Revista Catarinense da CIÊNCIA CONTÁBIL
ISSN 2237-7662, v. 25, 1-32, e3684, 2026
https://doi.org/10.16930/2237-7662202636842
Available at http://revista.crcsc.org.br

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), Graduate Program in Business Administration (PPGA), Center for Social and Applied Sciences (CCSA) - Mackenzie Business School, São Paulo, SP, Brazil
https://orcid.org/0009-0001-1756-2937
victorrotte@gmail.com

PEDRO PAULO BRAMBILLA PENTEADO
Universidade Presbiteriana Mackenzie (UPM), Bachelor's Degree in Accounting (Dual Study), Center for Social and Applied Sciences (CCSA) - Mackenzie Business School, São Paulo, SP, Brazil
https://orcid.org/0009-0004-2196-7998
pedrobrambilla12@gmail.com

JOSÉ EDUARDO RICCIARDI FAVARETTO[1]
Universidade Presbiteriana Mackenzie (UPM), Graduate Program in Controllership, Finance, and Management Technologies (PPGCFTG), Center for Social and Applied Sciences (CCSA) - Mackenzie Business School, São Paulo, SP, Brazil
https://orcid.org/0000-0002-0143-0809
jose.favaretto@mackenzie.br

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.

Edited in Portuguese and English. Original version in Portuguese.

[1] **Correspondence Address:** Rua da Consolação, 930 | Higienópolis | CEP 01302-000 | São Paulo, SP | Brazil.

Received on December 6, 2025. **Latest version received on** June 1, 2026. **Accepted on** July 30, 2026, by Rogério João Lunkes (Editor-in-Chief). **Published on** August 31, 2026.

(cc) BY

1 INTRODUCTION

The Financial Statements (FS) are communication instruments between companies and their stakeholders, such as investors, creditors, and managers. FS provide a snapshot of the economic and financial position of companies during the fiscal year, thereby supporting strategic decision-making (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a).

In Brazil and worldwide, publicly traded companies and corporations are required to publish their FS. In Brazil, these obligations are regulated by the Brazilian Corporations Law (Governo Federal do Brasil, 1976) and by the Brazilian Accounting Standards (NBC).

The academic literature reports several studies (Figari *et al.*, 2021; Ribeiro *et al.*, 2024; Silva & Diógenes, 2022) that use FS through traditional accounting and financial analysis methods, involving calculations, mathematical models, and/or comparisons of indicators.

One of the sectors in which the use of FS to support stakeholder decision-making can be observed is agribusiness, an economic sector encompassing all stages of the production chain, from crop cultivation and livestock farming to the industrialization and commercialization of products. In recent years, Brazil has stood out globally as one of the largest producers and exporters of agricultural commodities such as soybeans, corn, coffee, and beef (Cepea, 2025a). Brazilian agribusiness is responsible for food production and the export of these commodities, thereby contributing to food security and playing a significant role in the country's economy (Guedes *et al.*, 2021).

According to the Confederação da Agricultura e Pecuária do Brasil (CNA), an organization that brings together associations and political and rural leaders throughout the country, Brazil's agricultural production has developed over the past 50 years to such an extent that the country has become one of the world's leading producers and suppliers of food, fiber, and energy (CNA, 2025).

By disclosing their results and economic data through FS, companies enable their primary users, such as investors, creditors, or partners (stakeholders), to conduct transactions with or invest in them. Therefore, studying and understanding how the Brazilian agribusiness sector and its companies are structured from an accounting and financial perspective is important for market analysis and can also support stakeholder decision-making, given that many of these companies seek financing from BNDES and credit funds (Fossá *et al.*, 2024).

By adopting the conceptual framework of Signaling Theory (Connelly *et al.*, 2025; Ross, 1977; Spence, 1973), this study addresses the informational role of FS as one of the main communication instruments used by agribusiness companies (senders) when disclosing financial indicators, such as profitability, liquidity, and efficiency, among others (signals), with the aim of reducing information asymmetry with investors, creditors, and other interested parties (receivers).

Complementarily, the accounting profession is facing a new paradigm as Artificial Intelligence (AI) creates connections between data flows and learning algorithms, transforming the capacity to perform analyses and comply with regulatory standards (Codesso, 2025). In this digital revolution, statistical modeling techniques based on Machine Learning (ML) are available and can be applied to FS analysis (Bojer, 2022).

This study adopted a multivariate perspective in its development, focusing on financial statement analysis and using, as its context, the results of a study published in October 2024 by *Forbes Brasil*, a business and economics media outlet, known as the *Forbes Agro100 Brasil* ranking (Rodrigues *et al.*, 2025), which listed the 100 largest companies in the Brazilian agribusiness sector based on their 2023 net revenues.

A company's net revenue is an excellent determinant for quantifying the company's value and relevance, as it measures the amount of revenue generated from its operations; therefore, it can justify a ranking that positions the companies according to their relative influence on Brazilian agribusiness (Figari *et al.*, 2021).

However, this ranking reports only net revenue, which, according to the CPC and IFRS (Comitê de Pronunciamentos Contábeis, 2012; International Financial Reporting Standards, 2001c), is defined as the gross inflow of economic benefits during the period arising in the ordinary course of an entity's activities that increases equity, excluding increases related to contributions from owners, net of taxes withheld at source and other discounts.

Accordingly, this study sought to answer the following question: How can the FS of companies in the Brazilian agribusiness sector included in the *Forbes Agro100 2024* ranking be analyzed using ML techniques?

The general objective of this study was to analyze the financial results of the companies listed in this ranking based on data extracted from their FS (Balance Sheet and Income Statement) – financial values obtained from these documents and the calculation of other indicators (*financial ratios*) – using analytical techniques with ML models (Favaretto, 2024; Morelli *et al.*, 2023), including *Clustering* (Altuntas *et al.*, 2022; Dinh *et al.*, 2025) and *Principal Component Analysis* (PCA) (Marukatat, 2023; Mbona & Yusheng, 2019; Zhu *et al.*, 2023).

The expected contribution of this study, considering the informational role of FS and their indicators as means of communicating and transmitting signals to the market, is to identify patterns in these signals that reinforce the notion that characterizing a company ranking based on a single indicator (net revenue in this study) is limited and does not reflect the actual economic/financial condition of these companies.

2 THEORETICAL FRAMEWORK

2.1 Signaling Theory

Signaling Theory, introduced by Spence (1973, 1974), explains how market participants transmit information to reduce information asymmetry, with the signaling process essentially consisting of three components: senders, receivers, and the signal itself. The signal is information, whether positive or negative, private, that senders (companies) choose whether or not to communicate to receivers (the market and stakeholders) (Connelly *et al.*, 2025).

According to the theory (Connelly *et al.*, 2011; Ross, 1977; Spence, 1973), company managers (senders) possess information (signals) that the market and *stakeholders* (receivers) do not have, and, to communicate the quality of the company's results (its stability, governance quality, financial health, among others), senders (managers) emit signals to the receivers (investors, creditors, and regulators).

Researchers have applied Signaling Theory in various contexts to understand the influence of signals conveyed by financial indicators and organizational reports, as well as the subsequent reception of these signals by markets. For example, Al-Sartawi & Reyad (2018) used the theory to examine and report the extent and characteristics of institutions that determine online financial disclosure (OFD) practices, based on financial information from 48 Islamic banks listed on the stock exchanges of Gulf Cooperation Council (GCC) countries, from 2015 to 2017, using their own websites and stock exchange websites; López-Santamaría *et al.* (2021) studied sustainability disclosure practices in the reports of 43 companies listed on the Colombian stock exchange, and the most common signal observed was intention, related to future actions; Arhinful *et al.* (2025) studied how sustainable finance strengthened corporate reputation, using 17 years of data from 143 United Kingdom non-financial companies listed on the London Stock Exchange, and concluded that green bond issuance significantly increased ESG scores, suggesting that companies can signal their commitment to sustainability through visible actions.

Other studies have reinforced the broad application of Signaling Theory in the field, such as Valencia (2025), who analyzed how financial performance and corporate governance variables affected firm value in 91 small and medium-sized enterprises in Spain, with results indicating that

profitability, measured by Return on Equity (ROE), had a positive and statistically significant impact on firm value at the 10% level, thereby demonstrating that the market responds favorably to profitability and that investors use this information (signal) when making investment decisions; Eldomiaty *et al.* (2024), who examined the effects of financial signaling by companies on stock return synchronicity, which generally reflects how companies manage their operations to communicate signals to shareholders, empirically confirmed that financial indicators from the balance sheet and income statement acted as signals with specific informational content regarding stock return synchronicity; and Dumitrescu *et al.* (2025), who, by analyzing financial and governance data from 60 non-financial companies listed on the Bucharest Stock Exchange, found that the use of AI/ML techniques significantly improved the predictive accuracy and reliability of classifying and signaling financial distress among the companies evaluated.

