes técnicas de aprendizado de máquina na previsão do risco de insolvência em empresas não financeiras listadas na B3. Ao considerar tanto indicadores financeiros tradicionais quanto dinâmicos, esta pesquisa busca contribuir para o avanço do conhecimento científico, bem como oferecer subsídios práticos para gestores, investidores e formuladores de políticas que necessitam de ferramentas mais precisas e adaptadas à realidade brasileira para a tomada de decisões estratégicas e preventivas, bem como na identificação de variáveis que influenciam na insolvência empresarial, para ações preventivas.

3 METODOLOGIA

Neste tópico será apresentado o detalhamento da metodologia utilizada para execução do trabalho. Nesse sentido, será apresentado os modelos de aprendizado de máquina, a amostra coletada e as variáveis utilizadas.

Para gerar os resultados obtidos neste estudo, foi utilizado o programa *Python*, versão padrão 3.10, sendo gerada as saídas no *Google Colab*, versão 14 de outubro de 2025.

3.1 Definição da Amostra

A amostra deste estudo compreende empresas brasileiras não financeiras listadas na Brasil, Bolsa, Balcão (B3), no período de 2010 a 2024, com informações extraídas das Demonstrações Financeiras anuais por meio da plataforma Economatica. A base de dados apresenta estrutura longitudinal (dados em painel), na qual cada observação corresponde a uma empresa em um determinado ano, totalizando 5.247 registros referentes a 501 empresas distintas.

Embora a base possua estrutura em painel, os modelos de aprendizado de máquina empregados neste estudo não utilizam a abordagem econométrica tradicional de dados em painel, uma vez que tais algoritmos não requerem a especificação de efeitos fixos ou aleatórios, nem pressupõem estrutura estocástica temporal. Assim, para fins preditivos, cada observação empresa-ano é tratada como uma instância independente, respeitando-se, contudo, a separação temporal no processo de validação do modelo.

A amostra apresenta desbalanceamento de classes, sendo 4.451 observações classificadas como solventes e 796 como insolventes. Observações com dados ausentes foram excluídas para garantir a consistência da base. A estratégia de coleta e estruturação dos dados é compatível com estudos recentes que aplicam técnicas de aprendizado de máquina à previsão de insolvência empresarial, como Pinto (2021), Santos (2021), Garcia (2022), Jesus e Besarria (2023), Cunha, Penedo e Barboza (2024) e Tavares et al. (2025).

Para a definição da variável dependente, foram classificadas como insolventes as observações correspondentes a empresas que, em determinado período, apresentaram patrimônio líquido negativo ou protocolaram pedido de recuperação judicial ou extrajudicial. Nessas situações, a variável dependente assume valor igual a 1. As demais observações receberam valor 0, indicando condição de solvência. Dessa forma, uma mesma empresa pode

ser classificada como solvente em determinados períodos e insolvente em outros, refletindo a dinâmica temporal do risco de insolvência.

3.2 Definição dos Indicadores

Para os objetivos desta investigação, foram empregadas as variáveis delineadas por Prado *et al.* (2018) e Costa (2023), classificando indicadores financeiros obtidos a partir de estruturas comumente utilizadas, denominadas indicadores tradicionais, em conjunto com o modelo dinâmico de Fleuriet, resultando em um total agregado de 35 variáveis independentes.

Com a finalidade de aprimorar a qualidade dos modelos preditivos e garantir maior estabilidade na análise da importância das variáveis, realizou-se a análise do VIF (*Variance Inflation Factor*), onde variáveis que possuíam valor maior do que 10 foram retiradas, bem como verificado por meio da matriz de correlação, as variáveis que possuem valores maiores do que 0,8 (Fávero e Belfiore, 2017). O conjunto final de variáveis utilizado no estudo, selecionado após este processo de saneamento, está detalhado na Tabela 26.

Tabela 26 – Variáveis Utilizadas na Construção dos Modelos (continua).

Nome	Indicador Financeiro
X2	$\frac{\text{Capital de Giro}}{\text{Receita Líquida}}$
X4	$\frac{\text{Necessidade de Capital de Giro}}{\text{Receita Líquida}}$
X6	$\frac{\text{Saldo de Tesouraria}}{\text{Receita Líquida}}$
X7	$\frac{\text{Passivo Financeiro}}{\text{Ativo Circulante}}$
X9	$\frac{\text{Saldo de Tesouraria}}{\text{Necessidade de Capital de Giro}}$
X11	<i>Necessidade de Capital de Giro</i>
X12	<i>Capital de Giro</i>
X14	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$
X15	$\frac{\text{Receita Líquida Operacional}}{\text{Ativo Total}}$
X16	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Ativo Total}}$
X19	$\frac{\text{Disponibilidade}}{\text{Ativo Circulante}}$ + <i>Realizável a Longo Prazo</i>

Tabela 26 – Variáveis Utilizadas na Construção dos Modelos (conclusão).

Nome	Indicador Financeiro
X20	$\frac{\text{Estoque}}{\text{Ativo Total}}$
X21	$\frac{\text{Fornecedores}}{\text{Ativo Total}}$
X27	$\frac{\text{Financiamentos e Empréstimos Bancários}}{\text{Ativo Circulante}}$
X30	$\frac{\text{Patrimônio Líquido} - \text{Capital Social}}{\text{Ativo Total}}$
X32	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Impostos sobre Despesas Financeiras}}$
X34	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Dívida Líquida}}$

Fonte: Elaborado pelo autor (2025).

Por meio da Tabela 26, observa-se as variáveis utilizadas no estudo que possuem a ausência de colinearidade severa, permitindo maior robustez na previsão do modelo e restando, portanto, 17 variáveis a serem analisadas e estimadas nos modelos.

Embora modelos de aprendizado de máquina não lineares não apresentem pressupostos de ausência de multicolinearidade, a literatura aponta que a presença de variáveis altamente correlacionados pode aumentar a complexidade do modelo, elevar sua variância e reduzir a estabilidade da generalização dos dados. De acordo com James *et al.* (2023), métodos mais flexíveis estão particularmente sujeitos a esse *trade-off*, enquanto Kuhn e Johnson (2013) enfatizam que, em contextos de modelagem preditiva aplicada, a redundância entre variáveis pode favorecer *overfitting* devido a redundância de informações aplicadas na construção dos modelos, além de comprometer a interpretação de métricas de importância das variáveis para cada modelo.

Portanto, a filtragem por correlação e VIF mantém-se pertinente também para modelos de aprendizado de máquina, sobretudo quando o objetivo envolve não apenas desempenho preditivo, mas também explicabilidade e consistência dos resultados.

3.3 Técnicas de Aprendizado de Máquina

O surgimento das técnicas de aprendizado de máquina remonta a meados do século XX, com desenvolvimentos significativos ocorrendo nas últimas décadas. Inicialmente, o campo da

inteligência artificial (IA) se concentrou na criação de máquinas capazes de realizar tarefas intelectuais semelhantes às habilidades humanas. Essa mudança foi caracterizada pelo desenvolvimento de técnicas inspiradas na inteligência humana, como redes neurais, sistemas difusos e computação evolutiva, que foram aplicadas para resolver problemas do mundo real em vários domínios, incluindo medicina, economia e engenharia (Bosse; Hoogendoorn, 2014; Oja, 1999).

De acordo com Khusein (2023), a inteligência computacional é um subcampo da IA que visa imitar a cognição humana por meio de simulações computacionais. Ele se concentra em projetar e aprimorar modelos inteligentes capazes de processar e avaliar *big data*. Com isso, a inteligência computacional se distingue por permitir que os sistemas evoluam e aprendam com dados limitados, utilizando informações imperfeitas para gerar ações de controle. Essa abordagem incorpora vários métodos, incluindo inteligência de enxame e seleção natural, para permitir que os sistemas de computação exibam capacidades de tomada de decisão sem intervenção humana.

A técnicas de aprendizado de máquina podem ser classificadas em três categorias principais com base na interpretabilidade e no acesso ao funcionamento interno dos modelos, sendo elas a caixa branca, caixa cinza e caixa preta. Modelos de caixa branca (*white-box*), como árvores de decisão e regressões lineares, são transparentes, permitindo que se compreenda claramente como cada entrada influencia a saída. Os modelos de caixa preta (*black-box*), como redes neurais profundas e métodos ensemble complexos, fornecem alta precisão, mas sua lógica interna é opaca. Já os modelos de caixa cinza (*grey-box*) combinam elementos dos dois anteriores, buscando equilíbrio entre precisão e interpretabilidade (Grau *et al.*, 2016; Pintelas; Livieris; Pintelas, 2020; Seddik; Ahlem; Hocine, 2022).

No campo de estudos de aprendizado de máquina, frequentemente existe uma tensão entre precisão e interpretabilidade dos modelos. De forma geral, modelos mais simples e compreensíveis, como regressões ou árvores rasas, são mais interpretáveis, mas tendem a ser menos precisos. Por outro lado, modelos mais complexos como redes neurais profundas alcançam alta precisão às custas da transparência. Essa relação é chamada de *trade-off* entre interpretabilidade e precisão. Recentemente, estudos têm buscado formalizar esse dilema e desenvolver métricas para equilibrá-lo. Por exemplo, foi demonstrado que é possível aumentar a interpretabilidade de modelos com uma perda de apenas 2% na acurácia, evidenciando que decisões informadas podem ser tomadas mesmo com modelos levemente menos precisos. (Nazir *et al.*, 2021; Bersimas *et al.*, 2022).

Para este estudo, foram utilizadas as técnicas de aprendizado de máquina, a saber, *Random Forest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree* (DT) e *Extreme Gradient Boosting* (XGBoost) para análise comparativa de desempenho entre eles na previsão de insolvência empresarial.

As técnicas baseadas em *ensemble methods* têm se destacado na literatura recente por sua capacidade de aumentar a precisão e a estabilidade dos modelos preditivos. Conforme James *et al.* (2023), o princípio fundamental dessas abordagens é a combinação de múltiplos modelos individuais, de forma a reduzir a variância e aprimorar a generalização dos resultados. Em vez de depender de uma única árvore de decisão ou classificador, o método ensemble constrói várias versões do modelo e as integra por meio de votação ou média, o que torna o processo mais robusto a ruídos e variações na base de dados.

Entre as principais técnicas ensemble, destacam-se o *Bagging* (*Bootstrap Aggregating*), o *Random Forest* (RF) e o *Boosting*, do qual o *Extreme Gradient Boosting* (XGBoost) é uma das implementações mais eficientes. O *Bagging*, segundo Kuhn e Johnson (2013), consiste em gerar diversas amostras *bootstrap* da base original e treinar um modelo em cada uma delas, agregando as previsões para reduzir a variância e mitigar o *overfitting*. O *Random Forest*, por sua vez, aprimora o *Bagging* ao introduzir aleatoriedade na escolha dos preditores em cada divisão das árvores, resultando em modelos mais diversos e menos correlacionados entre si. Essa estratégia melhora a capacidade de generalização e reduz o impacto de preditores dominantes.

Já o *Boosting* adota uma abordagem sequencial, em que cada novo modelo é ajustado para corrigir os erros do anterior. O XGBoost, amplamente utilizado em estudos financeiros, refina esse processo ao empregar funções de perda diferenciáveis e regularização dos parâmetros, equilibrando viés e variância e aumentando a eficiência computacional. Como ressaltam James *et al.* (2023), essa técnica é especialmente adequada para bases com alta dimensionalidade e relações não lineares, características típicas dos dados contábeis e financeiros analisados em previsões de insolvência.

Além da capacidade preditiva, as técnicas ensemble oferecem a vantagem de mensurar a importância das variáveis (*feature importance*). Esse recurso possibilita identificar quais indicadores financeiros exercem maior influência sobre a classificação de solvência e insolvência, permitindo não apenas o aprimoramento do modelo, mas também a interpretação econômica dos resultados. Essa característica torna o uso das técnicas *ensemble* particularmente relevante no contexto da presente pesquisa, pois combina alto desempenho preditivo com aplicabilidade prática para gestores financeiros e analistas de risco.

3.3.1 Decision Tree (DT)

O modelo de DT é uma técnica de aprendizado supervisionado usada para tarefas de classificação e regressão. Ele funciona segmentando iterativamente os dados com base em atributos, criando uma estrutura hierárquica onde cada nó interno representa uma decisão baseada em uma variável, e cada nó terminal (ou "folha") representa um resultado. Sua principal vantagem é a interpretabilidade, já que o processo de decisão pode ser seguido de forma visual e lógica, facilitando o entendimento mesmo por não-especialistas (Izza; Ignatiev; Marques-Silva, 2020; Souza *et al.*, 2022).

O modelo foi desenvolvido por Quinlan (1986), sendo amplamente utilizado em tarefas de classificação, baseando-se na divisão sucessiva do conjunto de dados a partir de atributos que melhor separam as classes, segundo um critério de ganho de informação. A Equação 14 apresenta a equação geral do modelo conforme informado pelo autor, baseando-se no conceito de entropia da informação.

$$I(p, n) = -\left(\frac{p}{p+n}\right) \ln 2 \left(\frac{p}{p+n}\right) - \left(\frac{n}{p+n}\right) \ln 2 \left(\frac{n}{p+n}\right) \quad (14)$$

Em que:

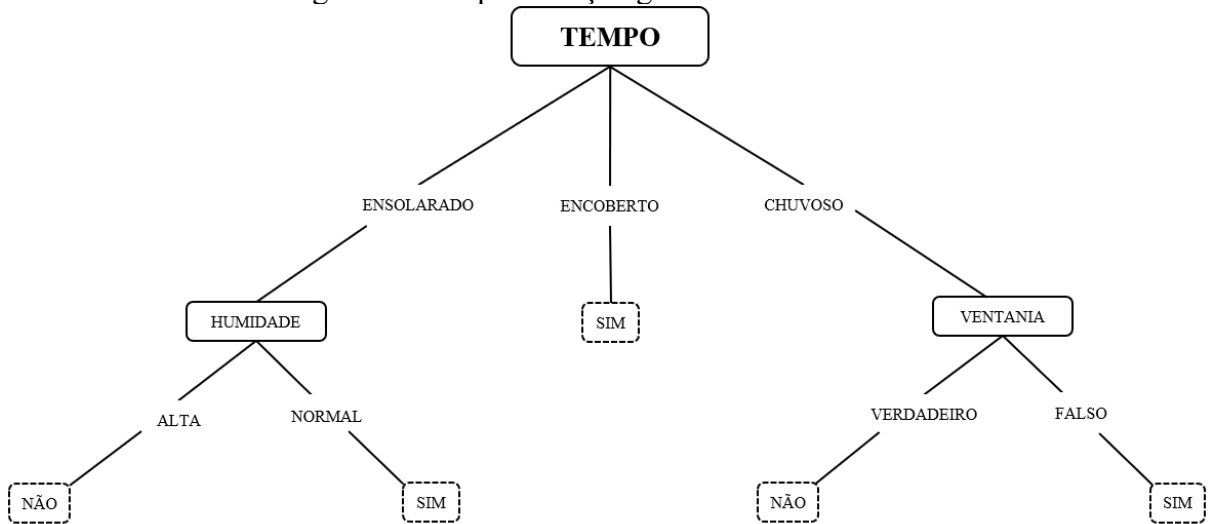
p representa o número de objetos da classe positiva (P);

n representa o número de objetos da classe negativa (N);

$I(p, n)$ é a medida da incerteza (ou entropia) associada à classificação de um objeto arbitrário do conjunto.

A entropia avalia a quantidade média de informação necessária para classificar um objeto. Quanto maior a entropia, maior a desordem ou incerteza na classificação. A Figura 13 apresenta a representação gráfica da árvore de decisão.

Figura 13 – Representação gráfica do Modelo DT.



Fonte: Adaptado de Quinlan (1986, p. 87).

Segundo Balcan e Sharda (2024), o modelo DT avalia quão bem um atributo separa as classes, escolhendo o que reduz mais a incerteza. A interpretação da árvore completa é feita lendo do nó raiz até uma folha, acumulando as regras de decisão ao longo do caminho. Segundo Moreira *et al.* (2022), o modelo é classificado como caixa branca devido a sua transparência na compreensão de como o modelo realizou as previsões e quais variáveis independentes foram importantes na interpretação do modelo.

3.3.2 Random Forest (RF)

De acordo com Nguyen, Viviani e Jabeur (2023), a RF consiste em um método desenvolvido por Breiman (2001). O método consiste em um conjunto de modelos de DT, sendo que cada árvore gera uma previsão diferente. Os autores afirmam que “o resultado da floresta aleatória é a previsão de classe que recebe mais votos de uma árvore individual”, sendo considerado como uma caixa preta. O modelo é amplamente utilizado em economia e finanças devido sua eficácia. A Equação 15 apresenta a construção da RF.

$$RF(x) = \frac{1}{K} \sum_{k=1}^K h(x, \theta_k) \quad (15)$$

Em que:

x é o vetor de entrada (*input*);

$h(x, \theta_k)$ é a predição do classificador (árvore de decisão) k -ésimo, construído a partir de uma amostra aleatória com *bootstrapping* do conjunto de dados e com um subconjunto aleatório de variáveis em cada divisão do nó;

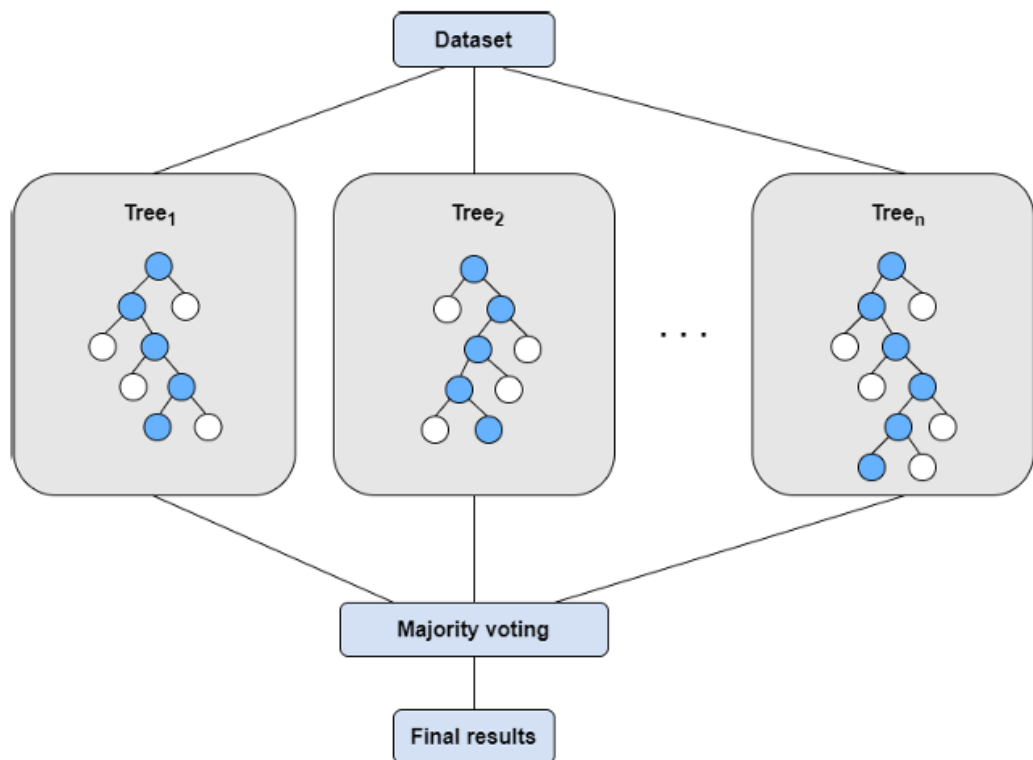
θ_k representa o vetor aleatório que determina como a k -ésima árvore é construída;

K é o número de árvores na floresta;

majority_vote é a função que retorna a classe mais votada entre todas as árvores.

Este modelo funciona bem devido à média de previsões (no caso de regressão) ou votação (no caso de classificação), o que reduz a variância sem aumentar muito o viés. A Figura 14 evidencia a funcionalidade do modelo.

Figura 14 – Representação gráfica do modelo RF.



Fonte: Aljawazneh *et al.* (2021, p. 97017).

A estratégia de construir as árvores de decisão com subconjuntos aleatórios de dados e variáveis em cada divisão ajuda a garantir a diversidade entre as árvores da floresta, evitando que todas elas se tornem muito semelhantes e, conseqüentemente, reduzindo a probabilidade de erro.

