PREDIÇÃO DE FALÊNCIA EMPRESARIAL COM SPARSE PLS-DA E SMOTE: UMA ABORDAGEM CONTABILOMÉTRICA PARA DADOS DESBALANCEADOS

Alan Cristian Maçulo de Carvalho[1], Erik Silva dos Santos[1], Igor Campos de Almeida Lima[1], Yuri Duarte Porto[2], Nelson Cevidanes Nascimento de Assis[3], Soraya de Oliveira Bandeira[4]

[1]Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil
(alanuerj@hotmail.com)
[2]Universidade Federal de Mato Grosso, Cuiabá, Brasil
[3]Universidade Federal de Juiz de Fora, Juiz de Fora, Brasil
[4]COPPE/Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil

Resumo: A falência empresarial é um evento raro, mas de alto impacto para credores, investidores e trabalhadores, tornando sua identificação antecipada analiticamente relevante. Avaliou-se a sparse PLS-DA combinada ao SMOTE na base Taiwanese Bankruptcy Prediction. Com duas variáveis latentes, seleção esparsa de indicadores e regra de decisão por distância de Mahalanobis, o modelo identificou 59 de 66 falências no conjunto de teste (sensibilidade = 0,894; AUC = 0,936; F2 = 0,482), indicando utilidade como ferramenta de triagem contabilométrica.

Palavras-chave: Sparse PLS-DA. SMOTE. Taiwanese Bankruptcy Prediction. Predição de falência empresarial. Contabilometria.

INTRODUÇÃO

A Contabilometria emprega métodos estatísticos, matemáticos e computacionais na formulação e avaliação de problemas contábeis, econômico-financeiros e gerenciais. Na formulação de Iudícibus (1982), publicada na Revista Brasileira de Contabilidade sob o título "Existirá a Contabilometria?", o termo foi proposto em analogia à Econometria, mas sem reduzir a área à mera aplicação mecânica de técnicas numéricas a registros contábeis. O ponto central é articular método quantitativo e teoria contábil, de modo que os resultados possam ser avaliados à luz do problema substantivo investigado. Silva, Chacon e Santos (2005) reforçam essa compreensão ao apresentar a Contabilometria como metodologia científica apoiada em Matemática, Estatística e Informática aplicadas à Contabilidade. A área alcança problemas de mensuração, análise das demonstrações contábeis, custos, controladoria, planejamento, simulação, avaliação de desempenho, apoio à decisão e modelagem de riscos econômico-financeiros. Estudos posteriores também destacam seu papel na identificação de relações entre variáveis empresariais, na análise de sensibilidade, em aplicações econométricas e na produção de evidências quantitativas úteis à gestão e à pesquisa contábil (Francischetti, Poker Júnior e Padoveze, 2017; Castro, Silva e Ribeiro, 2022).

Entre os problemas de interesse contabilométrico, a falência empresarial ocupa posição relevante por combinar baixa frequência relativa e elevado impacto econômico. A descontinuidade de uma firma afeta credores, investidores, trabalhadores, fornecedores, instituições financeiras e outros agentes vinculados à atividade empresarial. Por isso, sua detecção antecipada tem sido tratada como problema clássico em Contabilidade, Finanças e modelagem preditiva, tradicionalmente baseado em indicadores extraídos das demonstrações contábeis. Altman (1968) mostrou que índices financeiros poderiam ser combinados por análise discriminante para prever falência corporativa, enquanto Ohlson (1980) propôs uma abordagem probabilística apoiada em informações contábeis e financeiras. Shumway (2001), por sua vez, argumentou que a falência deve ser modelada como processo associado à deterioração financeira ao longo do tempo, defendendo modelos de duração em lugar de modelos estáticos de período único. Métodos de aprendizado de máquina têm sido utilizados para capturar relações não lineares e padrões complexos em dados empresariais, como observado em Barboza, Kimura e Altman (2017) e Wang e Liu (2021). Prever falências, portanto, não se restringe a separar empresas em dois grupos: trata-se de apoiar a

identificação precoce de risco econômico-financeiro, especialmente quando a classe de interesse é rara e sua não detecção pode gerar perdas relevantes.

Ferramentas de inteligência artificial, mineração de dados e aprendizado de máquina tornaram-se frequentes na análise de eventos raros em Contabilidade e Finanças. Na detecção de fraude, árvores de decisão, redes neurais, redes bayesianas, máquinas de vetores de suporte, algoritmos evolutivos e métodos de seleção de variáveis têm sido empregados para reconhecer padrões anômalos em demonstrações financeiras e transações, superando, em parte, procedimentos exclusivamente manuais ou baseados em regras fixas (Kirkos, Spathis e Manolopoulos, 2007; Ravisankar et al., 2011; Ngai et al., 2011; West e Bhattacharya, 2016). Na predição de falência, classificadores supervisionados exploram interações e relações não lineares entre indicadores econômico-financeiros e o estado futuro da firma, com desempenho frequentemente competitivo em relação a modelos estatísticos tradicionais (Barboza, Kimura e Altman, 2017). O problema torna-se mais exigente em bases desbalanceadas, nas quais a falência é minoritária e o custo de não detecção é elevado.

