

## Proposta de classificação de perfis de decisão: um estudo gerencial

### Proposed classification of decision profiles: a managerial study

DOI:10.34117/bjdv7n8-662

Recebimento dos originais: 07/07/2021

Aceitação para publicação: 30/08/2021

#### **Rodrigo Speckhahn Soares da Silva**

Mestrado em Administração pela Universidade Federal de Santa Catarina (UFSC)  
Instituição: Universidade Federal de Santa Catarina (UFSC)  
Endereço: UFSC - Campus Universitário Trindade - Centro Sócio Econômico (CSE)  
E-mail: speckhahn@gmail.com

#### **Claudelino Martins Dias Junior**

Doutorado em Co-Tutela pela Universidade Federal de Santa Catarina (UFSC) e  
Universidade Nova de Lis-boa (UNL)  
Instituição: Universidade Federal de Santa Catarina (UFSC)  
Endereço: UFSC - Campus Universitário Trindade - Centro Sócio Econômico (CSE)  
E-mail: claudelino@gmail.com

#### **RESUMO**

Este estudo propõe um modelo de classificação de perfis de decisão utilizando o método de agrupamento (*clustering*) através de aprendizado de máquina (*Machine Learning* - ML *K-means*). Para tanto, foram utilizados dados de um ambiente simulado do tipo industrial, que representam as decisões de 8 (oito) empresas ao longo de 12 (doze) trimestres de atuação. A técnica *K-means* segmentou o conjunto de empresas industriais tendo em conta as decisões quantitativas de todas, em momentos de decisão específicos representados por trimestres. Os resultados sugerem a existência 2 (dois) *clusters*, sendo que as empresas do primeiro *Cluster* (C1) apresentaram um resultado financeiro e evolução patrimonial superior, sendo reflexo de um conjunto de decisões mais assertivas quando comparadas àquelas tomadas pelas empresas do *Cluster* C2.

**Palavras-chave:** Tomada de decisão, *Clusters*, Machine Learning, *K-means*, Simulação Gerencial.

#### **ABSTRACT**

This study proposes a decision profile classification model using the clustering method through machine learning (Machine Learning - ML *K-means*). For this purpose, data from a simulated industrial environment were used, representing the decisions of 8 (eight) companies throughout 12 (twelve) quarters of operation. The *K-means* technique segments the set of industrial companies taking into account the quantitative decisions of all of them, in specific decision moments represented by quarters. The results suggest the existence of 2 (two) clusters, with the companies in the first cluster (C1) presenting a higher financial result and equity evolution, reflecting a set of more assertive decisions when compared to those made by the companies in Cluster C2.

**Keywords:** Decision Making, Clusters, Machine Learning, *K-means*, Managerial Simulation.

## 1 INTRODUÇÃO

A tomada de decisão constitui há muito tema complexo e de difícil de avaliação, decorrente de complexas relações entre variáveis (internas, externas, contextuais e ainda circunstanciais), posto que tais variáveis são atreladas a uma interpretação muito própria dos gestores e, nesse sentido, condicionadas a análises subjetivas. Ademais, seu processo de avaliação acontece a posteriori e condicionado ainda por eventos fora de controle.

O uso de ambientes simulados constitui uma oportunidade de melhor visualização de como algumas variáveis condicionam o comportamento de decisão dos gestores. Contudo, o uso desse tipo de framework exige uma definição clara dos objetivos a que se propõe bem como do tipo de público a que se destina.

Através do uso das Tecnologias da Informação (TICs), um conjunto cada vez maior e diversificado de dados e informações podem ser devidamente aproveitados no processo de tomada de decisão. Os gestores precisam condicionar-se a essa nova realidade e estarem devidamente preparados, de forma que possam decidir apoiados em elementos que, por vezes, se distanciam de suas próprias percepções e opiniões.

Neste estudo, a partir de dados retirados de um exercício de gestão simulado com a participação de alunos de graduação e pós-graduação, propõe-se categorizar a atuação de empresas virtuais tendo como referência os conjuntos de decisões tomadas por seus gestores.