3.3.3 Support Vector Machine (SVM)

O modelo de aprendizado de máquina conhecido como *Support Vector Machine* (SVM) foi introduzido por Cortes e Vapnik (1995), sendo uma das mais influentes abordagens de aprendizado supervisionado. A SVM busca encontrar o hiperplano ótimo que separa duas classes de dados com a maior margem possível, o que reduz o erro de generalização (Wang *et al.*, 2024). De forma complementar, Joachims (2002) informa que esse conceito se baseia no princípio da Minimização do Risco Estrutural de Vapnik, cujo objetivo é garantir a menor probabilidade de erro na classificação de novos dados. A Equação 16 apresenta a equação padrão para o caso linear.

$$h_{w,b}(x) = \text{sign}(w^T x + b) \quad (16)$$

Em que:

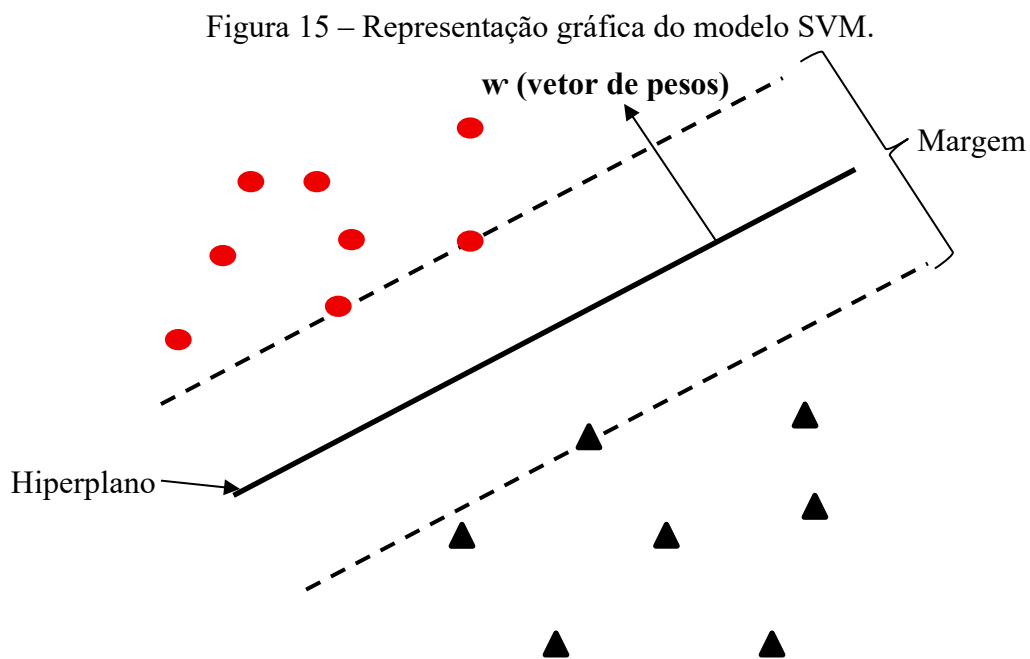
w é o vetor de pesos, que define a orientação do hiperplano;

x é o vetor de características de entrada;

b é o viés que define o deslocamento do hiperplano em relação à origem;

sign é a função que define a classe do dado de entrada: retorna +1 ou -1 dependendo do lado do hiperplano em que x se encontra.

A Figura 15 apresenta a forma de classificação do modelo.



Fonte: Adaptado de Hamel (2006).

De acordo com Hammer *et al.* (2012), quanto à interpretação do modelo, a SVM é classificada como modelo de caixa preta. Essa classificação surge porque os mecanismos subjacentes das SVMs não são facilmente interpretáveis por humanos, dificultando a compreensão de como as decisões são tomadas.

3.3.4 K-Nearest Neighbors (KNN)

Segundo Qiu *et al.* (2022), o modelo KNN é um dos algoritmos de aprendizado supervisionado mais simples e intuitivos, originalmente apresentado por Fix e Hodges (1951) no contexto de reconhecimento de padrões. Ele se baseia na ideia de que uma amostra deve ser classificada de acordo com as classes majoritárias de suas k amostras mais próximas no espaço de características. A proximidade entre os pontos é geralmente medida por uma métrica de distância, como a Euclidiana. O modelo não requer treinamento propriamente dito, onde todo o conjunto de dados é armazenado e consultado no momento da previsão. A Equação 17 apresenta o modelo mais comum usada no KNN para medir a proximidade que é a distância Euclidiana.

$$d(x, x_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2} \quad (17)$$

Em que:

x é o vetor de características de nova amostra;

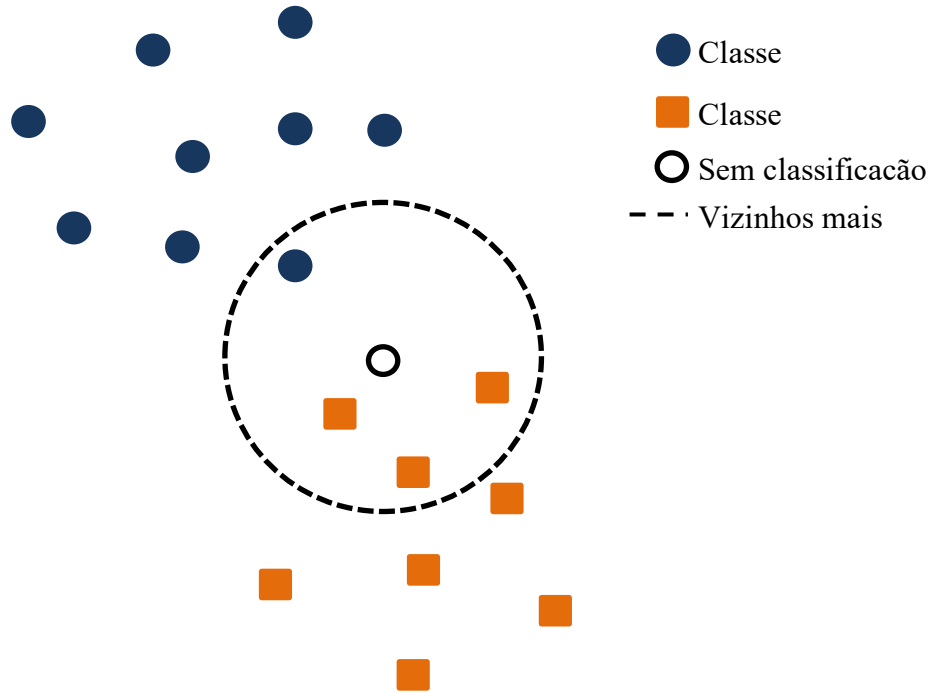
x_i representa uma amostra do conjunto de treinamento;

x_j e x_{ij} são os valores da j -ésima característica das amostras x e x_i , respectivamente;

n é o número de características (dimensão do vetor).

Após o cálculo dessa distância para todas as amostras, o modelo seleciona os k exemplos mais próximos e classifica a nova entrada com base na classe mais comum entre eles. A Figura 16 apresenta como o modelo classifica as amostras.

Figura 16 – Representação gráfica do modelo KNN.



Fonte: Elaborado pelo autor (2025).

Do ponto de vista interpretativo, o KNN é considerado um modelo de caixa branca, já que suas decisões podem ser facilmente explicadas com base nas amostras vizinhas. A classificação de uma entrada é diretamente atribuída às classes de seus vizinhos mais próximos, o que permite uma justificativa clara e rastreável para cada previsão (Barceló *et al.*, 2025).

3.3.5 Multilayer Perceptrons (MLP)

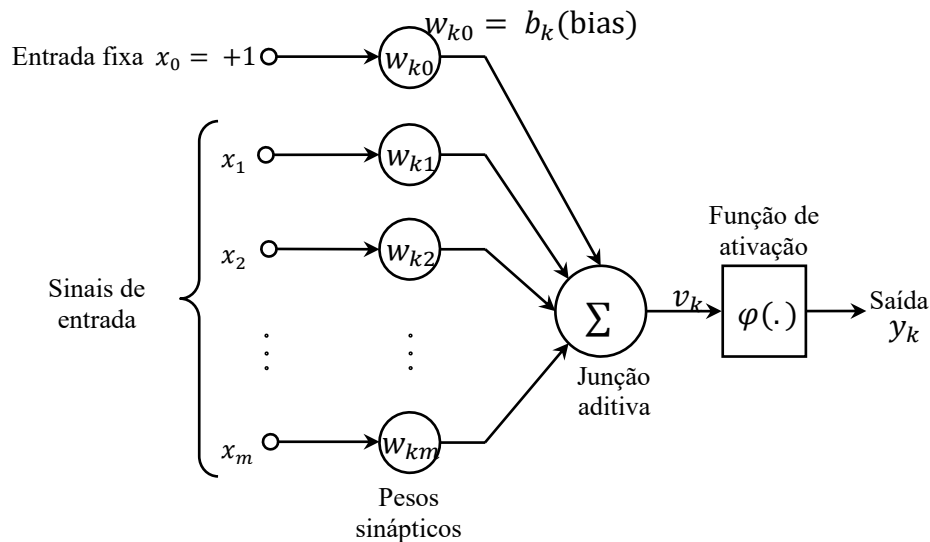
O *Multilayer Perceptrons* (MLP) é uma técnica de aprendizado de máquina inspirada no funcionamento do cérebro humano, fundamentada nos conceitos de redes neurais artificiais apresentados inicialmente por McCulloch e Pitts (1943). Eles introduziram o conceito de um neurônio artificial como uma forma de representar o comportamento lógico do cérebro.

Com base no artigo de Prado *et al.* (2020), redes neurais artificiais (RNAs), categoria à qual pertence o MLP, podem ser explicadas como sistemas computacionais inspirados no funcionamento do cérebro humano. Elas são compostas por camadas de unidades de processamento chamadas "neurônios", organizadas em três tipos principais de camadas, sendo elas as camadas de entrada, camadas ocultas e camadas de saídas. O funcionamento de uma rede neural envolve múltiplas camadas de unidades chamadas neurônios artificiais. Cada

neurônio realiza uma operação matemática simples, onde recebe entradas, aplica pesos, soma os resultados e passa por uma função de ativação não linear (Haykin, 2001).

Além disso, Prado (2016) reforça que essas redes operam a partir da organização de neurônios distribuídos em camadas interconectadas, nas quais os sinais são propagados, ponderados e transformados por funções de ativação, permitindo a modelagem de relações não lineares complexas entre variáveis. A Figura 17 apresenta a funcionalidade do modelo.

Figura 17 – Representação gráfica do modelo não linear.



Fonte: Prado (2016, p. 79).

A Figura 17 representa uma estrutura do MLP, composta por três tipos de camadas, sendo elas a camada de entrada, camada oculta e camada de saída. Cada círculo representa um neurônio artificial, e as linhas que os conectam simbolizam as sinapses, ou seja, os pesos que transmitem os sinais de uma camada para a outra. Os dados entram pela camada de entrada, são processados nas camadas ocultas com o uso de funções de ativação não lineares (sigmoide), e produzem uma saída com base no aprendizado anterior. O ajuste dos pesos é feito durante o processo de treinamento, geralmente usando o algoritmo de retropropagação do erro.

A interpretação desse modelo é complexa, principalmente em arquiteturas profundas com muitas camadas, o que torna o MLP um modelo considerado de caixa-preta. Isso significa que, embora o modelo possa gerar previsões precisas, é difícil entender como ele chega a essas conclusões internamente. Alguns métodos de interpretação, como LIME, SHAP e *Permutation Importance* vêm sendo utilizados para tentar explicar parcialmente o funcionamento interno dessas redes, mas sua interpretabilidade ainda é limitada (Dal’Bello e Izawa, 2022).

3.3.6 Extreme Gradient Boosting (XGBoost)

O *Extreme Gradient Boosting*, conhecido como XGBoost, é um modelo de aprendizado de máquina que impulsiona a performance de modelos baseados em árvores de decisão, se tornando popular devido à sua alta eficiência computacional e precisão preditiva (Chen; Guestrin, 2016). Segundo os autores, o sistema roda mais de dez vezes do que outras soluções para escalar bilhões de exemplos. A Equação 18 apresenta a equação geral do modelo.

$$\mathcal{L}(\phi) = \sum_{i=1}^n l(Y_i + \hat{Y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (18)$$

Em que:

$L(Y_i, \hat{Y}_i^{(t)})$ é a função de perda que mede o erro entre o valor real e a predição no estágio t ; e

$\Omega(f_k)$ é o termo de regularização que penaliza árvores mais complexas.

O funcionamento do XGBoost baseia-se na construção sequencial de árvores de decisão, onde cada nova árvore corrige os erros das anteriores. A função objetivo geral do XGBoost combina uma função de perda, que mede o erro entre as previsões e os valores reais, com um termo de regularização que penaliza a complexidade do modelo para evitar *overfitting*, que é, segundo Xu, Coen-Pirani e Jiang (2022), definido como um modelo que se ajusta perfeitamente a um conjunto de dados específico, o que enfraquece sua capacidade de generalização e afeta a precisão da previsão de dados futuros.

De acordo com Teodoro, Monaci e Palagi (2023), o modelo geralmente é classificado como caixa preta, devido à dificuldade de interpretar a forma como o modelo realiza a previsão.

3.4 Procedimentos de Estimção e Avaliação dos Modelos

Para o desenvolvimento dos modelos preditivos, foi aplicado o método *Hold-out* por empresa, respeitando a dependência temporal dos dados. A amostra foi dividida em 70% para treinamento e 30% para teste, de forma estratificada por empresa, garantindo que cada companhia estivesse integralmente em apenas um dos conjuntos de treino ou teste e preservando a ordem cronológica das observações. Essa abordagem evita o vazamento de dados de mesma empresa em grupos diferentes e assegura maior validade empírica, simulando um cenário real de previsão.

As variáveis independentes passaram por uma winsorização a 5%, sendo 2,5% em cada cauda da distribuição, reduzindo o impacto de outliers extremos. Em seguida, foi aplicada uma padronização *Z-score* que transforma as variáveis para média zero e desvio padrão igual a um, sendo essencial para o desempenho de modelos sensíveis à escala, como o SVM, KNN e a MLP.

Além disso, considerando que o conjunto de dados apresenta desequilíbrio entre empresas solventes e insolventes, todos os algoritmos foram ajustados com pesos de classe balanceados, priorizando a correta identificação das empresas insolventes, minimizando o erro Tipo II (Falsos Negativos). O *cut-off* utilizado para todos os modelos é de 0,5.

O modelo DT utiliza o critério Gini para medir a qualidade das divisões dos nós por meio do poder de cada variável em reduzir as impurezas dos grupos, ou seja, medir a qualidade da separação dos grupos solventes e insolventes, buscando maximizar a pureza das decisões em cada ramificação. Com relação a profundidade da árvore, foi permitido que ela se expanda até que todas as folhas contenham menos de duas amostras, criando uma árvore que se ajusta aos dados de treinamento.

O modelo Random Forest (RF) foi construído com 600 árvores de decisão, empregando o mesmo critério Gini. Em cada árvore, amostras e subconjuntos de variáveis foram selecionados aleatoriamente, mitigando o risco de *overfitting* e aumentando a robustez do ensemble. O número elevado de árvores proporcionou estabilidade nos resultados sem comprometer a generalização.

Para o modelo SVM, foi empregado o critério de *Kernel Radial Basis Function*, adequado para capturar relações não lineares entre os indicadores financeiros. Esse kernel projeta os dados em um espaço de maior dimensão, permitindo encontrar um hiperplano ótimo de separação entre as classes. As probabilidades foram posteriormente calibradas, uma vez que o SVM não fornece saídas probabilísticas de forma nativa.

Com relação ao modelo KNN, este estudo implementou como parâmetro a análise dos 15 vizinhos mais próximos no conjunto de treinamento, atribuindo pesos inversamente proporcionais à distância de cada observação em relação ao ponto avaliado, de modo que vizinhos mais próximos exerçam maior influência na decisão de classificação. A distância entre as observações foi calculada pelo método de Minkowski, uma métrica geral que abrange outras formas amplamente conhecidas, como as distâncias euclidiana e Manhattan, permitindo ajustar o parâmetro p para controlar a ênfase em diferenças absolutas ou quadráticas entre as variáveis. De acordo com Aljawazneh *et al.* (2021), o método de Minkowski apresenta elevada flexibilidade para mensurar similaridades em conjuntos de dados financeiros, o que o torna

adequado para problemas de classificação binária como a previsão de insolvência, por permitir que o modelo se adapte às diferentes escalas e distribuições dos indicadores analisados.

O modelo MLP foi configurado com duas camadas ocultas, compostas por 64 e 32 neurônios, respectivamente, e função de ativação ReLU (*Rectified Linear Unit*), escolhida por sua capacidade de introduzir não linearidade e mitigar o problema do desvanecimento do gradiente, favorecendo a estabilidade e a eficiência do aprendizado. O processo de treinamento foi conduzido pelo otimizador *Adam*, com taxa de aprendizado adaptativa (10^{-3}), que ajusta dinamicamente o passo de atualização dos pesos conforme a taxa de erro, acelerando a convergência. Para controlar o sobreajuste, foi aplicada regularização L2 ($\alpha=0,001$), penalizando pesos excessivamente elevados e promovendo melhor capacidade de generalização. O número máximo de iterações foi definido em 2.000 iterações máximas, com o recurso de parada antecipada, interrompendo o treinamento quando não houvesse melhoria adicional no desempenho.

Com as variáveis independentes padronizadas pelo método *Z-score*, é assegurado que os neurônios pudessem receber entradas em mesma escala, o que é fundamental para modelos sensíveis à magnitude das variáveis. De acordo com Aljawazneh *et al.* (2021), o modelo de Redes Neuras de Múltiplas Camadas (MLP) objeto deste estudo, é composto por uma estrutura de múltiplas camadas interconectadas, sendo elas de entrada, ocultas e de saída, nas quais cada neurônio aplica uma transformação ponderada das variáveis, permitindo capturar relações não lineares complexas e padrões subjacentes entre os dados financeiros e o risco de insolvência.

O modelo XGBoost foi configurado com 800 árvores de decisão e profundidade máxima de quatro níveis, limitando a complexidade individual de cada árvore e reduzindo o risco de sobreajuste. A taxa de aprendizado foi definida em 0,05, controlando o peso com que os ajustes de cada árvore são incorporados ao modelo final, garantindo aprendizado gradual e estável ao longo das iterações.

Além disso, foram aplicadas regularizações L1 e L2, que penalizam modelos excessivamente complexos e evitam a superparametrização. O modelo também utilizou subamostragem de 80% das observações e das variáveis em cada iteração, aumentando a diversidade das árvores e melhorando a capacidade de generalização. Foi utilizado o parâmetro de pesos para ajustar conforme o desequilíbrio entre empresas solventes e insolventes, atribuindo maior peso à classe minoritária de empresas insolventes. De acordo com Kuhn e Johnson (2013), os algoritmos baseados em *boosting* aprimoram iterativamente o desempenho das árvores de decisão ao reduzir o erro de previsão via descida do gradiente, combinando simplicidade interpretativa com elevado poder preditivo, que são características que tornam o

XGBoost um dos métodos mais robustos e eficientes para problemas de classificação financeira.

Após o treinamento, os modelos DT, RF, XGBoost e KNN foram submetidos a um processo de calibração isotônica, com o objetivo de ajustar as probabilidades previstas para refletirem de forma mais realista a frequência observada de insolvência. De acordo com James *et al.* (2023), muitos classificadores supervisionados produzem escores de decisão que não correspondem diretamente à probabilidade real de ocorrência de um evento, sendo necessário aplicar métodos de calibração para alinhar as previsões às frequências empíricas observadas. Esse ajuste é particularmente importante em modelos cujas saídas representam estimativas relativas, e não probabilidades propriamente ditas, como ocorre com o KNN e o SVM.

Conforme destacam Kuhn e Johnson (2013), técnicas de calibração como a regressão isotônica e o ajuste sigmoide são amplamente utilizadas para converter escores contínuos em probabilidades calibradas, corrigindo distorções nas estimativas produzidas pelos modelos. O processo busca estabelecer uma função monotônica crescente que relacione os valores preditos à frequência observada do evento, produzindo probabilidades mais confiáveis e comparáveis entre diferentes classificadores. Assim, a calibração isotônica contribui na interpretação e aplicabilidade das previsões de insolvência, tornando os resultados mais robustos e representativos do risco real das empresas analisadas.

Com relação a medição do desempenho do modelo, a principal métrica de análise foi a Área Sob a Curva ROC (AUC), que quantifica a habilidade geral do modelo em distinguir corretamente entre as classes solvente e insolvente, em que um valor de 1 representa um classificador perfeito e 0,5 um classificador aleatório.

Adicionalmente, avaliou-se o *F1-Score*, que representa a média harmônica entre a Precisão e a Sensibilidade, sendo útil para cenários onde o equilíbrio entre falsos positivos e falsos negativos é importante. A análise foi complementada pela Sensibilidade (*recall*), que mede a proporção de empresas insolventes que o modelo conseguiu identificar corretamente, e pela Especificidade, que mensura a proporção de empresas solventes corretamente classificadas.