Este trabalho avalia o uso da análise discriminante por mínimos quadrados parciais esparsa (sparse PLS-DA ou sPLS-DA), combinada com a técnica de reamostragem sintética (*Synthetic Minority Over-sampling Technique*, SMOTE), na predição de falência empresarial a partir de indicadores econômico-financeiros fornecidos na base Taiwanese Bankruptcy Prediction, do UCI Machine Learning Repository. Optamos por priorizar a detecção da classe falência, com ênfase em métricas sensíveis à classe minoritária e à redução de falsos negativos, por entendermos que o custo assimétrico entre esses dois tipos de erro justifica esse foco analítico.

MATERIAL E MÉTODOS

A análise foi conduzida em linguagem R (R Core Team, 2025) com a base Taiwanese Bankruptcy Prediction, disponibilizada pelo UCI Machine Learning Repository (2020). A base reúne informações econômico-financeiras de firmas de Taiwan, coletadas no Taiwan Economic Journal no período de 1999 a 2009, e está estruturada como problema de classificação binária. Foram analisadas 6.819 observações empresariais e 95 variáveis preditoras, sem valores ausentes na base original. A variável resposta Bankrupt? foi recodificada em duas classes: falencia, correspondente a Bankrupt? = 1, e nao_falencia, correspondente a Bankrupt? = 0. A distribuição das classes é fortemente desbalanceada, com 220 registros de falência e 6.599 registros de não falência.

Valores extremos foram tratados por winsorização nos quantis de 1% e 99%, procedimento que limita a influência de observações atípicas sem excluir registros completos da base (Dixon e Yuen, 1974). Os limites, as medianas, os parâmetros de centralização e os desvios de escala foram sempre estimados no conjunto de treino correspondente e aplicados aos subconjuntos de validação ou teste.

A base foi dividida em conjuntos de treino e teste por partição estratificada, preservando a proporção original entre empresas falidas e não falidas. Destinamos 70% dos registros ao treinamento e 30% ao teste retido. No treino, havia 154 empresas falidas e 4.620 não falidas; no teste, 66 empresas falidas e 1.979 não falidas. O conjunto de teste permaneceu separado durante toda a seleção de hiperparâmetros e não recebeu SMOTE, de modo a representar uma avaliação fora da amostra sob a distribuição original das classes. A seleção do modelo foi conduzida exclusivamente no conjunto de treino por validação cruzada 5-fold repetida 20 vezes (Kuhn, 2008). Em cada fold, os parâmetros de winsorização, remoção de variáveis de baixa variabilidade e padronização por escore-z foram estimados apenas na parcela de treinamento correspondente. A parcela de validação recebeu somente as transformações aprendidas nessa parcela de treinamento. O SMOTE foi aplicado exclusivamente à parcela de treinamento de cada fold, sem geração de observações sintéticas na validação nem no conjunto de teste retido. Esse procedimento evita vazamento de informação entre treino, validação e teste e preserva a avaliação fora da amostra sob a distribuição original das classes. Como a base é descrita como proveniente do período de 1999 a 2009, a partição estratificada aleatória deve ser interpretada como avaliação preditiva sob reamostragem, não como validação temporal. Uma validação cronológica — com treino em anos anteriores e teste em anos posteriores — seria mais próxima de um cenário operacional de previsão, caso a estrutura temporal completa estivesse disponível para esse desenho. Em todas as métricas binárias, adotamos a classe falência como positiva, por representar o evento minoritário e o principal interesse analítico deste estudo.

A sparse PLS-DA combina discriminação supervisionada e seleção de variáveis, permitindo restringir o modelo a subconjuntos de indicadores relevantes para a separação entre a classe de interesse e a classe de não interesse. O classificador foi implementado no pacote mixOmics (Rohart et al., 2017). A formulação pertence à família PLS/PLS-DA, na qual o bloco de preditores econômico-financeiros padronizados é representado por $\mathbf{X} \in \mathbb{R}^{n \times p}$, enquanto a resposta categórica é codificada por uma matriz indicadora de classes $\mathbf{Y} \in \{0,1\}^{n \times G}$, em que n é o número de observações, p é o número de preditores e

G é o número de classes. A PLS-DA projeta **X** e **Y** em variáveis latentes supervisionadas, extraídas de modo a maximizar a associação entre as projeções dos blocos preditor e resposta (Geladi e Kowalski, 1986; Wold, Sjöström e Eriksson, 2001; Barker e Rayens, 2003). A versão esparsa acrescenta seleção de variáveis à formulação, impondo esparsidade aos pesos associados ao bloco **X**. Na parametrização utilizada, o hiperparâmetro keepX definiu o número de preditores com peso não nulo em cada variável latente, permitindo discriminação supervisionada com redução dimensional e seleção de indicadores (Lê Cao, Boitard e Besse, 2011).