2.2 Financial Statements (FS) and Financial Indicators

Economic globalization affected organizational competitiveness, leading companies to create value for their stakeholders. In this context, major market frauds occurred, such as the Enron case. To prevent such fraud, the Sarbanes-Oxley Act (SOX) was enacted (Estados Unidos, 2002), with the aim of protecting investors and increasing the reliability of financial information. As a result, companies began disclosing their accounting policies and preparing their FS in accordance with *International Accounting Standards* (IAS) (Silva *et al.*, 2015).

In Brazil, accounting standards aimed at harmonizing Brazilian accounting with international standards established by the *International Financial Reporting Standards* (IFRS) have had a strong influence, affecting even small companies (Belmonte & Santos, 2015).

Disclosure of FS is mandatory for companies listed on stock exchanges; however, voluntary disclosure may convey optimism and a positive image to *stakeholders* (Rufino & Monte, 2014).

FS are mandatorily composed of the Balance Sheet, Income Statement, Statement of Cash Flows, and Notes to the Financial Statements. They may also include the Statement of Comprehensive Income and the Statement of Changes in Equity, which provide a structured representation of a company's financial position, financial performance, and cash flows (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a).

Based on various financial figures and calculations derived from information disclosed in FS, it is possible to develop a comprehensive analysis of companies, which can support more informed decision-making or even the forecasting of a company's future developments (Ribeiro *et al.*, 2024).

Net income is the company's result after all revenues, costs, expenses, taxes, and participations have been accounted for (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a). It indicates whether the company generated a profit or incurred a loss during the period and serves as a basis for calculating dividends, evaluating financial performance, and supporting decisions by investors and creditors.

Net revenue is calculated as gross sales revenue less deductions such as returns, allowances, discounts granted, and sales taxes (ICMS, PIS, Cofins) (Comitê de Pronunciamentos Contábeis, 2012a; International Financial Reporting Standards, 2001b) and reflects the amount effectively received or receivable by the company.

Current assets comprise assets and rights expected to be realized by the end of the following fiscal year (Comitê de Pronunciamentos Contábeis, 2011, 2019; International Financial Reporting Standards, 2001a, 2010) and indicate the company's short-term liquidity, including items such as cash, accounts receivable, and inventories. Non-current assets refer to assets and rights expected to be realized over the long term, that is, after the end of the following fiscal year (Comitê de Pronunciamentos Contábeis, 2009, 2010; International Financial Reporting Standards, 2001d,

2001b), and represent long-term resources important to the company's ongoing operations, such as machinery, property, brands, and investments, among others. Total assets are the sum of current and non-current assets (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a) and indicate the company's economic size. They are used in analyses such as profitability and asset valuation indicators.

Current liabilities comprise obligations that the company must settle by the end of the following fiscal year (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a) and indicate short-term commitments, such as accounts payable and wages payable. Non-current liabilities refer to obligations due over the long term (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a) and represent long-term debts and commitments, such as debentures and provisions for contingencies. Total liabilities are the sum of current and non-current liabilities (Comitê de Pronunciamentos Contábeis, 2011; International Financial Reporting Standards, 2001a) and indicate everything the company owes to third parties. They can be used to assess the degree of indebtedness and conduct solvency and capital structure analyses.

Equity is the difference between total assets and total liabilities and represents the owners' or shareholders' own resources (Comitê de Pronunciamentos Contábeis, 2011, 2019; International Financial Reporting Standards, 2001a, 2010). It can be used as the basis for calculating return on equity (ROE) and indicates the company's long-term financial soundness. The literature reports a study by Melo *et al.* (2024) that analyzed the effect of financial slack on the timeliness of accounting reports in companies listed on B3 and, among several variables, used current liquidity and return on assets (ROA) indicators in the empirical analysis.

International studies using data extracted from FS and organizational reports of companies in Bangladesh (Hassan *et al.*, 2020), China (Mbona & Yusheng, 2019), Colombia (López-Santamaría *et al.*, 2021), Spain (Valencia, 2025), Romania (Dumitrescu *et al.*, 2025), and Türkiye (Özari *et al.*, 2025) exemplify and confirm that this data-exploration practice, based on financial indicators and/or Signaling Theory, is frequent and relevant in the scientific literature, particularly for discovering new knowledge related to organizational management.

As FS constitute one of the main formal means of communication (signaling) between managers and stakeholders, such as investors, creditors, and regulators, by representing a company's economic and financial position during an accounting period, the use of multivariate analysis with ML techniques can capture patterns in these signals that make it possible to distinguish groups of companies according to their signaling profiles.

2.3 Brazilian Agribusiness

Agribusiness is a fundamental sector in several countries, as it encompasses the cultivation of soybeans, wheat, corn, and other commodities. The sector contributes to job creation, generates trade surpluses (positive trade balances), and drives economic growth (Bornhofen *et al.*, 2019).

Agribusiness in Brazil is one of the country's main economic sectors (Governo Federal do Brasil, 2025b), with a significant share of Brazilian GDP. In 2024, it accounted for 23.2%, extending beyond the primary sector to activities "beyond the farm gate", including industries and services (Cepea, 2025b).

This sector is also prominent in exports, accounting for 49% of everything Brazil exported, equivalent to US$164.4 billion in 2024, demonstrating its resilience and sustained prominence over the years (Cepea, 2025a).

Brazilian agribusiness is a global benchmark in terms of production and trade, leading the markets for soybeans, cotton, beef, and orange juice due to factors such as abundant natural resources, crop diversification, extensive territory, and the increasing adoption of production

technologies, positioning Brazil as one of the largest players in the global market (Cardoso *et al.*, 2025).

Given the great importance of agribusiness in Brazil, reflected in its share of GDP, global market leadership, predominant contribution to the Brazilian trade balance, and impact on the Brazilian economy (Moreira *et al.*, 2016), the need to develop studies on this sector is evident and relevant.

2.4 Multivariate Perspectives using Machine Learning (ML) techniques

Although various approaches can be used to evaluate a group of companies based on a body of knowledge in the financial/accounting fields, including financial firm taxonomies (Altman *et al.*, 2017; Fodor *et al.*, 2021), company segmentation by performance and capital structure (Correa *et al.*, 2013; Machado *et al.*, 2016), and comparisons with industry studies (Perressim *et al.*, 2026; Silva *et al.*, 2025; Souza *et al.*, 2022), this study was guided by a multivariate analysis perspective of financial statements, with a focus on the application of ML techniques.

Multivariate analytical ML techniques, including *Clustering* and PCA, are part of a subset of the AI field (Favaretto, 2024), represented by computational methods that learn patterns from datasets. In accounting and auditing, ML techniques can be applied to various tasks, such as transaction analysis, risk assessment, fraud prediction, bankruptcy prediction, detection of material misstatements, and accounting estimates, among others (Cho *et al.*, 2020). For example, Assis & Decoster (2025) used an ML technique called Gradient Boosting Decision Tree (GBDT) to develop and implement a predictive credit risk management model, combining financial and non-financial variables to achieve greater decision-making accuracy and improve risk mitigation.

The research by Dzuba & Krylov (2021) demonstrated that Clustering methods can categorize companies based on their financial strategies, while Fodor *et al.* (2021) also used *Clustering* to identify groups of large U.S. companies based on their financial statements, including the percentage composition of balance sheets and income statements.

Pooja *et al.* (2024) used *Clustering* techniques combined with the Capital Asset Pricing Model (CAPM) to identify different levels of risk appetite among customers, enabling personalized investment recommendations that enhance portfolio satisfaction and performance.

Therefore, the ML technique known as Clustering encompasses methods that group observations, objects, or items that are similar to one another based on observed or measured characteristics (Jain, 2010) and is widely used in studies in this field.

Agglomerative hierarchical Clustering (Altuntas *et al.*, 2022; Lu *et al.*, 2008) is a *Clustering* method that combines pairs at each iteration, with the objective of achieving maximum similarity within each cluster and minimum similarity between clusters. The result of this type of grouping is presented as a tree-like structure called a dendrogram (Gonzales & Hargreaves, 2022), facilitating the visualization of the relationships established through the integration of observations.

Complementarily, the k-means partitioning Clustering method is one of the most widely used due to its ease of use and efficiency in its results (Jain, 2010; Sinaga & Yang, 2020; Steinley, 2006; Ybañez *et al.*, 2023). In this method, observations are randomly divided into k groups and repeatedly reorganized to form cohesive groups.

The use of both Clustering methods, such as agglomerative hierarchical and k-means partitioning, is common in academic research to enable comparison and complementarity between their results. While the former is useful for visually "suggesting" the number of clusters through exploration of the dendrogram, the latter, using metrics such as the highest Silhouette coefficient for a given number of k groups, "indicates" the quality of cluster formation and the practical applicability of the model.

On the other hand, the literature also includes studies using the PCA analytical technique, such as Hu (2025) in accounting system risk management, with the aim of reducing the dimensionality of large financial datasets while preserving essential patterns and relationships and thereby identifying significant variables capturing risk factors, such as fraud detection, anomalies, or irregular transactions; Ongbali *et al.* (2024), who used PCA to identify critical variables influencing the growth of Small and Medium-sized Enterprises (SMEs), determining the main factors from the many variables available; Mbona & Yusheng (2019), who used PCA and financial indicators to analyze three Chinese telecommunications companies and found that financial performance could be assessed using a smaller and more manageable set of indicators rather than a broad and complex range that would make interpretation itself more difficult; and Gurbaxani *et al.* (2023), who investigated variables affecting investors' perceptions of Digital Gold and Gold ETFs as investment options and, through the application of PCA, identified factors influencing investment decisions.