Todas essas métricas foram calculadas a partir da Matriz de Confusão, que sumariza os acertos e erros das previsões. Os componentes da matriz são os Verdadeiros Positivos (VP), que representam as empresas insolventes corretamente previstas; os Falsos Positivos (FP), empresas solventes incorretamente classificadas como insolventes (Erro Tipo I) os Verdadeiros Negativos (VN), empresas solventes corretamente identificadas; e os Falsos Negativos (FN), que são as empresas insolventes que o modelo incorretamente classificou como solventes (Erro

Tipo II). A Precisão do modelo, por sua vez, foi calculada para indicar, dentre todas as empresas classificadas como insolventes, à proporção que de fato era.

Por fim, foi utilizada a biblioteca *Permutation Importance*, com o objetivo de identificar quais variáveis independentes mais influenciam o desempenho de cada modelo construído, permitindo verificar o impacto na AUC decorrente da permuta (embaralhamento) de determinada variável independente do modelo. Neste estudo, foram apresentadas as cinco principais variáveis de cada modelo que apresentaram maior relevância em sua construção, sendo realizada a interpretação a partir do conjunto desses indicadores financeiros identificados.

4 RESULTADOS E DISCUSSÃO

4.1 Análise de multicolinearidade e das estatísticas descritivas

A Tabela 27 apresenta os resultados do teste VIF para as variáveis selecionadas, evidenciando a ausência de multicolinearidade severa, o que permite analisar a partir dos modelos construídos a importância das variáveis distintas que, em conjunto, influenciaram na determinação de insolvência empresarial.

Tabela 27 – VIF das variáveis financeiras (continua).

Nome	Indicador Financeiro	VIF
X2	$\frac{\text{Capital de Giro}}{\text{Receita Líquida}}$	3,28
X4	$\frac{\text{Necessidade de Capital de Giro}}{\text{Receita Líquida}}$	2,24
X6	$\frac{\text{Saldo de Tesouraria}}{\text{Receita Líquida}}$	2,38
X7	$\frac{\text{Passivo Financeiro}}{\text{Ativo Circulante}}$	3,16
X9	$\frac{\text{Saldo de Tesouraria}}{\text{Necessidade de Capital de Giro}}$	1,11
X11	<i>Necessidade de Capital de Giro</i>	1,52
X12	<i>Capital de Giro</i>	1,45
X14	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$	2,55
X15	$\frac{\text{Receita Líquida Operacional}}{\text{Ativo Total}}$	4,18
X16	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Ativo Total}}$	2,50
X19	$\frac{\text{Disponibilidade}}{\text{Ativo Circulante}}$ + <i>Realizável a Longo Prazo</i>	2,09
X20	$\frac{\text{Estoque}}{\text{Ativo Total}}$	2,41
X21	$\frac{\text{Fornecedores}}{\text{Ativo Total}}$	2,85
X27	$\frac{\text{Financiamentos e Empréstimos Bancários}}{\text{Ativo Circulante}}$	4,59
X30	$\frac{\text{Patrimônio Líquido}}{\text{Capital Social}}$ – $\frac{\text{Capital Social}}{\text{Ativo Total}}$	2,08

Tabela 27 – VIF dos indicadores financeiros (conclusão).

Nome	Indicador Financeiro	VIF
X33	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Impostos sobre Despesas Financeiras}}$	1,83
X34	$\frac{\text{Lucro Antes dos Juros e Impostos}}{\text{Dívida Líquida}}$	1,12

Fonte: Elaborado pelo autor (2025).

Na sequência, a caracterização da amostra e das variáveis de pesquisa é apresentada na Tabela 28 por meio de uma análise estatística descritiva antes da transformação dos dados. Foram calculadas as medidas de tendência central (média), de dispersão (mínimo, máximo, desvio padrão e variância) e de formato da distribuição (assimetria e curtose) para cada variável do modelo.

Tabela 28 – Estatísticas descritivas dos indicadores financeiros.

Variável	Média	Mínimo	Máximo	D. Padrão	Variância	Assimetria	Curtose
X2	-0,081	-6,175	2,918	1,297	1,682	-2,569	11,658
X4	0,368	-1,222	4,460	0,916	0,840	2,867	9,809
X6	-3,983E-07	-9,714E-06	8,716E-07	1,686E-06	2,843E-12	-4,810	22,873
X7	0,217	0,000	2,035	0,396	0,157	3,166	10,735
X9	-0,251	-10,852	12,230	3,319	11,013	0,718	6,680
X11	1,766E+06	-5,896E+05	1,940E+07	4,126E+06	1,702E+13	3,063	8,995
X12	-9,244E+04	-1,078E+07	7,619E+06	2,827E+06	7,995E+12	-1,224	5,944
X14	1,340	1,00e-06	9,427	1,829	3,346	2,696	8,621
X15	0,411	1,00e-06	1,727	0,463	0,214	1,073	0,429
X16	0,018	-0,644	0,253	0,144	0,021	-2,750	10,536
X19	0,118	1,00e-06	0,771	0,178	0,032	2,039	3,996
X20	0,054	1,00e-06	0,310	0,086	0,007	1,537	1,227
X21	0,051	1,00e-06	0,326	0,074	0,005	2,093	4,358
X27	0,836	1,00e-06	7,265	1,432	2,050	2,982	9,680
X30	-0,539	-14,183	0,467	2,515	6,326	-4,709	21,644
X33	1,335	-7,111	18,102	3,897	15,188	2,334	8,367
X34	0,186	-5,913	7,447	1,808	3,271	0,823	8,516

Fonte: Elaborado pelo autor (2025).

A análise exploratória dos dados revela uma heterogeneidade expressiva entre as variáveis, caracterizada por alta dispersão das médias e distribuições com assimetria e curtose acentuadas, indicando a presença de caudas pesadas. Este perfil estatístico, que reflete a diversidade estrutural e de desempenho entre as empresas da amostra, é plenamente compatível com a abordagem baseada em aprendizado de máquina.

Diferentemente de técnicas estatísticas tradicionais, os algoritmos como *Decision Tree*, *Random Forest* e *XGBoost* não exigem normalidade dos dados, enquanto a padronização aplicada previamente assegura o bom desempenho de modelos como SVM, KNN e Redes Neurais. Portanto, a variabilidade observada é tratada como uma fonte valiosa de informação, permitindo que os modelos aprendam os padrões complexos que distinguem empresas solventes de insolventes.

4.2 Análise da construção dos modelos

Nesta seção, são apresentados os resultados do desenvolvimento dos modelos DT, RF, SVM, KNN, MLP e XGBoost. Para o desenvolvimento do modelo, foram divididos os dados para treino e teste, sendo utilizadas 121 observações para treino, correspondendo 70% da amostra, e 53 observações para teste (30% da amostra), verificando o aprendizado do modelo para novos conjuntos de dados. O desempenho foi mensurado por um conjunto de métricas, com foco na área sob a Curva ROC (AUC) como principal indicador de classificação de empresas solventes e insolventes no conjunto de dados de teste.

Nesse sentido, a Tabela 29 evidencia o desempenho consolidado de todos os modelos construídos neste estudo na amostra de teste, permitindo a comparação direta das performances por meio das métricas de avaliação.

Tabela 29 – Comparativo do desempenho dos modelos de aprendizado de máquina.

Modelo	AUC	Acurácia	Precisão	Sensibilidade	Especificidade	F1-Score	VP	FP	VN	FN
XGBoost	0,981	0,954	0,811	0,769	0,977	0,789	133	31	1342	40
RF	0,978	0,952	0,773	0,809	0,970	0,791	140	41	1332	33
SVM	0,959	0,917	0,590	0,85	0,926	0,697	147	102	1271	26
DT	0,956	0,946	0,816	0,665	0,981	0,732	115	26	1347	58
MLP	0,951	0,936	0,680	0,809	0,952	0,739	140	66	1307	33
KNN	0,888	0,914	0,670	0,457	0,972	0,543	79	39	1334	94

Fonte: Elaborado pelo autor (2025).

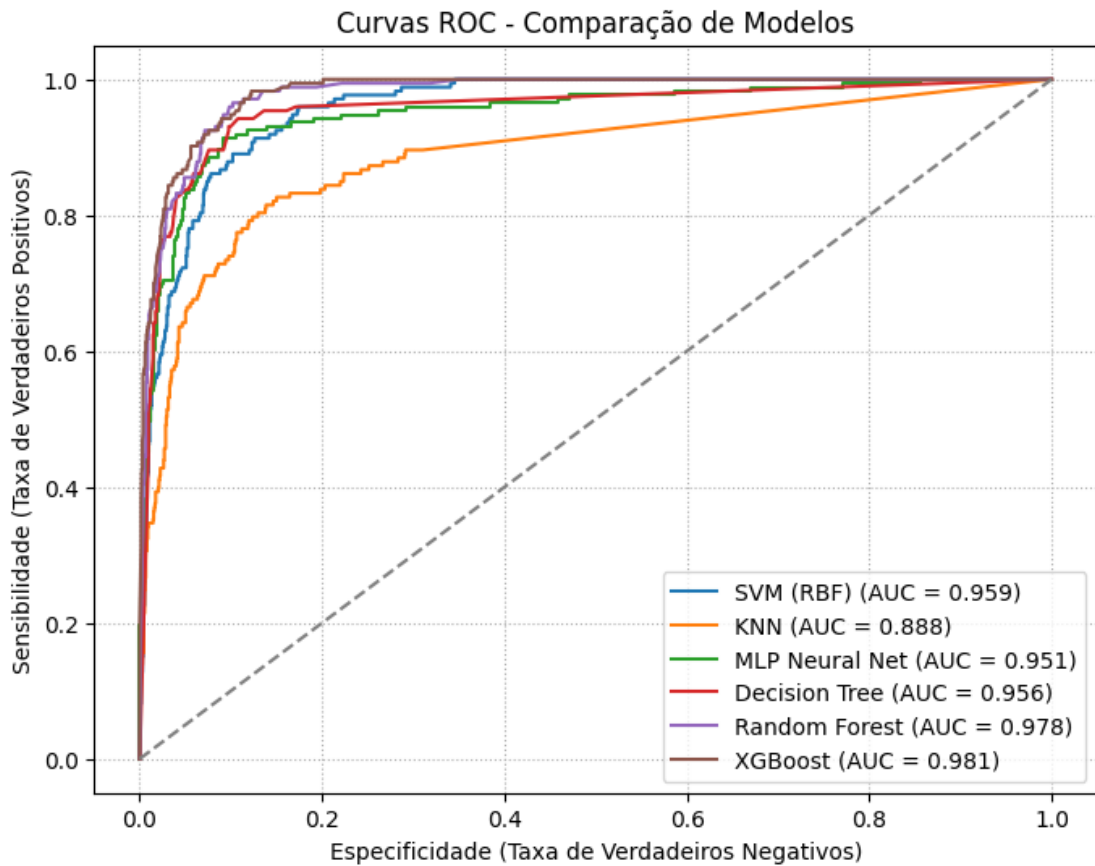
Os resultados apresentados na Tabela 29 mostram que o modelo XGBoost demonstrou melhor performance global, alcançando o maior valor de acurácia, 95,4%, e de AUC, 0,981, evidenciando sua superioridade em classificar tanto empresas solventes quanto insolventes.

De forma complementar, os modelos RF, SVM e DT apresentaram ótimo desempenho, com AUC de 0,978, 0,959 e 0,956, respectivamente. Ainda que valores elevados, a acurácia geral desses modelos foi inferior ligeiramente inferior ao XGBoost. Esses resultados

confirmam os achados na literatura como nos estudos Barboza, Kimura e Altman (2017), Aljawazneh *et al.* (2021), Pinto (2021), Santos (2021), Garcia (2022), Jesus e Bessaria (2023), Noh (2023) e Penedo e Barboza (2024) ao verificar que modelos *ensemble* que são construídos a partir de árvores de decisão como XGBoost, RF. Já os modelos MLP e SVM são eficazes conforme demonstra a literatura por Prado *et al.* (2020), Idhmad *et al.* (2024) e Tavares *et al.* (2025) para classificação binária na temática do risco de crédito empresarial.

Por outro lado, modelos de menor complexidade, como DT e KNN, obtiveram desempenho satisfatório, mas apresentaram métricas inferiores às técnicas mais robustas, como evidenciada a queda na sensibilidade (0,665 e 0,457, respectivamente), evidenciando maior limitação na captura de relações não lineares entre as variáveis. O XGBoost, após calibração, destacou-se como o modelo de melhor desempenho, alcançando AUC de 0,981 e acurácia de 95,4%, superando ligeiramente o Random Forest (AUC de 0,978 e acurácia de 95,2%). Esses resultados confirmam a elevada capacidade do XGBoost em aprender padrões complexos e otimizar o processo de classificação, especialmente em bases de dados com múltiplas dimensões e interações entre variáveis. De forma geral, os achados reforçam que, embora algoritmos de *ensemble* tradicionais, como o RF, continuem altamente eficientes, modelos mais avançados como o XGBoost podem oferecer ganhos marginais de desempenho quando bem calibrados e ajustados aos dados.

De forma complementar, para uma análise visual do desempenho dos modelos construídos neste estudo, A Figura 18 apresenta a Curva ROC de todos os modelos.



Fonte: Elaborado pelo autor (2025).

A análise visual da Figura 18 corrobora os resultados quantitativos. A curva do modelo XGBoost calibrado se posiciona mais próxima do canto superior esquerdo, confirmando seu desempenho ligeiramente superior em relação ao Random Forest. Nota-se também uma grande proximidade entre as curvas dos modelos MLP, SVM e Random Forest, que igualmente demonstram excelente poder classificatório. As curvas da DT e do KNN, embora revelem uma boa performance, situam-se em um patamar inferior, coerente com os valores de AUC observados na Tabela 29.

Em síntese, os resultados indicam que todos os modelos testados apresentaram bom desempenho preditivo, com destaque para o XGBoost, que superou marginalmente as demais técnicas nas principais métricas. Ainda que SVM, MLP e RF tenham demonstrado alta capacidade de classificação de empresas solventes e insolventes, o XGBoost confirmou-se como o modelo mais eficiente e consistente para a previsão neste estudo. Tais evidências reforçam a importância da comparação entre diferentes algoritmos, possibilitando identificar aquele que melhor se ajusta à natureza dos dados analisados.

Além de avaliar a acurácia preditiva, este estudo investigou quais variáveis exerceram maior influência na previsão e classificação de empresas solventes e insolventes. Para isso, foi

extraída a importância das variáveis por meio da biblioteca *Permutation Importance*, avaliando a perda de AUC após embaralhar a variável, sendo as que obtêm maiores quedas são consideradas as mais importantes para os modelos construídos. A Tabela 30 apresenta os indicadores financeiros mais relevantes.

Tabela 30 – Importância das cinco principais variáveis.

Ordenação	DT	KNN	MLP	RF	SVM	XGBoost
	X30	X19	X30	X30	X30	X30
1ª Variável	(PL-CS/AT)	(D/AC+RLP)	(PL-CS/AT)	(PL-CS/AT)	(PL-CS/AT)	(PL-CS/AT)
	0,3667	0,0427	0,1106	0,1637	0,0805	0,1808
	X14	X6	X14	X14	X19	X14
2ª Variável	(AC/PC)	(T/RecL)	(AC/PC)	(AC/PC)	(D/AC+RLP)	(AC/PC)
	0,0590	0,0293	0,0242	0,0104	0,0251	0,0159
	X19	X30	X19	X19	X14	X19
3ª Variável	(D/AC+RLP)	(PL-CS/AT)	(D/AC+RLP)	(D/AC+RLP)	(AC/PC)	(D/AC+RLP)
	0,0300	0,0269	0,0228	0,0030	0,0216	0,0058
	X12	X14	X2	X12	X33	X12
4ª Variável	(CDG)	(AC/PC)	(CDG/RecL)	(CDG)	(EBIT/ImpDF)	(CDG)
	0,0282	0,0256	0,0195	0,0027	0,0116	0,0042
	X15	X20	X20	X6	X6	X2
5ª Variável	(RLO/AT)	(Estoque/AT)	(Estoque/AT)	(T/RecL)	(T/RecL)	(CDG/RecL)
	0,0108	0,0211	0,0130	0,0007	0,0097	0,0021

Fonte: Elaborado pelo autor (2025).

A Tabela 30 evidencia a consistência de alguns indicadores-chave na explicação do risco de insolvência. Observa-se que o indicador X30 (Patrimônio Líquido menos o Capital Social sobre o Ativo Total), foi o mais importante em quase todos os modelos, representando a solidez patrimonial e a capacidade da reserva de lucros como fonte de financiamento próprio. Em seguida, os indicadores X14 (Ativo Circulante sobre Passivo Circulante) e X19 (Disponibilidade sobre o Ativo Circulante e Realizável a Longo Prazo) também se destacaram nos modelos, reforçando a importância da liquidez imediata na capacidade das empresas de pagar suas dívidas.

O indicador X12 (Capital de Giro) e o X6 (Saldo de Tesouraria sobre Receita Líquida) surgem com relevância em modelos como DT, KNN, RF, SVM e XGBoost, refletindo a gestão operacional de recursos de curto e longo prazo e o equilíbrio entre fontes e aplicações de capital de giro. Por fim, a variável X20 (Estoque sobre Ativo Total) demonstra o papel do ciclo operacional e da eficiência no uso dos ativos como determinantes da sustentabilidade financeira. Em conjunto, os resultados indicam que os modelos priorizam dimensões de

estrutura patrimonial, liquidez e eficiência operacional, convergindo na importância da geração interna de recursos e na gestão equilibrada entre capital próprio e de terceiros como variáveis explicativas centrais na previsão do risco de insolvência.

Com isso, os resultados obtidos nesta pesquisa confirmam a literatura recente, ao destacar de forma recorrente a centralidade de indicadores de rentabilidade, alavancagem e liquidez na previsão de insolvência. Estudos como os de Prado *et al.* (2020) e Santos (2021) já apontavam a relevância de métricas de rentabilidade e de estrutura de capital, enquanto Pinto (2021), ao aplicar técnicas de interpretabilidade em modelos baseados em XGBoost, evidenciou a importância do Patrimônio Líquido sobre o Exigível Total, da Liquidez Seca e da Necessidade de Capital de Giro. De forma complementar, Cunha, Penedo e Barboza (2024) confirmaram, no contexto brasileiro, que variáveis ligadas ao crescimento de ativos, liquidez e margem operacional eram as mais explicativas, e Costa (2023) demonstrou que ROE (Retorno sobre o Patrimônio Líquido), alavancagem e liquidez corrente se destacam entre os métodos de aprendizado de máquina.

Desta forma, a contribuição do presente estudo reforça essa convergência ao evidenciar, em uma base de dados ampla e recente de empresas brasileiras listadas na B3, que a interação entre rentabilidade, estrutura de capital e liquidez constitui o núcleo mais robusto para explicar a insolvência, bem com a integração de indicadores financeiros tradicionais e os do modelo dinâmico de Fleuriet. Nesse contexto, o modelo XGBoost destacou-se entre os demais métodos de aprendizado de máquina por apresentar o maior poder discriminante e a melhor capacidade de generalização na previsão do risco de insolvência. Essa técnica, baseada em um processo iterativo de construção sequencial de árvores de decisão ponderadas, aprimora continuamente o desempenho preditivo ao corrigir os erros das iterações anteriores, o que permite capturar relações não lineares e interações complexas entre os indicadores financeiros.

A métrica de Gini e a área sob a curva ROC (AUC) obtidas pelo XGBoost demonstraram que os indicadores associados à estrutura patrimonial, à liquidez corrente, e à eficiência operacional foram os mais determinantes para a classificação das empresas solventes e insolventes. Assim, além de confirmar achados anteriores, este estudo avança ao evidenciar empiricamente que o XGBoost é capaz de integrar de forma mais eficiente múltiplas dimensões financeiras, oferecendo uma visão abrangente e acurada da sustentabilidade econômico-financeira das empresas, com elevada aplicabilidade para gestores, credores e investidores.

Com isso, os resultados deste estudo reforçam a contribuição ao demonstrar o *trade-off* entre interpretabilidade e desempenho preditivo nos modelos de aprendizado de máquina aplicados à previsão de insolvência. Modelos como a DT e KNN oferecem maior transparência

e facilidade de compreensão das regras de decisão, sendo considerados de caixa branca, uma vez que permitem rastrear diretamente os critérios utilizados na classificação. Contudo, a acurácia obtida por modelos mais complexos, como o RF e o XGBoost, supera a dos modelos mais interpretáveis, ainda que sua estrutura de decisão seja menos acessível, caracterizando-os como modelos de caixa preta. Nesse contexto, a análise de importância das variáveis e os métodos de interpretabilidade global atuam como mecanismos mitigadores dessa limitação, permitindo identificar quais variáveis mais contribuíram para a precisão dos modelos, equilibrando o desempenho preditivo elevado com uma compreensão mais clara das variáveis determinantes da insolvência empresarial.

