Revista Catarinense da CIÊNCIA CONTÁBIL

ISSN 2237-7662, v. 25, 1-33, e3684, 2026
https://doi.org/10.16930/2237-7662202636841
Disponível em http://revista.crcsc.org.br

ANÁLISE DAS DEMONSTRAÇÕES FINANCEIRAS DAS MAIORES EMPRESAS DO SETOR DE AGRONEGÓCIOS BRASILEIRO UTILIZANDO TÉCNICAS DE MACHINE LEARNING

ANALYSIS OF THE FINANCIAL STATEMENTS OF THE LARGEST COMPANIES IN THE BRAZILIAN AGRIBUSINESS SECTOR USING MACHINE LEARNING TECHNIQUES

VICTOR ROTTE PEREIRA
Universidade Presbiteriana Mackenzie (UPM), Programa de Pós-Graduação em Administração de Empresas (PPGA), Centro de Ciências Sociais e Aplicadas (CCSA) - Mackenzie Business School, São Paulo, SP, Brasil
https://orcid.org/0009-0001-1756-2937
victorrotte@gmail.com

PEDRO PAULO BRAMBILLA PENTEADO
Universidade Presbiteriana Mackenzie (UPM), Graduação em Ciências Contábeis (Dual Study), Centro de Ciências Sociais e Aplicadas (CCSA) - Mackenzie Business School, São Paulo, SP, Brasil
https://orcid.org/0009-0004-2196-7998
pedrobrambilla12@gmail.com

JOSÉ EDUARDO RICCIARDI FAVARETTO[1]
Universidade Presbiteriana Mackenzie (UPM), Programa de Pós-Graduação em Controladoria, Finanças e Tecnologias de Gestão (PPGCFTG), Centro de Ciências Sociais e Aplicadas (CCSA) - Mackenzie Business School, São Paulo, SP, Brasil
https://orcid.org/0000-0002-0143-0809
jose.favaretto@mackenzie.br

RESUMO

As Demonstrações Financeiras (DFs) de empresas constituem um dos principais meios formais de comunicação (sinalização) entre gestores e stakeholders, investidores, credores e reguladores, para representar a situação econômica e financeira da empresa em um período contábil. Esta pesquisa baseou-se na Teoria da Sinalização para estudar empresas do segmento do agronegócio brasileiro, um dos principais setores da economia do país, com participação de 23,2% no PIB em 2024 e responsável por 49% das exportações, no valor de US$ 164,4 bilhões, a partir do *ranking* Forbes Agro100 Brasil, que reúne as maiores empresas do setor do agronegócio brasileiro, com base em suas receitas líquidas de 2023. O objetivo geral deste estudo foi analisar os resultados financeiros das empresas listadas neste *ranking*, com base em dados extraídos das DFs – valores financeiros obtidos a partir destes documentos e o cálculo de outros indicadores – utilizando técnicas analíticas/modelagens de Machine Learning (ML), como *Clustering* e *Principal Component Analysis* (PCA). As 80 empresas analisadas formaram quatro clusters (C1 a C4), nomeados: "Lucrativas" (C1, 8 empresas), "Encolhidas" (C2, 7 empresas), "Saudáveis" (C3, 18 empresas) e "Diversificadas" (C4, 47 empresas). A decomposição da variância dos dados em dimensões

---

Editado em português e inglês. Versão original em português.

[1] **Endereço para correspondência:** Rua da Consolação, 930 | Higienópolis | CEP 01302-000 | São Paulo, SP | Brasil.

Recebido em 06/12/2025. **Última versão recebida em** 01/06/2026. **Aceito em** 30/07/2026 por Rogério João Lunkes (Editor-Chefe). **Publicado em** 31/08/2026.

(cc) BY

principais ortogonais de "Eficiência" (PC1), "Escala" (PC2) e "Risco-Retorno" (PC3) apresentou explicações com potencial de aplicação na gestão organizacional. Foi possível compreender que, para ser mais completa, a estruturação de um *ranking* de empresas não deveria se limitar apenas aos valores de suas receitas líquidas, pois essa única mensuração não é capaz de captar toda a situação financeira/econômica de uma companhia.

Palavras-chave: demonstrações financeiras, balanço patrimonial, agronegócio brasileiro. aprendizado de máquina, *ranking* Forbes.

ABSTRACT

The Financial Statements (FS) of companies are one of the main formal means of communication (signaling) between managers and stakeholders, investors, creditors, and regulators, representing the company's economic and financial situation during an accounting period. This research was based on Signaling Theory to study companies in the Brazilian agribusiness sector, one of the main sectors of the country's economy, with a 23.2% participation in GDP in 2024 and responsible for 49% of exports, valued at US$164.4 billion, from the Forbes Agro100 Brazil ranking, which gathers the largest companies in the Brazilian agribusiness sector, based on their 2023 net revenues. The main objective of this study was to analyze the financial results of the companies listed in this ranking, based on data extracted from the FS – financial values obtained from these documents and the calculation of other indicators – employing analytical techniques/Machine Learning (ML) modeling, such as Clustering and Principal Component Analysis (PCA). The 80 companies analyzed formed four clusters (C1 to C4), named "Profitable" (C1, 8 companies), "Shrinking" (C2, 7 companies), "Healthy" (C3, 18 companies), and "Diversified" (C4, 47 companies). The decomposition of the data variance into the main orthogonal dimensions of "Efficiency" (PC1), "Scale" (PC2), and "Risk-Return" (PC3) provided explanations with potential applications in organizational management. It was possible to understand that, to be more complete, the structure of a company ranking should not be limited to net revenue, as this single measure cannot capture a company's entire financial/economic situation.

***Keywords**: financial statements, balance sheet, Brazilian agribusiness, machine learning, Forbes ranking.*

1 INTRODUÇÃO

As Demonstrações Financeiras (DFs) são instrumentos de comunicação entre as empresas e seus *stakeholders,* como investidores, credores e gestores. As DFs são o retrato da situação econômica e financeira do que aconteceu no exercício – ano contábil – das empresas, auxiliando-as para a tomada de decisão estratégica (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a).

No Brasil e no mundo, as empresas de capital aberto ou de sociedade anônima devem publicar suas DFs. No Brasil essas obrigações são regulamentadas pela Lei das Sociedades Anônimas (Governo Federal do Brasil, 1976) e pelas Normas Brasileiras de Contabilidade (NBC).

A literatura acadêmica relata diversos estudos (Figari *et al.*, 2021; Ribeiro *et al.*, 2024; Silva & Diógenes, 2022) que apresentam a utilização das DFs em modos tradicionais de análise contábil e financeira, com cálculos, modelos matemáticos e/ou comparações de indicadores.

Um dos setores do país em que é possível observar o uso das DFs no auxílio à decisão do stakeholder é o agronegócio – setor da economia que abrange todas as etapas da cadeia produtiva, desde o cultivo e a criação de animais até a industrialização e comercialização dos produtos. Nos últimos anos, o Brasil vem se destacando no mundo como um dos maiores produtores e

exportadores de algumas commodities agrícolas, como soja, milho, café e carne bovina (Cepea, 2025a). O agronegócio brasileiro é responsável pela geração de alimentos e exportação destas commodities, garantindo a segurança alimentar e relevância para a economia do país (Guedes *et al.*, 2021).

Segundo a Confederação da Agricultura e Pecuária do Brasil (CNA), entidade responsável por congregar associações e lideranças políticas e rurais em todo o país, a produção agropecuária do Brasil se desenvolveu nos últimos 50 anos de tal forma que tornou o país um dos principais produtores e fornecedores de alimentos, fibras e energia do mundo (CNA, 2025).

As empresas, ao divulgar seus resultados e dados econômicos via DFs, permitem que seus principais usuários – investidores, credores ou parceiros (stakeholders) – negociem com elas ou invistam nelas. Logo, estudar e compreender como o setor do agronegócio brasileiro e suas empresas estão estruturados contabilmente / financeiramente, é um importante quesito para análise de mercado, bem como possibilita auxiliar a tomada de decisão de *stakeholders*, visto que muitas destas empresas buscam créditos via BNDES e de fundos creditórios (Fossá *et al.*, 2024).

Ao adotar o corpo conceitual da Teoria da Sinalização (Connelly *et al.*, 2025; Ross, 1977; Spence, 1973), esse trabalho abordou o papel informacional das DFs como um dos principais instrumentos de comunicação entre as empresas do agronegócio (emissores), ao divulgar indicadores financeiros, como rentabilidade, liquidez e eficiência, entre outros (sinais), com o objetivo de reduzir a assimetria informacional com investidores, credores e demais partes interessadas (receptores).

De forma complementar, o cenário da profissão contábil está diante de um novo paradigma à medida que a Inteligência Artificial (IA) cria conexões entre fluxos de dados e algoritmos de aprendizado, transformando a capacidade de realizar análises e o cumprimento de padrões regulatórios (Codesso, 2025). Nesta revolução digital, há técnicas de modelagem estatística com uso de *Machine Learning* (ML) disponíveis, que podem ser aproveitadas nas análises das DFs (Bojer, 2022).

Esta pesquisa adotou a perspectiva multivariada em seu desenvolvimento, focando nas análises das demonstrações financeiras e aproveitando, como seu contexto, os resultados obtidos de um estudo publicado em outubro de 2024 pela revista Forbes Brasil – um veículo de comunicação com ênfase em negócios e economia, conhecido como *ranking* Forbes Agro100 Brasil (Rodrigues *et al.*, 2025), que listou as 100 maiores empresas do setor do agronegócio brasileiro com base em suas receitas líquidas de 2023.

A receita líquida de uma empresa é um excelente determinante para conseguir quantificar o valor da empresa e sua relevância, visto que mede o quanto entrou de dinheiro proveniente de sua operação; logo, é possível justificar uma listagem posicionando quem foi mais influente para o agronegócio (Figari *et al.*, 2021).

No entanto, este *ranking* informa apenas a receita líquida, que tem em sua definição, segundo o CPC e IFRS (Comitê de Pronunciamentos Contábeis, 2012; International Financial Reporting Standards, 2001c), de o ingresso bruto de benefícios econômicos durante o período observado, no curso das atividades ordinárias da entidade, que aumentam seu patrimônio líquido, excluindo os aumentos relacionados às contribuições dos proprietários, descontados impostos retidos na fonte e outros descontos.

Dessa forma, esta pesquisa buscou responder à seguinte pergunta: Como analisar as DFs de empresas do setor de agronegócios brasileiro, pertencentes ao *ranking* Forbes Agro100 2024, utilizando técnicas de ML?

O objetivo geral deste estudo foi analisar os resultados financeiros das empresas listadas neste *ranking*, com base em dados extraídos das DFs (Balanço Patrimonial e Demonstração do Resultado do Exercício) – valores financeiros obtidos destes documentos e cálculo de outros indicadores (*financial ratios*) – utilizando técnicas analíticas com modelagens de ML (Favaretto,

2024; Morelli *et al.*, 2023), entre elas, a técnica de *Clustering* (Altuntas *et al.*, 2022; Dinh *et al.*, 2025) e de *Principal Component Analysis* (PCA) (Marukatat, 2023; Mbona & Yusheng, 2019; Zhu *et al.*, 2023).

A contribuição esperada desta pesquisa, considerando o papel informacional das DFs e seus indicadores de emissão e de envio de sinais ao mercado, é a identificação de padrões nesses sinais que reforçassem a ideia de que a caracterização de um *ranking* de empresas, com base em um único indicador (receita líquida neste estudo), é limitada e não reflete a situação econômica/financeira real dessas empresas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Teoria da Sinalização

A Teoria da Sinalização, introduzida por Spence (1973, 1974), explica como os participantes de um mercado transmitem informações para reduzir a assimetria informacional, com o processo de sinalização consistindo essencialmente em três componentes: os sinalizadores, os receptores e o próprio sinal. O sinal é uma informação, sendo esta positiva ou negativa, privada, que os sinalizadores (empresas) escolhem comunicar ou não aos receptores (mercado e *stakeholders*) (Connelly *et al.*, 2025).

Segundo a teoria (Connelly *et al.*, 2011; Ross, 1977; Spence, 1973), os gestores das empresas (sinalizadores) possuem informações (sinais) que o mercado e os *stakeholders* (receptores) não têm e, para comunicar ao mercado a qualidade dos resultados da empresa (sua estabilidade, qualidade de governança, saúde financeira, entre outros), os sinalizadores (gestores) emitem sinais aos receptores (investidores, credores, reguladores).

Pesquisadores têm utilizado a Teoria da Sinalização em diversos contextos para compreender a influência dos sinais emitidos por indicadores financeiros e por relatórios organizacionais, bem como a consequente recepção desses sinais pelos mercados. Por exemplo, Al-Sartawi & Reyad (2018) usaram para examinar e relatar a extensão e as características das instituições que determinam as práticas de divulgação financeira on-line (OFD), com base em informações financeiras de 48 bancos islâmicos listados nas bolsas de valores dos países do Conselho de Cooperação do Golfo (GCC), de 2015 a 2017, utilizando seus próprios sites e os sites das bolsas de valores; López-Santamaría *et al.* (2021) estudaram as práticas de divulgação de sustentabilidade nos relatórios de 43 empresas listadas na bolsa de valores da Colômbia, e o sinal mais comum observado foi o da intenção, relacionada a ações futuras; Arhinful *et al.* (2025) estudaram como as finanças sustentáveis fortaleceram a reputação corporativa, usando dados de 17 anos de 143 empresas não financeiras do Reino Unido listadas na bolsa de Londres, e concluíram que a emissão de títulos verdes aumentou significativamente as pontuações ESG, sugerindo que as empresas podem sinalizar seu compromisso com a sustentabilidade por meio de ações visíveis.