PCA is considered a dimensionality-reduction technique used in multivariate statistics and ML studies (Abdi & Williams, 2010). It seeks to transform a set of potentially correlated variables into a set of uncorrelated variables known as Principal Components (PCs).

By reducing the dimensionality of the original data (Garcia-Alvarez *et al.*, 2023) while preserving as much of its variability (total variance) as possible, this process generates PCs ordered according to the amount of information (explained variance) they contain (Marukatat, 2023). This makes the analysis simpler, visually interpretable, and computationally efficient, allowing researchers to analyze only the first components without significantly compromising the essence of the dataset.

Clustering and PCA techniques are frequently applied together because, while Clustering seeks to identify homogeneity among observations within groups and heterogeneity between groups, PCA reduces the dimensionality of the variables, capturing most of the data variance in two or three PCs. By projecting the original observations onto a graphical representation, with the first two components as axes (a scatterplot), it is possible to visualize and compare patterns of dispersion among organizations, as well as identify the groups formed.

3 METHODOLOGICAL PROCEDURES

3.1 Data collection and structuring of the initial dataset

The initial dataset for this study was constructed based on the ranking publicly released by *Forbes Brasil* magazine in October 2024, the *Forbes Agro100 Brasil* (Rodrigues *et al.*, 2025), which brought together the 100 largest companies in the Brazilian agribusiness sector based on their 2023 net revenues.

It was found that the companies ranked in this ranking generated more than R$1.7 trillion in net revenue and were classified into seven categories, with the respective subtotal of revenue generated and number of companies: Agroenergy, with R$270.07 billion and 23 companies; Agrochemicals and Inputs, with R$61.54 billion and 10 companies; Food and Beverages, with R$416.02 billion and 15 companies; Pulp, Wood and Paper, with R$81.18 billion and 8 companies; Commerce and Trading Companies, with R$125.26 billion and 16 companies; Cooperatives, with R$140.82 billion and 12 companies; and Animal Protein, with R$617.60 billion and 16 companies, the latter category accounting for the highest volume of net revenue.

3.2 Financial Statements and expansion of the dataset

To expand the number of financial measures to be analyzed, the authors also consulted the companies' FS from public sources, such as annual reports available on the companies' websites,

the Securities and Exchange Commission (CVM) platform (CVM, 2025; Governo Federal do Brasil, 2025a), and newspapers published at the time of their disclosure (Valor Econômico, 2025).

The original dataset was expanded to include not only net revenue (NR_bi), but also net income (NI_bi), current assets (CA_bi) and non-current assets (NCA_bi), current liabilities (CL_bi) and non-current liabilities (NCL_bi), and equity (EQ_bi), with all values measured in billions of Brazilian reais (R$ billion).

Of the 100 companies included in the *Forbes Agro 100* ranking, the FS of 16 companies could not be identified. These companies are identified here according to three criteria: original ranking position, name, and category: #9 – Grupo Amaggi – Food and Beverages; #10 – Louis Dreyfus Company Brasil – Commerce and Trading Companies; #17 – Aurora Alimentos – Food and Beverages; #22 – Inpasa Brasil – Agroenergy; #29 – UPL do Brasil – Agrochemicals and Inputs; #31 – Laticínios Bela Vista Piracanjuba – Animal Protein; #34 – Três Corações – Food and Beverages; #50 – Cotrijal – Cooperatives; #55 – Fiagril – Commerce and Trading Companies; #59 – ICL – Agrochemicals and Inputs; #74 – Delta Sucroenergia – Agroenergy; #80 – Atvos Bioenergia Brenco – Agroenergy; #81 – Fazendão Agronegócio – Commerce and Trading Companies; #82 – Prima Foods – Animal Protein; #89 – SJC Bioenergia – Agroenergy; and #99 – Barra Mansa Carnes – Animal Protein. Therefore, the researchers decided to remove these companies from the original dataset, which, after this stage, comprised a total of 84 companies.

Table 1

Data excerpt with financial measures extracted from the FS of the Top 10 companies

Company Names	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

Note. Values in billions of Brazilian reais (R$), for 2023, for: Net Revenue (NR_bi), Net Income (NI_bi), Current Assets (CA_bi), Non-Current Assets (NCA_bi), Current Liabilities (CL_bi), Non-Current Liabilities (NCL_bi), and Equity (EQ_bi).

Source: Prepared by the authors based on the FS of the 84 companies in the initial sample of this study.

Table 1 presents an excerpt from the dataset containing the seven financial measures (NR_bi, NI_bi, CA_bi, NCA_bi, CL_bi, NCL_bi, EQ_bi), in billions of Brazilian reais (R$), extracted from the FS of the Top 10 companies among the 84 that publicly disclosed their data, of which 47 were identified as “unlisted” on the stock exchange and 37 as “listed”.

3.3 Calculated indicators and added to the research dataset

After reviewing the literature on financial indicators used in studies in the field (Correa-Mejía & Lopera-Castaño, 2020; Vukčević *et al.*, 2024), the researchers selected 7 additional indicators to be calculated for each of the 84 companies: Return on Assets (ROA), Return on

Equity (ROE), Net Profit Margin (NPM), Current Ratio (CR), Overall Liquidity Ratio (OLR), Debt-to-Assets Ratio (DAR), and Debt-to-Equity Ratio (DER).

For each of the 14 financial measures used in the study, their role in the financial statements of the organizations analyzed and the respective calculation formulas were identified, as presented in Table 2.

Table 2

List of financial measures as variables explored in this study

Variable label	Measurement description	Data source / Calculation
NR_bi	Net Revenue (in billions of R$): gross sales revenue less deductions, such as returns, allowances, discounts granted, and sales taxes.	Values extracted directly from the respective Financial Statements (FS) of the companies studied.
NI_bi	Net Income (in billions of R$): the company's result after accounting for all revenues, costs, expenses, taxes, and participations.	
CA_bi	Current Assets (in billions of R$): the assets and rights that will be realized by the end of the next fiscal year.	
NCA_bi	Non-Current Assets (in billions of R$): the assets and rights expected to be realized over the long term, that is, after the end of the following fiscal year.	
CL_bi	Current Liabilities (in billions of R$): obligations that the company must settle by the end of the following fiscal year.	
NCL_bi	Non-Current Liabilities (in billions of R$): long-term obligations, including debts and commitments such as debentures and provisions for contingencies.	
EQ_bi	Equity (in billions of R$): the difference between total assets and total liabilities, representing the owners' or shareholders' own resources.	
ROA	Return on Assets measures the company's ability to generate profit from its total assets.	$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}}$
ROE	Return on Equity indicates the amount of net income generated by the company for each monetary unit invested by its owners.	$\text{ROE} = \frac{\text{Net Income}}{\text{Equity}}$
NPM	Net Profit Margin shows how much profit remains for each monetary unit of net revenue and assesses the company's operating and financial efficiency.	$\text{Net Profit Margin} = \frac{Net\ Income}{\text{Net revenue}}$
CR	Current Ratio: indicates the company's ability to pay its short-term debts using its current assets.	$\text{Current Ratio} = \frac{\text{Current Assets}}{\text{Current Liabilities}}$
OLR	Overall Liquidity Ratio assesses the company's overall ability to meet its obligations by relating total realizable assets to total liabilities.	$\text{Overall Liquidity Ratio} = \frac{(\text{Currents Assets} + \text{No Current Assets})}{(\text{Current Liabilities} + \text{No Currents Liabilities})}$
DAR	Debt-to-Assets Ratio: used as a solvency indicator, as it shows the relationship between total debt and	$\text{Debt to Assets Ratio} = \frac{\text{Total Liabilities}}{\text{Total Assets}}$

	the company's total assets, assessing its level of financial risk.	
DER	Debt-to-Equity Ratio compares debt capital with equity and measures financial leverage, indicating the extent to which the company uses debt capital relative to equity.	$\text{Debt to Equity Ratio} = \frac{\text{Total Liabilities}}{\text{Equity}}$

Source: Prepared by the authors.

Table 2 presents a total of 14 variables (financial measures) explored in this study: 7 variables (absolute values, in billions of R$) collected directly from the companies' financial statements (as described in Subsection 3.2), plus 7 additional variables (relative values) calculated according to the formula provided.

3.4 Rationale for the methodological decisions

Based on the availability of financial measures for the 84 organizations comprising the initial sample of this study, with 14 numerical variables and no missing values, the researchers performed Pearson's linear correlation tests (Schober *et al.*, 2018) among the 14 variables to avoid overfitting, multicollinearity, and/or the "duplication of weight" of certain variables in the results obtained from the subsequent ML models. The purpose was to identify and remove highly correlated variables, resulting in a set of 5 variables (NR_bi, NI_bi, ROA, ROE, and CR) with low or no correlation with one another.