A identificação dos indicadores financeiros mais relevantes extrapola o campo teórico e oferece aplicações práticas diretas tanto para a gestão financeira quanto para a análise de crédito. A predominância das variáveis relacionadas à estrutura patrimonial (X30), à liquidez corrente (X14) e à eficiência operacional (X15) indica que a solidez do capital próprio, a capacidade de pagamento de curto prazo e o uso produtivo dos ativos constituem o núcleo central para compreender o risco de insolvência. Na prática, esse achado reforça que um painel gerencial baseado nesses três pilares (capital próprio, liquidez e eficiência operacional) pode funcionar como um sistema de alerta precoce altamente eficaz. Para analistas de crédito, a utilização de modelos como o XGBoost calibrado permite não apenas estimar a probabilidade de insolvência com alto grau de precisão, mas também justificar essa probabilidade com base em variáveis financeiramente interpretáveis, tornando o processo de avaliação de risco mais transparente, técnico e sustentado por evidências empíricas.

Embora a análise da importância das variáveis apresentada na Tabela 27 permita identificar os indicadores financeiros mais relevantes para a classificação das empresas solventes e insolventes, os resultados observados não decorrem exclusivamente da seleção das variáveis explicativas, mas também das características metodológicas inerentes a cada técnica de aprendizado de máquina empregada. Assim, a compreensão dos desempenhos obtidos requer uma análise complementar das propriedades dos modelos e da forma como cada abordagem processa e combina dos dados da amostra para realizar a previsão do risco de insolvência empresarial.

4.2.1 Interpretação dos modelos desenvolvidos

A análise comparativa dos modelos de aprendizado de máquina evidencia que as diferenças de desempenho preditivo estão associadas não apenas às variáveis utilizadas, mas

também às distintas capacidades dos algoritmos em lidar com padrões não lineares, heterogeneidade dos dados e estabilidade das estimativas. Conforme destacado na literatura de aprendizado estatístico aplicada, modelos diferem quanto ao equilíbrio entre flexibilidade, interpretabilidade e capacidade de generalização, aspectos particularmente relevantes em problemas complexos de previsão financeira, nos quais a insolvência empresarial resulta da interação simultânea de múltiplos fatores econômicos e contábeis (Kuhn; Johnson, 2013; James et al., 2023).

No caso do DT, o desempenho inferior observado pode ser explicado por sua elevada sensibilidade à variabilidade dos dados e à heterogeneidade estrutural dos indicadores financeiros. Embora as variáveis tenham sido previamente selecionadas de modo a mitigar problemas de multicolinearidade linear, o modelo permanece suscetível a relações não lineares, interações complexas e ruídos econômicos característicos de bases contábeis amplas. Por se tratar de uma única estrutura de decisão, pequenas variações na amostra de treinamento podem resultar em mudanças substanciais na árvore estimada, o que compromete sua capacidade de generalização em problemas complexos como a previsão do risco de insolvência.

O modelo RF apresentou desempenho substancialmente superior ao modelo de árvore de decisão individual, evidenciando as vantagens da combinação de múltiplas árvores de decisão na construção do modelo preditivo. Ao integrar diversas estruturas de decisão independentes, o RF reduz a instabilidade típica de árvores únicas e aumenta a robustez frente a ruídos e variações nos dados financeiros. Essa característica é especialmente relevante em bases contábeis amplas e heterogêneas, nas quais a interação entre indicadores de liquidez, rentabilidade e estrutura de capital tende a ser complexa. Além disso, a consistência observada na identificação das variáveis mais importantes reforça a capacidade do modelo de capturar padrões econômicos relevantes associados ao risco de insolvência.

O desempenho intermediário do SVM pode ser atribuído à sua dificuldade em representar adequadamente fenômenos econômicos de natureza gradual, como a insolvência empresarial. Em bases financeiras, é comum a sobreposição entre empresas solventes e insolventes, o que limita a eficiência de modelos baseados em fronteiras rígidas de separação. Ademais, a menor interpretabilidade do SVM restringe a análise econômica direta das relações entre variáveis e a probabilidade de insolvência.

O modelo KNN apresentou desempenho inferior em comparação às técnicas ensemble, o que pode ser explicado por sua elevada sensibilidade à métrica de distância e ao problema da alta dimensionalidade. Mesmo após a padronização das variáveis, a diluição das distâncias em espaços multivariados dificulta a identificação de padrões econômicos latentes, tornando o

KNN menos adequado para capturar a complexidade estrutural do risco de insolvência em bases financeiras extensas e heterogêneas.

O MLP demonstrou desempenho competitivo, refletindo sua capacidade de modelar relações não lineares entre os indicadores financeiros. Contudo, a necessidade de maior calibração dos hiperparâmetros e o risco de sobreajuste podem limitar sua eficiência em amostras caracterizadas por elevada heterogeneidade econômica. Além disso, a menor transparência do processo de aprendizagem dificulta a interpretação econômica direta das saídas do modelo, especialmente quando comparado a abordagens baseadas em estruturas de decisão.

Por fim, o XGBoost apresentou o melhor desempenho preditivo entre os modelos avaliados, evidenciando a superioridade das técnicas ensemble na previsão do risco de insolvência empresarial. Sua abordagem sequencial permite capturar interações complexas e não lineares entre os indicadores financeiros, corrigindo progressivamente os erros das estimativas anteriores. A convergência entre as variáveis mais importantes identificadas e aquelas tradicionalmente associadas à solvência empresarial reforça que a superioridade do XGBoost decorre de sua capacidade estrutural de aprendizado e generalização, e não de efeitos aleatórios da amostra.

De forma integrada, os resultados indicam que modelos baseados na combinação de múltiplas estruturas decisórias apresentam maior capacidade de capturar a complexidade do risco de insolvência empresarial em comparação aos modelos individuais. Essa evidência reforça a adequação dessas técnicas para aplicações preditivas em finanças corporativas, ao mesmo tempo em que confirma a relevância econômica dos indicadores financeiros selecionados para a antecipação de situações de deterioração financeira.

5 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo avaliar a efetividade comparativa de diferentes técnicas de aprendizado de máquina, sendo eles *Random Forest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree* (DT), *Multilayer Perceptrons* (MLP) e *Extreme Gradient Boosting* (XGBoost), na previsão do risco de insolvência de empresas não financeiras listadas na B3, considerando indicadores financeiros tradicionais e dinâmicos.

Os resultados evidenciaram que todos os modelos apresentaram desempenho satisfatório na classificação de empresas solventes e insolventes, com destaque para o XGBoost, que obteve acurácia de 95,4% e AUC de 0,981, confirmando sua robustez preditiva e superioridade frente às demais técnicas. Na sequência, RF, SVM e MLP, apresentaram métricas elevadas, embora ligeiramente inferiores, reforçando que métodos de maior complexidade nem sempre superam algoritmos ensemble tradicionais. Os modelos DT e KNN, apesar de satisfatórios, mostraram menor capacidade de capturar padrões não lineares, corroborando suas limitações em cenários complexos.

Além da performance preditiva, uma contribuição central deste estudo foi a identificação das principais variáveis para a previsão do risco de insolvência empresarial. A análise de importância revelou que os indicadores relacionados à estrutura patrimonial (X30), à liquidez corrente (X14), à eficiência operacional (X15), ao capital de giro (X12) e à disponibilidade imediata (X21) constituem as variáveis mais importantes para explicar o comportamento financeiro das empresas. Esses resultados reforçam que a combinação entre solidez do capital próprio, capacidade de pagamento e eficiência na gestão dos recursos operacionais forma o núcleo mais robusto de predição de insolvência. Assim, os achados não apenas confirmam pressupostos da teoria financeira, mas também demonstram a utilidade prática dos modelos de aprendizado de máquina como ferramentas de apoio à decisão gerencial, análise de crédito e monitoramento de risco, ampliando a capacidade de previsão e de intervenção preventiva em contextos corporativos.

Como toda pesquisa empírica, este trabalho apresenta limitações. A principal refere-se à utilização exclusiva de dados de empresas brasileiras não financeiras listadas na B3, o que restringe a generalização dos achados para outros contextos econômicos e institucionais. Além disso, a base foi composta unicamente por indicadores financeiros, não contemplando variáveis macroeconômicas ou qualitativas, como fatores de governança corporativa, aspectos regulatórios ou elementos de comportamento gerencial, que poderiam enriquecer a explicação dos modelos, bem como a ausência de métricas diretamente associadas ao fluxo de caixa pode

limitar a identificação de situações de insolvência iminente, uma vez que a incapacidade de pagamento se manifesta, em última instância, no regime de caixa e não apenas nos resultados apurados pelo regime de competência.

Diante dessas limitações, recomenda-se que futuras pesquisas ampliem o escopo da análise, seja por meio da utilização de bases de dados de outros países para fins comparativos, seja pela inclusão de variáveis macroeconômicas, qualitativas ou setoriais que possam impactar a insolvência. Sugere-se ainda a aplicação de modelos mais recentes de aprendizado de máquina e *deep learning*.

Em síntese, conclui-se que os resultados obtidos demonstram que o aprendizado de máquina constitui ferramenta valiosa para a previsão de insolvência no contexto brasileiro, oferecendo suporte relevante para gestores, investidores e formuladores de políticas. Contudo, a evolução da pesquisa na área exige o aprofundamento de metodologias e variáveis analisadas, a fim de consolidar modelos mais precisos, adaptáveis e aplicáveis a diferentes realidades empresariais.

REFERÊNCIAS

ALJAWAZNEH, H., MORA, A. M., GARCIA-SANCHEZ, P.; CASTILLO-VALDIVIESO, P. A. **Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure**. IEEE Access, v. 9, p. 97010–97038, 2021. DOI: <https://doi.org/10.1109/ACCESS.2021.3093461>.

ALMEIDA, L. **Risk and Bankruptcy Research: Mapping the State of the Art**. *Journal of Risk and Financial Management*, v. 16, n. 8, p. 361, 2023. DOI: <https://doi.org/10.3390/jrfm16080361>.

ALTMAN, E. I. **Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy**. *The Journal of Finance*, v. 23, n. 4, p. 589–609, 1968. DOI: <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>.

ALTMAN, E. I., BAIDYA, T. K.; DIAS, L. M. R. **Previsão de problemas financeiros em empresas**. *Revista de administração de empresas*, v. 19, p. 17-28, 1979. DOI: <https://doi.org/10.1590/S0034-75901979000100002>.

BALCAN, M.-F.; SHARMA, D. **Learning Accurate and Interpretable Decision Trees**. No. arXiv, v. 240, p. 5, 2024. DOI: <https://doi.org/10.48550/arXiv.2405.15911>.

BARBOZA, F.; KIMURA, H.; ALTMAN, E. (2017). **Machine learning models and bankruptcy prediction**. *Expert Systems with Applications*, v. 83, p. 405–417. DOI: <https://doi.org/10.1016/j.eswa.2017.04.006>.

BARCELÓ, P., KOZACHINSKIY, A., ORTH, M. R., SUBERCASEAUX, B.; VERSCHAE, J. **Explaining k-Nearest Neighbors: Abductive and Counterfactual Explanations**. No. arXiv, v. 250, p. 1, 2025. DOI: <https://doi.org/10.48550/arXiv.2501.06078>.

BERSIMAS, D.; KING, A.; NATHAN, A. **Statistical Learning via Mixed Integer Optimization: A Modern Optimization Lens on Statistical Learning**. *Journal of Machine Learning Research*, v. 23, p. 1-61, 2022. DOI: <https://doi.org/10.1007/s11071-022-08178-9>.

BILLIOS, D., SERETIDOU, D.; STAVROPOULOS, A. **The Power of Numerical Indicators in Predicting Bankruptcy: A Systematic Review**. *Journal of Risk and Financial Management*, v. 17, n. 10, p. 433, 2024. DOI: <https://doi.org/10.3390/jrfm17100433>.

BOSSE, T.; HOOGENDOORN, M. **Special issue on advances in applied artificial intelligence**. *Applied Intelligence*, v. 42, n. 1, p. 1–2, 2015. DOI: <https://doi.org/10.1007/s10489-014-0588-z>.

BREIMAN, L. **Random Forests**. *Machine Learning*, 45, 5–32, 2001.

CHEN, T.; GUESTRIN, C. **XGBoost: A Scalable Tree Boosting System**. *KDD: Knowledge Discovery and Data Mining*, v. 16, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794, 2016. DOI: <https://doi.org/10.1145/2939672.2939785>.

CORTES, C.; VAPNIK, V. **Support-vector networks**. *Machine Learning*, v. 20, n. 3, p. 273–297, 1995. DOI: <https://doi.org/10.1007/BF00994018>.

COSTA, T. R. D. **Risco De Insolvência: Uma Análise de Diferentes Métodos de Previsão e do Efeito da Estrutura de Capital**. Dissertação de Mestrado. Universidade Federal de Lavras., 2023. Disponível em: <http://repositorio.ufla.br/jspui/handle/1/57037>. Acesso em: 30 de abril de 2025.

CUNHA, M. A.; PENEDO, A. S. T.; BARBOZA, F. L. M. **Predicting Insolvency In Brazilian Companies Using Machine Learning: Na Anlalysis of the Random Forest Technique**. 2024. DOI: <https://doi.org/10.18624/etech.v17i1.1342>.

DAL'BELLO, L. R.; IZAWA, J. **Computational role of exploration noise in error-based de novo motor learning**. *Neural Networks*, v. 153, p. 349–372, 2022. DOI: <https://doi.org/10.1016/j.neunet.2022.06.011>.

DASILAS, A.; RIGANI, A. **Machine learning techniques in bankruptcy prediction: A systematic literature review**. *Expert Systems with Applications*, v. 255, p. 124761, 2024. DOI: <https://doi.org/10.1016/j.eswa.2024.124761>.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata**. Rio de Janeiro: Elsevier, 2017.

FIX, E.; HODGES, J. L. **Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties**. *International Statistical Review / Revue Internationale de Statistique*, v. 57, n. 3, p. 238, 1989. DOI: <https://doi.org/10.2307/1403797>.

FLEURIET, M. **A Dinâmica Financeira das Empresas Brasileiras**. 1. ed. 2003.

FLEURIET, M; KEHDYE, R; BLANC, G. **A Dinâmica Financeira das Empresas Brasileiras: Um Novo Método de Análise, Orçamento e Planejamento Financeiro**. 1. ed. 1978.

GARCIA, J. **Bankruptcy prediction using synthetic sampling**. *Machine Learning with Applications*, v. 9, p. 100343, 2022. DOI: <https://doi.org/10.1016/j.mlwa.2022.100343>.

GÓIS, Alan Diógenes; DE LUCA, Márcia Martins Mendes; DE LIMA, Gerlando Augusto Sampaio Franco; MEDEIROS, Jislene Trindade. **Corporate reputation and bankruptcy risk**. *Brazilian Administration Review*, Rio de Janeiro, v. 17, n. 2, e180159, 2020. DOI: <https://doi.org/10.1590/1807-7692bar2020180159>.

GOOGLE. **Google Colaboratory**. 2025. Disponível em: <https://colab.research.google.com>.

GRAU, I.; SENGUPTA, D.; LORENZO, M. M. G.; NOWE, A. **Grey-box model: an ensemble approach for addressing semi-supervised classification problems**. *Proceedings of Benelearn 2016*, p. 1-3, 2016.

HAMEL, L. **Visualization of Support Vector Machines with Unsupervised Learning**. v. 200, p. 6, 2006. DOI: <https://doi.org/10.1109/CIBCB.2006.330984>.

HAMMER, B., MOKBEL, B., SCHLEIF, F.-M.; ZHU, X. **White Box Classification of Dissimilarity Data**. Em E. Corchado, V. Snášel, A. Abraham, M. Woźniak, M. Graña, & S.-B. Cho (Orgs.), *Hybrid Artificial Intelligent Systems* (Vol., v. 720, p. 8, 2012. DOI: https://doi.org/10.1007/978-3-642-28942-2_28).

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900 p.

IDHMAD, A., KAICER, M., NEJJAR, C.; BENJOUAD, A. **Intelligent Bankruptcy Prediction Models Involving Corporate Governance Indicators, Financial Ratios and SMOTE**. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, v. 12, n. 1, p. 233–244, 2024. DOI: <https://doi.org/10.52549/ijeei.v12i1.5241>.

IZZA, Y., IGNATIEV, A.; MARQUES-SILVA, J. **On Explaining Decision Trees** (No. arXiv, v. 201, 2020. DOI: <https://doi.org/10.48550/arXiv.2010.11034>).

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An introduction to statistical learning: with applications in Python**. Nova York: Springer, 2023. DOI: <https://doi.org/10.1007/978-3-031-38747-0>.

JESUS, D. P.; BESARRIA, C. D. N. **Machine learning and sentiment analysis: Projecting bank insolvency risk**. *Research in Economics*, v. 77, n. 2, p. 226–238, 2023. DOI: <https://doi.org/10.1016/j.rie.2023.03.001>.

JOACHIMS, T. **Learning to Classify Text Using Support Vector Machines**. Springer US., 2002. DOI: <https://doi.org/10.1007/978-1-4615-0907-3>.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. New York: Springer, 2013. DOI: <https://doi.org/10.1007/978-1-4614-6849-3>.

KHUSEIN, A. **A Survey on Design, Applications and Limitations of Computational Intelligence**. *Journal of Computing and Natural Science*, p. 124, 2023. DOI: <https://doi.org/10.53759/181X/JCNS202303012>.

MCCULLOCH, W. S.; PITTS, W. **A logical calculus of the ideas immanent in nervous activity**. *The bulletin of mathematical biophysics*, v. 5, p. 115-133, 1943. DOI: <https://doi.org/10.1007/BF02478259>.

MOREIRA, C., CHOU, Y.-L., HSIEH, C., OUYANG, C., PEREIRA, J. M.; JORGE, J. Benchmarking **Instance-Centric Counterfactual Algorithms for XAI: From White Box to Black Box**. *ACM Computing Surveys*, v. 57, n. 6, p. 1–37, 2025. DOI: <https://doi.org/10.1145/3672553>.

NAZIR, Z., KALDYKHANOV, D., TOLEP, K.-K.; PARK, J.-G. **A Machine Learning Model Selection considering Tradeoffs between Accuracy and Interpretability**. v. 2021, n. 1, p. 3, 2021. DOI: <https://doi.org/10.1109/ICITEE53064.2021.9611872>.

NGUYEN, H. H., VIVIANI, J.-L.; BEN JABEUR, S. **Bankruptcy prediction using machine learning and Shapley additive explanations**. *Review of Quantitative Finance and Accounting*, 2023. DOI: <https://doi.org/10.1007/s11156-023-01192-x>.

NOH, S.-H. **Comparing the Performance of Corporate Bankruptcy Prediction Models Based on Imbalanced Financial Data**. Sustainability, v. 15, n. 6, p. 4794, 2023. DOI: <https://doi.org/10.3390/su15064794>.

ODOM, M. D.; SHARDA, R. **A neural network model for bankruptcy prediction**. v. 199, p. 0, 1990. DOI: <https://doi.org/10.3390/su15064794>.

OJA, E. **Adaptive and intelligent systems applications**. SMCia/, v. 9, p. 9, 1999.

PEREIRA, V. S.; MARTINS, V. F. **Estudos de previsão de falências—uma revisão das publicações internacionais e brasileiras de 1930 a 2015**. Revista Contemporânea de Contabilidade, v. 12, n. 26, p. 163-196, 2015. DOI: <https://doi.org/10.1109/SMCIA.1999.782727>.

PERES, D. L. O. **A importância da gestão de riscos e controles internos como respostas a riscos empresariais**. 2010. 285 f. Dissertação (Mestrado em Ciências Contábeis) – Faculdade de Administração e Finanças, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2010.

PINTELAS, E., LIVIERIS, I. E.; PINTELAS, P. **A Grey-Box Ensemble Model Exploiting Black-Box Accuracy and White-Box Intrinsic Interpretability**. Algorithms, v. 13, n. 1, p. 17, 2020. DOI: <https://doi.org/10.3390/a13010017>.

PINTO, A. C. **O poder preditivo dos modelos com aprendizado de máquina é superior aos modelos tradicionais para análise do risco de crédito?** Revista Debates em Economia Aplicada–REDEA, v. 1, p. 1, 2021.

PRADO, José Willer do. **Risco de crédito: uma abordagem utilizando análise discriminante, regressão logística e redes neurais artificiais**. 2016. 228 f. Dissertação (Mestrado em Administração) – Universidade Federal de Lavras, Lavras, 2016.