Outros trabalhos reforçaram a ampla aplicação da Teoria de Sinalização em estudos da área, como o de: Valencia (2025), que analisou como variáveis de desempenho financeiro e governança corporativa afetavam o valor da empresa em 91 pequenas e médias empresas na Espanha, com resultados indicando que a lucratividade, medida pelo Retorno sobre o Patrimônio Líquido (ROE), teve um impacto positivo e estatisticamente significativo no valor da empresa, ao nível de 10%, e, dessa forma, demonstrou que o mercado reage de maneira favorável à lucratividade, com os investidores utilizando essa informação (sinalização) para tomar decisões de investimento; Eldomiaty *et al.* (2024), ao examinarem os efeitos da sinalização financeira nas empresas sobre a sincronização dos retornos das ações, normalmente refletindo como as empresas gerenciavam suas operações para comunicar sinais aos acionistas, confirmaram empiricamente que indicadores financeiros do balanço patrimonial e da DRE atuaram como sinais com conteúdo

informacional específico sobre a sincronicidade dos retornos acionários; além de Dumitrescu *et al.* (2025) que, ao analisarem dados financeiros e de governança de 60 empresas não financeiras listadas na bolsa de valores de Bucareste, revelaram que o uso de técnicas de IA/ML melhorou significativamente a precisão preditiva e a confiabilidade na classificação e na sinalização de dificuldades financeiras das empresas avaliadas.

2.2 Demonstrações Financeiras (DFs) e Indicadores Financeiros

A globalização econômica afetou a competitividade das organizações, levando-as a criar valor para seus stakeholders. Nesse contexto, ocorreram grandes fraudes no mercado, como o caso da Enron. Para prevenir essas fraudes, foi instituída a Lei Sarbanes-Oxley (SOX) (Estados Unidos, 2002), com o objetivo de proteger os investidores e aumentar a confiabilidade das informações financeiras. Como resultado, as empresas passaram a divulgar suas políticas contábeis e a elaborar suas DFs de acordo com as normas da *International Accountig Standards* (IAS) (Silva *et al.*, 2015).

No Brasil, há uma forte influência das normas visando à harmonização da contabilidade brasileira com os padrões internacionais estabelecidos pela *International Financial Reporting Standards* (IFRS), ou Normas Internacionais de Relatório Financeiro, afetando até mesmo as pequenas empresas (Belmonte & Santos, 2015).

A divulgação das DFs para as empresas que estão listadas na bolsa de valores tem caráter obrigatório, contudo a divulgação voluntária pode refletir em otimismo e boa imagem aos *stakeholders* (Rufino & Monte, 2014).

As DFs são obrigatoriamente compostas por Balanço Patrimonial, Demonstração de Resultado do Exercício, Demonstração do Fluxo de Caixa e Notas Explicativas, podendo apresentar também a Demonstração do Resultado Abrangente, Demonstração da Mutação do Patrimônio Líquido, as quais oferecem uma representação estruturada da posição patrimonial e financeira, dos desempenhos e do fluxo de caixa da empresa (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a).

A partir de diversos valores financeiros e cálculos extraídos de informações publicadas nas DFs, é possível estruturar uma análise completa de empresas, que pode auxiliar na tomada de decisão mais embasada, ou mesmo na previsão dos próximos passos da empresa (Ribeiro *et al.*, 2024).

O lucro líquido é o resultado da empresa após todas as receitas, custos, despesas e tributos e participações terem sido contabilizados (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a). Ele indica se a empresa teve ganho ou prejuízo no período, servindo de base para o cálculo de dividendos, a avaliação do desempenho financeiro e a tomada de decisão por investidores e credores.

A receita líquida é calculada a partir da receita bruta de vendas menos deduções como devoluções, abatimentos, descontos concedidos e impostos sobre vendas (ICMS, PIS, Cofins) (Comitê de Pronunciamentos Contábeis, 2012a; International Financial Reporting Standards, 2001b) e serve para refletir o valor efetivamente recebido ou a receber pela empresa.

O ativo circulante é o conjunto de bens e direitos que serão realizados até o fim do próximo exercício (Comitê de Pronunciamentos Contábeis, 2011, 2019; International Financial Reporting Standards, 2001a, 2010) e demonstra a liquidez imediata da empresa contendo itens como caixa, contas a receber, estoques etc. O ativo não circulante refere-se aos bens e direitos, dessa vez, a longo prazo, ou seja, após o fim do próximo exercício (Comitê de Pronunciamentos Contábeis, 2009, 2010; International Financial Reporting Standards, 2001d, 2001b), e representa os recursos de longo prazo importantes para a operação contínua da empresa, como máquinas, imóveis, marcas, investimentos, entre outros. O ativo total é a soma do ativo circulante e não circulante (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards,

2001a) e mostra a dimensão econômica da empresa, ele é utilizado em análises, como indicadores de rentabilidade e de avaliação patrimonial.

O passivo circulante engloba as obrigações que a empresa deve pagar até o fim do próximo exercício (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a) e indica os compromissos de curto prazo, como fornecedores, salários a pagar etc. O passivo não circulante diz respeito às obrigações exigíveis no longo prazo (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a) e representa as dívidas e os compromissos de longo prazo, como debêntures e provisões por contingências. O passivo total é a soma do passivo circulante e não circulante (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a) e indica tudo o que a empresa deve a terceiros. Pode ser utilizado para avaliar o grau de endividamento e realizar análises de solvência e de estrutura de capital.

O patrimônio líquido é a diferença entre o ativo total e o passivo total, e representa os recursos próprios dos sócios ou acionistas (Comitê de Pronunciamentos Contábeis, 2011, 2019; International Financial Reporting Standards, 2001a, 2010), ele pode ser utilizado na base de cálculo do retorno sobre o patrimônio líquido (ROE) e demonstra quão sólida a empresa é a longo prazo. A literatura reporta estudo de Melo *et al.* (2024) que analisou o efeito da folga financeira sobre a tempestividade dos demonstrativos contábeis em companhias listadas na B3 e, dentre diversas variáveis, utilizou indicadores de liquidez corrente e de retorno sobre os ativos (ROA) na etapa empírica.

Estudos internacionais que utilizaram dados extraídos de DFs e de relatórios organizacionais de empresas de Bangladesh (Hassan *et al.*, 2020), da China (Mbona & Yusheng, 2019), da Colômbia (López-Santamaría *et al.*, 2021), da Espanha (Valencia, 2025), da Romênia (Dumitrescu *et al.*, 2025) e da Turquia (Özari *et al.*, 2025) exemplificam e confirmam que essa prática de exploração de dados, fundamentada em indicadores financeiros e/ou na Teoria da Sinalização, é frequente e relevante na literatura científica, especialmente na descoberta de novos conhecimentos relacionados à gestão organizacional.

Como as DFs constituem um dos principais meios formais de comunicação (sinalização) entre gestores e *stakeholders*, como investidores, credores e reguladores, para representar a situação econômica e financeira da empresa em um período contábil, a utilização de análise multivariada com técnicas de ML pode captar padrões nesses sinais que permitem distinguir grupos de empresas conforme seus perfis sinalizadores.

2.3 Agronegócio do Brasil

O agronegócio é um setor fundamental em vários países, pois envolve o cultivo de soja, trigo, milho e outras commodities. O setor contribui para a geração de empregos, promove superávits comerciais (balança comercial positiva) e impulsiona o crescimento econômico (Bornhofen *et al.*, 2019).

O agronegócio no Brasil é um dos principais setores da economia do país (Governo Federal do Brasil, 2025b), com participação relevante no PIB brasileiro: em 2024 representou 23,2 %, indo além do setor primário, para 'fora da porteira', sendo que está presente nas indústrias e serviços (Cepea, 2025b).

Este mesmo setor também tem destaque nas exportações, representando 49% de tudo o que o país enviou ao exterior, equivalente a US$ 164,4 bilhões no ano de 2024, o que mostra a sua resiliência, mantida em destaque ao longo dos anos (Cepea, 2025a).

O agronegócio brasileiro é referência mundial quando se trata de produção e comércio, sendo líder no mercado de soja, algodão, carne bovina, suco de laranja, devido a fatores como a abundância de recursos naturais, diversificação de culturas, vasto território, aumento das

tecnologias implantadas na produção, colocando o Brasil como um dos maiores *players* do mercado global (Cardoso *et al.*, 2025).

Devido à grande importância do agronegócio no Brasil, a partir da sua participação no PIB, sua liderança no mercado global, sua participação majoritária na balança comercial brasileira, e no impacto da economia brasileira (Moreira *et al.*, 2016), fica evidente e relevante a necessidade de desenvolver estudos sobre este setor.

2.4 Perspectivas Multivariadas com técnicas de Machine Learning (ML)

Apesar da existência de variadas formas de avaliação de um agrupamento de empresas, com base em um corpo de conhecimento na área financeira/contábil, dentre elas, aquelas que abrangem taxonomias financeiras de firmas (Altman *et al.*, 2017; Fodor *et al.*, 2021), segmentação de empresas por desempenho e estrutura de capital (Correa *et al.*, 2013; Machado *et al.*, 2016), e comparações com estudos setoriais (Perressim *et al.*, 2026; Silva *et al.*, 2025; Souza *et al.*, 2022), esta pesquisa foi orientada pela perspectiva da análise multivariada das demonstrações financeiras, com foco na aplicação de técnicas de ML.

As técnicas analíticas multivariadas de ML, incluindo clustering e PCA, fazem parte de um subconjunto do campo de estudos de IA (Favaretto, 2024), representado por métodos computacionais que aprendem padrões a partir de conjuntos de dados. Na área de contabilidade e auditoria, técnicas de ML podem ser aplicadas na resolução de diversas tarefas, como análise de transações, avaliação de riscos, previsão de fraudes, falências, distorções relevantes e estimativas contábeis, dentre outras (Cho *et al.*, 2020). Por exemplo, estudo de Assis & Decoster (2025) usou uma técnica de ML chamada *Gradient Boosting Decision Tree* (GBDT) na criação e na execução de um modelo preditivo de gestão de risco de crédito, combinando variáveis financeiras e não financeiras, com o intuito de obter maior acurácia na tomada de decisões e melhorar a mitigação de riscos.

A pesquisa de Dzuba & Krylov (2021) demonstrou que os métodos de clustering podem categorizar empresas com base nas estratégias financeiras adotadas por elas, assim como o trabalho de Fodor *et al.* (2021) também usou clustering para identificar grupos de grandes empresas dos EUA com base em suas demonstrações financeiras, incluindo a divisão percentual dos balanços patrimoniais e das demonstrações de resultados.

O estudo de Pooja *et al.* (2024) utilizou técnicas de clustering, combinadas com o Modelo de Precificação de Ativos de Capital (CAPM), para identificar diferentes níveis de apetite ao risco entre os clientes, permitindo recomendações de investimento personalizadas que aumentam a satisfação e o desempenho das carteiras.

Portanto, a técnica de ML conhecida como Clustering abrange métodos que agruparam observações, objetos ou itens semelhantes entre si, com base em características observadas ou mensuradas (Jain, 2010), e é amplamente utilizada em estudos da área.

O Clustering hierárquico aglomerativo (Altuntas *et al.*, 2022; Lu *et al.*, 2008) é um método de *Clustering* que combina pares em cada iteração, com o objetivo de alcançar a máxima similaridade dentro de cada cluster e a mínima entre os clusters. O resultado desse tipo de agrupamento é apresentado com os dados formando uma estrutura em árvore chamada *dendrograma* (Gonzales & Hargreaves, 2022), sendo um modo de facilitar a visualização das relações estabelecidas na integração das observações.

De forma complementar, o método de *Clustering particional k-means* é um dos mais popularmente utilizados, devido à sua facilidade de uso e eficiência em seus resultados (Jain, 2010; Sinaga & Yang, 2020; Steinley, 2006; Ybañez *et al.*, 2023), em que as observações são aleatoriamente divididas em *k* grupos e reorganizadas repetidamente para formar grupos coesos.

É usual em pesquisas acadêmicas a utilização dos dois métodos de *Clustering*, como por exemplo, *hierárquico aglomerativo* e *particional k-means,* para permitir a

comparação/complementação dos resultados obtidos em cada um deles – enquanto o primeiro se torna útil para "sugerir" visualmente o número de *clusters* com a exploração do dendrograma, o segundo, com a utilização de métricas como o maior coeficiente *Silhouette* para a quantidade de *k* grupos, "indica" a qualidade da formação dos *clusters* e aplicabilidade prática do modelo.

Por outro lado, a literatura também menciona estudos que utilizaram a técnica analítica de PCA, como o de Hu (2025) na gestão de riscos em sistemas contábeis, com o objetivo de reduzir a dimensionalidade de grandes conjuntos de dados financeiros, preservando padrões e relações essenciais, e, assim, identificando variáveis significativas que capturavam fatores de risco, como detecção de fraudes, anomalias ou transações irregulares; a pesquisa de Ongbali *et al.* (2024) utilizou PCA para identificar as variáveis críticas que influenciavam o crescimento de Pequenas e Médias Empresas (PMEs), determinando os fatores principais a partir das muitas variáveis disponibilizadas; o estudo de Mbona & Yusheng (2019) que utilizou PCA e indicadores financeiros para analisar três empresas chinesas de telecomunicações, constatou que o desempenho financeiro podia ser avaliado com um conjunto menor e mais gerenciável de indicadores, em vez de uma gama vasta e complexa, o que dificultaria a própria interpretação; assim como o estudo de Gurbaxani *et al.* (2023) que investigou variáveis que afetavam as percepções de investidores sobre Ouro Digital e ETF de Ouro como opções de investimento e, por meio da aplicação da técnica de PCA, identificou fatores que influenciavam as decisões de investimento.

Considerada uma técnica de redução de dimensionalidade utilizada em estudos da estatística multivariada e de ML (Abdi & Williams, 2010), a técnica de PCA busca transformar um conjunto de variáveis possivelmente correlacionadas em um conjunto de variáveis não correlacionadas, chamadas de Componentes Principais (PCs).

Ao reduzir a dimensionalidade dos dados originais (Garcia-Alvarez *et al.*, 2023), preservando o máximo possível de sua variabilidade (variância total), esse processo gera PCs ordenadas em função da quantidade de informação (variância explicada) que elas carregam (Marukatat, 2023). Isso torna a análise mais simples, visualmente interpretável e computacionalmente eficiente, permitindo que pesquisadores analisem apenas os primeiros componentes, sem comprometer significativamente a essência do conjunto de dados.

As técnicas de *Clustering* e PCA são frequentemente aplicadas em conjunto, pois, enquanto o *Clustering* busca identificar homogeneidade de observações dentro dos grupos e heterogeneidade entre eles, a PCA reduz a dimensionalidade das variáveis, capturando a maior parte da variância dos dados em duas ou três PCs. Ao projetar as observações originais em uma visualização gráfica, tendo como eixos as duas primeiras componentes (gráfico do tipo scatterplot), é possível visualizar e comparar padrões de dispersão entre as organizações, além de identificar grupos formados.