Next, the Mahalanobis test (Cabana *et al.*, 2021; Mahalanobis, 1930) was also performed to detect multivariate outlier observations (discrepant measures). Four outlier observations were identified, corresponding to the following ranking positions and companies: 1 JBS, #5 Ambev, #9 Suzano, and #43 Fertilizantes Heringer.

After a detailed examination of these observations in the study dataset, atypical values that were substantially discrepant from those of the other companies in the sample were identified in at least one of the five measured variables. For Fertilizantes Heringer, the ROE value was −48.0249; for Suzano and Ambev, the NI_bi values were 14.1064 and 14.9605, respectively; and for JBS, the ROE value was −0.0155 and the NR_bi value was 363.8165.

Thus, since the purpose of the study was to collectively examine the largest possible number of organizations in the sector included in the ranking, the researchers decided to exclude these 4 outlier observations from the sample to prevent them from influencing the subsequent stages of the study.

Accordingly, the final sample consisted of 80 companies, of which 47 were unlisted on the stock exchange (58.8% of the total) and 33 were listed (41.2% of the total). Table 3 presents the descriptive statistics of the five variables used in the ML models.

Table 3
Descriptive statistics of the variables used in the ML models studied

Variable	Mean	Standard Deviation	Minimum	Q1	Median	Q3	Maximum
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

Note. Final dataset sample for this study (n = 80 organizations), with no missing data; absolute values in billions of Brazilian reais (R$) for Net Revenue (NR_bi) and Net Income (NI_bi); relative values (*ratios*) for Return on Assets (ROA), Return on Equity (ROE), and Current Ratio (CR).
Source: Prepared by the authors.

As extensively discussed in subsection 2.4, the researchers drew on studies from the accounting and finance literature to conduct the subsequent stages of this research, focusing on multivariate analysis through the application of ML techniques, namely Clustering (Dang *et al.*, 2022) and PCA (Mbona & Yusheng, 2019).

4 RESULTS AND ANALYSES

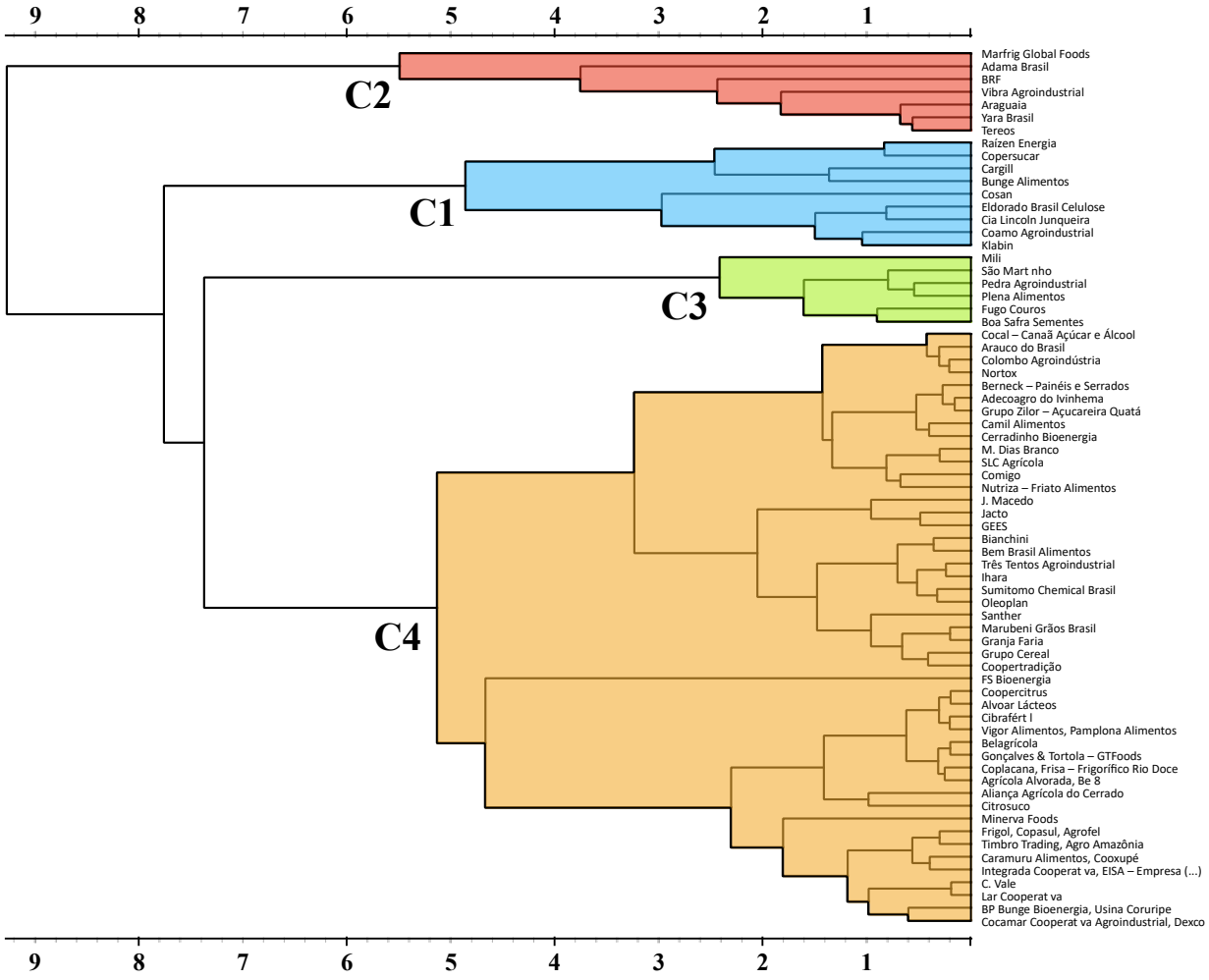
To prevent variables with larger magnitudes from dominating the results of the ML models, the five variables selected by the researchers were statistically standardized by subtracting the mean and dividing by the Standard Deviation. This resulted in columns with a mean = 0 and Standard Deviation = 1, using Orange Data Mining software (Demšar *et al.*, 2013) and R/RStudio softwares (Posit Team, 2025; R Core Team, 2025).

4.1 Clustering Models

Based on the final study sample, consisting of n = 80 organizations, ML Clustering techniques were applied (Sinaga & Yang, 2020) with the aim of identifying and interpreting the characteristics that the companies had in common based on this set of elements, including both indicators generated in the study and absolute financial values extracted directly from the FS.

The first technique applied was hierarchical Clustering (Lu *et al.*, 2008), using normalized Euclidean distance and Ward's linkage method, which resulted in Figure 1, a dendrogram, a visual representation that brings together each of the layers formed during the cluster generation procedure.

Figure 1
Dendrogram with the identification of four clusters (C1, C2, C3, and C4)



Note. after applying the hierarchical Clustering method to the final dataset of this study (n = 80 organizations) and forming four clusters (C1, C2, C3, and C4).
Source: Prepared by the authors using Orange Data Mining software (Demšar *et al.*, 2013).

When generating the *dendrogram* structure in Figure 1, it was identified that the set of 80 organizations (observations) was divided into four clusters (C1 = 9 observations, C2 = 7 observations, C3 = 6 observations, and C4 = 58 observations). The literature considers that researchers should assess whether the resulting clusters are meaningful in relation to the research problem under investigation and, accordingly, decide on the appropriate number of clusters (Kar & Varsha, 2023) to proceed with the remaining stages of the study.

To assist in this decision and assess the stability of the solution obtained from the hierarchical Clustering model, a second model was run, this time using the k-means Clustering technique (Jain, 2010). This model identified 0.336 as the highest value of the Silhouette coefficient metric (Bagirov *et al.*, 2023) and suggested the same number of four clusters, but proposed changes in the assignment of some observations among the clusters.

The proposed changes resulted in an increase in the number of observations in cluster C3, mostly from 11 observations in cluster C4. These observations corresponded to the following companies and their positions in the original ranking: 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), and Nortox

(97). In addition, one observation was reassigned from cluster C1 to cluster C3, corresponding to Cia Lincoln Junqueira (53).