O ferramental metodológico é representado por técnicas de aprendizado de máquina para a identificação das semelhanças e das diferenças entre os players (empresas virtuais), bem como pela utilização do algoritmo de aprendizado não supervisionado *K-means*.

## 2 FUNDAMENTAÇÃO TEÓRICA

March e Simon (1958) afirmam que em função da limitação da capacidade intelectual humana diante de problemas que as pessoas e organizações enfrentam, o comportamento racional exige a elaboração de modelos simplificados que capturem as principais nuances de um problema sem levar em conta toda sua complexidade. Nesse sentido, o uso de ambientes simulados pode contribuir no estudo de determinados elementos, considerando que esses últimos possam ser isolados e avaliados separadamente.

Papadakis e Barwise (2002) em seu estudo afirmaram que 2 (dois) dos principais temas de estudos em estratégia são: o papel da alta administração e a tomada de decisão

em nível estratégico. Segundo Summer et al (1990) e Ketchen Jr. e Shook (1996) a pesquisa de gestão estratégica foca relações entre estratégia, ambiente, liderança organizacional e desempenho sendo que cada uma desses 4 (quatro) elementos é multidimensional.

A estratégia, segundo Ansoff (1965), é composta de preocupações de processo e conteúdo; para Hofer e Schendel (1980) como escopo e distribuição de recursos; e para Andrews (1971) como questões corporativas, comerciais e funcionais. Thompson (1967) divide o ambiente em tarefas e elementos gerais. Para Summer et al. (1990) a liderança organizacional encapsula uma variedade de características firmes, incluindo estrutura e cultura. Finalmente, o desempenho para Venkatraman e Ramanujam (1986) é composto ao menos por 3 (três) categorias: finanças, operações e eficácia geral. Assim, a multidimensionalidade dessas construções cria um desafio conceitual na medida em que uma vasta gama de combinações específicas poderia ser desenvolvida ao longo dessas dimensões para descrever as organizações (KETCHEN JR.; SHOOK, 1996).

Hambrick e Mason (1984) ao tratarem da perspectiva da alta administração ou “alto escalão” de uma organização empresarial afirmam que o desempenho e o rumo estratégico são, em parte, reflexos das características e das escolhas dos gestores desse nível. Nesse contexto, Serra, Serra e Tomei (2014) identificaram 3 (três) *clusters* teóricos como características do Top Management Team (TMT) e como essas mesmas características influenciam a tomada de decisão, sendo: (i) fatores ambientais, (ii) exposição de modelos e processos e, (iii) conflito e consenso.

Meyer, Tsui e Hinings (1993) utilizam o termo “configuração organizacional” para designar um agrupamento multidimensional de características conceitualmente distintas, mas que, habitualmente, ocorrem em conjunto. Essas configurações podem ser representadas em tipologias desenvolvidas conceitualmente ou capturadas em taxonomias derivadas empiricamente. Podem ser situadas em múltiplos níveis de análise, representando padrões comuns entre indivíduos, grupos, departamentos, organizações ou ainda redes de organizações.

De um número infinito de possibilidades, o que se observa é que a variedade potencial limita-se pela tendência de os atributos agruparem-se em padrões coerentes. Ou seja, apenas uma fração de todas as possibilidades teóricas concebíveis são viáveis e aptas a serem observadas empiricamente. Os investigadores na seara de configurações procuram gerar tipologias e taxonomias, ou seja, grupos distintos que esgotam coletivamente uma

parte significativa da população alvo em análise (MEYER; TSUI; HININGS, 1993; MILLER; FRIESEN; MINTZBERG, 1984).

A geração dessas configurações ocorre em função da atuação de muitas forças capazes de fazer com que os atributos organizacionais se agrupem sistematicamente ou aparentemente fazê-lo. Dessa forma, uma resposta popular a este desafio foi identificar “configurações organizacionais”, ou seja, conjuntos de empresas que possuem perfis comuns em relação a variáveis conceitualmente distintas. Independentemente de quaisquer rótulos específicos, algumas configurações podem capturar a complexidade da realidade organizacional (MEYER; TSUI; HININGS, 1993).