3 PROCEDIMENTOS METODOLÓGICOS

3.1 Coleta de dados e estruturação do conjunto de dados inicial

O conjunto de dados (*dataset*) inicial desta pesquisa foi construído a partir do *ranking* divulgado publicamente pela revista Forbes Brasil em outubro de 2024 – o Forbes Agro100 Brasil (Rodrigues *et al.*, 2025), que reuniu as cem maiores empresas do setor de agronegócio brasileiro com base em suas receitas líquidas de 2023.

Foi possível identificar que as empresas classificadas neste *ranking* totalizaram mais de R$ 1,7 trilhão em suas receitas líquidas e foram segmentadas em sete categorias, com o subtotal de geração de receita e de empresas integrantes, sendo: Agroenergia com R$ 270,07 bilhões e 23 empresas; Agroquímica e Insumos com R$ 61,54 bilhões e 10 empresas; Alimentos e Bebidas com R$ 416,02 bilhões e 15 empresas; Celulose, Madeira e Papel com R$ 81,18 bilhões e 8 empresas; Comércio e Tradings com R$ 125,26 bilhões e 16 empresas; Cooperativas com R$ 140,82 bilhões

e 12 empresas; e Proteína Animal com R$ 617,60 bilhões e 16 empresas – com essa última categoria apresentando o maior volume em receitas líquidas.

3.2 Demonstrações Financeiras e ampliação do conjunto de dados

Para ampliar o número de mensurações financeiras a serem analisadas, os autores decidiram também consultar as DFs das empresas a partir de fontes públicas, tais como: relatórios anuais disponíveis no site das empresas, plataforma da Comissão de Valores Mobiliários (CVM) (CVM, 2025; Governo Federal do Brasil, 2025a), e jornais publicados na época de suas divulgações (Valor Econômico, 2025).

O *dataset* original da pesquisa foi ampliado, passando a incluir não só a receita líquida (NR_bi), mas também o lucro líquido (NI_bi), o ativo circulante (CA_bi) e não circulante (NCA_bi), o passivo circulante (CL_bi) e o não circulante (NCL_bi), e o patrimônio líquido (EQ_bi) – todos os valores mensurados em bilhões de reais (R$ bilhões).

Das cem empresas pertencentes ao *ranking* Forbes Agro 100, não foi possível identificar as DFs de 16 empresas, sendo estas identificadas aqui conforme três quesitos (posição no *ranking* original – nome – categoria que pertence): #9 - Grupo Amaggi - Alimentos e Bebidas; #10 - Louis Dreyfus Company Brasil - Comércio e Tradings; #17 - Aurora Alimentos - Alimentos e Bebidas; #22 - Inpasa Brasil – Agroenergia; #29 - UPL do Brasil - Agroquímica e Insumos; #31 - Laticínios Bela Vista Piracanjuba - Proteína Animal; #34 - Três Corações - Alimentos e Bebidas; #50 - Cotrijal – Cooperativas, #55 - Fiagril - Comércio e Tradings; #59 - ICL - Agroquímica e Insumos, #74 - Delta Sucroenergia – Agroenergia; #80 - Atvos Bioenergia Brenco – Agroenergia; #81 - Fazendão Agronegócio - Comércio e Tradings; #82 - Prima Foods - Proteína Animal; #89 - SJC Bioenergia – Agroenergia; #99 - Barra Mansa Carnes - Proteína Animal; e por esse motivo, os pesquisadores decidiram remover tais empresas do *dataset* original da pesquisa, o qual, após esta etapa, passou a totalizar um número de 84 empresas.

Tabela 1

Recorte dos dados com as mensurações financeiras extraídas das DFs das Top 10 empresas

Nome das Empresas	NR_bi	NI_bi	CA_bi	NCA_bi	CL_bi	NCL_bi	EQ_bi
JBS	363,817	-0,729	78,605	127,527	47,914	111,220	46,998
Marfrig Global Foods	132,218	-3,621	52,101	78,854	42,171	63,894	24,890
Cargill	124,134	2,428	27,797	11,432	21,932	6,551	10,745
Bunge Alimentos	81,704	2,619	14,878	9,966	9,463	3,390	11,991
Ambev	79,737	14,961	36,563	96,081	41,005	11,495	80,144
Raízen Energia	78,450	0,645	42,646	39,538	25,776	34,236	22,172
Copersucar	70,144	0,679	12,150	3,461	7,036	6,395	2,181
BRF	53,615	-1,869	25,954	31,318	19,390	22,238	15,644
Suzano	39,756	14,106	38,569	105,024	14,795	83,988	44,810
Cosan	39,468	4,885	28,610	111,255	16,159	72,725	50,981

Nota. valores em bilhões de reais (R$), referentes ao ano de 2023, para: Receita Líquida (NR_bi), Lucro Líquido (NI_bi), Ativo Circulante (CA_bi), Ativo Não Circulante (NCA_bi), Passivo Circulante (CL_bi), Passivo Não Circulante (NCL_bi), e Patrimônio Líquido (EQ_bi).

Fonte: elaborada pelos autores a partir das DFs das 84 empresas da amostra inicial deste trabalho.

A Tabela 1 apresenta um recorte da amostra de dados com as 7 mensurações financeiras (NR_bi, NI_bi, CA_bi, NCA_bi, CL_bi, NCL_bi, EQ_bi), em bilhões de reais (R$), extraídas das DFs das Top 10 empresas entre as 84 que disponibilizaram publicamente seus dados, sendo 47 identificadas como ‘não listadas’ na bolsa de valores e 37 como ‘listadas’.

3.3 Indicadores calculados e incluídos no conjunto de dados da pesquisa

Após a consulta da literatura sobre indicadores financeiros utilizados em estudos da área (Correa-Mejía & Lopera-Castaño, 2020; Vukčević *et al.*, 2024), por escolha dos pesquisadores, foram selecionados para cálculo outros 7 indicadores relacionados ao desempenho de cada uma das 84 empresas: o Retorno sobre os Ativos (ROA), o Retorno sobre o Patrimônio Líquido (ROE), a Margem de Lucro Líquido (NPM), a Liquidez Corrente (CR), a Liquidez Geral (OLR), o Índice de Endividamento (DAR) e o Índice de Endividamento de Terceiros (DER).

Para cada uma destas 14 mensurações financeiras utilizadas na pesquisa, foram identificados o seu papel nas DFs das organizações estudadas e as respectivas formulações de cálculo, conforme informado na Tabela 2.

Tabela 2

Lista das mensurações financeiras como variáveis exploradas neste estudo

Rótulo da variável	Descrição da mensuração	Fonte dos dados / Cálculo
NR_bi	Receita Líquida (em bilhões de R$): é a receita bruta de vendas menos deduções, como devoluções, abatimentos, descontos concedidos e impostos sobre as vendas.	Valores extraídos diretamente das respectivas Demonstrações Financeiras (DFs) das empresas estudadas.
NI_bi	Lucro Líquido (em bilhões de R$): é o resultado da empresa após a contabilização de todas as receitas, custos, despesas, tributos e participações.	
CA_bi	Ativo Circulante (em bilhões de R$): é o conjunto de bens e direitos que serão realizados até o fim do próximo exercício.	
NCA_bi	Ativo Não Circulante (em bilhões de R$): são bens e direitos de longo prazo, ou seja, após o fim do próximo exercício.	
CL_bi	Passivo Circulante (em bilhões de R$): são as obrigações que a empresa deve pagar até o fim do próximo exercício.	
NCL_bi	Passivo Não Circulante (em bilhões de R$): são as obrigações exigíveis no longo prazo e incluem dívidas e compromissos, como debêntures e provisões para contingências.	
EQ_bi	Patrimônio Líquido (em bilhões de R$): é a diferença entre o ativo total e o passivo total e representa os recursos próprios dos sócios ou acionistas.	
ROA	Retorno sobre os Ativos: mensura a capacidade da empresa de gerar lucro com o total de seus ativos.	$\text{ROA} = \frac{\text{Lucro Líquido}}{\text{Ativo Total}}$
ROE	Retorno sobre o Patrimônio Líquido: indica quanto de lucro líquido a empresa gera a cada unidade monetária investida pelos sócios.	$\text{ROE} = \frac{\text{Lucro Líquido}}{\text{Patrimônio Líquido}}$
NPM	Margem de Lucro Líquida: mostra o quanto sobra de lucro para cada unidade monetária de receita líquida e avalia a eficiência operacional e financeira da empresa.	$\text{Margem Líquida} = \frac{\text{Lucro Líquido}}{\text{Receita Líquida}}$
CR	Liquidez Corrente: indica a capacidade da empresa de pagar suas dívidas de curto prazo com seus ativos de curto prazo.	$\text{Liquidez Corrente} = \frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$

OLR	Liquidez Geral: avalia a capacidade de pagamento global, na relação entre os ativos totais realizáveis e os passivos totais exigíveis.	$\text{Liquidez Geral} = \frac{(\text{Ativo Circulante} + \text{Ativo não Circulante})}{(\text{Passivo Circulante} + \text{Passivo não Circulante})}$
DAR	Índice de Endividamento: utilizado como indicador de solvência, pois mostra a relação entre o total de dívidas e o ativo total da empresa, avaliando o nível de risco financeiro.	$\text{Endividamento} = \frac{\text{Passivo Total}}{\text{Ativo Total}}$
DER	Índice de Endividamento de Terceiros: compara o capital de terceiros ao capital próprio e mensura a alavancagem financeira, indicando o quanto a empresa utiliza capital de terceiros em relação ao capital próprio.	$\text{Endividamento de Terceiros} = \frac{\text{Passivo Total}}{\text{Patrimônio Líquido}}$

Fonte: elaborada pelos autores.

A Tabela 2 apresenta o total de 14 variáveis (mensurações financeiras) exploradas neste estudo: 7 variáveis (em valores absolutos, em bilhões de R$) coletadas diretamente a partir das demonstrações financeiras das empresas (como descrito na subseção 3.2), além de outras 7 variáveis (em valores relativos), calculadas conforme a fórmula informada.

3.4 Justificativas das decisões metodológicas

A partir da disponibilidade das mensurações financeiras das 84 organizações que compuseram a amostra inicial deste estudo, sendo 14 variáveis numéricas sem valores faltantes, para evitar a ocorrência de *overfitting*, multicolinearidade e/ou mesmo a "duplicação de peso" de determinadas variáveis nos resultados que seriam obtidos após a execução dos modelos de ML, os pesquisadores realizaram testes de correlação linear de Pearson (Schober *et al.*, 2018) entre essas 14 variáveis, com a finalidade de escolher e remover variáveis altamente correlacionadas, o que resultou na identificação de um conjunto de 5 variáveis (NR_bi, NI_bi, ROA, ROE, CR) com baixa ou nenhuma correlação entre si.

Em seguida, foi realizado também o teste de Mahalanobis (Cabana *et al.*, 2021; Mahalanobis, 1930) para a detecção de observações *outliers* multivariadas (medidas discrepantes), identificando-se 4 observações *outliers*, com as posições de *ranking* e os nomes das empresas sendo: #1 JBS, #5 Ambev, #9 Suzano e #43 Fertilizantes Heringer.

Após a checagem detalhada dessas observações no conjunto de dados da pesquisa, foram percebidos valores atípicos, bastante discrepantes em relação às demais empresas da amostra, em pelo menos uma das 5 variáveis mensuradas, sendo elas: para Fertilizantes Heringer, o valor de ROE de -48,0249; para Suzano e Ambev, o valor de NI_bi de 14,1064 e 14,9605, respectivamente; e para JBS, o valor de ROE de -0,0155 e de NR_bi de 363,8165.

Assim, como o intuito da pesquisa era compreender, coletivamente, o maior número possível de organizações do setor participante do *ranking*, por decisão dos pesquisadores, essas 4 mensurações consideradas *outliers* foram excluídas da amostra para evitar que influenciassem as próximas etapas do estudo.

Dessa forma, a amostra final do estudo totalizou 80 empresas, sendo 47 não listadas na bolsa (58,8% do total) e 33 listadas (41,2% do total). A Tabela 3 apresenta as estatísticas descritivas das 5 variáveis utilizadas nos modelos de ML.

Tabela 3
Estatísticas descritivas das variáveis utilizadas nos modelos de ML estudados

Variável	Média	Desvio Padrão	Mínimo	Q1	Mediana	Q3	Máximo
NR_bi	13,166	24,301	1,858	3,087	5,178	8,565	132,218
NI_bi	0,363	0,975	-3,621	0,042	0,230	0,507	4,885
ROA	0,045	0,063	-0,157	0,012	0,042	0,077	0,236
ROE	0,128	0,296	-1,511	0,068	0,144	0,221	1,462
CR	1,859	0,993	0,170	1,260	1,580	2,172	5,267

Nota. amostra final de dados desta pesquisa (n = 80 organizações), sem dados faltantes; valores absolutos em bilhões de reais (R$) para Receita Líquida (NR_bi) e Lucro Líquido (NI_bi); valores relativos (*ratios*) para o Retorno sobre os Ativos (ROA), o Retorno sobre o Patrimônio Líquido (ROE) e a Liquidez Corrente (CR).
Fonte: elaborada pelos autores.

Como amplamente fundamentado na subseção 2.4, os pesquisadores se apoiaram em estudos da literatura de ciências contábeis e finanças para a condução das demais etapas desta pesquisa com foco na análise multivariada com aplicação das técnicas de ML de clustering (Dang *et al.*, 2022) e PCA (Mbona & Yusheng, 2019).