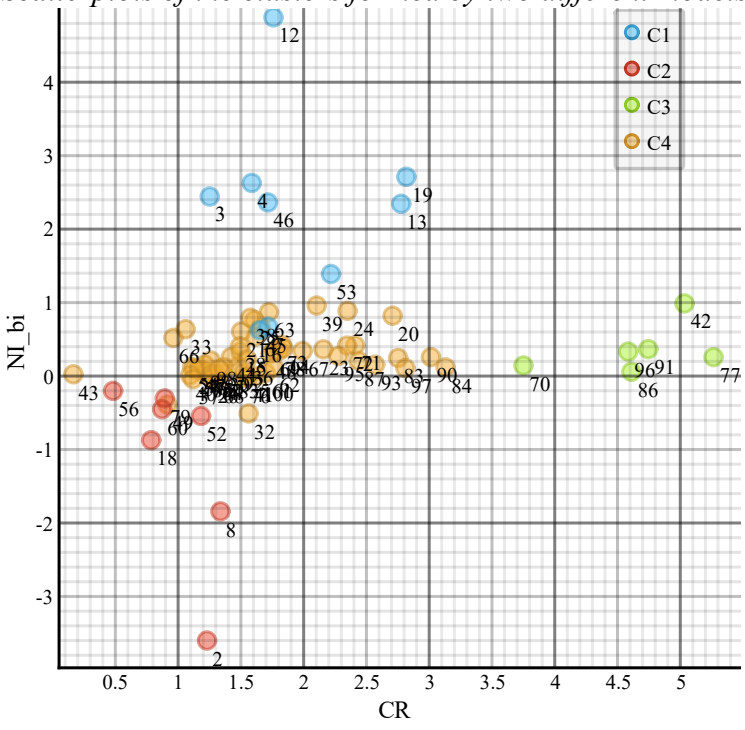
Therefore, after the formation of the clusters by the hierarchical model and the reassignment of 12 observations, the researchers chose to use the k-means model results in their entirety, as it was considered more appropriate for the research problem, demonstrating high stability and reinforcing the reliability of the selected solution.

Thus, the final distribution of observations across the clusters according to the k-means model was as follows: cluster C1 = 8 companies, cluster C2 = 7 companies, cluster C3 = 18 companies, and cluster C4 = 47 companies.

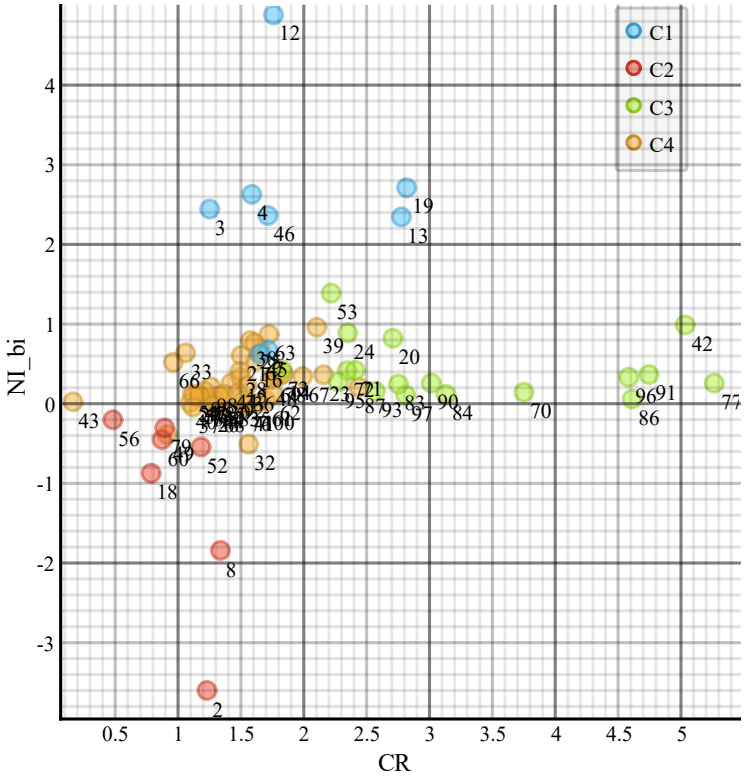
To illustrate the reassignment of observations (companies) among the clusters from the hierarchical model to the k-means model, Figure 2 was prepared.

Figure 2

Scatterplots of the clusters formed by two different models (hierarchical and k-means)



(a) hierarchical model



(b) k-means model

Note. n = 80 organizations. Part (a) of the Figure presents the result of the hierarchical Clustering model; part (b) presents the result of the k-means Clustering model.
Source: Prepared by the authors using Orange Data Mining software (Demšar *et al.*, 2013).

In the visual analysis of the Scatterplot in Figure 2, prepared using the measurements of two original variables, Current Ratio (CR) on the x-axis and Net Income (NI_bi) on the y-axis, the formation of four clusters (C1, C2, C3, and C4) among the 80 organizations analyzed was clearly identified. Part (a) of the Figure shows the results obtained with the hierarchical Clustering model, while part (b) of the same Figure presents the results obtained with the k-means Clustering model. When simultaneously observing the central region of parts (a) and (b) of Figure 2, a concentration of observations can be perceived, including the companies ranked at positions 53, 24, 20, 84, and 90. This confirms that the k-means model, which included these observations in cluster C3, represented in green, provided a more efficient separation of the clusters than the hierarchical model, which had kept these observations distributed across two other clusters (C1 and C4).

Of the four clusters formed, Cluster C1, shown in blue and named “Profitable,” consisted of 8 observations (Cargill, Bunge Alimentos, Raízen Energia, Copersucar, Cosan, Coamo Agroindustrial, Klabin, and Eldorado Brasil Celulose). It presented positive Current Ratio (CR) values ranging from 1.267 to 2.808 and the highest Net Income (NI_bi) value among all companies in the sample, at R$4.885 billion, reported by Cosan, ranked 12th in the ranking, belonging to the Agroenergy category and listed on the stock exchange. It also included five other companies with net income values above R$2.323 billion.

Cluster C2, shown in red and named “Shrinking,” consisted of 7 companies (Marfrig Global Foods, BRF, Yara Brasil, Tereos, Araguaia, Adama Brasil, and Vibra Agroindustrial), including three companies listed on the stock exchange. It presented negative values for both ROE and ROA across all companies, indicating business operations with losses and low profitability during the period evaluated. All Net Income (NI_bi) values were negative, with the largest losses identified at R$−3.621 billion (Marfrig Global Foods, ranked 2nd in the ranking) and R$−1.869 billion (BRF, ranked 8th in the ranking).

Cluster C3, shown in green and named "Healthy," consisted of 18 companies (see the list containing all names in Table 7). These companies presented positive Net Income (NI_bi), with a median of R$0.310 billion, and two of them, both belonging to the Agroenergy category, reported profits above R$1 billion: R$1.404 billion for Cia Lincoln Junqueira (ranked 53rd in the ranking) and R$1.016 billion for São Martinho (ranked 42nd in the ranking). Regarding the Current Ratio (CR), only one company presented a ratio below 2.219, namely Bem Brasil Alimentos, with a ratio of 1.832. This company was unlisted on the stock exchange, belonged to the Food and Beverages category, and ranked 73rd in the ranking. Conversely, Pedra Agroindustrial, ranked 77th in the ranking, listed on the stock exchange and belonging to the Agroenergy category, presented the highest Current Ratio (5.267) among all companies in the sample. All companies in cluster C3 also presented positive ROE (median of 0.172) and ROA (median of 0.09) indicators, reflecting financial health, profitability, and soundness.

Finally, Cluster C4, shown in orange or ochre and the largest of all the clusters, was named "Diversified" and consisted of 47 companies, as listed in Table 7. Eighteen companies (38.3%) were listed on the stock exchange, while 29 companies (61.7%) were unlisted. This cluster included seven companies with negative ROA and Net Income (NI_bi) values (FS Bioenergia, Belagrícola, Cibrafértil, Citrosuco, Gonçalves & Tortola – GTFoods, Vigor Alimentos, and Frisa – Frigorífico Rio Doce), with FS Bioenergia presenting the most negative value in this cluster (R$−0.489 billion). Five of these seven companies also had negative ROE values. Unlike cluster C1, comprising companies referred to as "Profitable," the Net Income (NI_bi) values in cluster C4 did not exceed R$1 billion. The company with the highest result in this group was SLC Agrícola, with net income of R$0.938 billion, belonging to the Food and Beverages category, listed on the stock exchange, and ranked 39th in the ranking. The Current Ratio (CR) for this group ranged from 0.167 (Minimum) to 2.441 (Maximum), with values generally lower than those of nearly all companies in cluster C3. This cluster brought together an eclectic set of companies, characterized by the absence of large profits, the presence of both positive and negative financial indicators, and a mix of listed and unlisted companies, as well as disparities in Current Ratio levels.

4.2 Principal Component Analysis (PCA) Model

Prior to using and executing the PCA analytical technique (Greenacre *et al.*, 2022), the authors evaluated two statistical criteria: (1) Bartlett's Test of Sphericity, which resulted in χ^2 = 59.04691, with df = 10 and p-value = 5.487361e-09; and (2) the Kaiser-Meyer-Olkin (KMO) test, a statistical measure used to assess the adequacy of the sample for PCA, which resulted in an overall KMO value above 0.6.

After validating and meeting both statistical criteria, the PCA model was executed, identifying five interpretable PCs. In interpreting each PC, the first component typically captures the dominant common factor, and it is recommended to examine the magnitude and direction of the coefficients of the original variables, with higher absolute coefficient values indicating greater importance of the respective original variable in calculating the PC. Researchers should use their own knowledge, together with findings from the literature, to determine which levels of correlation among variables should be accepted in the analyses.

It was identified that the first component (PC1) explained 40.80% of the variance in the data, capturing overall profitability (ROA, ROE) and Current Ratio (CR).