Ketchen Jr. e Shook (1996) destacam que, em função da ênfase da administração estratégica na identificação de grupos de organizações similares, a análise de agrupamento tem se popularizado. Hoskisson et al. (1999) afirmam que, inicialmente, essa busca por configurações centrava-se na literatura ligada à economia industrial/organizacional e os grupos eram definidos a partir de conjuntos limitados de variáveis, frequentemente 2 (duas) ou 3 (três). Essas análises, portanto, possibilitavam mapear estruturas, mas eram demasiado grosseiras para capturar a multidimensionalidade dos construtos.

## 2.1 CLASSIFICAÇÕES DE ANÁLISES ORGANIZACIONAIS

Meyer, Tsui e Hinings (1993) afirmam que, tradicionalmente, na análise organizacional existem inúmeras tentativas de classificar organizações e que, portanto, tais tentativas estão na raiz das teorizações. Citam: as noções de Weber (1947) de carisma, tradicionalismo e burocracia; a distinção de Burns e Stalker (1961) entre as estruturas mecanicistas e orgânicas; as distinções de Mintzberg (1979) entre estrutura simples, burocracia de máquina, burocracia profissional (formas divionalizada e “adhocracia”). Essas classificações têm por objetivo possibilitarem abstrações, explorando ideias teóricas centrais de forma sistemática e servir de apoio a um princípio central da teoria da organização: a existência de diferentes tipos de organização e os seus muitos (ou todos) aspectos de funcionamento.

As configurações organizacionais que incorporam múltiplas dimensões são capazes de se mostrarem mais valiosas, tanto em aplicações teóricas como empíricas. Contudo, na medida em que as dimensões são acrescentadas para aumentar a congruência com a realidade, as configurações, necessariamente, tornam-se mais complexas e desajeitadas. Nenhuma taxonomia, por melhor que seja, seria capaz de replicar perfeitamente a realidade, no entanto, ainda que pudesse ser construída, sua

especificidade derrotaria seu propósito de generalizar e abstrair. Estudos dessa natureza, com abordagens configuracionais, são agrupados em 2 (duas) vertentes: a linha dos tipologistas e a dos taxonomistas (Quadro 1). No primeiro grupo, majoritariamente, referem-se a conjuntos de configurações conceitualmente derivadas; e no segundo grupo, a conjuntos de configurações empiricamente derivadas (MEYER; TSUI; HININGS, 1993).

**Quadro 1** – Abordagens configuracionais em administração.

Tipologistas	Taxonomistas
Seguem, em geral, a lógica weberiana de tipos ideais, acentuando características consideradas chave de modo a traçar distinções a priori entre organizações.	Sua lógica e abordagem configuradora alternativa, está na classificação empírica baseada na análise multivariada de múltiplas dimensões que podem cobrir estruturas, processos, estratégias e contextos.

**Fonte:** Adaptado de Meyer, Tsui e Hinings (1993).

## 2.2 ANÁLISE DE *CLUSTERS*

Segundo Ketchen Jr e Shook (1996), a análise de *clusters* (de conglomerados, classificação ou agrupamento) toma uma amostra de elementos como uma organização e os agrupa de tal forma que a variância estatística entre os elementos é minimizada, enquanto a variância entre os grupos é maximizada. Estudos dessa natureza foram inicialmente abordados na literatura de economia industrial e organizacional com grupos definidos através de conjuntos com poucas variáveis. Como consequências, tinham dificuldades em captar a multidimensionalidade das construções de interesse na investigação estratégica.

A análise de *clusters* é um dos métodos exploratórios no contexto da Análise Multivariada que permite compreender o comportamento dos dados, admitindo que haja um grau de similaridade ou de diferença entre eles. Assim, pode ser utilizada na classificação de elementos em grupos homogêneos de acordo com uma característica de interesse. Essa metodologia é muito comum em disciplinas que envolvam a análise de dados multivariados, muito utilizada, inclusive no contexto gerencial, pois permite resumir uma grande quantidade de dados a partir de uma visualização simplificada de interpretação. No contexto da pesquisa gerencial essa metodologia pode ser utilizada na segmentação de mercados e na identificação de diferentes padrões e perfis de comportamentos (JAIN, 2010; KAUFMAN; ROUSSEEUW, 2005; LOPES; GOSLING, 2021).