4 RESULTADOS E ANÁLISES

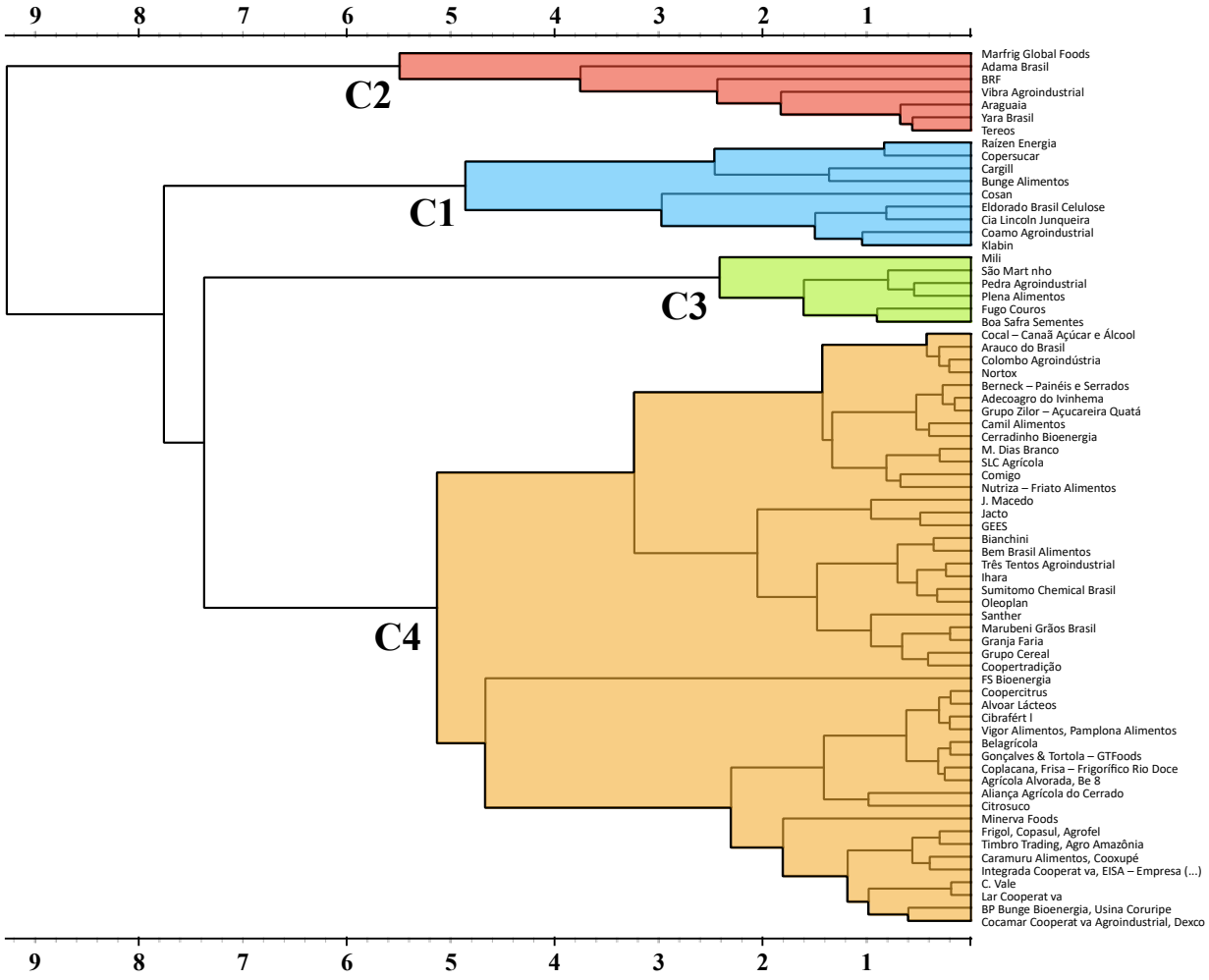
Para evitar que variáveis com magnitudes maiores dominassem os resultados dos modelos de ML, foi realizada a padronização estatística do conjunto das 5 variáveis escolhidas pelos pesquisadores – por meio da subtração da média e da divisão pelo desvio padrão, o que resultou em colunas com média = 0 e desvio padrão = 1, com a utilização do *software* Orange Data Mining (Demšar *et al.*, 2013) e dos *softwares* R / RStudio (Posit Team, 2025; R Core Team, 2025).

4.1 Modelos de *Clustering*

A partir da amostra final do estudo, composta por n=80 organizações, foi possível aplicar técnicas de ML de *Clustering* (Sinaga & Yang, 2020) com a intenção de identificar e interpretar o que as empresas tinham em comum, a partir desse conjunto de elementos, tanto de indicadores gerados pelo estudo quanto de valores financeiros absolutos extraídos diretamente das DFs.

A primeira técnica executada foi a de *clustering hierárquico* (Lu *et al.*, 2008) usando a distância euclidiana normalizada e o método de ligação de Ward, o que resultou na geração da Figura 1, um *dendrograma* – uma construção visual que reúne cada uma das camadas formadas durante o procedimento de geração dos *clusters*.

Figura 1
Dendrograma com a identificação de quatro clusters (C1, C2, C3 e C4)



Nota. após a execução do método de *clustering hierárquico* com a amostra de dados final desta pesquisa (n=80 organizações) e a formação de quatro *clusters* (C1, C2, C3 e C4).
Fonte: elaborada pelos autores com a utilização do *software* Orange Data Mining (Demšar *et al.*, 2013).

Ao gerar a estrutura do *dendrograma* da Figura 1, identificou-se que o conjunto de 80 organizações (observações) foi dividido em quatro *clusters* (C1 = 9 observações, C2 = 7 observações, C3 = 6 observações e C4 = 58 observações). A literatura considera que cabe aos pesquisadores observarem se os *clusters* gerados fazem sentido com o problema de pesquisa que está sendo estudado, e dessa forma, tomarem a decisão da escolha do número adequado de *clusters* (Kar & Varsha, 2023) para dar continuidade à condução das etapas restantes.

Com a intenção de auxiliar nesta decisão e avaliar a estabilidade da solução obtida da modelagem de *clustering hierárquico*, uma segunda modelagem foi executada, agora com a técnica de *clustering k-means* (Jain, 2010), que identificou o valor de 0,336 como o maior valor da métrica do coeficiente *Silhouette* (Bagirov *et al.*, 2023) e sugeriu a mesma quantidade de quatro clusters, mas propôs a troca de posicionamento de algumas observações entre os *clusters*.

As trocas propostas geraram um incremento do número de observações no cluster C3, vindos em sua maioria de 11 observações do cluster C4, observações essas representadas pelas empresas e posição do *ranking* original: Comigo (20), M. Dias Branco (24), Jacto (71), J. Macedo (72), Bem Brasil Alimentos (73), Colombo Agroindústria (83), Arauco do Brasil (84), Cocal – Canaã Açúcar e Álcool (90), Nutriza – Friato Alimentos (93), GEES (95), e Nortox (97); além da

mudança do posicionamento de uma observação do cluster C1 para o cluster C3: a observação referente à empresa Cia Lincoln Junqueira (53).

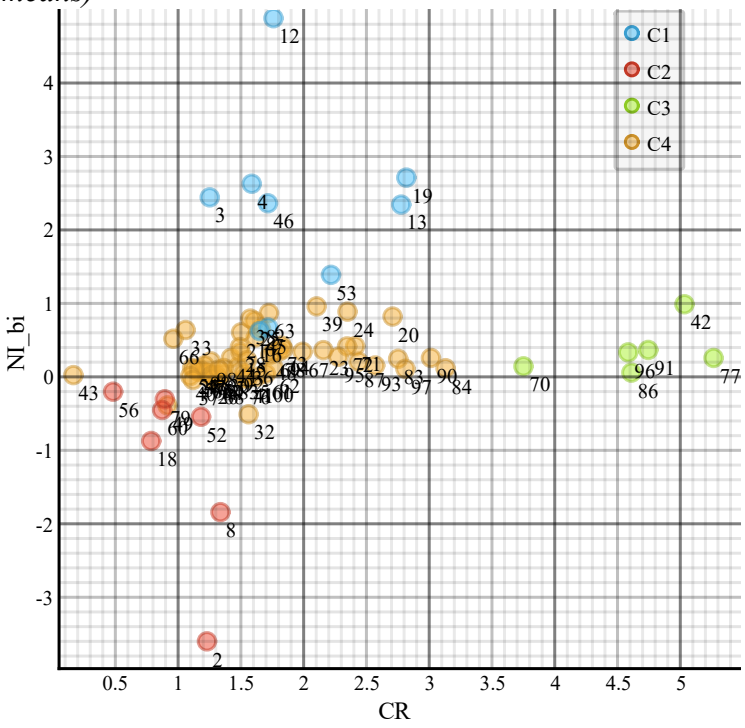
Portanto, após a formação dos *clusters* pelo modelo *hierárquico* e as trocas de posição de 12 observações, os pesquisadores optaram por utilizar integralmente os resultados do modelo *k-means*, pois este se mostrou mais adequado ao problema de pesquisa, apresentando alta estabilidade e reforçando a confiabilidade da solução escolhida.

Assim, a composição final do número de observações em cada *cluster* pelo modelo *k-means* ficou distribuída da seguinte forma: o *cluster* C1=8 empresas, o *cluster* C2=7 empresas, o *cluster* C3=18 empresas e o *cluster* C4=47 empresas.

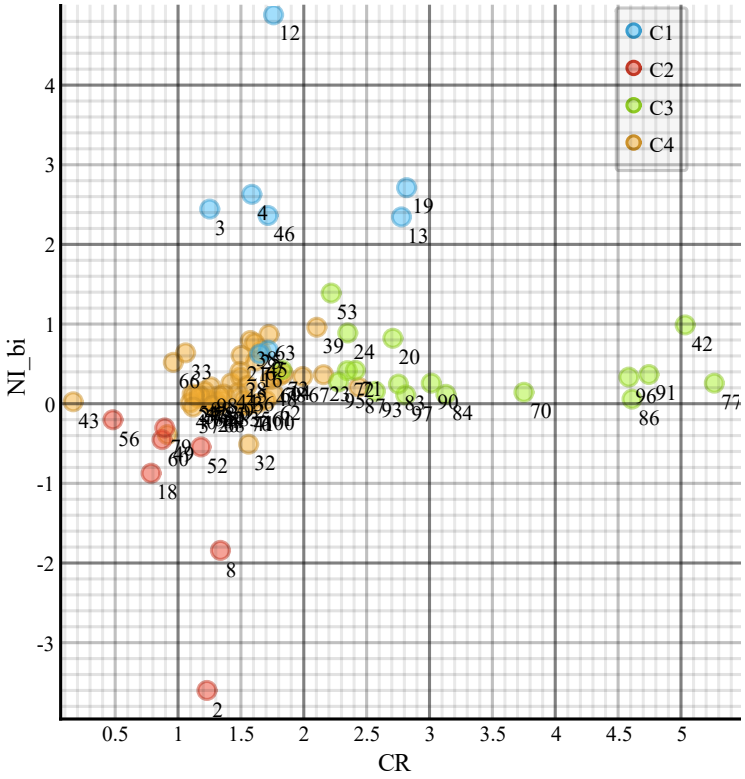
Para ilustrar as trocas de posicionamento das observações (empresas) entre os clusters, do modelo *hierárquico* para o modelo *k-means*, foi elaborada a Figura 2.

Figura 2

Gráficos de Scatterplot com os clusters formados por dois modelos diferentes (hierárquico e k-means)



(a) modelo *hierárquico*



(b) modelo *k-means*

Nota. n=80 organizações. Na parte (a) da figura, é apresentado o resultado da modelagem de *clustering hierárquico*; na parte (b), o resultado da modelagem de *clustering k-means*.
Fonte: elaborada pelos autores com a utilização do *software* Orange Data Mining (Demšar *et al.*, 2013).

Na análise visual do gráfico de *Scatterplot* da Figura 2, elaborado com as mensurações de duas variáveis originais – Liquidez Corrente (CR) no eixo x e Lucro Líquido (NI_bi) no eixo y –, identificou-se claramente a formação de quatro clusters (C1, C2, C3 e C4) entre as 80 organizações analisadas. A parte (a) da figura mostra os resultados obtidos com a modelagem *hierárquica*, e a parte (b) desta mesma figura, exibe os resultados obtidos com a modelagem *k-means*. Ao observar simultaneamente a região central das partes (a) e (b) da Figura 2, percebe-se que há uma concentração de observações incluindo as empresas posicionadas no *ranking* com os números 53, 24, 20, 84, e 90, confirmando que a separação dos *clusters* conduzida pela modelagem *k-means*, a qual incluiu estas observações mencionadas no *cluster* C3 representada pela cor verde, foi mais eficiente que a modelagem *hierárquica*, que havia mantido tais observações distribuídas em dois outros *clusters* (C1 e C4).

Dos quatro clusters formados, o Cluster C1, na cor azul, que recebeu o nome de "Lucrativas", foi formado por 8 observações (Cargill, Bunge Alimentos, Raízen Energia, Copersucar, Cosan, Coamo Agroindustrial, Klabin, Eldorado Brasil Celulose), trouxe o indicador de Liquidez Corrente (NR) com valores positivos variando de 1,267 até 2,808, e o maior valor de Lucro Líquido (NI_bi) de R$ 4,885 bilhões entre todas as empresas da amostra, obtido pela empresa Cosan – posicionada na 12ª colocação do *ranking*, integrando a categoria de Agroenergia e listada na bolsa de valores, além de conter outras cinco empresas com valores de lucro líquido acima de 2,323 bilhões de reais.

O Cluster C2, na cor vermelha, que recebeu o nome de "Encolhidas", foi composto por 7 empresas (Marfrig Global Foods, BRF, Yara Brasil, Tereos, Araguaia, Adama Brasil, Vibra Agroindustrial), contendo três empresas listadas na bolsa, trazendo todos os valores negativos para os indicadores ROE e ROA, o que indica operações empresariais com prejuízos e baixa rentabilidade no período avaliado, com todos os valores de Lucro Líquido (NI_bi) negativos, sendo

os maiores identificados em R$ -3,621 bilhões (Marfrig Global Foods, 2ª colocação do *ranking*) e o valor de R$ -1,869 bilhões (BRF, 8ª colocação do *ranking*).

O Cluster C3, na cor verde, que recebeu o nome de "Saudáveis", ficou constituído por 18 empresas (ver a lista com todos os nomes na Tabela 7), com estas apresentando Lucro Líquido (NI_bi) positivo (mediana de R$ 0,310 bilhões), e duas delas, participantes da categoria de Agroenergia, obtiveram lucro acima de um bilhão de reais, sendo de R$ 1,404 bilhões para a empresa Cia Lincoln Junqueira (53ª colocação do *ranking*) e de R$ 1,016 bilhões para a empresa São Martinho (42ª colocação do *ranking*). Ao comparar o indicador de Liquidez Corrente (CR), apenas uma empresa teve o indicador mensurado abaixo de 2,219 – a Bem Brasil Alimentos, com liquidez de 1,832, não listada na bolsa, participante da categoria de Alimentos e Bebidas, e na 73ª colocação do *ranking*. Por outro lado, a empresa Pedra Agroindustrial, ocupando a 77ª posição do *ranking*, listada na bolsa de valores e participante da categoria de Agroenergia, foi a que apresentou a maior liquidez corrente (mensuração de 5,267) entre todas as empresas da amostra estudada. Todas as empresas do *cluster* C3 também apresentaram indicadores positivos de ROE (mediana de 0,172) e de ROA (mediana de 0,09), o que reflete saúde financeira, rentabilidade e solidez.