The second component (PC2) was predominantly composed of Net Revenue (NR_bi) (loading = 0.88), which can be interpreted as representing the organization's scale or size, and accounted for an additional 21.16% of the variance. Together, the first and second components, PC1 and PC2, explained 61.96% of the total variance in the data, meeting the minimum criterion of 60% recommended by Hair *et al.* (2019).

The third component (PC3) explained 15.84% of the variance, while the remaining components (PC4 and PC5) contributed 14.97% and 7.22% of the variance in the data, respectively.

Table 4 and Figure 3 provide a detailed illustration of the composition of the five PCs (PC1, PC2, PC3, PC4, and PC5) obtained after execution of the PCA model.

Table 4

Explained Variance by each of the five PCs

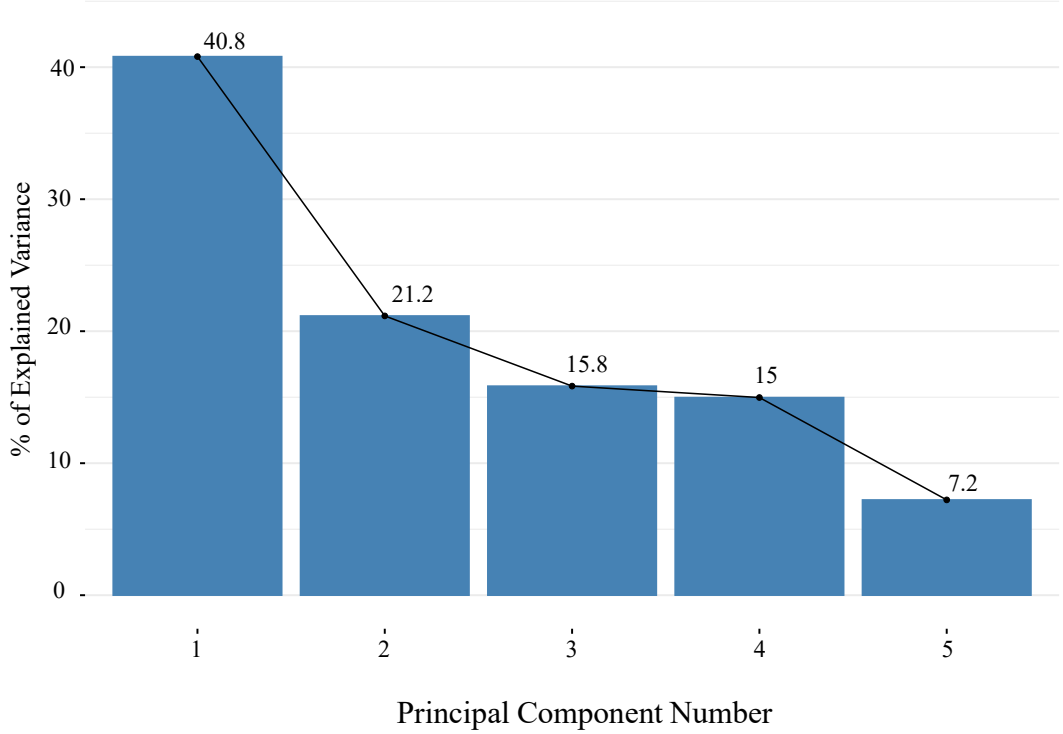
Component	Standard Deviation	Variance / Eigenvalue	Variance Proportion (%)	Cumulative Variance (%)
PC1	1,428360	2.040211	40.804220	40.80
PC2	1.028556	1.057928	21.158565	61.96
PC3	0.889970	0.792047	15.840949	77.80
PC4	0.865317	0.748774	14.975484	92.78
PC5	0.600865	0.361039	7.220782	100

Source: Prepared by the authors after executing the PCA model.

The literature reports that, generally, the first three PCs are the most relevant for interpretation and analysis, as they account for most of the cumulative variance. In this study, as shown in Table 4, the first three components together explained 77.80% of the variance in the data of the organizations evaluated.

Figure 3

Scree plot showing the explained variance by the five PCs



Note. In the graph, the numbers highlighted above the bars are expressed as percentages.
Source: Prepared by the authors after executing the PCA model.

Figure 3 presents a Scree plot (Cattell, 1966) showing the explained variance by each PC (numbers 1, 2, 3, 4, and 5) in descending order. The *x-axis* represents the five PCs, while the *y-axis* shows the proportion of the total variance in the data captured by each component.

Table 5 presents the loadings (or factor loadings) of each original variable, which are coefficients that multiply the original variables to form the PCs and represent part of the correlation between each original variable and each principal component.

Table 5
Factor loadings (or loadings) of the original variables for each PC

Original Variables	Principal Components				
	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

Source: Prepared by the authors after executing the PCA model.

Based on Table 5, when a PC has an original variable with a positive factor loading (loading), this indicates a direct relationship between the original variable and the respective PC: as the variable increases, the PC also tends to increase. Conversely, when the factor loading is negative, this indicates an inverse relationship: as the value of those specific variable increases, the PC tends to decrease. For example, in PC1, the loading of -0.102735 for the original variable NR_bi is a coefficient indicating the contribution or weight of NR_bi to that PC, and its negative sign ("-") indicates that an increase in this original variable is associated with a decrease in the value of PC1, and vice versa. Negative loadings are very common and should not be interpreted as having a "bad" or problematic meaning, but simply as indicating the inverse direction of the relationship (correlation) between an original variable and the respective PC, or even between the original variables comprising the same PC.

Furthermore, when original variables within the same PC have opposite signs, they contribute inversely to the composition of that PC. For example, because the value of PC2 is generated by a linear combination of its original variables (PC2 = 0.877409×NR_bi + 0.351813 ×NI_bi + 0.135779×ROE + 0.018261×ROA + -0.295993×CR), in the context of PC2, Net Revenue (NR_bi) and Current Ratio (CR) have opposite signs in the composition of PC2 – positive value 0.877409 and negative value -0.295993, respectively − and therefore move in opposite directions along this PC axis.

Thus, the formation of the first component, PC1, comprising the original variables with rounded values, NI_bi (+0.42), CR (+0.45), and ROE (+0.47), with moderate positive loadings; ROA (+0.62), with a somewhat higher positive loading; and NR_bi (−0.1), with a very low negative loading, was interpreted as the component of the Efficiency and Solvency Index (Performance Quality), or, more concisely, as the organizational "*Efficiency*" component.

PC1 is influenced by efficiency in asset utilization (ROA) and moderately influenced by shareholder returns (ROE), net income (NI_bi), and Current Ratio (CR), which indicates the ability to meet short-term obligations. This component groups variables related to profitability relative to assets and equity, the ability to generate profit, and current liquidity. In other words, relative operating efficiency (efficiency in generating returns) and earnings quality contribute most strongly to PC1.

The low loading (nearly zero) of Net Revenue (NR_bi) on PC1 indicates that revenue volume is not the determining factor in this component, which assesses the quality of financial performance, that is, how profitable and liquid a company is, regardless of its size (revenue volume). Companies with high PC1 scores are regarded as efficient and maintain financial flexibility.

The second component, PC2, which is orthogonal (uncorrelated) to PC1, comprises the variables with rounded values NR_bi (+0.88), with a very strong positive loading; NI_bi (+0.35) and ROE (+0.14), with weak positive loadings; ROA (+0.02), with a positive loading close to zero; and CR (-0.3), with a weak negative loading. It was interpreted as the component of the Scale and Size Index (Operating Volume), or, more concisely, as the organizational "*Scale*" component.

In PC2, the dominant loading is that of Net Revenue (NR_bi), which is positively correlated with Net Income (NI_bi), meaning that absolute revenue and profit vary together (both positively), reflecting the capacity to generate revenue in absolute terms, while varying independently of relative efficiency (ROA ≈ 0) and inversely with Current Ratio (CR, negative).

For PC3 (the third principal component), which is most strongly influenced by ROE (+0.76), with a strong positive loading, it can be named the Financial Leverage Index (Risk/Return), or, more concisely, the organizational "Risk-Return" component. ROA (+0.05) and NR_bi (−0.06) can be identified as variables with virtually zero correlation, while Net Income (NI_bi = −0.49) and Current Ratio (CR = −0.42) have moderate negative *loadings*. The latter two values indicate that both variables move in the opposite direction as this PC increases.

High PC3 scores (positive) differentiate companies primarily by ROE (capital structure), rather than by operating performance (ROA, which is essentially zero). This isolates the financial leverage component from profitability, indicating a leveraged capital structure.