Na *h*-ésima variável latente, considerando os blocos residuais $\mathbf{X}_{h-1}$ e $\mathbf{Y}_{h-1}$, os vetores de pesos $\mathbf{w}_h$ e $\mathbf{c}_h$ são estimados de forma a maximizar a covariância quadrática entre os escores dos blocos preditor e resposta (eq. 1):

$$\max_{\mathbf{w}_h,\mathbf{c}_h} \operatorname{cov}^2\left(\mathbf{X}_{h-1}\mathbf{w}_h, \mathbf{Y}_{h-1}\mathbf{c}_h\right) \tag{1}$$

sujeito a $\|\mathbf{w}_h\|_2 = 1$, $\|\mathbf{c}_h\|_2 = 1$ e $\|\mathbf{w}_h\|_0 \leq k_h$. As restrições $\|\mathbf{w}_h\|_2 = 1$ e $\|\mathbf{c}_h\|_2 = 1$ normalizam os vetores de pesos, enquanto $\|\mathbf{w}_h\|_0 \leq k_h$ impõe esparsidade ao vetor de pesos do bloco preditor, limitando a no máximo k_h o número de preditores com peso não nulo na *h*-ésima variável latente. No pacote mixOmics, esse controle é operacionalizado pelo parâmetro keepX. Os escores correspondentes são definidos por (eq. 2 e 3):

$$\mathbf{t}_h = \mathbf{X}_{h-1}\mathbf{w}_h \tag{2}$$

$$\mathbf{u}_h = \mathbf{Y}_{h-1}\mathbf{c}_h \tag{3}$$

em que $\mathbf{t}_h$ é o vetor de escores do bloco preditor e $\mathbf{u}_h$ é o vetor de escores associado ao bloco resposta. Após a extração de *H* variáveis latentes, os escores do bloco preditor são reunidos na matriz $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_H]$. Sendo $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_H]$ a matriz de pesos esparsos estimada pelo modelo, a projeção dos preditores no espaço latente pode ser escrita como (eq. 4):

$$\mathbf{T} = \mathbf{X}\mathbf{W} \tag{4}$$

Nessa representação, cada coluna de **W** corresponde ao vetor de pesos associado a uma variável latente. A esparsidade decorre da restrição imposta pelo parâmetro keepX, que limita o número de preditores com peso não nulo em cada dimensão latente.

O desbalanceamento da classe falência foi tratado por SMOTE, através do pacote themis (Hvitfeldt, 2025), com razão de sobreamostragem igual a 1, restrito às amostras de treinamento. As observações sintéticas foram geradas por interpolação entre vizinhos da classe minoritária e não foram introduzidas no teste retido (Chawla et al., 2002). A regra de decisão adotada foi a distância quadrática de Mahalanobis no espaço dos escores. Para uma observação com vetor de escores **t**, a distância à classe *g* foi definida por (eq. 5):

$$d_g^2(\mathbf{t}) = \left(\mathbf{t} - \bar{\mathbf{t}}_g\right)^T \mathbf{S}_g^{-1} \left(\mathbf{t} - \bar{\mathbf{t}}_g\right) \tag{5}$$

em que $\bar{\mathbf{t}}_g$ é o vetor médio dos escores da classe *g* e $\mathbf{S}_g$ é a matriz de covariâncias dos escores dessa classe. A classe predita correspondeu à menor distância quadrática de Mahalanobis.

Os hiperparâmetros foram definidos por validação cruzada 5-fold repetida 20 vezes (Kuhn, 2008). A grade examinada combinou 1 a 15 variáveis latentes com 5, 10, 15, 20, 30, 40, 50, 75 e 95 variáveis retidas por variável latente. Em cada fold, o pré-processamento e o SMOTE foram estimados somente na parcela de treinamento. O critério primário de seleção foi o escore F_2, adotado por atribuir maior peso à sensibilidade da classe falência do que à precisão (Powers, 2011). Empates foram resolvidos, sucessivamente, por escore F_1, área sob a curva (AUC), coeficiente de correlação de Matthews, índice de Youden, sensibilidade, especificidade e menor complexidade do modelo.

A busca em grade selecionou duas variáveis latentes e dez variáveis retidas por variável latente. A execução final fixou essa configuração para viabilizar 99.999 permutações no teste de significância, sem repetir a grade completa em cada aleatorização. A inferência permutacional deve ser interpretada como avaliação condicional do classificador final previamente selecionado.

O modelo final foi reajustado no treino completo sob o mesmo protocolo de pré-processamento. O desempenho foi avaliado no treino original, na validação cruzada e no teste retido por métricas de acerto global, ordenação, detecção da classe positiva e concordância entre classes. Os escores F_1, F_2, o índice de Youden e o MCC foram interpretados conforme a síntese de Powers (2011), enquanto MCC, FMI e kappa foram associados às formulações originais de Matthews (1975), Fowlkes e Mallows (1983) e Cohen (1960), respectivamente. A área sob a curva ROC foi calculada com apoio do pacote pROC (Robin et al., 2011), a partir do escore discriminante contínuo associado à classe falência obtido nas predições da sparse PLS-DA. Esse escore avaliou a ordenação das observações entre falência e não falência, enquanto a distância quadrática de Mahalanobis definiu a classe predita final e, consequentemente, as matrizes de confusão e as métricas baseadas em contagens.