PRADO, J. W. DO; CASTRO ALCÂNTARA, V. DE; MELO CARVALHO, F. DE; VIEIRA, K. C.; MACHADO, L. K. C.; TONELLI, D. F. **Multivariate analysis of credit risk and bankruptcy research data: A bibliometric study involving different knowledge fields (1968–2014)**. Scientometrics, v. 106, n. 3, p. 1007–1029, 2016. DOI: <https://doi.org/10.1007/s11192-015-1829-6>.

PRADO, J. W. D., CARVALHO, F. D. M., BENEDICTO, G. C. D., ALCÂNTARA, V. D. C.; SANTOS, A. C. D. **Uma abordagem para análise do risco de crédito utilizando o modelo Fleuriet**. Revista de Educação e Pesquisa em Contabilidade (REPeC), v. 12, n. 3, p. 341–363, 2018. DOI: <https://doi.org/10.17524/repec.v12i3.1816>.

PRADO, J. W. D., VILAMAIOR, A. G., CAMPOS, A. C.; NASCIMENTO, T. B. P. D. **Previsão da Insolvência Empresarial Utilizando Redes Neurais Artificiais**. Revista Gestão e Desenvolvimento, v. 17, n. 2, p. 136, 2020. DOI: <https://doi.org/10.17524/repec.v12i3.1816>.

PYTHON SOFTWARE FOUNDATION. **Python: versão 3.10**. Python Software Foundation, 2021. Disponível em: <https://www.python.org>.

QIU, Q., SHEN, S., GUAN, S., LI, M.; DENG, S. **An Attribute Contribution-Based K-Nearest Neighbor Classifier**. v. 2022, p. 6, 2022. DOI: <https://doi.org/10.1109/ACAIT56212.2022.10137909>.

QUINLAN, J. R. **Induction of decision trees**. *Machine Learning*, v. 1, n. 1, p. 81–106, 1986. DOI: <https://doi.org/10.1007/BF00116251>.

ROY, J. K.; VASA, L. **Transforming credit risk assessment: A systematic review of AI and machine learning applications**. *Journal of Infrastructure, Policy and Development*, v. 9, n. 1, p. 9652, 2025. DOI: <https://doi.org/10.24294/jipd9652>.

SANTOS, ALLISSON SILVA DOS. **Previsão de Insolvência Corporativa: Uma Análise de Empresas Brasileiras de Capital Aberto Por Meio de Aprendizado de Máquina**. 2021. 73 F. Dissertação de Mestrado (Mestrado em Administração). Universidade Federal da Paraíba, João Pessoa, 2021.

SEDDIK, B., AHLEM, D.; HOCINE, C. **An Explainable Self-Labeling Grey-Box Model**. v. 2022, p. 4, 2022. DOI: <https://doi.org/10.1109/PAIS56586.2022.9946912>.

SHI, Y.; LI, X. **A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms**. *Heliyon*, v. 5, n. 1, p. 2, 2019. DOI: <https://doi.org/10.1016/j.heliyon.2019.e02997>.

SOUZA, V. F., CICALESE, F., LABER, E.; MOLINARO, M. **Decision trees with short explainable rules**. *Advances in neural information processing systems*, v. 35, p. 12365-12379, 2022. DOI: <https://doi.org/10.1016/j.tcs.2025.115344>.

TAM, K. **Neural network models and the prediction of bank bankruptcy**. *Omega*, v. 19, n. 5, p. 429–445, 1991. DOI: [https://doi.org/10.1016/0305-0483\(91\)90060-7](https://doi.org/10.1016/0305-0483(91)90060-7).

TAVARES, L., MAZZON, J., PALETTA, F.; BARROS, F. **Bankruptcy analysis using images and convolutional neural networks (CNN)**, 2025. DOI: <https://doi.org/10.48550/arXiv.2502.15726>.

TEODORO, G. D., MONACI, M.; PALAGI, L. **Unboxing Tree Ensembles for interpretability: A hierarchical visualization tool and a multivariate optimal re-built tree**. *EURO Journal on Computational Optimization*, v. 12, p. 100084, 2024. DOI: <https://doi.org/10.1016/j.ejco.2024.100084>.

VÁSQUEZ-SERPA, L.-J., RODRÍGUEZ, C., PÉREZ-NÚÑEZ, J.-R.; NAVARRO, C. **Challenges of Artificial Intelligence for the Prevention and Identification of Bankruptcy Risk in Financial Institutions: A Systematic Review**. *Journal of Risk and Financial Management*, v. 18, n. 1, p. 26, 2025. DOI: <https://doi.org/10.3390/jrfm18010026>.

WANG, Y., LIAO, W., SHEN, H., JIANG, Z.; ZHOU, J. **Some notes on the basic concepts of support vector machines**. *Journal of Computational Science*, v. 82, p. 102390, 2024. DOI: <https://doi.org/10.1016/j.jocs.2024.102390>.

XU, C., COEN-PIRANI, P.; JIANG, X. **Empirical Study of Overfitting in Deep FNN Prediction Models for Breast Cancer Metastasis**. arXiv preprint arXiv:, v. 220, p. 8, 2022. <https://doi.org/10.48550/arXiv.2208.02150>.

TERCEIRA PARTE – CONSIDERAÇÕES FINAIS

A presente dissertação teve como objetivo a analisar o desempenho preditivo de técnicas estatísticas multivariadas e de aprendizado de máquina na previsão do risco de insolvência de empresas brasileiras não financeiras listadas na Brasil, Bolsa, Balcão (B3), durante o período de 2010 a 2024. Por meio da condução de três estudos empíricos, buscou-se não apenas apresentar o desempenho dos modelos utilizando diferentes abordagens metodológicas no contexto das empresas brasileiras não financeiras, mas também identificar os indicadores financeiros, tanto tradicionais quanto dinâmicos, com maior poder discriminante, preenchendo lacunas na literatura nacional e oferecendo subsídios práticos para a tomada de decisão no ambiente corporativo.

O primeiro artigo investigou a capacidade do modelo desenvolvido por meio da Análise Discriminante, uma técnica estatística consolidada, para classificar empresas como solventes ou insolventes. O modelo demonstrou um elevado poder preditivo, com uma acurácia de 93,3% na amostra de teste e uma AUC de 0,955.

Os resultados destacaram a relevância de indicadores por meio do modelo discriminante, que combinam a estrutura de capital liquidez e capacidade de geração de disponibilidade financeira, sendo as variáveis X6 (Saldo de Tesouraria sobre a Receita Líquida), X13 (Passivo Circulante Financeiro e Passivo Não Circulante Financeiro sobre o Ativo Total), ambas do modelo dinâmico de Fleuriet, bem como a X14 (Ativo Circulante sobre o Passivo Circulante), X16 (Lucro Antes dos Juros e Impostos sobre o Ativo Total) e X33 (Lucro Antes dos Juros e Impostos sobre os Impostos Sobre Despesas Financeiras), relacionado aos indicadores tradicionais. Tais resultados reforçam a validade de modelos estatísticos interpretáveis como ferramentas robustas para a análise de risco de crédito.

O segundo artigo aprofundou a análise por meio da Regressão Logística aplicada a dados em painel, uma abordagem que permite capturar a dinâmica temporal e a heterogeneidade entre as empresas. O modelo final de Efeitos Aleatórios alcançou uma excelente Área Sob a Curva ROC (AUC) de 0,915 e acurácia de 90,2%.

As variáveis mais influentes no modelo longitudinal logístico para dados em painel foram as relacionadas à estrutura de capital, gestão do capital de giro e disponibilidade financeira, como as variáveis X30 (Patrimônio Líquido – Capital Social sobre o Ativo Total), X19 (Disponibilidade sobre o Ativo Circulante e Realizável a Longo Prazo), X14 (Ativo Circulante sobre o Passivo Circulante), X11 (Necessidade de Capital de Giro), X2 (Capital de Giro sobre a Receita Líquida), X27 (Dívida Bruta sobre o Ativo Circulante), X21 (Fornecedores

sobre o Ativo Total) e X6 (Saldo de Tesouraria sobre a Receita Líquida). Os achados evidenciam a importância integração entre os modelos tradicionais e dinâmicos de Fleuriet, bem como a importância de uma boa gestão da estrutura de capital, capital de giro, lucratividade e disponibilidade financeira para honrar com as obrigações de curto e longo prazo.

Por fim, o terceiro artigo realizou uma análise comparativa entre seis diferentes métodos construídos de aprendizado de máquina, sendo eles *Decision Tree (DT)*, *Random Forest (RF)*, *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, *Multilayer Perceptrons (MLP)* e *Extreme Gradient Boosting (XGBoost)*. Os resultados demonstraram a superioridade das técnicas de ensemble, com o XGBoost se destacando como o modelo de melhor desempenho, atingindo uma acurácia de 95,4% e uma AUC de 0,981, e na sequência o modelo RF com 95,2% de acurácia e AUC de 0,978, destacando, portanto, as técnicas *ensemble* como as que possui melhor poder preditivo.

A análise de importância das variáveis dos modelos desenvolvidos de aprendizado de máquina no artigo 3 reforçam os achados dos artigos anteriores, apontando indicadores de relacionados a estrutura de capital, como X30 (Patrimônio Líquido menos Capital Social sobre o Ativo Total), liquidez e fluxo de caixa, como a X6 (Saldo de Tesouraria sobre a Receita Líquida) e a X14 (Ativo Circulante sobre o Passivo Circulante) e a gestão dos capital de giro, como a X2 (Capital de Giro sobre a Receita Líquida) e X12 (Capital de Giro), sendo tais variáveis analisadas como as mais relevantes para a previsão da insolvência. A evidência das principais variáveis que influenciaram no desenvolvimento de tais modelos permitiu mitigar a característica de "caixa-preta" do modelo, oferecendo interpretabilidade aos seus resultados de alta precisão.

Em conjunto, os três artigos confirmam que tanto as abordagens estatísticas clássicas quanto os modelos de aprendizado de máquina possuem alta capacidade de previsão do risco de insolvência no contexto brasileiro. A comparação entre os métodos revelou que as técnicas de aprendizado de máquina, em especial o XGBoost e o RF, oferecem ganhos marginais de acurácia e melhor equilíbrio entre sensibilidade e especificidade, validando sua superioridade em bases de dados complexas, como dados em painel, captando de forma precisa as características de solvência e insolvência empresarial. Ainda assim, a Análise Discriminante mostrou-se uma ferramenta eficaz e de mais fácil interpretação, representando uma alternativa valiosa e acessível.

De forma complementar, as variáveis que mais contribuíram na construção dos modelos construídos e apresentados nesta dissertação foram a X30 (Patrimônio Líquido menos Capital Social sobre o Ativo Total), X14 (Ativo Circulante sobre o Passivo Circulante) do modelo

tradicional, enquanto as variáveis independentes do modelo dinâmico foram a X6 (Saldo de Tesouraria sobre a Receita Líquida), X2 (Capital de Giro sobre a Receita Líquida).

A contribuição teórica desta dissertação reside na construção de modelos utilizando diferentes abordagens de previsão na classificação do risco de insolvência em um contexto amplo e atualizado do mercado brasileiro, preenchendo lacunas na literatura nacional que carecia de estudos com maior abrangência de técnicas, de empresas de diferentes setores e de um período de análise temporal maior. A incorporação de indicadores do Modelo Dinâmico de Fleuriet ao lado de métricas tradicionais também enriqueceu a análise, demonstrando o valor de uma visão integrada da gestão financeira.

Do ponto de vista prático e gerencial, os resultados oferecem subsídios concretos para aprimorar políticas de crédito, gestão financeira e estratégias de investimento. Os modelos desenvolvidos podem ser utilizados como sistemas de alerta precoce, permitindo que gestores ajam preventivamente para mitigar riscos, e que credores e investidores tomem decisões mais embasadas. A identificação consistente de um núcleo de indicadores-chave, como lucratividade, endividamento e liquidez, simplifica o monitoramento da saúde financeira das organizações.

Como limitações da pesquisa, destaca-se o foco em empresas brasileiras não financeiras de capital aberto, o que restringe a generalização dos resultados para outros países, setores ou para empresas de capital fechado. Além disso, a análise se baseou exclusivamente em indicadores financeiros quantitativos, não incorporando variáveis macroeconômicas ou qualitativas, como governança corporativa e métricas ESG, que poderiam agregar poder explicativo aos modelos, assim como a não utilização de medidas diretamente associadas ao fluxo de caixa, o que pode limitar a identificação de eventos de insolvência relacionados a restrições de liquidez. Por fim, a dificuldade em atender plenamente a premissa de normalidade multivariada na Análise Discriminante é uma limitação metodológica inerente a dados financeiros.

Para pesquisas futuras, sugere-se a expansão do escopo amostral para incluir empresas de outros países, permitindo análises comparativas internacionais, bem como a aplicação dos modelos ao setor financeiro. Recomenda-se também a incorporação de variáveis não financeiras, como indicadores de governança, sustentabilidade e dados macroeconômicos, para a construção de modelos mais holísticos. Por fim, a aplicação de técnicas mais avançadas de *deep learning* e a exploração contínua de métodos de interpretabilidade para modelos complexos representam avenidas promissoras para o avanço do conhecimento na área.

Em síntese, esta dissertação demonstrou a elevada eficácia de modelos preditivos na identificação do risco de insolvência, contribuindo para o avanço teórico e prático no campo das finanças e oferecendo ferramentas analíticas de alta precisão para a promoção da estabilidade e sustentabilidade no ambiente corporativo.

Conclui-se, assim, que esta dissertação amplia o conhecimento científico sobre a previsão do risco de insolvência ao demonstrar, de forma comparativa e integrada com variáveis financeiras tradicionais e do modelo dinâmico de Fleuriet, a efetividade de modelos estatísticos e de aprendizado de máquina aplicados ao contexto das empresas brasileiras de capital aberto listadas na Brasil, Bolsa, Balcão (B3), com exceção do setor financeiro. Os achados fornecem bases consistentes para que pesquisadores avancem na construção de modelos mais precisos e interpretáveis, ao mesmo tempo em que oferecem aos profissionais de mercado ferramentas sólidas para aprimorar a análise de crédito e a análise por meio dos indicadores financeiros identificados para mitigação do risco de insolvência empresarial de forma preventiva.

REFERÊNCIAS

- ACOSTA-GONZÁLEZ, E., FERNÁNDEZ-RODRÍGUEZ, F.; GANGA, H. **Predicting Corporate Financial Failure Using Macroeconomic Variables and Accounting Data**. *Computational Economics*, v. 53, n. 1, p. 227–257, 2019. DOI: <https://doi.org/10.1007/s10614-017-9737-x>.
- ADIPUTRA, I. G.; RUSLIM, H. **The Influence Of Exchangeability, Leverage, Inflation, and Interest Rate on Financial Problems Among Manufacturing Companies Listed on Indonesia Stock Exchange During 2018-2021**. *International Journal of Application on Economics and Business*, v. 1, n. 3, p. 1062–1073, 2023. DOI: <https://doi.org/10.24912/ijaeb.v1i3.1062-1073>.
- AGRAWAL, N. **An Exploratory Study on Compulsory Liquidation of a Company and Its Liability**. *International Journal For Multidisciplinary Research.*, 2023.
- AFFES, Z.; HENTATI-KAFFEL, R. **Predicting US Banks Bankruptcy: Logit Versus Canonical Discriminant Analysis**. *Computational Economics*, v. 54, n. 1, p. 199–244, 2017. DOI: <https://doi.org/10.1007/s10614-017-9698-0>.
- AKIAMA, S. R. **Probabilidade de Inadimplência de Grandes Empresas no Sistema Financeiro Nacional**. Dissertação (Mestrado em Ciências Contábeis) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo, 2008.
- ALHARAHSEH, Husam Helmi; PIUS, Abraham. **A review of key paradigms: positivism vs interpretivism**. *Global Academic Journal of Humanities and Social Sciences*, v. 2, n. 3, p. 39–43, jun. 2020.
- ALJAWAZNEH, H., MORA, A. M., GARCIA-SANCHEZ, P.; CASTILLO-VALDIVIESO, P. A. **Comparing the Performance of Deep Learning Methods to Predict Companies' Financial Failure**. *IEEE Access*, v. 9, p. 97010–97038, 2021. DOI: <https://doi.org/10.1109/ACCESS.2021.3093461>.
- ALMEIDA, L. **Risk and Bankruptcy Research: Mapping the State of the Art**. *Journal of Risk and Financial Management*, v. 16, n. 8, p. 361, 2023. DOI: <https://doi.org/10.3390/jrfm16080361>.
- ALMEIDA, R.; SALES, G.; NUNES, R. V. N. **Predição de Insolvência de Empresas Através da Inteligência Artificial-Técnicas de Aprendizado de Máquina**. *REPAE-Revista de Ensino e Pesquisa em Administração e Engenharia*, v. 9, n. 1, p. 39-58, 2023. DOI: <https://doi.org/10.51923/repae.v9i1.292>.
- ALTMAN, E. I. **Financial Ratios, Discriminant Analysis And The Prediction Of Corporate Bankruptcy**. *The Journal of Finance*, v. 23, n. 4, p. 589–609, 1968. DOI: <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>.
- ALTMAN, E. I., BAIDYA, T. K.; DIAS, L. M. R. **Previsão de problemas financeiros em empresas**. *Revista de administração de empresas*, v. 19, p. 17-28, 1979. DOI: <https://doi.org/10.1590/S0034-75901979000100002>.

AMARO, A. **Consciência e cultura do risco nas organizações**. Territorium, v. 12, p. 5–9, 2005.

ARANHA, J. A. M.; GONDRIGE, E. D. O. **Estatística multivariada aplicada: Construção do Modelo de Previsão de Insolvência Aranha e Gondrige**. CONTRIBUCIONES A LAS CIENCIAS SOCIALES, v. 17, p. 8, 2024. DOI: <https://doi.org/10.55905/revconv.17n.8-279>.

AREOSA, J. **O risco nas ciências sociais: Uma visão crítica ao paradigma dominante**. Revista Angolana de Sociologia, 2010. DOI: <https://hdl.handle.net/1822/16117>.

ARORA, I; SINGH, N. **Prediction of Corporate Bankruptcy using Financial Ratios and News**. International Journal of Engineering and Management Research, v. 10, n. 5, p. 82–87, 2020. DOI: <https://doi.org/10.31033/ijemr.10.5.15>.

ASSAF NETO, A; SILVA, C. A. T. **Administração do Capital de Giro**. 4. ed. Atlas, 2012.

BARBOSA, A. A. D. S.; NOBRE, F. C. **Mapeamento da produção científica sobre os modelos preditivos de insolvência no Brasil**. Revista de Gestão e Secretariado, v. 14, n. 1, p. 645–661, 2023. DOI: <https://doi.org/10.7769/gesec.v14i1.1540>.

BARBOZA, F., BASSO, L. F. C.; KIMURA, H. **New metrics and approaches for predicting bankruptcy**. Communications in Statistics - Simulation and Computation, v. 52, n. 6, p. 2615–2632, 2023. DOI: <https://doi.org/10.1080/03610918.2021.1910837>.

BEAVER, W. H. **Financial Ratios as Predictors of Failure-Reply**. Journal of Accounting Research, v. 4, p. 123-127, 1966. Disponível em: <https://www.jstor.org/stable/10.2307/2490174?origin=crossref>.

BILLIOS, D., SERETIDOU, D.; STAVROPOULOS, A. **The Power of Numerical Indicators in Predicting Bankruptcy: A Systematic Review**. Journal of Risk and Financial Management, v. 17, n. 10, p. 433, 2024. DOI: <https://doi.org/10.3390/jrfm17100433>.

BRANCHER, Pedro Txai. **Agência e estrutura em sistemas internacionais: a ontologia do realismo científico e suas implicações para uma agenda de pesquisa**. Mural Internacional, v. 9, n. 2, p. 266–277, jul./dez. 2018. DOI: <https://doi.org/10.12957/rmi.2018.39008>.

BRASIL. **Decreto-Lei nº 7.661, de 21 de junho de 1945**. Lei de Falências. Disponível em: Acesso em: 4 maio 2025. Disponível em: https://www.planalto.gov.br/ccivil_03/Decreto-Lei/Del7661.htm. Acesso em: 4 de maio de 2025.

BRASIL. **Lei nº 11.101, de 9 de fevereiro de 2005**. Regula a recuperação judicial, a extrajudicial e a falência do empresário e da sociedade empresária. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2005/Lei/L11101.htm. Acesso em: 4 de maio de 2025.