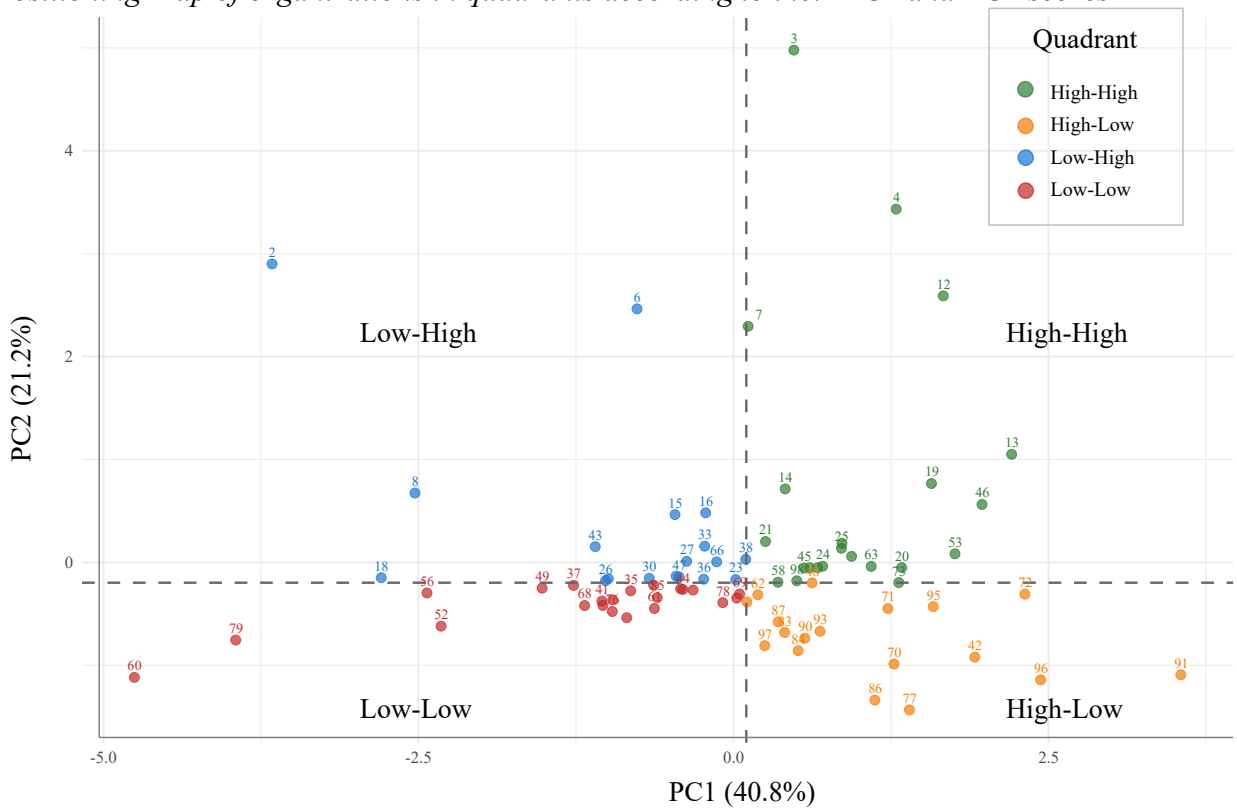
Figure 4 shows the Positioning Map (scatterplot) of the organizations in two-dimensional space, based on the specific measurements of two of their PCs, PC1 and PC2, which together explain 62.0% of the total variance in the data (40.8% and 21.2%, respectively), meeting the minimum criterion of 60% recommended by Hair *et al.* (2019).

Quadrant analysis is a technique used in accounting, finance, and organizational management to visually group company characteristics based on two key indicators. In this study, the quadrants (high-high, high-low, low-high, low-low) are formed using the medians of these measurements (PC1 Median = 0.1020348 and PC2 Median = −0.1972229), represented by dotted lines and used as dividing axes (PC1 on the *x-axis* versus PC2 on the *y-axis*), corresponding to four distinct financial performance profiles among the 80 organizations in the study sample.

On the horizontal *x-axis*, the principal component PC1 can represent the quality of returns and financial soundness, with positive values characterizing high operating efficiency and liquidity ("*Efficiency*"). On the vertical *y-axis*, the principal component PC2 can represent absolute company size and revenue-generating capacity, with positive values indicating greater organizational scale ("*Scale*").

Figure 4

Positioning Map of organizations in quadrants according to their PC1 and PC2 scores



Note. the number shown in the label of each point corresponds to the organization's position in the original Forbes Agro 100 *ranking*; the dotted lines represent the medians of the PCs, with a PC1 median of 0.1020348 and a PC2 median of −0.1972229.

Source: Prepared by the authors after executing the PCA model.

Table 6

Characteristics and profile of organizations according to their position in the quadrants

Quadrant Position	Number of Companies	PC1 and PC2 Measurements	Profile Assigned to Organizations	Map Labels (Examples)	Quadrant Characteristics
High–High	22	PC1 > 0.102 and PC2 > -0.197	Performance and Market Leaders	3, 4, 7, 12	High Efficiency and High Scale
High–Low	18	PC1 > 0.102 and PC2 < -0.197	Smaller and Profitable	77, 86, 91, 96	High Efficiency and Low Scale
Low–High	18	PC1 < 0.102 and PC2 > -0.197	Larger and Less Efficient or Experiencing a Temporary Downturn	2, 6, 8, 15	Low Efficiency and High Scale
Low–Low	22	PC1 < 0.102 and PC2 < -0.197	Smaller and Less Efficient	52, 56, 60, 79	Low Efficiency and Low Scale

Source: Prepared by the authors after executing the PCA model.

Table 6 also complements Figure 4 by presenting the summarized characteristics of the interpretation of each quadrant and the abbreviated profiles of the organizations occupying the respective positions. This interpretation is based on the factors loadings of the original variables comprising the PCs (PC1 and PC2), which define the economic meaning of each axis in the Figure.

The visual analysis of Figure 4 revealed four distinct business profiles: (i) the *High–High* quadrant (green) positions large, profitable companies characterized by high ROA, ROE, and CR

indicators combined with high NR_bi; (ii) the *High–Low* quadrant (orange) positions smaller companies with high relative profitability but lower operating scale than the previous quadrant. These may represent organizations with growth opportunities and established operating efficiency; (iii) the *Low–High* quadrant (blue) positions large companies with low profitability, indicating possible operating efficiency challenges; and (iv) the *Low–Low* quadrant (red) concentrates companies characterized by low profitability/performance and smaller size, and is the quadrant containing the largest number of companies (14) with negative Net Income (NI_bi).

Table 7

List of organizations in the final study sample with the results of the ML models

Rank	Organization Name	Cluster k	Silhouette	Cluster h	Quadrant
2	Marfrig Global Foods	C2	0,531288	C2	Low–High
3	Cargill	C1	0.597275	C1	High–High
4	Bunge Alimentos	C1	0.611179	C1	High–High
6	Raízen Energia	C1	0.500019	C1	Low–High
7	Copersucar	C1	0.493918	C1	High–High
8	BRF	C2	0.498727	C2	Low–High
12	Cosan	C1	0.583040	C1	High–High
13	Coamo Agroindustrial	C1	0.486515	C1	High–High
14	Minerva Foods	C4	0.587184	C4	High–High
15	C. Vale	C4	0.654965	C4	Low–High
16	Lar Cooperativa	C4	0.646484	C4	Low–High
18	Yara Brasil	C2	0.512805	C2	Low–High
19	Klabin	C1	0.498610	C1	High–High
20	Comigo	**C3**	0.567341	C4	High–High
21	Cocamar Cooperativa Agroindustrial	C4	0.646039	C4	High–High
23	Camil Alimentos	C4	0.617098	C4	Low–High
24	M. Dias Branco	**C3**	0.460072	C4	High–High
25	Três Tentos Agroindustrial	C4	0.573489	C4	High–High
26	Agrícola Alvorada	C4	0.673283	C4	Low–High
27	Timbro Trading	C4	0.675374	C4	Low–High
28	Bianchini	C4	0.556512	C4	High–High
30	Integrada Cooperativa	C4	0.679409	C4	Low–High
32	FS Bioenergia	C4	0.538078	C4	High–High
33	BP Bunge Bioenergia	C4	0.667061	C4	Low–High
35	Coopercitrus	C4	0.670562	C4	Low-Low
36	Caramuru Alimentos	C4	0.671576	C4	Low–High
37	Belagrícola	C4	0.664333	C4	Low-Low
38	Dexco	C4	0.647157	C4	Low–High
39	SLC Agrícola	C4	0.577705	C4	High–High
40	Be 8	C4	0.676662	C4	Low–High
41	Cibrafértil	C4	0.658270	C4	Low-Low
42	São Martinho	C3	0.607044	C3	High-Low
43	Aliança Agrícola do Cerrado	C4	0.657335	C4	Low–High
44	Cooxupé	C4	0.678605	C4	Low-Low