Os diagnósticos gráficos reportados incluem a variância acumulada do bloco de preditores, as variáveis retidas pela sparse PLS-DA e a distribuição permutacional do F_2. Os resultados completos da

busca em grade, da seleção do número de variáveis latentes e da taxa de classificação incorreta por dimensão latente foram preservados como controle computacional, mas não foram inseridos no corpo principal do artigo.

A significância empírica do classificador final foi avaliada por permutação dos rótulos de classe no treino. A matriz de preditores foi mantida fixa, enquanto os rótulos falência/não falência foram aleatorizados. Em cada permutação, o modelo foi reajustado com duas variáveis latentes e dez variáveis retidas por variável latente, com SMOTE aplicado aos rótulos permutados, e avaliado no mesmo teste retido. As estatísticas observadas foram comparadas com as distribuições permutadas de AUC, F1, F2, MCC e índice de Youden, seguindo a avaliação permutacional de classificadores discutida por Ojala e Garriga (2010). O valor-p empírico foi calculado por (eq. 6):

$$p = \frac{r+1}{B+1} \qquad (6)$$

em que r é o número de permutações com desempenho igual ou superior ao observado e B é o número de permutações válidas, evitando valores-p iguais a zero em simulações de Monte Carlo (Phipson e Smyth, 2010). Na rotina final, foram calculadas 99.999 permutações. Como a busca em grade não foi repetida dentro de cada permutação, a inferência obtida refere-se ao desempenho da configuração final selecionada, e não de todo o procedimento de seleção de hiperparâmetros.

RESULTADOS E DISCUSSÃO

A base apresentou o padrão típico de eventos raros em modelagem econômico-financeira: a classe falência correspondeu a 3,23% das observações. A partição estratificada preservou essa prevalência no treino e no teste retido, mantendo a avaliação fora da amostra sob a distribuição original das classes. Nessa configuração, a exatidão isolada é pouco informativa (um classificador orientado apenas à classe majoritária produziria elevado acerto global sem recuperar adequadamente o evento de interesse).

A validação cruzada da grade selecionou o modelo com duas variáveis latentes e dez variáveis retidas por variável latente, mantendo SMOTE e decisão por distância quadrática de Mahalanobis. A execução final fixou essa configuração para concentrar o custo computacional no teste de permutação com 99.999 aleatorizações. O desempenho obtido no treino original, na validação cruzada e no teste retido foi compatível entre as três avaliações. A AUC (0,936) e a sensibilidade (0,894) foram levemente superiores no teste retido em relação ao treino original (0,932 e 0,877) e à validação cruzada (0,929 e 0,868), padrão coerente com a variabilidade amostral esperada na partição aleatória estratificada com minoria reduzida, sem evidência de degradação nessa partição retida, como mostra a Tabela 1.

Tabela 1. Desempenho do modelo sparse PLS-DA+SMOTE nos conjuntos de treino original, validação cruzada e teste retido.

Métrica / Conj.	Treino original	Validação cruzada	Teste retido
Exatidão	0,862	0,861	0,855
AUC	0,932	0,929	0,936
Sens.	0,877	0,868	0,894
Esp.	0,862	0,861	0,854
F_2	0,486	0,48	0,482
MCC	0,354	0,349	0,352

A Tabela 2 apresenta a matriz de confusão do conjunto de teste retido. O classificador demonstrou orientação clara à recuperação da classe rara: apenas 7 das 66 falências não foram sinalizadas, enquanto 289 registros de não falência foram classificados como falência. Esse padrão é coerente com a função de perda implícita na escolha do F2 como critério primário e com o balanceamento sintético aplicado durante o treinamento. Em termos operacionais, o modelo privilegia a triagem de risco em detrimento de uma regra decisória altamente específica para confirmação de falência.

Tabela 2. Matriz de confusão.

Valor real	Valor predito	
	Falência	Não falência
Falência	59	7
Não falência	289	1690

As métricas do conjunto de teste retido estão reunidas na Tabela 3. O escore F_2 = 0,482 reflete o peso conferido à recuperação da classe falência pela métrica, enquanto a baixa precisão (0,17) evidencia que uma parcela expressiva das sinalizações positivas corresponde a registros não falidos. Essa combinação não invalida o modelo, mas delimita sua função: a sparse PLS-DA com SMOTE opera melhor como filtro inicial de risco do que como diagnóstico conclusivo. Em bases com prevalência próxima de 3%, mesmo uma especificidade adequada produz número absoluto elevado de falsos positivos.

Tabela 3. Métricas do modelo no conjunto de teste retido.

Métrica	Resultado
Exatidão	0,855
AUC	0,936
Sensibilidade	0,894
Especificidade	0,854
Precisão	0,17
Valor preditivo negativo	0,996
Escore F_1	0,285
Escore F_2	0,482
Índice de Youden	0,748
MCC	0,352
FMI	0,389
Kappa	0,244

A AUC de 0,936 foi calculada a partir do escore discriminante contínuo associado à classe falência. No conjunto de teste, esse escore foi mais elevado entre os registros efetivamente pertencentes à classe falência, com média de 0,800 e mediana de 0,730, enquanto os registros de não falência apresentaram média de 0,239 e mediana de 0,236. Essa separação entre as distribuições dos escores explica a elevada capacidade de ordenação do modelo, embora não elimine a sobreposição entre os grupos nem substitua a regra final de classificação por distância de Mahalanobis.