BRASIL. **Lei nº 11.638, de 28 de dezembro de 2007**. Altera e revoga dispositivos da Lei nº 6.404, de 15 de dezembro de 1976, e da Lei nº 6.385, de 7 de dezembro de 1976, e estende às sociedades de grande porte disposições relativas à elaboração e divulgação de demonstrações financeiras. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/lei/11638.htm. Acesso em: 31 de janeiro de 2026.

BRASIL. **Lei nº 11.941, de 27 de maio de 2009**. Altera as normas tributárias para adequação as normas internacionais de contabilidade e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2009/lei/11941.htm. Acesso em: 31 de janeiro de 2026.

BRASIL. **Lei nº 14.112, de 24 de dezembro de 2020**. Altera a Lei nº 11.101, de 9 de fevereiro de 2005, que regula a recuperação judicial, a extrajudicial e a falência do empresário e da sociedade empresária. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2020/Lei/L14112.htm. Acesso em: 4 maio 2025.

BRITO, G. A. S., ASSAF NETO, A.; CORRAR, L. J. **Sistema de classificação de risco de crédito: uma aplicação a companhias abertas no Brasil**. Revista Contabilidade e Finanças, v. 20, n. 5, p. 1, 2009. DOI: <https://doi.org/10.1590/S1519-70772009000300003>.

CAMAYD, Yohandra Rad; FREIRE, Eudaldo Enrique Espinoza. **Estratégias metodológica de investigação nas ciências sociais**. Revista Conrado, Cienfuegos, v. 16, n. 77, p. 65–73, nov./dez. 2020.

CARVALHO, Leonardo Lana de; VARENNE, Franck; BRAGA, Elayne de Moura. **Ontologias para a modelagem multiagente de sistemas complexos em ciências cognitivas**. Ciências e Cognição, v. 19, n. 1, p. 58–75, 2014. Disponível em: <http://www.cienciasecognicao.org>. Acesso em: 5 nov. 2025.

CHAGLLA, J. A. C. **El riesgo de insolvencia de las cooperativas de ahorro y crédito segmento 2 en el Ecuador**. 2019.

CLOSSET, F.; GROSSMANN, C.; KASERER, C.; URBAN, D. **Corporate Restructuring and Creditor Power: Evidence from European Insolvency Law Reforms**. 2021. DOI: <https://doi.org/10.1016/j.jbankfin.2022.106756>.

COSTA, Flaviano; ESPEJO, Márcia Maria dos Santos Bortolucci; ALMEIDA, Lauro Brito de; CRUZ, Ana Paula Capuano da. **Identificação da base intelectual das publicações científicas em controladoria no Brasil por meio da polifonia discursiva: um estudo inspirado na arqueologia foucaultiana**. CONGRESSO USP DE CONTROLADORIA E CONTABILIDADE, 11., 2011, São Paulo. *Anais*. São Paulo: FEA/USP, 2011.

CRISÓSTOMO, V. L. MELO JÚNIOR, C. V DE. **Uma Análise Comparativa de Indicadores Financeiros Entre Empresas Listadas em Segmentos Diferenciados Bovespa e as do Mercado Tradicional**. V. L., v. 10, p. 3, 2015.

CUNHA, M. A.; PENEDO, A. S. T.; BARBOZA, F. L. M. **Predicting Insolvency In Brazilian Companies Using Machine Learning: Na Anlalysis of the Random Forest Technique**. 2024. DOI: <https://doi.org/10.18624/etech.v17i1.1342>.

DASILAS, A.; RIGANI, A. **Machine learning techniques in bankruptcy prediction: A systematic literature review**. Expert Systems with Applications, v. 255, p. 124761, 2024. DOI: <https://doi.org/10.1016/j.eswa.2024.124761>.

DIONNE, G. **Risk Management: History, Definition, and Critique**. Risk Management and Insurance Review, v. 16, n. 2, p. 147–166, 2013. DOI: <https://doi.org/10.1111/rmir.12016>.

EIFERT, D. S. **Análise quantitativa na concessão de crédito versus inadimplência: um estudo empírico**. Dissertação de mestrado, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil, , 2003. DOI: <http://hdl.handle.net/10183/3533>.

ELIZABETSKY, R. **Um modelo matemático para decisão no banco comercial**. Universidade de São Paulo, 1976.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata**. Rio de Janeiro: Elsevier, 2017.

FERREIRA, Cássia Oliveira; FERREIRA, Priscila Oliveira; LAMOUNIER, Wagner Moura; AVELAR, Ewerton Alex. **Fluxos de caixa livres e endividamento em operadoras de planos de saúde**. ForScience, Formiga, v. 9, n. 1, e00865, jan./jun. 2021. DOI: <https://doi.org/10.29069/forscience.2021v9n1.e865>.

FIRMANSYAH, Muhammad Raihan; ASTUTI, Yani Parti. **Stroke classification comparison with KNN through standardization and normalization techniques**. Advance Sustainable Science, Engineering and Technology (ASSET), v. 6, n. 1, p. 02401012-01–02401012-08, jan. 2024. DOI: <https://doi.org/10.26877/asset.v6i1.17685>.

FONTALVO-HERRERA, T. J.; DE LA HOZ-GRANADILLO, E. **Método conglomerado-análisis discriminante-análisis envolvente de datos para clasificar y evaluar eficiencia empresarial**. Entramado, v. 16, n. 2, p. 46–55, 2020. DOI: <https://doi.org/10.18041/1900-3803/entramado.2.6437>.

GAJDOSIKOVA, D.; VALASKOVA, K. **Bankruptcy Prediction Model Development and its Implications on Financial Performance in Slovakia**. Economics and Culture, v. 20, n. 1, p. 30–42, 2023. DOI: <https://doi.org/10.2478/jec-2023-0003>.

GARCIA, J. **Bankruptcy prediction using synthetic sampling**. Machine Learning with Applications, v. 9, p. 100343, 2022. DOI: <https://doi.org/10.1016/j.mlwa.2022.100343>.

GOMEZ, Maria Nélide Gonzalez. **Validade científica: da epistemologia à política e à ética**. *Liinc em Revista*, Rio de Janeiro, v. 11, n. 2, p. 339–359, nov. 2015. DOI: <https://doi.org/10.18617/liinc.v11i2.816>.

GONZALES, Katia Guerchi; NEVES, Tatiani Garcia; SANTOS, Cintia Melo dos. **Abordagens metodológicas de pesquisa: algumas notas**. Revista Ensaios Pedagógicos em Educação, Ciências e Humanidades, v. 19, n. 2, p. 217–226, 2018. DOI: <https://doi.org/10.17921/2447-8733.2018v19n2p217-226>.

GRISHINA, Elena Valerievna. **Ideal-realism as a religious ontology in the philosophy of religion of B. P. Vysheslavtsev**. *Vestnik PSTGU. Series I: Theology. Philosophy. Religious Studies*, issue 113, p. 98–120, 2024. DOI: <https://doi.org/10.15382/sturI2024112.98-120>.

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria básica**. 5. ed. Porto Alegre: AMGH, 2011.

HORTA, R. A. M. **Uma metodologia de mineração de dados para a previsão de insolvência de empresas brasileiras de capital aberto**. Tese de doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brasil, 2010.

HUO, Y., CHAN, L. H.; MILLER, D. **Bankruptcy prediction for restaurant firms: A comparative analysis of multiple discriminant analysis and logistic regression**. *Journal of Risk and Financial Management*, v. 17, n. 39, p. 9, 2024. DOI: <https://doi.org/10.3390/jrfm17090399>.

IDHMAD, A., KAICER, M., NEJJAR, C.; BENJOUAD, A. **Intelligent Bankruptcy Prediction Models Involving Corporate Governance Indicators, Financial Ratios and SMOTE**. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, v. 12, n. 1, p. 233–244, 2024. DOI: <https://doi.org/10.52549/ijeei.v12i1.5241>.

JAAFAR, M. N.; MUHAMAT, A. A.; ALWI, S. F. S.; KARIM, N. A.; RAHMAN, S. B. A. **Determinants of Financial Distress among the Companies Practise Note 17 Listed in Bursa Malaysia**. *International Journal of Academic Research in Business and Social Sciences*, v. 8, 800-811, 2018. DOI: <https://doi.org/10.6007/IJARBS/v8-i11/4956>.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An introduction to statistical learning: with applications in Python**. Nova York: Springer, 2023. DOI: <https://doi.org/10.1007/978-3-031-38747-0>.

JESUS, D. P.; BESARRIA, C. D. N. **Machine learning and sentiment analysis: Projecting bank insolvency risk**. *Research in Economics*, v. 77, n. 2, p. 226–238, 2023. DOI: <https://doi.org/10.1016/j.rie.2023.03.001>.

JONNARDI, J., BANGUN, N.; NATSIR, K. **The Determinants of Company'S Financial Distress**. *International Journal of Application on Economics and Business*, v. 1, n. 3, p. 1613–1624, 2023. DOI: <https://doi.org/10.24912/ijaeb.v1i3.1613-1624>.

KAMALOV, Firuz; LEUNG, Ho Hon. **ROC curve model under Pareto distribution**. *Applied Mathematical Sciences*, v. 10, n. 10, p. 461–466, 2016. DOI: <https://doi.org/10.12988/ams.2016.512736>.

KANITZ, S. **Como Prever Falências de Empresas**. *Revista Negócios em Exame*, p. 95–102, 1974.

KOCHAŃSKI, Błażej. **The shape of an ROC curve in the evaluation of credit scoring models**. *Statistics in Transition new series*, v. 25, n. 2, p. 205–218, jun. 2024. Disponível em: <https://doi.org/10.59170/stattrans-2024-022>.

KOTHURU, V., JHA, R. K., RANJIT, S., KUMAR REDDY, B. V., ROY, S.; SUDHEER, S. **Prediction of Bankruptcy of a company using machine learning techniques**, 2022. DOI: <https://doi.org/10.1109/ICESC54411.2022.9885619>.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. New York: Springer, 2013. DOI: <https://doi.org/10.1007/978-1-4614-6849-3>.

LEE, S.; CHOI, K.; YOO, D. **Predicting the Insolvency of SMEs Using Technological Feasibility Assessment Information and Data Mining Techniques.** *Sustainability*, v. 12, n. 23, p. 9790, 2020. DOI: <https://doi.org/10.3390/su12239790>.

LIMA, F. G.; PAULINO, C. T.; FÁVERO, L. P. L. **ESG e Machine Learning: O Impacto na Previsão de Insolvência de Empresas Brasileiras.** F. G., *Contabilometria*, v. 11, p. 1, 2022.

LISBOA, J. C. V. **A importância da análise das demonstrações contábeis para suporte as decisões gerenciais das organizações.** *Administração de Empresas em Revista*, v. 1, n. 9, p. 159-170, 2014.

LUKASON, O.; ANDRESSON, A. **Tax Arrears Versus Financial Ratios in Bankruptcy Prediction.** *Journal of Risk and Financial Management*, v. 12, n. 4, p. 187, 2019. DOI: <https://doi.org/10.3390/jrfm12040187>.

MATIAS, A. B. **Contribuição às técnicas de análise financeira: um modelo de concessão de crédito.** Trabalho apresentado ao Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo. São Paulo, 1978.

MELO, A. C.; COUTINHO, E. S. **O modelo Fleuriet como indicador conjunto de solvência e rentabilidade.** *Anais do EnANPAD*, Rio de Janeiro, RJ, Brasil, v. 3, p. 1, 2007.

MINATTO, F.; BORBA, J. **Insolvency in Brazilian Football Clubs: Proposition of Models Based on Neural Networks.** *Brazilian Business Review*, v. 18, n. 6, p. 624–642, 2021. DOI: <https://doi.org/10.15728/bbr.2021.18.6.2>.

MINUSSI, J. A., DAMACENA, C.; NESS JR., W. L. **Um Modelo de Previsão de Solvência Utilizando Regressão Logística.** *Revista de Administração Contemporânea*, v. 6, p. 3, 2002. DOI: <https://doi.org/10.1590/S1415-6552002000300007>.

MORAES, F. B.; DONADELLO, D. F.; DE ANDRADE, C. R. **Possibilidade do Pedido de Recuperação Judicial da Empresa Individual de Responsabilidade Limitada–Eireli.** *Revista CEJ*, v. 23, n. 7, p. 7, 2019.

NGUYEN, H. H., VIVIANI, J.-L.; BEN JABEUR, S. **Bankruptcy prediction using machine learning and Shapley additive explanations.** *Review of Quantitative Finance and Accounting*, 2023. DOI: <https://doi.org/10.1007/s11156-023-01192-x>.

NOH, S.-H. **Comparing the Performance of Corporate Bankruptcy Prediction Models Based on Imbalanced Financial Data.** *Sustainability*, v. 15, n. 6, p. 4794, 2023. DOI: <https://doi.org/10.3390/su15064794>

ODOM, M. D.; SHARDA, R. **A neural network model for bankruptcy prediction.** v. 199, p. 0, 1990. DOI: <https://doi.org/10.3390/su15064794>.

OHLSON, J. A. **Financial ratios and the probabilistic prediction of bankruptcy.** *Journal of accounting research*, 109-131, 1980. DOI: <https://doi.org/10.2307/2490395>.

PEREIRA, V. S.; MARTINS, V. F. **Estudos de previsão de falências—uma revisão das publicações internacionais e brasileiras de 1930 a 2015**. Revista Contemporânea de Contabilidade, v. 12, n. 26, p. 163-196, 2015. DOI: <https://doi.org/10.1109/SMCIA.1999.782727>.

PRADO, J. W.; DE CASTRO ALCÂNTARA, V.; MELO CARVALHO, F.; VIEIRA, K. C.; MACHADO, L. K. C.; TONELLI, D. F. **Multivariate analysis of credit risk and bankruptcy research data: A bibliometric study involving different knowledge fields (1968–2014)**. Scientometrics, v. 106, n. 3, p. 1007–1029, 2016. DOI: <https://doi.org/10.1007/s11192-015-1829-6>.

PRADO, J. W. do. **Determinantes e implicações da estrutura de capital, da estrutura de propriedade e da governança corporativa: um modelo multiteórico de análise**. 2019. 194 p. Tese (Doutorado em Administração). Universidade Federal de Lavras, Lavras, 2019.

PRADO, J. W. D., CARVALHO, F. D. M., BENEDICTO, G. C. D., ALCÂNTARA, V. D. C.; SANTOS, A. C. D. **Uma abordagem para análise do risco de crédito utilizando o modelo Fleuriet**. Revista de Educação e Pesquisa em Contabilidade (REPeC), v. 12, n. 3, p. 341–363, 2018. DOI: <https://doi.org/10.17524/repec.v12i3.1816>.

PRADO, J. W. D., VILAMAIOR, A. G., CAMPOS, A. C.; NASCIMENTO, T. B. P. D. **Previsão da Insolvência Empresarial Utilizando Redes Neurais Artificiais**. Revista Gestão e Desenvolvimento, v. 17, n. 2, p. 136, 2020. DOI: <https://doi.org/10.17524/repec.v12i3.1816>.

PINTO, A. C. **O poder preditivo dos modelos com aprendizado de máquina é superior aos modelos tradicionais para análise do risco de crédito?** Revista Debates em Economia Aplicada–REDEA, v. 1, p. 1, 2021.

ROSA, Eliana Cristina. **O conhecimento científico da metodologia: com o olhar para o método hipotético dedutivo como ferramenta de pesquisa**. Revista Iniciação & Formação Docente, Dossiê do X Seminário de Leitura e Produção no Ensino Superior, v. 2, n. 2, p. 1–10, jul. 2015–jan. 2016. Disponível em: <https://ssrn.com/abstract=2989102>. Acesso em: 5 nov. 2025.

ROY, J. K.; VASA, L. **Transforming credit risk assessment: A systematic review of AI and machine learning applications**. Journal of Infrastructure, Policy and Development, v. 9, n. 1, p. 9652, 2025. DOI: <https://doi.org/10.24294/jipd9652>.

SANTOS, ALLISSON SILVA DOS. **Previsão de Insolvência Corporativa: Uma Análise de Empresas Brasileiras de Capital Aberto Por Meio de Aprendizado de Máquina**. 2021. 73 F. Dissertação de Mestrado (Mestrado em Administração). Universidade Federal da Paraíba, João Pessoa, 2021.

SANVICENTE, A. Z.; MINARDI, A. M. A. F. **Identificação de indicadores contábeis significativos para a previsão de concordata de empresas**. Instituto Brasileiro de Mercado de Capitais, Working Paper, v. 1968, p. 1-12, 1998.

SHI, Y.; LI, X. **A bibliometric study on intelligent techniques of bankruptcy prediction for corporate firms**. Heliyon, v. 5, n. 1, p. 2, 2019. DOI: <https://doi.org/10.1016/j.heliyon.2019.e02997>.

SCHIO, T. A.; SAMPAIO, A. V. **Previsão da Recuperação Judicial de Empresas no Brasil: Uma Investigação Empírica**. *Pensar Contábil*, v. 24, n. 8, p. 5, 2023.

SHETTY, S., MUSA, M.; BRÉDART, X. **Bankruptcy Prediction Using Machine Learning Techniques**. *Journal of Risk and Financial Management*, v. 15, n. 1, p. 35, 2022. DOI: <https://doi.org/10.3390/jrfm15010035>.

SILVA, J. **Modelos para classificação de empresas com vistas a concessão de crédito**. Dissertação (Mestrado em Administração) - Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas. São Paulo, v. 198, p. 2, 1982.

SILVA, V. B.; SAITO, R. **Corporate Financial Distress and Reorganization: A Survey of Theoretical and Empirical Contributions**. *Review of Business Management*, v. 2, p. 2, 2020. DOI: <https://doi.org/10.7819/rbgn.v22i0.4057>.

STEPPAN, A. I. B.; SANTIAGO, J. S.; SILVA, J. D. G.; CAVALCANTE, P. R. N.; SILVA, C. A. T. **Julgamento na auditoria: um estudo das características das pesquisas científicas**. *Ciências & Cognição*, Rio de Janeiro, v. 20, n. 1, p. 61–78, 2015.

STÜPP, D. R. **Previsão de insolvência a partir de indicadores contábeis: evidências de empresas listadas na BM&FBOVESPA nos anos 2004-2013**. 2015. 137 f. Dissertação (Mestrado em Contabilidade) – Universidade Federal de Santa Catarina, Florianópolis, 2015.

SULISTIANI, I., MUFIDA, E., YASSER, P. M.; ALAMSYAH, L. **Systematic Literature Review: Bankruptcy Prediction Menggunakan Teknik Machine Learning dan Deep Learning**. *INTECH*, v. 2, n. 1, p. 13–18, 2021. DOI: <https://doi.org/10.54895/intech.v2i1.824>.

TAM, K. **Neural network models and the prediction of bank bankruptcy**. *Omega*, v. 19, n. 5, p. 429–445, 1991. DOI: [https://doi.org/10.1016/0305-0483\(91\)90060-7](https://doi.org/10.1016/0305-0483(91)90060-7).

TAVARES, L., MAZZON, J., PALETTA, F.; BARROS, F. **Bankruptcy analysis using images and convolutional neural networks (CNN)**, 2025. DOI: <https://doi.org/10.48550/arXiv.2502.15726>.

TEIXEIRA, N.; ROSÁRIO, A.; BRITES, R. **As Estruturas de Capitais das Empresas Exportadoras para o Mercado Extracomunitário**. *GESTIN-Revista Internacional de Gestão, Direito e Turismo*, v. 24, p. 115-137, 2022.

UTSUMI, Miriam Cardoso; CAZORLA, Irene Mauricio; VENDRAMINI, Claudette Maria Medeiros; MENDES, Clayde Regina. **Questões metodológicas dos trabalhos de abordagem quantitativa apresentados no GT19-ANPEd**. *Educação Matemática em Pesquisa*, São Paulo, v. 9, n. 1, p. 83–101, 2007.

VALDES, M. M. L.; ALEAGA, A. M. L.; GARCÍA VIDAL, G. **Redes neuronales artificiales en la predicción de insolvencia. Un cambio de paradigma ante recetas tradicionales de prácticas empresariales**. *Enfoque UTE*, v. 5, n. 2, p. 38–58, 2014. DOI: <https://doi.org/10.29019/enfoqueute.v5n2.39>.

VÁSQUEZ-SERPA, L.-J., RODRÍGUEZ, C., PÉREZ-NÚÑEZ, J.-R.; NAVARRO, C. **Challenges of Artificial Intelligence for the Prevention and Identification of Bankruptcy Risk in Financial Institutions: A Systematic Review**. *Journal of Risk and Financial Management*, v. 18, n. 1, p. 26, 2025. DOI: <https://doi.org/10.3390/jrfm18010026>.

VIEIRA, M. V. **Administração estratégica do capital de giro**. (2ª ed.). São Paulo: Atlas, 2008.