E o Cluster C4, na cor laranja ou ocre, o maior entre todos os demais, recebeu o nome de "Diversificadas", e foi formado por um total de 47 empresas, conforme lista disponibilizada na Tabela 7, sendo dezoito empresas (38,3%) listadas na bolsa de valores, e vinte e nove empresas (61,7%) não listadas na bolsa, foi constituído por sete empresas que apresentaram ROA e Lucro Líquido (NI_bi) negativos (FS Bioenergia, Belagrícola, Cibrafértil, Citrosuco, Gonçalves & Tortola – GTFoods, Vigor Alimentos, Frisa – Frigorífico Rio Doce), com a empresa FS Bioenergia trazendo o valor mais negativo deste *cluster* (R$ -0,489 bilhões), assim como cinco destas sete empresas mencionadas também possuíam ROE negativos. Diferentemente do agrupamento C1 (empresas chamadas de "Lucrativas"), os valores de Lucro Líquido (NI_bi) deste agrupamento C4 de empresas não ultrapassaram o montante de R$ 1 bilhão, sendo que a empresa mais bem colocada neste resultado foi a SLC Agrícola, com lucro líquido de R$ 0,938, pertencente à categoria de Alimentos e Bebidas, além de estar listada na bolsa e ocupar a 39ª posição do *ranking*. O indicador de liquidez corrente (CR) deste grupo de empresas apresentou valores entre 0,167 (mínimo) e 2,441 (máximo), praticamente inferiores aos de quase todas as empresas do *cluster* C3. Este cluster reuniu um conjunto eclético de empresas, caracterizado por não apresentar grandes lucros, por apresentar indicadores financeiros positivos e negativos e por misturar empresas listadas e não listadas em bolsa, além de apontar disparidade nos níveis de liquidez corrente.

4.2 Modelo de *Principal Component Analysis* (PCA)

Previamente à utilização e execução da técnica analítica de PCA (Greenacre *et al.*, 2022), os autores avaliaram dois critérios estatísticos: 1) o Teste de Esfericidade de Bartlett – o qual resultou em: χ^2=59.04691 com df=10 e *p-value*=5.487361e-09 e, 2) e o teste Kaiser-Meyer-Olkin (KMO), uma medida estatística usada para avaliar a adequação da amostra para uso na PCA, a qual resultou no valor de um KMO global acima de 0,6.

Após a validação e o atendimento a ambos os testes estatísticos, a modelagem de PCA foi executada, identificando 5 PCs interpretáveis. Na interpretação de cada PC, tipicamente, a primeira componente obtida captura o fator comum dominante, e recomenda-se examinar a magnitude e a direção dos coeficientes das variáveis originais, verificando que, quanto maior o valor absoluto (módulo) do coeficiente, mais importante será a respectiva variável original no cálculo do PC. Cabe aos pesquisadores utilizar seus próprios conhecimentos, aliados aos estudos da literatura, para determinar quais níveis de correlação entre as variáveis devem ser aceitos nas análises.

Foi possível identificar que a primeira componente (PC1) explicou 40,80% da variância dos dados, capturando a rentabilidade geral (ROA, ROE) e a liquidez corrente (CR).

A segunda componente (PC2) foi majoritariamente composta pela receita líquida (NR_bi) (carga = 0.88), podendo ser interpretada como a representação do porte ou tamanho da organização e trouxe a explicação de 21,16% da variância adicional – conjuntamente, a primeira e segunda componente, PC1 e PC2 explicaram 61,96% da variância total dos dados, atendendo ao critério mínimo de 60% recomendado por Hair *et al.* (2019).

A terceira componente (PC3) explicou 15,84% da variância, e as demais componentes (PC4 e PC5) contribuíram, respectivamente, com 14,97% e 7,22% da variância dos dados.

A Tabela 4 e a Figura 3 ilustram detalhadamente a constituição das 5 PCs (PC1, PC2, PC3, PC4 e PC5) obtidas após a execução do modelo de PCA.

Tabela 4

Variância explicada por cada uma das cinco PCs

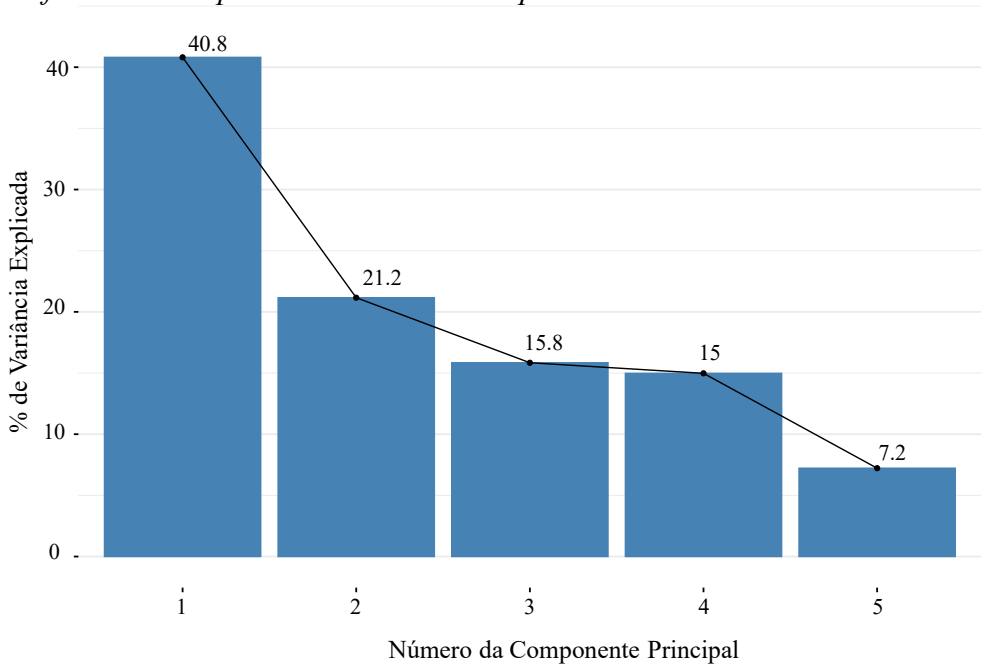
Componente	Desvio Padrão	Variância / Autovalor	Proporção Variância (%)	Variância Acumulada (%)
PC1	1,428360	2,040211	40,804220	40,80
PC2	1,028556	1,057928	21,158565	61,96
PC3	0,889970	0,792047	15,840949	77,80
PC4	0,865317	0,748774	14,975484	92,78
PC5	0,600865	0,361039	7,220782	100

Fonte: elaborada pelos autores após a execução do modelo de PCA.

A literatura relata que, geralmente, as três primeiras PCs são as mais relevantes para a interpretação e a análise, pois carregam a maior parte da variância acumulada – neste estudo, como mostrado na Tabela 4, as três primeiras componentes, somadas, explicaram 77,80% da variância dos dados das organizações avaliadas.

Figura 3

Gráfico de Scree plot com a variância explicada das cinco PCs



Nota. no gráfico, os números destacados sobre as barras estão em percentual.
Fonte: elaborada pelos autores após a execução do modelo de PCA.

A Figura 3 traz um gráfico de *Scree plot* (Cattell, 1966) que mostra a variância explicada de cada PC (números 1, 2, 3, 4 e 5) em ordem decrescente. No *eixo x*, está a representação das 5 PCs, e no *eixo y*, apresenta-se a proporção da variância total dos dados capturada por cada uma delas.

A Tabela 5 apresenta os *loadings* (ou cargas fatoriais) de cada variável original: coeficientes que multiplicam as próprias variáveis originais para formar as PCs e que representam parte da correlação entre cada variável original e cada componente principal.

Tabela 5

Cargas fatoriais (ou loadings) das variáveis originais em cada uma das PCs

Variáveis Originais	Componentes Principais				
	PC1	**PC2**	**PC3**	**PC4**	**PC5**
ROA	**0,615865**	0,018261	0,050339	0,043773	**0,784810**
ROE	**0,473717**	0,135779	**0,760370**	-0,016207	**-0,422767**
CR	**0,454229**	-0,295993	**-0,415828**	**0,636176**	**-0,358372**
NI_bi	**0,423584**	**0,351813**	**-0,492554**	**-0,615429**	-0,274668
NR_bi	-0,102735	**0,877409**	-0,061496	**0,462978**	0,038326

Fonte: elaborada pelos autores após a execução do modelo de PCA.

Com base na Tabela 5, quando uma PC tem em sua composição uma variável original com carga fatorial (*loading*) positiva, isso indica uma relação direta entre a variável original e a respectiva PC: à medida que a variável aumenta, a PC também tende a aumentar. Ou ainda, quando a carga fatorial é negativa nesta variável original, isso indica uma relação inversa: à medida que o valor desta variável específica aumenta, o da PC tende a diminuir. Por exemplo, na PC1, o valor do *loading* de -0,102735 da variável original NR_bi é um coeficiente que indica a contribuição ou peso da NR_bi para essa PC, e seu sinal negativo ("-") indica que o aumento desta variável original está associado à diminuição do valor da PC1, e vice-versa. É muito comum a ocorrência de *loadings* negativos, que não devem ser interpretados como algo de significado "ruim" ou problemático, mas apenas como a indicação da direção inversa da relação (correlação) entre uma variável original e a respectiva PC, ou mesmo entre as variáveis originais que compõem a mesma PC.

Complementando, quando variáveis originais, juntas na mesma PC, apresentam sinais opostos, contribuem de forma inversa para a composição desta PC. Por exemplo, como o valor da PC2 é gerado por uma combinação linear de suas variáveis originais (PC2 = 0,877409×NR_bi + 0,351813×NR_bi + 0,135779×ROE + 0,018261×ROA + -0,295993×CR), neste contexto da PC2, a Receita Líquida (NR_bi) e a Liquidez Corrente (CR) possuem sinais contrários na composição da PC2 – valor positivo de 0,877409 e valor negativo de -0,295993, respectivamente − e, assim, se movem em direções opostas ao longo desse eixo da PC.

Assim, a formação da primeira componente, a PC1, composta pelas variáveis originais com valores arredondados, NI_bi (+0,42), CR (+0,45) e ROE (+0,47), com cargas positivas moderadas, ROA (+0,62), com carga positiva um pouco mais alta, e NR_bi (-0,1), com carga negativa muito baixa, foi interpretada como a componente de Índice de Eficiência e Solvência (Qualidade do Desempenho) ou, resumidamente, em um único termo, a componente de "*Eficiência*" organizacional.

A PC1 é influenciada pela eficiência no uso dos ativos (ROA) e moderadamente influenciada pelo retorno aos acionistas (ROE), pelo resultado líquido (NI_bi) e pela liquidez corrente (CR), que indica a capacidade de honrar obrigações no curto prazo. Esta componente agrupa variáveis ligadas à rentabilidade relativa aos ativos e ao capital, à capacidade de gerar lucro

e à liquidez corrente, ou seja, a eficiência operacional relativa (eficiência na geração de retornos) e a qualidade do lucro agregam mais para PC1.

A baixa carga (quase nula) na receita líquida (NR_bi) na PC1 indica que o volume de receita não é o fator determinante nesta PC, que avalia a qualidade do desempenho financeiro – ou seja, quão lucrativa e líquida a empresa é, independentemente do seu tamanho (volume de receita). Empresas com altos escores na PC1 são vistas como eficientes e mantêm flexibilidade financeira.

A componente PC2, a segunda PC, que é ortogonal (não correlacionada) à PC1, e composta pelas variáveis com valores arredondados, NR_bi (+0,88) com carga positiva muito forte, NI_bi (+0,35) e ROE (+0,14) com cargas positivas fracas, ROA (+0,02) com carga positiva próximo de zero, e CR (-0,3) com carga negativa também fraca, foi interpretada como sendo a componente de Índice de Escala e Tamanho (Volume Operacional) ou resumidamente, em único termo, a componente de "*Escala"* organizacional.

Nesta componente PC2, a carga que prevalece é a da variável receita líquida (NR_bi), que se correlaciona positivamente com o lucro líquido (NI_bi), ou seja, receita e lucro absolutos variam juntos (ambos positivos) – capacidade de geração de receita em termos absolutos –, mas variam independentemente da eficiência relativa (ROA≈0) e inversamente com a liquidez corrente (CR negativa).

Para a componente PC3 (terceira componente principal), que é mais influenciada por ROE (+0,76), de forte carga positiva, pode-se atribuir o nome de Índice de Alavancagem Financeira (Risco/Retorno) ou, resumidamente, em um termo composto, a componente de "Risco-Retorno" organizacional, ao identificar ROA (0,05) e NR_bi (-0,06) como variáveis com correlação praticamente nula, o lucro líquido (NI_bi = -0,49) e a liquidez corrente (CR = -0,42) com *loadings* moderados de sinal negativo, indicando nestes dois últimos valores que ambas as variáveis se movem em direção oposta ao aumento desta PC.

Escores altos em PC3 (positivos) diferenciam as empresas principalmente pelo valor do ROE (estrutura de capital), e não pelo desempenho operacional (ROA essencialmente nulo); isolam o componente de alavancagem financeira da rentabilidade, indicando uma estrutura de capital alavancada.