O índice de Youden foi 0,748, compatível com discriminação global satisfatória entre sensibilidade e especificidade. O FMI de 0,389 corresponde à média geométrica entre precisão e sensibilidade, refletindo o efeito restritivo da baixa precisão sobre a métrica. O MCC de 0,352 e o kappa de 0,244 apontam associação positiva, mas ainda limitada quando os quatro elementos da matriz de confusão são considerados simultaneamente. Essa diferença entre AUC e sensibilidade, de um lado, e as métricas de concordância, de outro, é esperada em bases raras: mesmo com especificidade de 0,854, o grande volume de registros de não falência converte uma fração relativamente pequena de erros nessa classe em número absoluto elevado de falsos positivos.

As taxas de classificação incorreta da validação cruzada e do teste retido permaneceram próximas, o que é compatível com estabilidade entre ajuste interno e avaliação fora da amostra por partição aleatória. O erro global, contudo, deve permanecer como diagnóstico secundário, pois a exatidão é dominada pela classe majoritária. A interpretação principal do desempenho deve recair sobre sensibilidade, precisão, F_2, AUC, MCC e comportamento dos falsos negativos.

A variância acumulada do bloco **X** foi de aproximadamente 25% na primeira variável latente e 34% com duas variáveis latentes. Esse valor não deve ser tomado como critério de seleção, pois a sparse PLS-DA é supervisionada: as variáveis latentes são extraídas considerando sua relação com a classe de falência, não apenas a variabilidade interna dos preditores. A baixa dimensionalidade do modelo final deve ser lida, portanto, como ganho de parcimônia, e não como tentativa de explicar a maior parcela possível da variância total dos indicadores.

A formulação esparsa reduziu o bloco original de 95 indicadores a dez variáveis por variável latente. Variáveis redundantes ou com baixa contribuição discriminante no ajuste supervisionado receberam peso nulo, enquanto um subconjunto reduzido permaneceu ativo na construção dos escores.

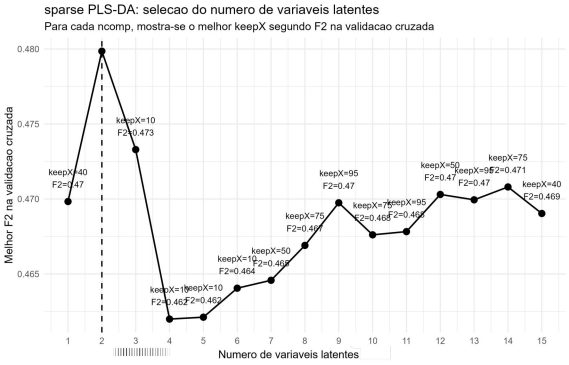

Figura 1. Seleção do número de variáveis latentes e do parâmetro keepX por validação cruzada repetida.

A Figura 1 sintetiza o desempenho médio da sparse PLS-DA na validação cruzada repetida em função do número de variáveis latentes e das combinações avaliadas para o parâmetro keepX. O melhor desempenho médio pelo critério F_2 foi obtido com duas variáveis latentes e dez variáveis retidas por variável latente, configuração adotada no ajuste final e no teste de permutação.

A Figura 2 complementa a etapa de seleção apresentada na Figura 1 ao mostrar quais indicadores econômico-financeiros permaneceram ativos no modelo final. As barras representam o maior valor absoluto de loading de cada variável entre as duas variáveis latentes selecionadas. Entre os indicadores com maior contribuição discriminante aparecem as variáveis X19 (*Persistent EPS in the Last Four Seasons*), X37 (*Debt ratio %*), X38 (*Net worth/Assets*), X40 (*Borrowing dependency*) e X86 (*Net Income to Total Assets*), sugerindo que a separação entre empresas falidas e não falidas foi conduzida principalmente por variáveis relacionadas à persistência de resultados, endividamento, estrutura

patrimonial, dependência de financiamento e rentabilidade.

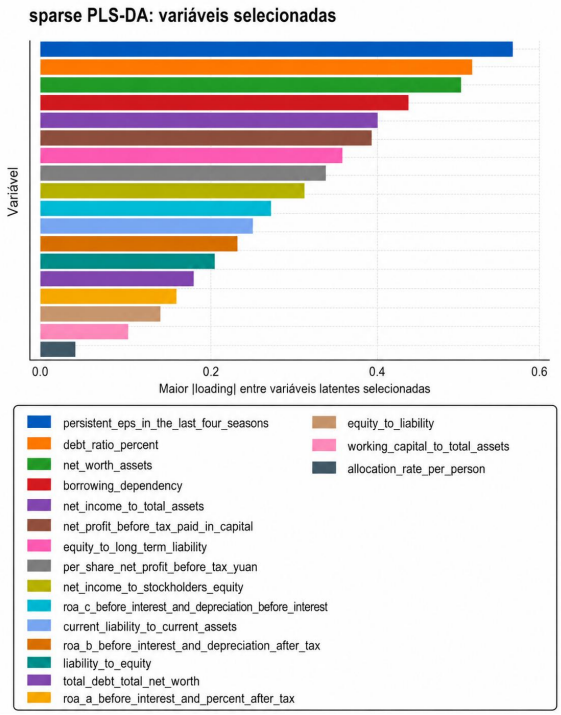

Figura 2. Indicadores econômico-financeiros selecionados pela sparse PLS-DA no modelo final. As barras representam o maior valor absoluto de loading de cada variável entre as duas variáveis latentes selecionadas.