XHINDI, T.; SHESTANI, K. **Financial Distress and Bankruptcy Prediction: An Empirical Analysis of the Manufacturing Industry in Albania**. *Wseas Transactions on Business and Economics*, v. 17, p. 33–40, 2020. DOI: <https://doi.org/10.37394/23207.2020.17.5>.

XIE, C.; LUO, C.; YU, X. **Financial distress prediction based on SVM and MDA methods: The case of Chinese listed companies**. *Quality & Quantity*, v. 45, n. 3, p. 671–686, 2011. DOI: <https://doi.org/10.1007/s11135-010-9376-y>.

WINTER, N. M. C.; OLIVEIRA, F. C. **O papel da recuperação judicial em face das crises da empresa e o pressuposto de sustentabilidade da atividade empresarial**. *Revista Juridica*. v. 4, n. 33, p. 380-406, 2013. DOI: <https://dx.doi.org/10.26668/revistajur.2316-753X.v4i33.764>.

WYROBEK, J. **Predicting Bankruptcy at Polish Companies: A Comparison of Selected Machine Learning and Deep Learning Algorithms**. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, v. 6, n. 978, p. 41–60, 2018. DOI: <https://doi.org/10.15678/ZNUEK.2018.0978.0603>.

YUVARAJ S.E; A. POONGODI. **Efficient Bankruptcy Prediction using Machine Learning**. *International Journal of Advanced Research in Science, Communication and Technology*, v. 94, p. –100, 2024. DOI: <https://doi.org/10.48175/IJARSCT-18618>.

APÊNDICE A – Código em R para construção da Regressão Logística para Dados em Painel do Artigo 2

Segue abaixo o código elaborado em R para a construção do modelo longitudinal logístico para dados em painel.

Pré-requisitos: importar pacotes

```
library(readxl); library(dplyr); library(tidyr); library(plm); library(lme4)
library(survival); library(lmtest); library(sandwich); library(pROC)
library(caret); library(broom); library(pscl); library(writexl); library(car)
library(ggplot2)
```

0) Diretório de trabalho

```
setwd("C:/Users/lefar/OneDrive/Desktop/Mestrado/9. Dissertação/6. Banco de
dados/Resultados")
cat("Diretório atual:", getwd(), "\n\n")
```

1) Parâmetros de entrada

```
file_path <- "C:/Users/lefar/OneDrive/Desktop/Mestrado/9. Dissertação/6. Banco de
dados/Winsorização 5%/Dados em
Painel/Dados_em_Painel_Winsorizada_variaveis_e_zeros_retirados_correlação_VIF.xlsx"
sheet_name <- "Dados_em_Painel"
dependent_var <- "Y"
id_var <- "Nome"
time_var <- "Data"
train_prop <- 0.7
set.seed(12345)
target_sens <- 0.80
```

2) Leitura e preparação da base de dados

```
df_original <- read_excel(file_path, sheet = sheet_name)
df <- as.data.frame(df_original)

x_cols <- grep("^X\\d+$", names(df), value = TRUE)

df[[id_var]] <- as.factor(df[[id_var]])
df[[time_var]] <- as.integer(df[[time_var]])
df[[dependent_var]] <- as.integer(df[[dependent_var]])
```

```
df_scaled <- df
df_scaled[x_cols] <- scale(df[, x_cols])
write_xlsx(df_scaled, "Dados_Painel_Padronizado_R_FINAL.xlsx")
```

3) Holdout por empresas (70% treino e 30% teste do modelo)

```
companies <- unique(df_scaled[[id_var]])
n_companies <- length(companies)
n_train_companies <- floor(n_companies * train_prop)
train_companies <- sample(companies, n_train_companies)

train_df <- df_scaled %>% filter(.data[[id_var]] %in% train_companies)
test_df <- df_scaled %>% filter(!.data[[id_var]] %in% train_companies)

cat("Empresas (treino/teste):", length(unique(train_df[[id_var]])), "/",
length(unique(test_df[[id_var]])), "\n")
```

4) Modelos: Pooled, RE e FE

```
formula_all <- as.formula(paste(dependent_var, "~", paste(x_cols, collapse = " + ")))
formula_null <- as.formula(paste(dependent_var, "~ 1"))

pooled_logit <- glm(formula_all, data = train_df, family = binomial(link = "logit"))
pooled_null <- glm(formula_null, data = train_df, family = binomial(link = "logit"))

cat("=== Pooled Logit ===\n"); print(summary(pooled_logit))
cat("\nLR Test:\n"); print(lmtest::lrtest(pooled_logit, pooled_null))
cat("\nMcFadden R²:", pscl::pR2(pooled_logit)["McFadden"], "\n")

panel_train <- pdata.frame(train_df, index = c(id_var, time_var))
lpm_fixed <- plm(formula_all, data = panel_train, model = "within")
lpm_random <- plm(formula_all, data = panel_train, model = "random")

plm::phtest(lpm_fixed, lpm_random)

cat("\nBreusch-Pagan (LM):\n"); print(plm::plmtest(lpm_pooling, type="bp"))
cat("\nHausman (FE vs RE):\n"); print(plm::phtest(lpm_fixed, lpm_random))

cat("\n=== Random Effects Logit (glmer) ===\n")
form_glmer <- as.formula(paste(dependent_var, "~", paste(x_cols, collapse = " + "), "+ (1 |",
id_var, ")"))
glmer_fit <- glmer(form_glmer, data = train_df, family = binomial(link="logit"),
control = glmerControl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
print(summary(glmer_fit))

cat("\n=== Fixed Effects Logit (clogit) ===\n")
clogit_formula <- as.formula(paste(dependent_var, "~", paste(x_cols, collapse=" + "), "+
strata(", id_var, ")"))
clogit_fit <- clogit(clogit_formula, data = train_df)
```

```
print(summary(clogit_fit))
```

5) Predição e Curva ROC

```
test_df$pred_prob <- predict(glmer_fit, newdata = test_df, type = "response",
allow.new.levels = TRUE)
roc_obj <- pROC::roc(test_df[[dependent_var]], test_df$pred_prob, quiet = TRUE)
auc_val <- pROC::auc(roc_obj)

roc_df <- data.frame(Especificidade = 1 - roc_obj$specificities,
Sensibilidade = roc_obj$sensitivities
)

gg_roc <- ggplot(roc_df, aes(x = Especificidade, y = Sensibilidade)) +
geom_line(color = "darkgreen", size = 1.2) +
geom_abline(intercept = 0, slope = 1, color = "orange", linetype = "dashed", size = 1) +
labs(
title = "Curva ROC - Regressão Logística para Dados em Painel",
subtitle = paste0("AUC = ", round(auc_val, 4)),
x = "Especificidade (Taxa de Verdadeiros Negativos)",
y = "Sensibilidade (Taxa de Verdadeiros Positivos)"
) +
coord_cartesian(xlim = c(0,1), ylim = c(0,1)) +
theme_minimal(base_size = 13) +
theme(
plot.title = element_text(face = "bold", hjust = 0.5),
plot.subtitle = element_text(hjust = 0.5),
panel.grid = element_blank(),
axis.line = element_line(color = "black")
)

print(gg_roc)
ggsave("Curva_ROC_Regressao_Logistica_Painel.png", plot = gg_roc, width = 8, height = 6,
dpi = 300)
cat("Gráfico da Curva ROC salvo com sucesso.\n\n")
```

6) Métricas de Desempenho

```
roc_coords <- coords(roc_obj, x = "all", ret = c("threshold", "sensitivity", "specificity", "ppv",
"npv"), transpose = FALSE)
roc_df_full <- as.data.frame(roc_coords) %>% mutate(fl =
2*(ppv*sensitivity)/(ppv+sensitivity))
chosen_cutoff <- roc_df_full %>% filter(sensitivity >= target_sens) %>%
arrange(desc(specificity)) %>% slice(1) %>% pull(threshold)
if(is.na(chosen_cutoff)) chosen_cutoff <- roc_df_full %>% mutate(youden = sensitivity +
specificity - 1) %>% arrange(desc(youden)) %>% slice(1) %>% pull(threshold)

test_df$pred_class <- ifelse(test_df$pred_prob >= chosen_cutoff, 1, 0)
```

```
cm_table <- table(Real = test_df[[dependent_var]], Predito = test_df$pred_class)
conf <- caret::confusionMatrix(as.factor(test_df$pred_class),
as.factor(test_df[[dependent_var]]), positive = "1")
```

7) Tabela de Desempenho Consolidada

```
VN <- cm_table["0","0"]; FP <- cm_table["0","1"]; FN <- cm_table["1","0"]; VP <-
cm_table["1","1"]
accuracy <- conf$overall["Accuracy"]
precision <- conf$byClass["Precision"]
recall <- conf$byClass["Recall"]
specificity <- conf$byClass["Specificity"]
f1 <- conf$byClass["F1"]

metrics_summary <- data.frame(
  Métrica = c("AUC", "Acurácia", "Precisão", "Sensibilidade", "Especificidade", "F1-Score",
"VP", "FP", "VN", "FN", "Cutoff"),
  Valor = c(round(as.numeric(auc_val),4),
            round(as.numeric(accuracy),4),
            round(as.numeric(precision),4),
            round(as.numeric(recall),4),
            round(as.numeric(specificity),4),
            round(as.numeric(f1),4),
            VP, FP, VN, FN,
            round(as.numeric(chosen_cutoff),4))
)
print(metrics_summary)

write_xlsx(list(
  "Resumo_Preditivo" = metrics_summary,
  "Matriz_Confusao" = as.data.frame.matrix(cm_table),
  "ROC_Thresholds" = roc_df_full
), "Tabela_Desempenho_Preditivo_RE_Logit_FINAL.xlsx")
```

8) Exportações de Coeficientes

```
coefs_glmer <- as.data.frame(coef(summary(glmer_fit)))
coefs_glmer$term <- rownames(coefs_glmer)
coefs_glmer <- coefs_glmer[, c("term", "Estimate", "Std. Error", "z value", "Pr(>|z|)")]

coef_list <- list(
  "pooled" = broom::tidy(pooled_logit, conf.int = TRUE),
  "glmer" = coefs_glmer,
  "clogit" = broom::tidy(clogit_fit, conf.int = TRUE)
)
write_xlsx(coef_list, "Coeficientes_Modelos_R_FINAL.xlsx")

cat("\n Todos os arquivos gerados com sucesso em:\n", getwd(), "\n")
```

```

cat("- Curva_ROC_Regressao_Logistica_Painel.png\n")
cat("- Tabela_Desempenho_Preditivo_RE_Logit_FINAL.xlsx\n")
cat("- Coeficientes_Modelos_R_FINAL.xlsx\n")
cat("- Resultados_Preditivos_R_FINAL.xlsx\n")

```

9) Tabela Comparativa: Pooled x RE x FE (logLik, AIC, McFadden R²)

```

suppressWarnings({
# --- Pooled (GLM)
ll_pooled <- as.numeric(logLik(pooled_logit))
ll_pooled0 <- as.numeric(logLik(pooled_null))
aic_pooled <- AIC(pooled_logit)
r2_pooled <- 1 - (ll_pooled / ll_pooled0)

# --- Modelo de Efeitos Aleatórios
glmer_null <- tryCatch(
  lme4::glmer(
    as.formula(paste(dependent_var, "~ 1 + (1 |", id_var, ")")),
    data = train_df, family = binomial(link = "logit"),
    control = glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e5))
  ),
  error = function(e) NULL
)

ll_re <- as.numeric(logLik(glmer_fit))
aic_re <- AIC(glmer_fit)
ll_re0 <- if (!is.null(glmer_null)) as.numeric(logLik(glmer_null)) else NA_real_
r2_re <- if (!is.na(ll_re0)) 1 - (ll_re / ll_re0) else NA_real_

# --- Modelo de Efeitos Fixos
if (!is.null(clogit_fit)) {
  ll_fe0 <- as.numeric(clogit_fit$loglik[1])
  ll_fe <- as.numeric(clogit_fit$loglik[2])
  aic_fe <- AIC(clogit_fit)
  r2_fe <- 1 - (ll_fe / ll_fe0)
} else {
  ll_fe0 <- ll_fe <- aic_fe <- r2_fe <- NA_real_
}

# --- Monta a tabela
comp_tbl <- data.frame(
  Modelo = c("Pooled Logit (GLM)", "Efeitos Aleatórios (GLMM)", "Efeitos Fixos
(Conditional Logit)"),
  logLik = round(c(ll_pooled, ll_re, ll_fe), 4),
  AIC = round(c(aic_pooled, aic_re, aic_fe), 3),
  McFadden_R2 = round(c(r2_pooled, r2_re, r2_fe), 4),
  Obs_Treino = rep(nrow(train_df), 3),
  Empresas_Treino = rep(length(unique(train_df[[id_var]])), 3),
  Nota = c(

```

```
"R2 de McFadden vs. GLM nulo",
"R2 de McFadden vs. GLMM nulo (intercepto + efeito aleatório)",
"R2 de McFadden vs. partial-likelihood nula (comparabilidade limitada)"
),
check.names = FALSE
)

print(comp_tbl)

# --- Salva em Excel
writexl::write_xlsx(
  list("Comparativo_Modelos" = comp_tbl),
  "Comparativo_Modelos_Logit_Painel.xlsx"
)

cat("\n Tabela comparativa salva em: 'Comparativo_Modelos_Logit_Painel.xlsx'\n")
})
```

APÊNDICE B – Código em Python para elaboração dos modelos de aprendizado de máquina do Artigo 3

Abaixo encontra-se o código em Python elaborado para a construção dos modelos de aprendizado de máquina, sendo eles a *Decision Tree*, *Random Forest*, *Support Vector Machine*, *K-Nearest Neighbors* e *Extreme Gradient Boosting*.

1) Instalação das bibliotecas adicionais (apenas no Colab)

```
!pip install -q xgboost
```

2) Importação de bibliotecas básicas e configuração de alertas

```
import os, warnings
warnings.filterwarnings("ignore")
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

3) Funções do scikit-learn para modelagem, pré-processamento e métricas

```
from sklearn.model_selection import TimeSeriesSplit, RandomizedSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.metrics import (
    accuracy_score, precision_score, recall_score, f1_score,
    roc_auc_score, roc_curve, log_loss, brier_score_loss,
    classification_report, confusion_matrix,
    precision_recall_curve, average_precision_score
)
from sklearn.utils.class_weight import compute_class_weight, compute_sample_weight
from sklearn.utils.validation import has_fit_parameter
from sklearn.inspection import permutation_importance
from scipy.stats import spearmanr

from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.svm import SVC
```

```

from sklearn.calibration import CalibratedClassifierCV
from sklearn.base import clone
from inspect import signature

```

4) Importação opcional de pacote do XGBoost

```

# - Se não estiver instalado, o código segue sem travar
try:
    from xgboost import XGBClassifier
    XGB_AVAILABLE = True
except Exception:
    XGB_AVAILABLE = False
    print("XGBoost não disponível. Instale com: pip install xgboost")

```

5) Configurações gerais do experimento

```

RANDOM_STATE = 42
TEST_GROUP_FRACTION = 0.30    # % de empresas destinadas ao conjunto de teste
N_SPLITS_CV = 5                # Número de folds para TimeSeriesSplit
FIGSIZE = (8, 6)

# Flags de plots/tuning
ENABLE_CLASS_DIST_PLOT = True
ENABLE_PR_CURVE = True
ENABLE_TUNING_NN_XGB = True

```

6) Parâmetros da base de dados (caminho, planilha e colunas-chave)

```

EXCEL_PATH = "r:/content/Dados_em_Painel_Winsorizada_variaveis_e_zeros_retirados_correlação_VIF.xlsx"
SHEET_NAME = "Dados_em_Painel"
TARGET_COL = "Y"
GROUP_COL = "Nome" # identifica a empresa
TIME_COL = "Data" # ano ou data (série temporal)

```

7) Funções auxiliares para tratamento de dados e métricas

```

def ensure_datetime_or_numeric(s: pd.Series) -> pd.Series:
    """
    Garante que a coluna seja convertida para datetime ou numérica,

```

```

evitando problemas de tipo na ordenação temporal.
"""
if np.issubdtype(s.dtype, np.number):
    return s
s_dt = pd.to_datetime(s, errors="coerce", dayfirst=True)
if s_dt.notna().any():
    return s_dt
return pd.to_numeric(s, errors="coerce")

def ks_statistic(y_true, y_score):
    """
    Calcula a estatística KS (Kolmogorov-Smirnov) entre as distribuições
    de score das classes positiva e negativa.
    """
    y_true = np.asarray(y_true); y_score = np.asarray(y_score)
    pos = y_score[y_true == 1]; neg = y_score[y_true == 0]
    vals = np.sort(np.unique(y_score))
    if len(pos) == 0 or len(neg) == 0 or len(vals) == 0:
        return np.nan
    cdf_pos = np.searchsorted(np.sort(pos), vals, side='right') / len(pos)
    cdf_neg = np.searchsorted(np.sort(neg), vals, side='right') / len(neg)
    return float(np.max(np.abs(cdf_pos - cdf_neg)))

def proba_or_score(model, X_):
    """
    Retorna probabilidade da classe positiva (quando possível) ou
    um score normalizado entre 0 e 1.
    """
    if hasattr(model, "predict_proba"):
        return model.predict_proba(X_)[:, 1]
    elif hasattr(model, "decision_function"):
        scores = model.decision_function(X_)
        s_min, s_max = scores.min(), scores.max()
        if s_max == s_min:
            return np.zeros_like(scores) + 0.5
        return (scores - s_min) / (s_max - s_min)
    else:
        return model.predict(X_).astype(float)

def split_holdout_by_group(df, group_col, test_frac=0.3, random_state=42):
    """
    Realiza a divisão holdout por empresa (group-wise), garantindo que
    todas as observações de uma mesma empresa fiquem em treino ou teste.
    """
    rng = np.random.RandomState(random_state)
    all_groups = df[group_col].dropna().unique()
    test_groups = rng.choice(all_groups, size=max(1, int(len(all_groups)*test_frac)),
replace=False)
    mask = df[group_col].isin(test_groups)
    return df.loc[~mask].copy(), df.loc[mask].copy(), test_groups

```

```

def time_series_cv_sorted(df_train, time_col, n_splits):
    """
    Ordena o conjunto de treino no tempo e cria um TimeSeriesSplit
    para validação cruzada respeitando a dimensão temporal.
    """
    df_sorted = df_train.sort_values(by=time_col)
    uniq = pd.Series(df_sorted[time_col].unique()).sort_values().values
    n_splits = max(2, min(n_splits, len(uniq)-1)) if len(uniq) > 1 else 2
    return df_sorted, TimeSeriesSplit(n_splits=n_splits)

def get_preproc_and_estimator(fitted_pipeline):
    """
    Separa o pipeline em parte de pré-processamento e estimador final (clf),
    facilitando o uso em importâncias nativas quando aplicável.
    """
    if isinstance(fitted_pipeline, Pipeline):
        steps = fitted_pipeline.named_steps
        preproc_steps = [(k,v) for k,v in fitted_pipeline.steps if k != "clf"]
        preproc = Pipeline(preproc_steps) if preproc_steps else None
        estimator = steps.get("clf", fitted_pipeline)
        return preproc, estimator
    return None, fitted_pipeline

def split_summary_groups(df_all, df_tr, df_te, group_col):
    """
    Gera um resumo da divisão treino/teste por quantidade de linhas,
    número de empresas e proporção de insolventes.
    """
    def counts(part):
        return pd.Series({
            "N_linhas": len(part),
            "N_empresas": part[group_col].nunique(),
            "Pos(1)": int((part[TARGET_COL] == 1).sum()),
            "Neg(0)": int((part[TARGET_COL] == 0).sum()),
            "Pct_Pos": (part[TARGET_COL] == 1).mean()
        })
    out = pd.concat({"TOTAL": counts(df_all), "TREINO": counts(df_tr), "TESTE":
counts(df_te)}, axis=1).T
    return out.round(4)

```

8) Leitura da base, tratamento inicial e seleção das variáveis X

```

df = pd.read_excel(EXCEL_PATH, sheet_name=SHEET_NAME)
df.rename(columns=lambda c: c.strip(), inplace=True)
df[TIME_COL] = ensure_datetime_or_numeric(df[TIME_COL])

```

Seleciona todas as variáveis X* (sem excluir X18/X19, pois já foram tratadas na base)

```

feature_cols = [c for c in df.columns if c.upper().startswith("X")]
if not feature_cols:
    raise ValueError("Nenhuma variável X* encontrada na base.")

print(f"Variáveis utilizadas ({len(feature_cols)}): {feature_cols}")

```

9) Análise exploratória básica da variável resposta (distribuição de classes)