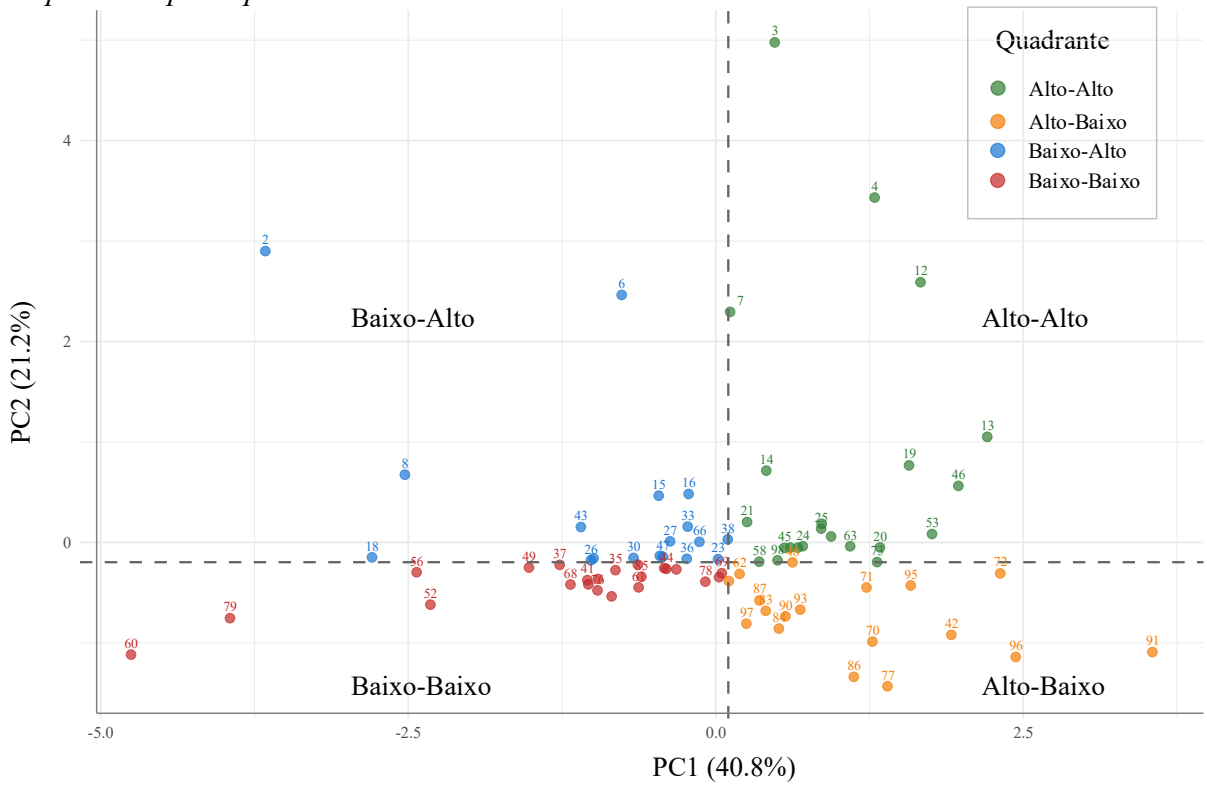
A Figura 4 mostra o Mapa de Posicionamento (diagrama de dispersão ou *scatterplot*) das organizações no espaço bidimensional, com base na mensuração específica de duas de suas PCs, PC1 e PC2, que, conjuntamente, explicam 62,0% da variância total dos dados (40,8% e 21,2%, respectivamente), atendendo ao critério mínimo de 60% recomendado por Hair *et al.* (2019).

A análise de quadrantes é uma técnica utilizada em contabilidade, finanças e gestão organizacional para reunir visualmente as características de empresas com base em dois indicadores-chave. Neste estudo, os quadrantes (alto-alto, alto-baixo, baixo-alto, baixo-baixo) são formados pelas medianas dessas mensurações (mediana de PC1 = 0,1020348 e mediana de PC2 = -0,1972229), representadas por linhas pontilhadas e usadas como eixos de divisão (PC1 no *eixo-x* versus PC2 no *eixo-y*), e correspondem a quatro perfis distintos de desempenho financeiro das 80 organizações da amostra do estudo.

No eixo-x (horizontal), a componente principal PC1 pode representar a qualidade dos retornos e a solidez financeira, com valores positivos caracterizando alta eficiência operacional e liquidez ("*Eficiência*"). No eixo-y (vertical), a componente principal PC2 pode representar o porte absoluto e a capacidade de geração de receita, com valores positivos indicando maior porte empresarial ("*Escala"*).

Figura 4

Mapa de Posicionamento das organizações em quadrantes conforme a mensuração de suas componentes principais PC1 e PC2



Nota. o número informado no rótulo de cada ponto corresponde à posição da organização no *ranking* original da Forbes Agro 100; as linhas pontilhadas representam as medianas das PCs, sendo a mediana da PC1 = 0,1020348 e a mediana da PC2 = -0,1972229.

Fonte: elaborada pelos autores após a execução do modelo de PCA.

Tabela 6

Características e perfil das organizações conforme posição ocupada nos quadrantes

Posição no Quadrante	Número de Empresas	Mensurações PC1 e PC2	Perfil atribuído às Organizações	Rótulos do Mapa (exemplos)	Características do Quadrante
Alto-Alto	22	PC1 > 0.102 e PC2 > -0.197	Líderes de desempenho e mercado	3, 4, 7, 12	Alta Eficiência e Alta Escala
Alto-Baixo	18	PC1 > 0.102 e PC2 < -0.197	Menores e rentáveis	77, 86, 91, 96	Alta Eficiência e Baixa Escala
Baixo-Alto	18	PC1 < 0.102 e PC2 > -0.197	Grandes e menos eficientes ou em crise momentânea	2, 6, 8, 15	Baixa Eficiência e Alta Escala
Baixo-Baixo	22	PC1 < 0.102 e PC2 < -0.197	Menores e menos eficientes	52, 56, 60, 79	Baixa Eficiência e Baixa Escala

Fonte: elaborada pelos autores após a execução do modelo de PCA.

A Tabela 6 também complementa a Figura 4, apresentando as características sintéticas da interpretação de cada quadrante e os perfis abreviados das organizações que ocupam as respectivas posições. Essa interpretação baseia-se nas cargas fatoriais (*loadings*) das variáveis originais que compõem as PCs (PC1 e PC2), as quais definem o significado econômico de cada eixo da figura.

A análise visual da Figura 4 revelou quatro perfis empresariais distintos: (i) Quadrante *Alto-Alto* (verde): posiciona empresas de grande porte rentáveis, caracterizadas por elevados indicadores de ROA, ROE e CR, combinados com alta NR_bi; (ii) Quadrante *Alto-Baixo* (laranja): posiciona empresas de menor porte com alta rentabilidade relativa, mas menor escala operacional que o quadrante anterior – podem representar organizações com oportunidades de crescimento com eficiência operacional já estabelecida; (iii) Quadrante *Baixo-Alto* (azul): posiciona empresas de grande porte com baixa rentabilidade, evidenciando possíveis desafios de eficiência operacional; e (iv) Quadrante *Baixo-Baixo* (vermelho): concentra empresas que são caracterizadas por baixa rentabilidade/performance e menor porte, sendo o quadrante que posiciona o maior número de empresas (14) que possuem lucro líquido NI_bi negativo.

Tabela 7

Lista das organizações da amostra final do estudo com os resultados dos modelos de ML

Rank	Nome da Organização	Cluster k	Silhouette	Cluster h	Quadrante
2	Marfrig Global Foods	C2	0,531288	C2	Baixo-Alto
3	Cargill	C1	0,597275	C1	Alto-Alto
4	Bunge Alimentos	C1	0,611179	C1	Alto-Alto
6	Raízen Energia	C1	0,500019	C1	Baixo-Alto
7	Copersucar	C1	0,493918	C1	Alto-Alto
8	BRF	C2	0,498727	C2	Baixo-Alto
12	Cosan	C1	0,583040	C1	Alto-Alto
13	Coamo Agroindustrial	C1	0,486515	C1	Alto-Alto
14	Minerva Foods	C4	0,587184	C4	Alto-Alto
15	C. Vale	C4	0,654965	C4	Baixo-Alto
16	Lar Cooperativa	C4	0,646484	C4	Baixo-Alto
18	Yara Brasil	C2	0,512805	C2	Baixo-Alto
19	Klabin	C1	0,498610	C1	Alto-Alto
20	Comigo	**C3**	0,567341	C4	Alto-Alto
21	Cocamar Cooperativa Agroindustrial	C4	0,646039	C4	Alto-Alto
23	Camil Alimentos	C4	0,617098	C4	Baixo-Alto
24	M. Dias Branco	**C3**	0,460072	C4	Alto-Alto
25	Três Tentos Agroindustrial	C4	0,573489	C4	Alto-Alto
26	Agrícola Alvorada	C4	0,673283	C4	Baixo-Alto
27	Timbro Trading	C4	0,675374	C4	Baixo-Alto
28	Bianchini	C4	0,556512	C4	Alto-Alto
30	Integrada Cooperativa	C4	0,679409	C4	Baixo-Alto
32	FS Bioenergia	C4	0,538078	C4	Alto-Alto
33	BP Bunge Bioenergia	C4	0,667061	C4	Baixo-Alto
35	Coopercitrus	C4	0,670562	C4	Baixo-Baixo
36	Caramuru Alimentos	C4	0,671576	C4	Baixo-Alto
37	Belagrícola	C4	0,664333	C4	Baixo-Baixo
38	Dexco	C4	0,647157	C4	Baixo-Alto
39	SLC Agrícola	C4	0,577705	C4	Alto-Alto
40	Be 8	C4	0,676662	C4	Baixo-Alto
41	Cibrafértil	C4	0,658270	C4	Baixo-Baixo

42	São Martinho	C3	0,607044	C3	Alto-Baixo
43	Aliança Agrícola do Cerrado	C4	0,657335	C4	Baixo-Alto
44	Cooxupé	C4	0,678605	C4	Baixo-Baixo
45	Sumitomo Chemical Brasil	C4	0,603736	C4	Alto-Alto
46	Eldorado Brasil Celulose	C1	0,462455	C1	Alto-Alto
47	Copasul	C4	0,680429	C4	Baixo-Alto
48	Oleoplan	C4	0,591254	C4	Alto-Baixo
49	Citrosuco	C4	0,635172	C4	Baixo-Baixo
52	Tereos	C2	0,459307	C2	Baixo-Baixo
53	Cia Lincoln Junqueira	**C3**	0,527408	**C1**	Alto-Alto
54	Agro Amazônia	C4	0,680128	C4	Baixo-Alto
56	Araguaia	C2	0,449725	C2	Baixo-Baixo
57	EISA – Empresa Interagrícola	C4	0,681771	C4	Baixo-Baixo
58	Marubeni Grãos Brasil	C4	0,627120	C4	Alto-Alto
60	Adama Brasil	C2	0,568579	C2	Baixo-Baixo
61	Alvoar Lácteos	C4	0,663464	C4	Baixo-Baixo
62	Grupo Cereal	C4	0,640556	C4	Alto-Baixo
63	Ihara	C4	0,534639	C4	Alto-Alto
64	Coplacana	C4	0,673491	C4	Baixo-Baixo
65	Agrofel	C4	0,679382	C4	Baixo-Baixo
66	Usina Coruripe	C4	0,666473	C4	Baixo-Alto
67	Adecoagro do Ivinhema	C4	0,635740	C4	Alto-Baixo
68	Gonçalves & Tortola – GTFoods	C4	0,663482	C4	Baixo-Baixo
69	Grupo Zilor – Açucareira Quatá	C4	0,654304	C4	Baixo-Baixo
70	Fugo Couros	C3	0,618109	C3	Alto-Baixo
71	Jacto	**C3**	0,549145	C4	Alto-Baixo
72	J. Macedo	**C3**	0,562016	C4	Alto-Baixo
73	Bem Brasil Alimentos	**C3**	0,479976	C4	Alto-Alto
75	Frigol	C4	0,675731	C4	Baixo-Baixo
76	Vigor Alimentos	C4	0,664209	C4	Baixo-Baixo
77	Pedra Agroindustrial	C3	0,598865	C3	Alto-Baixo
78	Berneck – Painéis e Serrados	C4	0,648997	C4	Baixo-Baixo
79	Vibra Agroindustrial	C2	0,572208	C2	Baixo-Baixo
83	Colombo Agroindústria	**C3**	0,490017	C4	Alto-Baixo
84	Arauco do Brasil	**C3**	0,553020	C4	Alto-Baixo
85	CMAA – Cia. Mineira de Açúcar e Álcool	C4	0,679348	C4	Baixo-Baixo
86	Plena Alimentos	C3	0,606592	C3	Alto-Baixo
87	Cerradinho Bioenergia	C4	0,554057	C4	Alto-Baixo
88	Santher	C4	0,601910	C4	Alto-Alto
90	Cocal – Canaã Açúcar e Álcool	**C3**	0,518812	C4	Alto-Baixo
91	Mili	C3	0,604391	C3	Alto-Baixo
92	Coopertradição	C4	0,653179	C4	Baixo-Baixo
93	Nutriza – Friato Alimentos	**C3**	0,519902	C4	Alto-Baixo
94	Frisa – Frigorífico Rio Doce	C4	0,670756	C4	Baixo-Baixo

95	GEES	**C3**	0,549034	C4	Alto-Baixo
96	Boa Safra Sementes	C3	0,625329	C3	Alto-Baixo
97	Nortox	**C3**	0,494371	C4	Alto-Baixo
98	Granja Faria	C4	0,620105	C4	Alto-Alto
100	Pamplona Alimentos	C4	0,661480	C4	Baixo-Baixo

Nota. as informações do 'Rank' e 'Nome da Organização' foram obtidas do *Ranking Forbes Agro 100* do ano de 2024 (Rodrigues *et al.*, 2025) e as demais informações da tabela ('Cluster k', 'Silhouette', 'Cluster_h' e 'Quadrante') foram geradas pelos autores a partir dos resultados dos modelos de ML estudados.
Fonte: elaborada pelos autores.

A Tabela 7 mostra a lista das organizações que fizeram parte da amostra final deste estudo, sendo informadas a posição original do *ranking* (Rank), o nome abreviado da organização, as informações obtidas das modelagens – identificação no respectivo *clustering K-means* (Cluster k), o escore de coesão e separação da observação no *cluster* (Silhouette), a identificação no respectivo *clustering* Hierárquico (Cluster h), e a posição da organização (Quadrante) no gráfico de *Scatteplot* formado pelas componentes principais PC1 e PC2, conforme disponibilizado na Figura 4.

Tendo como pano de fundo a Teoria da Sinalização e a utilização das DFs como emissores de "sinais" sobre os indicadores de desempenho, qualidade financeira e perspectivas dos negócios das empresas do agronegócio, pesquisadores atualmente podem extrapolar o papel das DFs como integradores de métricas contábeis e financeiras observáveis, empregando técnicas estatísticas e modelagens multivariadas, com tais sinais sendo decodificados por meio de técnicas de ML, para explorar as variadas formas de comunicação dessas empresas com o mercado e auxiliar na tomada de decisões dos stakeholders.