A Figura 2 apresenta magnitudes absolutas dos loadings, não informando, isoladamente, a direção da associação de cada indicador com a classe falência, nem permite atribuição causal. A formulação esparsa reduziu o conjunto original de 95 preditores a um subconjunto menor de indicadores com contribuição para a construção dos escores discriminantes. Após essa definição da estrutura final do modelo, avaliou-se se o desempenho obtido poderia ser explicado por associação aleatória entre preditores e rótulos de classe, etapa resumida pela análise permutacional apresentada a seguir.

A Tabela 4 aprofunda a avaliação inferencial do classificador final ao comparar o desempenho observado no teste retido com a distribuição obtida após permutação dos rótulos de classe no conjunto de treino. Para todas as métricas avaliadas (AUC, escores F_1 e F_2, MCC e índice de Youden), o número de permutações com desempenho igual ou superior ao observado foi nulo ($r = 0$). Assim, os valores-p empíricos atingiram o limite inferior determinado pelo número de permutações válidas, aproximadamente $1{,}0 \times 10^{-5}$. Esse resultado indica que, mantendo fixa a matriz de preditores e aleatorizando a associação entre indicadores econômico-financeiros e falência, não se observou desempenho comparável ao do modelo ajustado com os rótulos reais.

A interpretação deve considerar o papel complementar das métricas. A AUC observada de 0,936 indica elevada capacidade de ordenação entre empresas falidas e não falidas; o índice de Youden de 0,748 resume o equilíbrio entre sensibilidade e especificidade; e o F2 de 0,482 confirma que o modelo preserva desempenho relevante quando se atribui maior peso à recuperação da classe falência. Já o MCC de 0,352 e o F1 de 0,285 são mais penalizados pelo número de falsos positivos, mas ainda assim permaneceram acima de todos os valores gerados sob permutação. Portanto, mesmo métricas mais exigentes em relação ao equilíbrio da matriz de confusão apresentaram desempenho incompatível com uma classificação aleatória dos rótulos.

Os valores-p próximos de $1{,}0 \times 10^{-5}$ não devem ser lidos como probabilidade exata de ocorrência do resultado sob a hipótese nula, mas como o menor nível de significância resolvível com o número de permutações executadas. O erro-padrão de Monte Carlo, também próximo de $1{,}0 \times 10^{-5}$, reflete essa resolução numérica no limite inferior do procedimento, pois nenhuma permutação superou ou igualou o desempenho observado. Dessa forma, a evidência permutacional é forte, mas permanece condicionada à configuração final previamente selecionada, com duas variáveis latentes e dez variáveis retidas por variável latente. Como a busca em grade não foi repetida em cada permutação, a Tabela 4 testa a significância empírica do classificador final fixado, e não de todo o procedimento de seleção de hiperparâmetros. O erro-padrão de Monte Carlo do valor-p empírico foi aproximado por (eq. 7):

$$\mathrm{EP}_{MC}(\hat{p}) \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{(B+1)}} \qquad (7)$$

onde $\hat{p}$ é o valor-p empírico estimado por permutação e B é o número de permutações válidas.

Tabela 4. Resultados do teste de permutação para o modelo final no conjunto de teste retido.

Métrica	AUC	Escore F_1	Escore F_2	MCC	Índice de Youden
Valor observado	0,936	0,285	0,482	0,352	0,748
B válido	99999	99988	99988	99999	99999
r	0	0	0	0	0
p empírico	$1{,}0 \times 10^{-5}$	$1{,}0 \times 10^{-5}$	$1{,}0 \times 10^{-5}$	$1{,}0 \times 10^{-5}$	$1{,}0 \times 10^{-5}$
Erro-padrão MC	$1{,}0 \times 10^{-5}$	$1{,}0 \times 10^{-5}$	$1{,}0 \times 10^{-5}$	$1{,}0 \times 10^{-5}$	$1{,}0 \times 10^{-5}$

Para os escores F_1 e F_2, 11 permutações produziram métricas indefinidas por denominadores nulos, sendo excluídas do cálculo empírico correspondente.

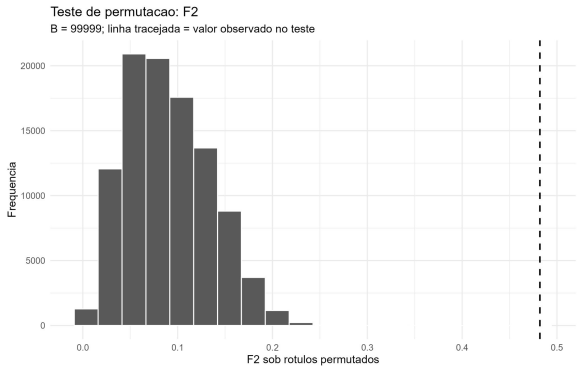

Figura 3. Distribuição do escore F_2 sob permutação dos rótulos no conjunto de treino. A linha tracejada indica o escore F_2 observado no teste retido.