```

print("=== Distribuição das classes (0=solvente, 1=insolvente) ===")
print(df[TARGET_COL].value_counts(dropna=False))
print("Proporção de positivos:", round(df[TARGET_COL].mean(), 4))

if ENABLE_CLASS_DIST_PLOT:
    plt.figure(figsize=(6, 4))
    df[TARGET_COL].value_counts().sort_index().plot(kind="bar")
    plt.title("Distribuição das Classes (0=Solvente, 1=Insolvente)")
    plt.xlabel("Classe"); plt.ylabel("Quantidade"); plt.grid(axis="y", linestyle=":")
    plt.show()

```

10) Criação do holdout por empresa (treino x teste) e resumo

```

df_train, df_test, test_groups = split_holdout_by_group(df, GROUP_COL,
TEST_GROUP_FRACTION, RANDOM_STATE)
split_table = split_summary_groups(df, df_train, df_test, GROUP_COL)
print("\n=== Divisão por grupos (empresa) ===")
print(split_table)
print(f"Empresas no teste ({len(test_groups)}):", list(test_groups)[:10], "... " if len(test_groups)
> 10 else "")

# Ordena treino por tempo
df_train_sorted, tss = time_series_cv_sorted(df_train, TIME_COL, N_SPLITS_CV)

# Criação dos dataframes finais de treino e teste
X_train_df = df_train[feature_cols].copy()
y_train_sr = df_train[TARGET_COL].astype(int).copy()
X_test_df = df_test[feature_cols].copy()
y_test_sr = df_test[TARGET_COL].astype(int).copy()

```

11) Cálculo dos pesos de classe (amostra desbalanceada)

```

classes = np.unique(y_train_sr)
class_weights = compute_class_weight(class_weight="balanced", classes=classes,
y=y_train_sr)
cw_dict = {int(c): w for c, w in zip(classes, class_weights)}

```

```
sample_weight_train = compute_sample_weight(class_weight="balanced", y=y_train_sr)
```

12) Definição dos modelos de classificação e pipelines

```
imputer = SimpleImputer(strategy="median")
```

```
# Decision Tree
```

```
dt = Pipeline([
    ("imputer", imputer),
    ("clf", DecisionTreeClassifier(
        random_state=RANDOM_STATE, class_weight="balanced",
        max_depth=None, min_samples_leaf=2
    ))
])
```

```
# Random Forest
```

```
rf = Pipeline([
    ("imputer", imputer),
    ("clf", RandomForestClassifier(
        n_estimators=600, random_state=RANDOM_STATE,
        class_weight="balanced_subsample", n_jobs=-1, min_samples_leaf=2
    ))
])
```

```
# Extreme Gradient Boosting com atribuição e pesos para classes insolventes
if XGB_AVAILABLE:
```

```
    neg = (y_train_sr == 0).sum()
    pos = (y_train_sr == 1).sum()
    spw = max(1.0, neg / max(1, pos))
    xgb = Pipeline([
        ("imputer", imputer),
        ("clf", XGBClassifier(
            objective="binary:logistic", eval_metric="logloss",
            n_estimators=800, learning_rate=0.05, max_depth=4,
            subsample=0.8, colsample_bytree=0.8,
            reg_lambda=1.0, random_state=RANDOM_STATE,
            scale_pos_weight=spw, n_jobs=-1, tree_method="hist"
        ))
    ])
```

```
else:
```

```
    xgb = None
```

```
# Support Vector Machine
```

```
svm = Pipeline([
    ("imputer", imputer),
    ("scaler", StandardScaler()),
    ("clf", SVC(
        kernel="rbf", probability=True,

```

```

        class_weight="balanced", random_state=RANDOM_STATE
    ))
])

# K-Nearest Neighbors
knn = Pipeline([
    ("imputer", imputer),
    ("scaler", StandardScaler()),
    ("clf", KNeighborsClassifier(n_neighbors=15, weights="distance", metric="minkowski"))
])

# Multilayer Perceptrons
mlp = Pipeline([
    ("imputer", imputer),
    ("scaler", StandardScaler()),
    ("clf", MLPClassifier(
        hidden_layer_sizes=(64, 32), activation="relu",
        solver="adam", alpha=1e-3, learning_rate="adaptive",
        learning_rate_init=1e-3, max_iter=2000,
        early_stopping=True, random_state=RANDOM_STATE
    ))
])

# Dicionário com todos os modelos
models = {
    "Decision Tree": dt,
    "Random Forest": rf,
    "SVM (RBF)": svm,
    "KNN": knn,
    "MLP Neural Net": mlp
}
if xgb is not None:
    models["XGBoost"] = xgb

```

13) Tuning da MLP e do XGBoost

```

df_train_sorted, tss = time_series_cv_sorted(df_train, TIME_COL, N_SPLITS_CV)
X_train_sorted = df_train_sorted[feature_cols]
y_train_sorted = df_train_sorted[TARGET_COL].astype(int)

best_params_log = {}
if ENABLE_TUNING_NN_XGB:
    # MLP – RandomizedSearch
    mlp_param_dist = {
        "clf__hidden_layer_sizes": [(64,), (64,32), (128,64), (100,)],
        "clf__alpha": np.logspace(-4, -1, 8),
        "clf__learning_rate_init": np.logspace(-4, -2, 8),
        "clf__activation": ["relu", "tanh"]
    }

```

```

}
mlp_rs = RandomizedSearchCV(
    estimator=mlp, param_distributions=mlp_param_dist,
    n_iter=20, scoring="roc_auc", cv=tss, n_jobs=-1,
    random_state=RANDOM_STATE, verbose=1
)
print("\n Tuning MLP (TimeSeriesSplit)...")
mlp_rs.fit(X_train_sorted, y_train_sorted)
models["MLP Neural Net"] = mlp_rs.best_estimator_
best_params_log["MLP Neural Net"] = mlp_rs.best_params_

# XGBoost – RandomizedSearch
if xgb is not None:
    xgb_param_dist = {
        "clf__n_estimators": np.arange(400, 1201, 100),
        "clf__learning_rate": np.linspace(0.02, 0.15, 8),
        "clf__max_depth": [3, 4, 5, 6],
        "clf__subsample": np.linspace(0.6, 0.9, 4),
        "clf__colsample_bytree": np.linspace(0.6, 0.9, 4),
        "clf__reg_lambda": np.linspace(0.5, 2.0, 8)
    }
    xgb_rs = RandomizedSearchCV(
        estimator=xgb, param_distributions=xgb_param_dist,
        n_iter=25, scoring="roc_auc", cv=tss, n_jobs=-1,
        random_state=RANDOM_STATE, verbose=1
    )
    print("\n Tuning XGBoost (TimeSeriesSplit)...")
    xgb_rs.fit(X_train_sorted, y_train_sorted)
    models["XGBoost"] = xgb_rs.best_estimator_
    best_params_log["XGBoost"] = xgb_rs.best_params_

```

14) Construção de versões calibradas (isotonic) para modelos ensemble

```

def build_calibrated_version(model_name, pipeline_model, tss_cv):
    """
    Constrói um pipeline calibrado (CalibratedClassifierCV) preservando
    os passos de pré-processamento do modelo original.
    """
    preproc, est = get_preproc_and_estimator(pipeline_model)
    est_cloned = clone(est)
    params = signature(CalibratedClassifierCV).parameters
    if "estimator" in params:
        cal = CalibratedClassifierCV(estimator=est_cloned, method="isotonic", cv=tss_cv)
    else:
        cal = CalibratedClassifierCV(base_estimator=est_cloned, method="isotonic", cv=tss_cv)

    steps = []
    if preproc is not None:

```

```

        steps.extend(preproc.steps)
        steps.append(("clf", cal))
        return model_name, Pipeline(steps)

models_cal = {}

targets_to_calibrate = ("Decision Tree", "Random Forest", "XGBoost")

for base_name in list(models.keys()):
    if base_name in targets_to_calibrate and base_name in models:
        clean_name, cal_pipe = build_calibrated_version(base_name, models[base_name], tss)
        models_cal[clean_name] = cal_pipe

# Remover as versões NÃO calibradas
for k in targets_to_calibrate:
    if k in models:
        del models[k]

# Adicionar as versões calibradas (nomes sem "(Cal)")
models.update(models_cal)

print("\nModelos calibrados mantidos (sem sufixo):")
for k in models_cal.keys():
    print(" -", k)

```

15) Funções de avaliação

```

def evaluate_model(name, model, X_train, y_train, X_test, y_test, sw_train=None):
    """
    Treina o modelo, gera predições no teste e calcula todas as métricas:
    Acurácia, Precisão, Recall, F1, AUC, Gini, KS, Brier, LogLoss, PR_AUC,
    Especificidade, matriz de confusão e relatório de classificação.
    """
    fit_kwargs = {}
    if sw_train is not None and has_fit_parameter(model, "sample_weight"):
        fit_kwargs["sample_weight"] = sw_train

    model.fit(X_train, y_train, **fit_kwargs)
    y_pred = model.predict(X_test)
    y_score = proba_or_score(model, X_test)

    acc = accuracy_score(y_test, y_pred)
    prec = precision_score(y_test, y_pred, zero_division=0)
    rec = recall_score(y_test, y_pred, zero_division=0)
    fl = f1_score(y_test, y_pred, zero_division=0)
    try:
        auc = roc_auc_score(y_test, y_score)
    except Exception:

```

```

    auc = np.nan
    gini = 2*auc - 1 if np.isfinite(auc) else np.nan
    try:
        ll = log_loss(y_test, y_score, labels=[0, 1])
    except Exception:
        ll = np.nan
    try:
        brier = brier_score_loss(y_test, y_score)
    except Exception:
        brier = np.nan
    ks = ks_statistic(y_test, y_score)

# Curvas ROC
try:
    fpr, tpr, _ = roc_curve(y_test, y_score)
except Exception:
    fpr, tpr = np.array([0, 1]), np.array([0, 1])

# Curvas Precisão-Revocação
try:
    prec_curve, rec_curve, _ = precision_recall_curve(y_test, y_score)
    ap = average_precision_score(y_test, y_score)
except Exception:
    prec_curve, rec_curve, ap = np.array([1, 0]), np.array([0, 1]), np.nan

cm = confusion_matrix(y_test, y_pred, labels=[0, 1])
tn, fp, fn, tp = cm.ravel()
specificity = tn / (tn + fp) if (tn + fp) > 0 else 0.0

report = classification_report(
    y_test, y_pred,
    target_names=["Solvente(0)", "Insolvente(1)"],
    zero_division=0
)

metrics_row = {
    "Modelo": name, "Accuracy": acc, "Precision": prec, "Recall": rec, "F1": f1,
    "AUC": auc, "Gini": gini, "KS": ks, "Brier": brier, "LogLoss": ll,
    "PR_AUC(AP)": ap, "Specificity": specificity, "VP": tp, "FP": fp, "VN": tn, "FN": fn
}
return metrics_row, (fpr, tpr), (prec_curve, rec_curve), cm, report, model

def plot_roc_single(name, fpr, tpr, auc_value):
    """Plota a curva ROC de um único modelo."""
    plt.figure(figsize=FIGSIZE)
    plt.plot(fpr, tpr, label=f" {name} (AUC = {auc_value:.3f})")
    plt.plot([0, 1], [0, 1], linestyle="--", label="Aleatório")
    plt.xlabel("FPR"); plt.ylabel("TPR")
    plt.title(f"Curva ROC - {name}")
    plt.legend(loc="lower right"); plt.grid(True, linestyle=":")

```

```

plt.show()

def plot_pr_single(name, precision, recall, ap_value):
    """Plota a curva Precisão–Revocação de um único modelo."""
    plt.figure(figsize=FIGSIZE)
    plt.plot(recall, precision, label=f" {name} (AP = {ap_value:.3f})")
    plt.xlabel("Recall"); plt.ylabel("Precision")
    plt.title(f"Curva Precisão–Revocação - {name}")
    plt.legend(loc="upper right"); plt.grid(True, linestyle=":")
    plt.show()

def plot_roc_combined(curves):
    """Plota as curvas ROC de todos os modelos em um único gráfico comparativo."""
    plt.figure(figsize=FIGSIZE)
    for (name, (fpr, tpr), auc_value) in curves:
        plt.plot(fpr, tpr, label=f" {name} (AUC = {auc_value:.3f})")
    plt.plot([0, 1], [0, 1], linestyle="--", color='gray', label='_nolegend_')
    plt.xlabel("Especificidade (Taxa de Verdadeiros Negativos)")
    plt.ylabel("Sensibilidade (Taxa de Verdadeiros Positivos)")
    plt.title("Curvas ROC - Comparação de Modelos")
    plt.legend(loc="lower right"); plt.grid(True, linestyle=":")
    plt.show()

```

16) Loop principal de treino, avaliação e explicabilidade (Permutation Importance e Importâncias nativas)

```

metrics_list, roc_curves = [], []
perm_rows, native_rows = [], []

print("\n=== Treinando e avaliando (holdout por empresa) ===")
for name, model in models.items():
    # Avalia cada modelo no hold-out
    mrow, (fpr, tpr), (prec_curve, rec_curve), cm, report, fitted = evaluate_model(
        name, model, X_train_df, y_train_sr, X_test_df, y_test_sr, sample_weight_train
    )
    metrics_list.append(mrow)
    auc_value = mrow["AUC"] if np.isfinite(mrow["AUC"]) else 0.0
    roc_curves.append((name, (fpr, tpr), auc_value))

    print(f"\n--- {name} ---")

print(pd.DataFrame([mrow]).set_index("Modelo")[['Accuracy','AUC','Gini','KS']].round(4))
print(pd.DataFrame(cm, index=["Real 0", "Real 1"], columns=["Pred 0", "Pred 1"]))
print(report)
plot_roc_single(name, fpr, tpr, auc_value)

```

```
if ENABLE_PR_CURVE:
    plot_pr_single(name, prec_curve, rec_curve, mrow.get("PR_AUC(AP)", np.nan))
```

16.1) Importância por permutação (Permutation Importance)

```
try:
    pi = permutation_importance(
        fitted, X_test_df, y_test_sr,
        scoring="roc_auc", n_repeats=10,
        random_state=RANDOM_STATE, n_jobs=-1
    )
    y_score_test = proba_or_score(fitted, X_test_df)

    for j, feat in enumerate(feature_cols):
        mean_imp = float(pi.importances_mean[j])
        std_imp = float(pi.importances_std[j])
        try:
            rho, _ = spearmanr(X_test_df[feat].values, y_score_test)
            sign = np.sign(rho) if np.isfinite(rho) else np.nan
        except Exception:
            sign = np.nan

        perm_rows.append({
            "Modelo": name, "Feature": feat,
            "PermImportance_mean": mean_imp, "PermImportance_std": std_imp,
            "Sign_via_Spearman(X,score)": int(sign) if np.isfinite(sign) else np.nan
        })
except Exception as e:
    print(f"[{name}] Permutation importance falhou: {e}")
```

16.2) Importâncias nativas de modelos baseados em árvore

```
try:
    preproc, est = get_preproc_and_estimator(fitted)
    if hasattr(est, "feature_importances_"):
        for feat, val in zip(feature_cols, est.feature_importances_):
            native_rows.append({"Modelo": name, "Feature": feat, "NativeImportance":
float(val)})
    except Exception:
        pass
```

17) Tabela comparativa de desempenho entre modelos

```
metrics_df
pd.DataFrame(metrics_list).set_index("Modelo").round(4).sort_values(by="AUC",
ascending=False)
print("\n=== Tabela Comparativa Geral (ordenada por AUC) ===")
print(metrics_df)
```

```

summary_df = pd.DataFrame(metrics_list)
summary_df.rename(columns={"Recall": "Sensibilidade", "Accuracy": "Acurácia",
"Precision": "Precisão", "F1": "F1-Score"}, inplace=True)
requested_cols = ["Modelo", "AUC", "Acurácia", "Precisão", "Sensibilidade", "Specificity",
"F1-Score", "VP", "FP", "VN", "FN"]
final_cols = [col for col in requested_cols if col in summary_df.columns]
summary_table = summary_df[final_cols].set_index("Modelo")
pd.options.display.float_format = '{:.4f}'.format
summary_table_sorted = summary_table.sort_values(by="AUC", ascending=False)

# Curva ROC combinada com todos os modelos
plot_roc_combined([(n, (fpr, tpr), auc) for (n, (fpr, tpr), auc) in roc_curves])

# DataFrames de importâncias
perm_df = pd.DataFrame(perm_rows)
native_df = pd.DataFrame(native_rows)

```

18) Impressão detalhada das importâncias de variáveis por modelo

```

print("\n\n" + "="*60)
print("==== IMPORTÂNCIA DAS VARIÁVEIS (PARA CADA MODELO) ====")
print("="*60)

for model_name in models.keys():
    print(f"\n\n{'-'*25} {model_name.upper()} {'-'*25}")
    print(f"\n>>> Permutation Importance (Agnóstico do Modelo)")
    if not perm_df.empty and model_name in perm_df["Modelo"].unique():
        perm_imp_subset = perm_df[perm_df["Modelo"] == model_name].copy()
        perm_imp_subset = perm_imp_subset.sort_values(by="PermImportance_mean",
ascending=False)
        print(perm_imp_subset[['Feature', 'PermImportance_mean',
'Sign_via_Spearman(X,score)']]
.reset_index(drop=True).round(5))
    else:
        print(f"\nNão foram encontradas importâncias de permutação para '{model_name}'.")

    print(f"\n>>> Importância Nativa (apenas árvores)")
    if not native_df.empty:
        native_subset = native_df[native_df["Modelo"] == model_name].copy()
        if not native_subset.empty:
            final_importance = native_subset["NativeImportance"].mean().reset_index()
            final_importance = final_importance.sort_values(by="NativeImportance",
ascending=False)
            print(final_importance.reset_index(drop=True).round(5))
        else:
            print(f"\nNão foram encontradas importâncias nativas para '{model_name}'.")

```

```

else:
    print("Nenhuma importância nativa foi calculada no script.")

```

19) Construção da tabela resumo Top-5 variáveis por modelo

```

print("\n\n" + "="*70)
print("==== Tabela Resumo: Top 5 Variáveis por Modelo (Permutation Importance) ====")
print("="*70)

if not perm_df.empty:
    positive_imp = perm_df[perm_df['PermImportance_mean'] > 0].copy()
    positive_imp = positive_imp.sort_values(by=['Modelo', 'PermImportance_mean'],
ascending=[True, False])
    top_vars_df = positive_imp.groupby('Modelo').head(5)

    top_vars_df['Variavel (Impacto na AUC)'] = top_vars_df.apply(
        lambda row: f'{row["Feature"]} ({row["PermImportance_mean"]:.4f})', axis=1
    )
    top_vars_df['Posição'] = top_vars_df.groupby('Modelo').cumcount() + 1

    posicao_map = {
        1: '1ª Principal',
        2: '2ª Principal',
        3: '3ª Principal',
        4: '4ª Principal',
        5: '5ª Principal'
    }
    top_vars_df['Posição'] = top_vars_df['Posição'].map(posicao_map)

    summary_top3 = top_vars_df.pivot_table(
        index='Posição', columns='Modelo', values='Variavel (Impacto na AUC)', aggfunc='first'
    ).reindex(posicao_map.values())

    summary_top3.fillna('-', inplace=True)
    print(summary_top3)
else:
    print("DataFrame de Permutation Importance ('perm_df') está vazio. Tabela não pôde ser gerada.")

```

20) Exportação dos resultados consolidados para arquivo Excel

```

out_xlsx = "resultado_modelos_credito_painel_forward_split.xlsx"
with pd.ExcelWriter(out_xlsx) as writer:
    split_table.to_excel(writer, sheet_name="Split_Grupos")
    metrics_df.to_excel(writer, sheet_name="Metricas_Gerais")
    summary_table_sorted.to_excel(writer, sheet_name="Metricas_Especificas")

```

```

if not perm_df.empty:
    perm_df["rank_desc"]
perm_df.groupby("Modelo")["PermImportance_mean"].rank(ascending=False,
method="dense")
    perm_df.sort_values(["Modelo", "rank_desc", "Feature"]).to_excel(
        writer, sheet_name="Importancias_Permutation", index=False
    )

if not native_df.empty:
    native_df["rank_desc"]
native_df.groupby("Modelo")["NativeImportance"].rank(ascending=False, method="dense")
    native_df.sort_values(["Modelo", "rank_desc", "Feature"]).to_excel(
        writer, sheet_name="Importancias_Nativas", index=False
    )

if ENABLE_TUNING_NN_XGB and best_params_log:
    pd.DataFrame(best_params_log).to_excel(writer, sheet_name="Best_Params")

if not perm_df.empty:
    summary_top3.to_excel(writer, sheet_name="Top5_Variaveis")

print(f"\nRelatório salvo em: {out_xlsx}")

```