5 LIMITAÇÕES DA PESQUISA E ESTUDOS FUTUROS

No desenvolvimento deste trabalho, foram identificadas algumas limitações metodológicas e operacionais que se refletiram na sua execução. Em primeiro lugar, houve dificuldade de acesso às DFs de determinadas empresas participantes do *ranking* utilizado no estudo – apesar de serem grandes players do setor, várias dessas empresas não as disponibilizaram em plataformas públicas ou mesmo em seus próprios sites institucionais, o que reduziu a amostra e restringiu a representatividade dos resultados.

Em segundo lugar, a não inclusão de métricas financeiras avançadas, como por exemplo o *Economic Value Added* (EVA) e o *Market Value Added* (MVA), também pode ser considerada uma limitação desta pesquisa, pois tais métricas demandariam informações complementares, como a consulta de relatórios de mercado, a identificação do *Weighted Average Cost of Capital* (WACC), além de valores de mercado das empresas, os quais extrapolariam o delimitado pelas DFs, que foi o foco principal da origem deste estudo.

Em terceiro lugar, para evitar inconsistências nos cálculos, a integração e a estruturação do conjunto de dados da pesquisa (*dataset*) também foram desafiadoras para os pesquisadores, exigindo a padronização dos dados financeiros e recorrentes processos de conversão, conferência e ajuste, o que demandou tempo adicional de dedicação na execução dessas rotinas.

De forma complementar, a análise multivariada com o uso de técnicas de ML para a análise de métricas financeiras enfrenta limitações inerentes à qualidade e à disponibilidade dos dados – os pesquisadores reiteram que foram empenhados amplos esforços para a execução de todas as etapas desta pesquisa, com foco no rigor científico e relevância aplicada, e que os resultados obtidos neste estudo não podem ser generalizados para outros contextos, setores, países, períodos ou mesmo outros *rankings* de empresas, sendo restritos apenas para a amostra de dados que foi utilizada na pesquisa.

Estudos futuros poderão sugerir a operacionalização de um novo *ranking*, proporcionando uma compreensão mais completa das 80 organizações que participaram deste estudo, por meio da combinação das cinco métricas financeiras analisadas (receita líquida, lucro líquido, ROA, ROE e índice de liquidez corrente), indo além do *ranking* original, que utilizou apenas uma métrica (receita líquida).

Por fim, os pesquisadores reiteram a relevância de avançar em pesquisas acadêmico-científicas, estruturadas por meio da colaboração entre acadêmicos e profissionais/práticos, que contribuam para a criação de métricas mais precisas e realistas no desenvolvimento de *rankings*, capazes de auxiliar empresas, associações e veículos de mídia do setor empresarial na promoção de *rankings* que reflitam melhor as necessidades de seus setores-alvo, como o próprio setor do agronegócios, fundamental tanto para a economia do Brasil quanto para a economia global.

6 CONSIDERAÇÕES FINAIS

Este estudo analisou as DFs (balanço patrimonial e demonstração do resultado do exercício) de empresas do setor de agronegócio brasileiro, pertencentes ao ranking Forbes Agro100 2024, utilizando mensurações financeiras (valores absolutos e *ratios*) e aplicando técnicas analíticas e modelagens de ML (*Clustering* e PCA).

As 80 empresas analisadas constituíram quatro clusters, nomeados: "Lucrativas" (*cluster* C1, com 8 empresas), "Encolhidas" (*cluster* C2, com 7 empresas), "Saudáveis" (*cluster* C3, com 18 empresas) e "Diversificadas" (*cluster* C4, com 47 empresas).

Achados relevantes sobre a estrutura do mercado do agronegócio brasileiro podem oferecer valor acadêmico e gerencial em pelo menos três aspectos: 1) a decomposição da variância dos dados em dimensões ortogonais principais de ”Eficiência” (PC1), “Escala” (PC2) e “Risco-Retorno” (PC3), trouxe explicações com potencial utilização na gestão organizacional a partir da análise das métricas financeiras estudadas; 2) a rentabilidade e porte organizacional são dimensões relativamente independentes (PC1 é ortogonal a PC2), indicando, para essa amostra de empresas atuantes no mercado brasileiro, que escala/porte não garante eficiência operacional; 3) ao gerar o mapa de posicionamento - diagrama de dispersão (*scatterplot*) formado por PC1 (eixo-x) e PC2 (eixo-y) – com a visualização dos quatro quadrantes, foi possível revelar características de desempenho financeiros das empresas agrupadas, abrindo novos caminhos para pesquisas futuras, especialmente para a estruturação de novas métricas de formação de *ranking* de empresas.

A Teoria de Sinalização evidenciou a importância de as empresas transmitirem suas informações às partes interessadas e ao mercado. Ao fundamentar o papel informacional das demonstrações financeiras no contexto do agronegócio, os indicadores financeiros extraídos das DFs não são apenas métricas contábeis, mas também veículos de informação assimétrica (que emitem sinais e os enviam ao mercado) sobre os compromissos, comportamentos e intenções das empresas, bem como sobre o desempenho financeiro dessas empresas.

Ao utilizar análise multivariada com técnicas de ML, este estudo capturou padrões nesses sinais e permitiu distinguir e justificar a constituição de diferentes grupos de empresas conforme seus perfis sinalizadores.

A partir dos resultados apresentados por esta pesquisa, como contribuição para a área de conhecimento de finanças e ciências contábeis, foi possível compreender que a estruturação de um *ranking* de empresas, para ser mais completo, não deveria ficar limitado apenas aos valores de suas receitas líquidas – tanto as composições dos *clusters* (C1, C2, C3 e C4) assim como a formação das três primeiras componentes principais (PC1, PC2 e PC3) indicaram que há empresas que estão pior posicionadas no *ranking*, mas com qualidade superior em seus valores absolutos/indicadores financeiros quando comparadas com outras em posição privilegiadas neste mesmo *ranking*.

REFERÊNCIAS

Abdi, H., & Williams, L. J. (2010). Principal Component Analysis. *WIREs Computational Statistics*, *2*(4), 433–459. https://doi.org/https://doi.org/10.1002/wics.101

Al-Sartawi, A. M., & Reyad, S. (2018). Signaling theory and the determinants of online financial disclosure. *Journal of Economic and Administrative Sciences*, *34*(3), 237–247. https://doi.org/10.1108/JEAS-10-2017-0103

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journal of International Financial Management and Accounting*, *28*(2), 131–171. https://doi.org/10.1111/jifm.12053

Altuntas, S., Selim, S., & Altuntas, F. (2022). A hierarchical clustering based panel data approach: A case study of regional incentives. *International Journal of Information Management Data Insights*, *2*(2). https://doi.org/10.1016/j.jjimei.2022.100098

Arhinful, R., Mensah, L., Amin, H. I. M., Obeng, H. A., & Gyamfi, B. A. (2025). The Strategic Role of Sustainable Finance in Corporate Reputation: A Signaling Theory Perspective. *Sustainability*, *17*(11), 5002. https://doi.org/10.3390/su17115002

Assis, A. G. de, & Decoster, S. R. A. (2025). Gerenciamento de risco de crédito por meio da utilização de aprendizado de máquina: o caso do Banco BS2. *Revista Catarinense da Ciência Contábil*, *24*, 1–31, e3526. https://doi.org/10.16930/2237-7662202535261

Bagirov, A. M., Aliguliyev, R. M., & Sultanova, N. (2023). Finding compact and well-separated clusters: clustering using silhouette coefficients. *Pattern Recognition*, *135*, 109144. https://doi.org/10.1016/j.patcog.2022.109144

Belmonte, T. G., & Santos, F. de A. (2015). The use of accounting statements by small and medium-sized enterprises. *ENIAC Pesquisa*, *4*(2), 143–161. https://doi.org/10.22567/rep.v4i2.268

Bojer, C. S. (2022). Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities. *International Journal of Forecasting*, *38*(4), 1555–1561. https://doi.org/10.1016/j.ijforecast.2021.11.003

Bornhofen, E., Ramires, T. G., Bergonci, T., Nakamura, L. R., & Righetto, A. J. (2019). Associations between global indices of risk management and agricultural development. *Agricultural Systems*, *173*, 281–288. https://doi.org/10.1016/j.agsy.2019.03.006

Cabana, E., Lillo, R. E., & Laniado, H. (2021). Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Statistical Papers*, *62*(4), 1583–1609. https://doi.org/10.1007/s00362-019-01148-1

Cardoso, V. M., Gilio, L., Jank, M. S., & Pioli, L. (2025, February 12). *Brasil torna-se o maior país exportador de commodities do agro no mundo*. Insper Agro Global – Global Agribusiness Center. Insper Agro Global - Working Paper n.2/2025. https://agro.insper.edu.br/storage/papers/February2025/EUA%20x%20Brasil.pdf

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Cepea. (2025a). *Índices - Exportação do agronegócio | Janeiro a Dezembro de 2024*. CEPEA - Centro de Estudos Avançados em Economia Aplicada - Esalq/USP. https://www.cepea.org.br/br/indicador/serie-indices-de-exportacao-do-agronegocio.aspx

Cepea. (2025b). *PIB do agronegócio brasileiro*. CEPEA - Centro de Estudos Avançados em Economia Aplicada - Esalq/USP. https://www.cepea.org.br/br/pib-do-agronegocio-brasileiro.aspx

Cho, S., Vasarhelyi, M. A., Sun, T., & Zhang, C. (2020). Learning from machine learning in accounting and assurance. *Journal of Emerging Technologies in Accounting*, *17*(1), 1–10. https://doi.org/10.2308/jeta-10718

CNA. (2025). *Panorama do Agro*. Confederação da Agricultura e Pecuária do Brasil (CNA). https://www.cnabrasil.org.br/cna/panorama-do-agro

Codesso, M. de M. (2025). Editorial - A Inteligência Artificial transforma os Sistemas de Informação Contábil em todo o mundo. *Revista Catarinense da Ciência Contábil*, *24*, 1–4, e3644. https://doi.org/10.16930/2237-7662202536442

Comitê de Pronunciamentos Contábeis. (2009). *CPC 27 - Ativo Imobilizado*. CPC - Comitê de Pronunciamentos Contábeis. https://www.cpc.org.br/CPC/Documentos-Emitidos/Pronunciamentos/Pronunciamento?Id=58

Comitê de Pronunciamentos Contábeis. (2010). *CPC 04 - Ativo Intangível*. CPC - Comitê de Pronunciamentos Contábeis. https://www.cpc.org.br/CPC/Documentos-Emitidos/Pronunciamentos/Pronunciamento?Id=35

Comitê de Pronunciamentos Contábeis. (2011). *CPC 26 - Apresentação das Demonstrações Financeiras*. CPC - Comitê de Pronunciamentos Contábeis. https://www.cpc.org.br/CPC/Documentos-Emitidos/Pronunciamentos/Pronunciamento?Id=57

Comitê de Pronunciamentos Contábeis. (2012). *CPC 30 - Receitas*. CPC - Comitê de Pronunciamentos Contábeis. https://www.cpc.org.br/CPC/Documentos-Emitidos/Pronunciamentos/Pronunciamento?Id=61

Comitê de Pronunciamentos Contábeis. (2019). *CPC 00 - Estrutura Conceitual para Relatório Financeiro*. CPC - Comitê de Pronunciamentos Contábeis. https://www.cpc.org.br/CPC/Documentos-Emitidos/Pronunciamentos/Pronunciamento?Id=80

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, *37*(1), 39–67. https://doi.org/10.1177/0149206310388419

Connelly, B. L., Certo, S. T., Reutzel, C. R., DesJardine, M. R., & Zhou, Y. S. (2025). Signaling Theory: State of the Theory and Its Future. *Journal of Management*, *51*(1), 24–61. https://doi.org/10.1177/01492063241268459

Correa, C. A., Basso, L. F. C., & Nakamura, W. T. (2013). A estrutura de capital das maiores empresas brasileiras: análise empírica das teorias de Pecking Order e trade-off, usando panel data. *Revista de Administração Mackenzie*, *14*(4), 106–133. https://doi.org/10.1590/S1678-69712013000400005

Correa-Mejía, D. A., & Lopera-Castaño, M. (2020). Financial ratios as a powerful instrument to predict insolvency; a study using boosting algorithms in Colombian firms. *Estudios Gerenciales*, *36*(155), 229–238. https://doi.org/10.18046/j.estger.2020.155.3588

CVM. (2025). *Central de Sistemas - CVMWeb*. Comissão de Valores Mobiliários (CVM). https://cvmweb.cvm.gov.br/swb/

Dang, C., Chen, X., Yu, S., Chen, R., & Yang, Y. (2022). Credit ratings of Chinese households using factor scores and K-means clustering method. *International Review of Economics and Finance*, *78*, 309–320. https://doi.org/10.1016/j.iref.2021.12.014

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., & Zupan, B. (2013). Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, *14*, 2349–2353. http://jmlr.org/papers/v14/demsar13a.html

Dinh, T., Wong, H., Lisik, D., Koren, M., Tran, D., Yu, P. S., & Torres-Sospedra, J. (2025). Data clustering: a fundamental method in data science and management. *Data Science and Management*. https://doi.org/10.1016/j.dsm.2025.08.001

Dumitrescu, D., Bobitan, N., Popa, A. F., Sahlian, D. N., & Stanila, C. A. (2025). Signaling Financial Distress Through Z-Scores and Corporate Governance Compliance Interplay: A Random Forest Approach. *Electronics*, *14*(11), 2151. https://doi.org/10.3390/electronics14112151

Dzuba, S., & Krylov, D. (2021). Cluster analysis of financial strategies of companies. *Mathematics*, *9*(24), 3192. https://doi.org/10.3390/math9243192

Eldomiaty, T., Azzam, I., Afifi, K. T. H., & Rashwan, M. H. (2024). An Alignment of Financial Signaling and Stock Return Synchronicity. *Journal of Risk and Financial Management*, *17*(4), 162. https://doi.org/10.3390/jrfm17040162