A Figura 3 ilustra especificamente o comportamento do escore F_2 sob a hipótese nula induzida pela permutação dos rótulos de classe no conjunto de treino. A distribuição dos valores permutados concentra-se em níveis substancialmente inferiores ao F_2 observado no teste retido, representado pela linha tracejada. Essa separação visual é coerente com o resultado apresentado na Tabela 4, na qual nenhuma permutação válida produziu F_2 igual ou superior ao observado. Assim, a figura reforça que o desempenho do classificador final, para uma métrica que privilegia a recuperação da classe falência, não foi reproduzido quando a associação entre os indicadores econômico-financeiros e os rótulos de falência foi aleatorizada.

Em conjunto, os resultados indicam que a sparse PLS-DA com SMOTE e regra de Mahalanobis funcionou como classificador sensível à classe falência, com boa capacidade de ordenação e comportamento consistente entre validação cruzada e teste retido. A limitação prática permanece na baixa precisão, decorrente do volume de falsos positivos em uma base com prevalência de falência próxima de 3%. Assim, o uso mais adequado do modelo é como ferramenta de triagem contabilométrica: reduzir a probabilidade de que registros de falência deixem de ser sinalizados, com posterior avaliação especializada dos casos classificados como falência.

CONCLUSÃO

A sparse PLS-DA combinada ao SMOTE apresentou desempenho compatível com uma aplicação de triagem de falência empresarial em base fortemente desbalanceada. No conjunto de teste retido, o modelo identificou 59 das 66 empresas falidas, com sensibilidade de 0,894 e AUC de 0,936. Esses resultados indicam boa capacidade de recuperação e ordenação da classe minoritária, que constitui o principal foco analítico do estudo.

A interpretação do modelo, entretanto, deve considerar a baixa precisão observada. Embora a maior parte das falências tenha sido sinalizada, 289 empresas não falidas foram classificadas como falência. Esse padrão é coerente com a escolha do F2 como critério primário de seleção, pois essa métrica atribui maior peso à sensibilidade do que à precisão. Assim, a abordagem não deve ser entendida como instrumento conclusivo de diagnóstico, mas como filtro inicial para reduzir a probabilidade de que empresas em risco deixem de ser encaminhadas a uma análise econômico-financeira mais detalhada.

A formulação esparsa contribuiu para reduzir a dimensionalidade do problema ao restringir o modelo a um subconjunto de indicadores por variável latente. Essa característica favorece a parcimônia e pode auxiliar a interpretação contabilométrica, desde que os indicadores selecionados sejam analisados como variáveis com contribuição discriminante no modelo supervisionado, e não como evidência causal de falência.

O teste de permutação com 99.999 aleatorizações forneceu evidência de que o desempenho observado não decorre apenas de associação aleatória entre preditores e rótulos, considerando a configuração final previamente selecionada. Essa inferência é condicional ao modelo fixado com duas variáveis latentes e dez variáveis retidas por variável latente, não à repetição de toda a busca em grade em cada permutação. Como limitação, a avaliação foi baseada em partição aleatória estratificada; uma validação cronológica seria mais próxima de um cenário operacional de previsão, caso as informações temporais necessárias estivessem disponíveis. Em conjunto, os resultados sustentam o uso da sparse PLS-DA com SMOTE como ferramenta de apoio à triagem contabilométrica de falência empresarial, com validação adicional recomendável antes de aplicação decisória.

AGRADECIMENTOS

Os autores agradecem à Universidade do Estado do Rio de Janeiro pelo apoio institucional. N.C.N.A. agradece o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES), Código de Financiamento 001, pela bolsa de doutorado. Y.D.P. agradece à Secretaria de Estado de Desenvolvimento Econômico de Mato Grosso (SEDEC-MT), Cuiabá, Brasil, pela bolsa de pós-doutorado concedida no âmbito do projeto SEDECPRO-2022/02316 (Projeto Número: 3.008.004). S.O.B. agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro concedido por meio de bolsa, processo nº 130848/2026-3.

REFERÊNCIAS

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: 10.1111/j.1540-6261.1968.tb00843.x.

BARBOZA, F.; KIMURA, H.; ALTMAN, E. Machine learning models and bankruptcy prediction. Expert Systems with Applications, v. 83, p. 405-417, 2017. DOI: 10.1016/j.eswa.2017.04.006.

BARKER, M.; RAYENS, W. Partial least squares for discrimination. Journal of Chemometrics, v. 17, n. 3, p. 166-173, 2003. DOI: 10.1002/cem.785.

CASTRO, C. A.; SILVA, L. G. F.; RIBEIRO, P. M. Contabilometria: sua importância e os métodos quantitativos aplicados no curso de Ciências Contábeis. Revista Científica FacMais, v. 19, n. 1, 2022.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, v. 16, p. 321-357, 2002. DOI: 10.1613/jair.953.