45	Sumitomo Chemical Brasil	C4	0.603736	C4	High–High
46	Eldorado Brasil Celulose	C1	0.462455	C1	High–High
47	Copasul	C4	0.680429	C4	Low–High
48	Oleoplan	C4	0.591254	C4	High-Low
49	Citrosuco	C4	0.635172	C4	Low-Low
52	Tereos	C2	0.459307	C2	Low-Low
53	Cia Lincoln Junqueira	**C3**	0.527408	**C1**	High–High
54	Agro Amazônia	C4	0.680128	C4	Low–High
56	Araguaia	C2	0.449725	C2	Low-Low
57	EISA – Empresa Interagrícola	C4	0.681771	C4	Low-Low
58	Marubeni Grãos Brasil	C4	0.627120	C4	High–High
60	Adama Brasil	C2	0.568579	C2	Low-Low
61	Alvoar Lácteos	C4	0.663464	C4	Low-Low
62	Grupo Cereal	C4	0.640556	C4	High-Low
63	Ihara	C4	0.534639	C4	High–High
64	Coplacana	C4	0.673491	C4	Low-Low
65	Agrofel	C4	0.679382	C4	Low-Low
66	Usina Coruripe	C4	0.666473	C4	Low–High
67	Adecoagro do Ivinhema	C4	0.635740	C4	High-Low
68	Gonçalves & Tortola – GTFoods	C4	0.663482	C4	Low-Low
69	Grupo Zilor – Açucareira Quatá	C4	0.654304	C4	Low-Low
70	Fugo Couros	C3	0.618109	C3	High-Low
71	Jacto	**C3**	0.549145	C4	High-Low
72	J. Macedo	**C3**	0.562016	C4	High-Low
73	Bem Brasil Alimentos	**C3**	0.479976	C4	High–High
75	Frigol	C4	0.675731	C4	Low-Low
76	Vigor Alimentos	C4	0.664209	C4	Low-Low
77	Pedra Agroindustrial	C3	0.598865	C3	High-Low
78	Berneck – Painéis e Serrados	C4	0.648997	C4	Low-Low
79	Vibra Agroindustrial	C2	0.572208	C2	Low-Low
83	Colombo Agroindústria	**C3**	0.490017	C4	High-Low
84	Arauco do Brasil	**C3**	0.553020	C4	High-Low
85	CMAA – Cia. Mineira de Açúcar e Álcool	C4	0.679348	C4	Low-Low
86	Plena Alimentos	C3	0.606592	C3	High-Low
87	Cerradinho Bioenergia	C4	0.554057	C4	High-Low
88	Santher	C4	0.601910	C4	High–High
90	Cocal – Canaã Açúcar e Álcool	**C3**	0.518812	C4	High-Low
91	Mili	C3	0.604391	C3	High-Low
92	Coopertradição	C4	0.653179	C4	Low-Low
93	Nutriza – Friato Alimentos	**C3**	0.519902	C4	High-Low
94	Frisa – Frigorífico Rio Doce	C4	0.670756	C4	Low-Low
95	GEES	**C3**	0.549034	C4	High-Low
96	Boa Safra Sementes	C3	0.625329	C3	High-Low
97	Nortox	**C3**	0.494371	C4	High-Low

98	Granja Faria	C4	0.620105	C4	High–High
100	Pamplona Alimentos	C4	0.661480	C4	Low-Low

Note. The information in the "Rank" and "Organization Name" columns was obtained from the *Forbes Agro 100 Ranking* for 2024 (Rodrigues *et al.*, 2025), while the remaining information in the Table ("Cluster k," "Silhouette," "Cluster_h," and "Quadrant") was generated by the authors based on the results of the ML models studied.
Source: Prepared by the authors.

Table 7 presents the list of organizations included in the final sample of this study, indicating their original position in the ranking (Rank), the abbreviated name of the organization, the information obtained from the models, including the identification in the respective k-means Clustering (Cluster k), the cohesion and separation score of the observation within the cluster (Silhouette), the identification in the respective hierarchical Clustering (Cluster h), and the organization's position (Quadrant) in the Scatterplot formed by the principal components PC1 and PC2, as presented in Figure 4.

Against the backdrop of Signaling Theory and the use of FS as issuers of "signals" regarding the performance indicators, financial quality, and business prospects of agribusiness companies, researchers can now extend the role of FS beyond the integration of observable accounting and financial metrics by employing statistical techniques and multivariate modeling. Such signals can be decoded through ML techniques to explore the various forms of communication between these companies and the market and support stakeholders' decision-making.

5 RESEARCH LIMITATIONS AND FUTURE STUDIES

During the development of this work, some methodological and operational limitations were identified, which affected the execution of the study. First, there were difficulties in accessing the FS of certain companies included in the ranking used in the study. Despite being major players in the sector, several of these companies did not make their FS available on public platforms or even on their own institutional websites, which reduced the sample size and constrained the representativeness of the results.

Second, the exclusion of advanced financial metrics, such as Economic Value Added (EVA) and Market Value Added (MVA), can also be considered a limitation of this study, as such metrics would require complementary information, including market reports, identification of the Weighted Average Cost of Capital (WACC), and companies' market values. These requirements would extend beyond the scope defined by the FS, which were the primary focus of this study.

Third, to avoid inconsistencies in the calculations, the integration and structuring of the research dataset also posed challenges for the researchers, requiring the standardization of financial data and recurring processes of conversion, verification, and adjustment, which demanded additional time and effort to perform these procedures.

Additionally, multivariate analysis using ML techniques to analyze financial metrics faces limitations inherent to data quality and availability. The researchers reiterate that extensive efforts were made throughout all stages of this study, with a focus on scientific rigor and applied relevance, and that the results obtained cannot be generalized to other contexts, sectors, countries, periods, or even other company rankings, as they are restricted to the dataset used in this research.

Future studies may propose the development of a new ranking, providing a more comprehensive understanding of the 80 organizations included in this study by combining the five financial metrics analyzed (Net Revenue, Net Income, ROA, ROE, and Current Ratio), thereby going beyond the original ranking, which used only one metric (Net Revenue).

Finally, the researchers reiterate the relevance of advancing academic and scientific research structured through collaboration between academics and professionals/practitioners,

contributing to the development of more precise and realistic metrics for the construction of rankings. Such metrics may assist companies, associations, and business media outlets in developing rankings that better reflect the needs of their target sectors, such as the agribusiness sector itself, which is fundamental to both the Brazilian and global economies.

6 FINAL CONSIDERATIONS

This study analyzed the FS (balance sheet and income statement) of companies in the Brazilian agribusiness sector included in the 2024 *Forbes Agro100 Ranking*, using financial measurements (absolute values and ratios) and applying analytical techniques and ML models (Clustering and PCA).

The 80 companies analyzed formed four clusters, named "Profitable" (cluster C1, with 8 companies), "Shrinking" (cluster C2, with 7 companies), "Healthy" (cluster C3, with 18 companies), and "Diversified" (cluster C4, with 47 companies).

Relevant findings regarding the structure of the Brazilian agribusiness market may offer academic and managerial value in at least three respects: (1) the decomposition of data variance into orthogonal principal dimensions of "Efficiency" (PC1), "Scale" (PC2), and "Risk-Return" (PC3) provided insights with potential applications in organizational management based on the analysis of the financial metrics studied; (2) profitability and organizational size are relatively independent dimensions (PC1 is orthogonal to PC2), indicating, for this sample of companies operating in the Brazilian market, that scale/size does not guarantee operating efficiency; and (3) by generating the positioning map, a scatterplot formed by PC1 (x-axis) and PC2 (y-axis), with the visualization of the four quadrants, it was possible to reveal financial performance characteristics of the grouped companies, opening new avenues for future research, particularly regarding the development of new metrics for constructing company rankings.

Signaling Theory highlighted the importance of companies communicating their information to stakeholders and the market. By establishing the informational role of financial statements in the agribusiness context, the financial indicators extracted from the FS are not merely accounting metrics, but also vehicles for communicating asymmetric information, emitting and transmitting signals to the market regarding companies' commitments, behaviors, and intentions, as well as their financial performance.

By applying multivariate analysis with ML techniques, this study captured patterns in these signals and made it possible to distinguish and substantiate the formation of different groups of companies according to their signaling profiles.

Based on the results presented in this study, as a contribution to the fields of finance and accounting, it was possible to understand that, for a company ranking to be more comprehensive, it should not be limited solely to Net Revenue values. Both the composition of the clusters (C1, C2, C3, and C4) and the formation of the first three principal components (PC1, PC2, and PC3) indicated that some companies positioned lower in the ranking nevertheless have higher-quality absolute values and financial indicators when compared with others occupying more prominent positions in the same ranking.

REFERENCES

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

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest.

DATA AVAILABILITY

The complete dataset supporting the findings of this study is available from the authors upon reasonable request.

AUTHOR CONTRIBUTIONS

Roles	1st author	2nd author	3rd author
Conceptualization	♦	♦	♦
Data Curation	♦	♦	♦
Formal Analysis			♦
Funding Acquisition			
Investigation	♦	♦	♦
Methodology	♦	♦	♦
Project Administration			♦
Resources			
Software			♦
Supervision			♦
Validation	♦	♦	♦
Visualization	♦	♦	♦
Writing – Original Draft Preparation	♦	♦	
Writing – Review & Editing	♦	♦	♦

COPYRIGHT

The authors retain the copyright to their articles and grant the Revista Catarinense da Ciência Contábil (RCCC) the right of first publication. The work is simultaneously licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).