Estados Unidos. (2002). *Sarbanes-Oxley Act of 2002. (Lei Pública 107-204).* https://www.dol.gov/agencies/oalj/PUBLIC/WHISTLEBLOWER/REFERENCES/STATUTES/SARBANES_OXLEY_ACT_OF_2002

Favaretto, J. E. R. (2024, October 21). *Artificial Intelligence (AI) explained by Venn diagram - revision 1.0*. Zenodo. https://doi.org/10.5281/zenodo.17064700

Figari, A. K. P., Lima, F. G., Gatsios, R. C., & Magnani, V. M. (2021). Economic value drivers for brazilian agricultural cooperatives. *Base - Revista de Administração e Contabilidade da Unisinos*, *18*(1), 56–77. https://doi.org/10.4013/base.2021.181.03

Fodor, A., Jorgensen, R. D., & Stowe, J. D. (2021). Financial clusters, industry groups, and stock return correlations. *Journal of Financial Research*, *44*(1), 121–144. https://doi.org/10.1111/jfir.12236

Fossá, J. L., Villwock, A. P. S., & Matte, A. (2024). Analysis of the distribution of rural credits among the federal units in the period 2013 to 2022. *Desenvolvimento Em Questão*, *22*(60), e14701. https://doi.org/10.21527/2237-6453.2024.60.14701

Garcia-Alvarez, D., Bregon, A., Pulido, B., & Alonso-Gonzalez, C. J. (2023). Integrating PCA and structural model decomposition to improve fault monitoring and diagnosis with varying operation points. *Engineering Applications of Artificial Intelligence*, *122*(106145). https://doi.org/10.1016/j.engappai.2023.106145

Gonzales, R. M. D., & Hargreaves, C. A. (2022). How can we use artificial intelligence for stock recommendation and risk management? A proposed decision support system. *International Journal of Information Management Data Insights*, *2*(2), 100130. https://doi.org/10.1016/j.jjimei.2022.100130

Governo Federal do Brasil. (1976). *Lei nº 6.404, de 15 de dezembro de 1976. Dispõe sobre as Sociedades por Ações (Lei das S.A.)*. https://www.planalto.gov.br/ccivil_03/leis/l6404consol.htm

Governo Federal do Brasil. (2025a). *Comissão de Valores Mobiliários (CVM)*. Governo Federal do Brasil. https://www.gov.br/cvm/pt-br

Governo Federal do Brasil. (2025b, January 8). *Balança Comercial de 2024 - Marca histórica do agronegócio brasileiro destaca protagonismo na segurança alimentar global*. Ministério da Agricultura e Pecuária. https://www.gov.br/secom/pt-br/assuntos/noticias/2025/janeiro/marca-historica-do-agronegocio-brasileiro-destaca-protagonismo-na-seguranca-alimentar-global

Greenacre, M., Groenen, P. J. F., Hastie, T., D'enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal Component Analysis. *Nature Reviews Methods Primers*, *2*(100). https://doi.org/10.1038/s43586-022-00184-w

Guedes, D. S., Oliveira, E. R., Santos, G. C. dos, & Ferreira, R. A. (2021). Relação do desempenho econômico-financeiro das empresas do agronegócio da B3 e os ciclos econômicos. *Redeca*, *8*, 128–143. https://doi.org/10.23925/2446-9513.2021v8i1p128-143

Gurbaxani, A., Thakkar, J., Pathak, S., Mathur, A., & Raees, S. (2023). Examining factors influencing investment in Digital Gold and Gold ETF using the PCA technique. *Investment Management and Financial Innovations*, *20*(2), 160–170. https://doi.org/10.21511/imfi.20(2).2023.14

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (Eighth edition). Cengage. https://worldcat.org/title/1082318713

Hassan, A., Elamer, A. A., Fletcher, M., & Sobhan, N. (2020). Voluntary assurance of sustainability reporting: evidence from an emerging economy. *Accounting Research Journal*, *33*(2), 391–410. https://doi.org/10.1108/ARJ-10-2018-0169

Hu, Y. (2025). Risk management of accounting systems: Using artificial intelligence for real-time analysis and decision-making. *Journal of Computational Methods in Sciences and Engineering*, *25*(5), 4802–4816. https://doi.org/10.1177/14727978251338006

International Financial Reporting Standards. (2001a). *IAS 1 - Presentation of Financial Statements*. The IFRS Foundation. https://www.ifrs.org/issued-standards/list-of-standards/ias-1-presentation-of-financial-statements/

International Financial Reporting Standards. (2001b). *IAS 16 - Property, Plant and Equipment*. The IFRS Foundation. https://www.ifrs.org/issued-standards/list-of-standards/ias-16-property-plant-and-equipment/

International Financial Reporting Standards. (2001c). *IAS 18 - Revenue*. The IFRS Foundation. https://www.ifrs.org/issued-standards/list-of-standards/ias-18-revenue/

International Financial Reporting Standards. (2001d). *IAS 38 - Intangible Assets*. The IFRS Foundation. https://www.ifrs.org/issued-standards/list-of-standards/ias-38-intangible-assets/

International Financial Reporting Standards. (2010). *IFRS - Conceptual Framework for Financial Reporting*. The IFRS Foundation. https://www.ifrs.org/issued-standards/list-of-standards/conceptual-framework/

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Kar, A. K., & Varsha, P. S. (2023). Unravelling the techno-functional building blocks of metaverse ecosystems – A review and research agenda. *International Journal of Information Management Data Insights*, *3*(2), 100176. https://doi.org/10.1016/j.jjimei.2023.100176

López-Santamaría, M., Amaya, N., Hinestroza, M. P. G., & Cuero, Y. A. (2021). Sustainability disclosure practices as seen through the lens of the signaling theory: A study of companies listed on the Colombian Stock Exchange. *Journal of Cleaner Production*, *317*, 128416. https://doi.org/10.1016/j.jclepro.2021.128416

Lu, J. F., Tang, J. B., Tang, Z. M., & Yang, J. Y. (2008). Hierarchical initialization approach for K-Means clustering. *Pattern Recognition Letters*, *29*(6), 787–795. https://doi.org/10.1016/j.patrec.2007.12.009

Machado, L. K. C., Prado, J. W. do, Vieira, K. C., Antonialli, L. M., & Santos, A. C. dos. (2016). A relevância da estrutura de capital no desempenho das firmas: uma análise multivariada das empresas brasileiras de capital aberto. *Revista de Educação e Pesquisa em Contabilidade (REPeC)*, *9*(4). https://doi.org/10.17524/repec.v9i4.1313

Mahalanobis, P. C. (1930). On tests and measures of group divergence. *Journal and Proceedings, Asiatic Society of Bengal*, *XXVI*(4), 541–588. http://dspace.isical.ac.in:8080/jspui/handle/10263/1639

Marukatat, S. (2023). Tutorial on PCA and approximate PCA and approximate kernel PCA. *Artificial Intelligence Review*, *56*, 5445–5477. https://doi.org/10.1007/s10462-022-10297-z

Mbona, R. M., & Yusheng, K. (2019). Financial statement analysis: Principal component analysis (PCA) approach case study on China telecoms industry. *Asian Journal of Accounting Research*, *4*(2), 233–245. https://doi.org/10.1108/AJAR-05-2019-0037

Melo, T. da S., Ribeiro, A. M., Rover, S., & Pires, P. A. da S. (2024). Tempestividade das demonstrações contábeis sob a ótica da folga financeira em companhias listadas na B3. *Revista Catarinense da Ciência Contábil*, *23*, 1–19, e3501. https://doi.org/10.16930/2237-7662202435011

Moreira, V. R., Kureski, R., & Veiga, C. P. Da. (2016). Assessment of the Economic Structure of Brazilian Agribusiness. *Scientific World Journal*, *2016*(7517806), 1–10. https://doi.org/10.1155/2016/7517806

Morelli, L., Araujo, F. C., & Favaretto, J. E. R. (2023). Transformação inteligente do Financial Planning and Analysis (FP&A) empresarial com a utilização de técnicas supervisionadas de Machine Learning. In *XLVII Encontro da ANPAD - EnANPAD 2023* (pp. 1–19).

Ongbali, S. O., Omotehinse, S. A., Adams, C. O., Salawu, E. Y., & Afolalu, S. A. (2024). Analysis of the key factors for small and medium-sized enterprises growth using principal component analysis. *Heliyon*, *10*, e33573. https://doi.org/10.1016/j.heliyon.2024.e33573

Özari, Ç., Can, E. N., & Demirkale, Ö. (2025). Financial Fraud Detection with Altman Z-Score and Beneish M-Score via Random Forest: Verified by Borsa Istanbul Fines (2018–2022). *SAGE Open*, *15*(4). https://doi.org/10.1177/21582440251386174

Perressim, W. S., Batalha, M. O., Carrer, M. J., & Franzotti, M. R. (2026). Technical efficiency in agribusiness cooperatives: insights into Brazilian organizations. *Agricultural Finance Review*, 1–18. https://doi.org/10.1108/AFR-05-2025-0078

Pooja, R., Kayal, P., & Maiti, M. (2024). Enhancing portfolio decision-making: a capital asset pricing model-based clustering analysis. *Journal of Economic Studies*, *51*(9), 358–379. https://doi.org/10.1108/JES-08-2024-0573

Posit Team. (2025). *RStudio: Integrated Development Environment (IDE) for R*. Posit Software, PBC, Boston, MA. https://posit.co/products/open-source/rstudio

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org

Ribeiro, L. M. da S., Brandalise, N., & Garcia, V. da S. (2024). Optimum selection of investments through economic-financial indicators and share prices. *Revista Foco*, *17*(5), e5117. https://doi.org/10.54751/revistafoco.v17n5-076

Rodrigues, C., Moitinho, F., & Ondei, V. (2025). *Agro100: Quais as Maiores Empresas do Brasil no Ranking Forbes*. https://forbes.com.br/escolhas-do-editor/2025/01/agro100-quais-as-maiores-empresas-do-brasil-no-ranking-forbes

Ross, S. A. (1977). The determination of financial structure: the incentive-signalling approach. *The Bell Journal of Economics*, *8*(1), 23–40. https://doi.org/10.2307/3003485

Rufino, M. A., & Monte, P. A. do. (2014). Factors that Explain the Voluntary Disclosure of 100 Firms with More Negotiated Shares at BM&FBOVESPA. *Sociedade, Contabilidade e Gestão*, *9*(3), 59–75. https://doi.org/10.21446/scg_ufrj.v9i3.13332

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Silva, T. S. da, & Diógenes, C. G. B. (2022). The managerial applicability of liquidity indicators in the dynamic analysis balance sheet. *Caderno Organização Sistêmica*, *2*(1), 156–170.

Silva, V. B. da, Campos, K. C., Campos, R. T., & Bezerra, F. N. R. (2025). Determinantes da insolvência de empresas do índice agronegócio B3 no Brasil. *Anais Do 63º Congresso Da Sociedade Brasileira de Economia, Administração e Sociologia Rural (SOBER)*. https://doi.org/10.29327/9786527214786.1101151

Silva, M. D. L. e, Albuquerque, F. H. F. de, Marcelino, M. M., & Quirós, J. T. (2015). Risk disclosure in the financial statements: an analysis of the notes of portuguese non-financial corporations. *Journal of Education and Research in Accounting (REPeC)*, *9*(3). https://doi.org/10.17524/repec.v9i3.1176

Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Souza, Â. R. L. de, Koeche, A. S., Allgaier, J. F., Santos, F. V. dos, & Oliveira, L. de. (2022). Desempenho econômico-financeiro de empresas do agronegócio da carne, listadas na B3, antes e após o início da Covid-19. *Revista de Gestão e Secretariado*, *13*(3), 1890–1908. https://doi.org/10.7769/gesec.v13i3.1450

Spence, A. M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, *87*, 355–374. https://doi.org/10.2307/1882010

Spence, A. M. (1974). *Market signaling: informational transfer in hiring and related screening processes*. Cambridge: Harvard University Press.

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1–34. https://doi.org/10.1348/000711005X48266

Valencia, L. R. (2025). Financial Performance and Corporate Governance on Firm Value: Evidence from Spain. *International Journal of Financial Studies*, *13*(3). https://doi.org/10.3390/ijfs13030123

Valor Econômico. (2025). *Valor RI | Central de Resultados - Pesquise pelo nome da empresa*. Editora Globo - Valor Econômico. https://valor.globo.com/valor-ri/central-de-resultados/

Vukčević, M., Lakićević, M., Melović, B., Backović, T., & Dudić, B. (2024). Modern models for predicting bankruptcy to detect early signals of business failure: Evidence from Montenegro. *PLoS ONE*, *19*(5), e0303793. https://doi.org/10.1371/journal.pone.0303793

Ybañez, A., Ancheta, R., Evangelista, S. S., Aro, J. L., Maturan, F., Atibing, N. M., Selerio, E., Yamagishi, K., & Ocampo, L. (2023). How can we use machine learning for characterizing organizational identification - a study using clustering with picture fuzzy datasets. *International Journal of Information Management Data Insights*, *3*(1). https://doi.org/10.1016/j.jjimei.2023.100157

Zhu, X., Chu, Q., Song, X., Hu, P., & Peng, L. (2023). Explainable prediction of loan default based on machine learning models. *Data Science and Management*, *6*(3), 123–133. https://doi.org/10.1016/j.dsm.2023.04.003

CONFLITO DE INTERESSES

Os autores afirmam não haver conflito de interesses.

DISPONIBILIDADE DOS DADOS

Todo o conjunto de dados que dá suporte aos resultados deste estudo pode ser disponibilizado a partir de solicitação aos autores.

CONTRIBUIÇÕES DOS AUTORES

Funções	1º autor	2º autor	3º autor
Conceituação	♦	♦	♦
Curadoria de dados	♦	♦	♦
Análise Formal			♦
Obtenção de financiamento			
Investigação	♦	♦	♦
Metodologia	♦	♦	♦
Administração do projeto			♦
Recursos			
Software			♦

Supervisão			♦
Validação	♦	♦	♦
Visualização	♦	♦	♦
Escrita – primeira redação	♦	♦	
Escrita – revisão e edição	♦	♦	♦

DIREITOS AUTORAIS

Os autores mantêm os direitos autorais relativos aos seus artigos e concedem à Revista Catarinense da Ciência Contábil (RCCC) o direito de primeira publicação, sendo a obra simultaneamente licenciada sob a Licença Creative Commons Atribuição 4.0 Internacional (CC BY 4.0).