COHEN, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, v. 20, n. 1, p. 37-46, 1960. DOI: 10.1177/001316446002000104.

DIXON, W. J.; YUEN, K. K. Trimming and winsorization: a review. Statistische Hefte, v. 15, n. 2-3, p. 157-170, 1974. DOI: 10.1007/BF02922904.

FOWLKES, E. B.; MALLOWS, C. L. A method for comparing two hierarchical clusterings. Journal of the American Statistical Association, v. 78, n. 383, p. 553-569, 1983. DOI: 10.1080/01621459.1983.10478008.

FRANCISCHETTI, C. E.; POKER JÚNIOR, J. H.; PADOVEZE, C. L. Contabilometria: análise bibliométrica, tendências e reflexões em publicações da base de dados Scopus de 1982 até 2014. Contabilometria: Brazilian Journal of Quantitative Methods Applied to Accounting, v. 4, n. 1, p. 31-44, 2017.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. Analytica Chimica Acta, v. 185, p. 1-17, 1986. DOI: 10.1016/0003-2670(86)80028-9.

HVITFELDT, E. themis: Extra Recipes Steps for Dealing with Unbalanced Data. R package version 1.0.3, 2025. Disponível em: https://github.com/tidymodels/themis.

IUDÍCIBUS, S. de. Existirá a contabilometria? Revista Brasileira de Contabilidade, Rio de Janeiro, n. 41, p. 44-60, 1982.

KIRKOS, E.; SPATHIS, C.; MANOLOPOULOS, Y. Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications, v. 32, n. 4, p. 995-1003, 2007. DOI: 10.1016/j.eswa.2006.02.016.

KUHN, M. Building predictive models in R using the caret package. Journal of Statistical Software, v. 28, n. 5, p. 1-26, 2008. DOI: 10.18637/jss.v028.i05.

LÊ CAO, K.-A.; BOITARD, S.; BESSE, P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics, v. 12, art. 253, 2011. DOI: 10.1186/1471-2105-12-253.

MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta, v. 405, n. 2, p. 442-451, 1975. DOI: 10.1016/0005-2795(75)90109-9.

NGAI, E. W. T.; HU, Y.; WONG, Y. H.; CHEN, Y.; SUN, X. The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. Decision Support Systems, v. 50, n. 3, p. 559-569, 2011. DOI: 10.1016/j.dss.2010.08.006.

OHLSON, J. A. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, v. 18, n. 1, p. 109-131, 1980. DOI: 10.2307/2490395.

OJALA, M.; GARRIGA, G. Permutation tests for studying classifier performance. Journal of Machine Learning Research, v. 11, p. 1833-1863, 2010.

PHIPSON, B.; SMYTH, G. K. Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. Statistical Applications in Genetics and Molecular Biology, v. 9, art. 39, 2010. DOI: 10.2202/1544-6115.1585.

POWERS, D. M. W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, v. 2, n. 1, p. 37-63, 2011.

R CORE TEAM. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2025.

RAVISANKAR, P.; RAVI, V.; RAGHAVA RAO, G.; BOSE, I. Detection of financial statement fraud and feature selection using data mining

techniques. Decision Support Systems, v. 50, n. 2, p. 491-500, 2011. DOI: 10.1016/j.dss.2010.11.006.

ROBIN, X.; TURCK, N.; HAINARD, A.; TIBERTI, N.; LISACEK, F.; SANCHEZ, J.-C.; MÜLLER, M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, v. 12, art. 77, 2011. DOI: 10.1186/1471-2105-12-77.

ROHART, F.; GAUTIER, B.; SINGH, A.; LÊ CAO, K.-A. mixOmics: an R package for omics feature selection and multiple data integration. PLOS Computational Biology, v. 13, n. 11, e1005752, 2017. DOI: 10.1371/journal.pcbi.1005752.

SHUMWAY, T. Forecasting bankruptcy more accurately: a simple hazard model. The Journal of Business, v. 74, n. 1, p. 101-124, 2001. DOI: 10.1086/209665.

SILVA, M. C.; CHACON, M. J. M.; SANTOS, J. O que é Contabilometria? Pensar Contábil, v. 7, n. 27, p. 40-43, 2005.

UCI MACHINE LEARNING REPOSITORY. Taiwanese Bankruptcy Prediction. 2020. DOI: 10.24432/C5004D. Disponível em: https://archive.ics.uci.edu/dataset/572/taiwanese+bankruptcy+prediction.

WANG, H.; LIU, X. Undersampling bankruptcy prediction: Taiwan bankruptcy data. PLOS ONE, v. 16, n. 7, e0254030, 2021. DOI: 10.1371/journal.pone.0254030.

WEST, J.; BHATTACHARYA, M. Intelligent financial fraud detection: a comprehensive review. Computers & Security, v. 57, p. 47-66, 2016. DOI: 10.1016/j.cose.2015.09.005.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, v. 58, n. 2, p. 109-130, 2001. DOI: 10.1016/S0169-7439(01)00155-1.