Segundo Jain (2010) o objetivo da análise de *cluster* é descobrir os agrupamentos naturais de um conjunto de padrões, pontos ou objetos. Do ponto de vista operacional, define-se a clusterização como numa representação de  $n$  objetos para encontrar  $K$  grupos com base em uma medida de similaridade tal que as semelhanças entre objetos do mesmo grupo sejam altas, enquanto as semelhanças entre objetos diferentes sejam baixas. Os *clusters* podem diferir em forma, tamanho e densidade e quando da presença de ruído sua detecção torna-se ainda mais difícil. Verifica-se, dessa forma, que um *cluster* é uma entidade subjetiva e que está nos olhos de quem vê e cujo significado e interpretação requerem conhecimento e domínio.

Sun et al (2017) afirmam que o agrupamento desempenha um papel importante nos contextos empresariais, sendo amplamente utilizado na segmentação de mercados, categorização de produtos, na análise de perfis de clientes, padrões comportamentais, níveis de desempenho, tipos de fluxo de trabalho e modelos empresariais. Não apenas as empresas grandes e tradicionais, mas as novas micro, pequenas e médias, que tem surgido de forma rápida e inovadora, têm se orientado e se beneficiando com decisões orientadas por dados (*data-driven*). Por conseguinte, tem-se exigido, por parte dos gestores, conhecimentos em tecnologias de dados (*data technologies*) e análise de negócios (*business analytics*) como competências centrais, sendo considerado um fator crítico de sucesso.

Jain (2010) apresenta 3 (três) objetivos principais no uso da análise de *clusters* de dados: (i) estrutura subjacente, ou seja, a obtenção de uma visão geral dos dados, para formulação de hipóteses, detecção de anomalias e identificação de características proeminentes; (ii) classificação natural através da identificação do grau de semelhança entre formas ou organismos (relação filogenética); e (iii) compressão através da organização dos dados com o objetivo de resumi-los através de protótipos de agrupamento.

Do ponto de vista metodológico os métodos de agrupamento podem ser divididos em 5 (cinco) categorias: agrupamento baseado em centroide (métodos de partição); agrupamento baseado em conectividade (ou agrupamentos hierárquicos); agrupamentos baseados em densidade; agrupamentos baseados em grade e agrupamento baseados em modelos (HAN; KAMBER; PEI, 2011). Essas metodologias e suas referências são apresentadas de forma sintética no Quadro 2.

**Quadro 2** – Métodos de *Clustering*.

<b>Método</b>	<b>Subtipos</b>	<b>Referências</b>
<b>Centroid-Based Clustering</b> (partitioning methods)	<i>K-means clustering</i>	(JAIN, 2010; KAUFMAN; ROUSSEEUW, 2005; LIKAS; VLASSIS; J. VERBEEK, 2003)
	<i>K-medoids clustering</i>	(KAUFMAN; ROUSSEEUW, 2005)
	<i>Affinity propagation clustering</i>	(MURPHY, 2012)
<b>Connectivity-Based Clustering</b> (hierarhical methods)	<i>Clustering Using REpresentatives (CURE)</i>	(LIKAS; VLASSIS; J. VERBEEK, 2003)
	<i>Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH)</i>	(HAN; KAMBER; PEI, 2011)
	<i>RObust Clustering using linKs (ROCK)</i>	(LIKAS; VLASSIS; J. VERBEEK, 2003)
	CHAMELEON	(HAN; KAMBER; PEI, 2011)
<b>Density-Based Clustering</b>	<i>Density-Based Spatial Clustering of Applications with Noise (DBSCAN)</i>	(ESTER et al., 1996)
	<i>Mean-shift</i>	(DERPANIS, 2005; WU; YANG, 2007)
	<i>Density-peaks-based clustering (DPC)</i>	(YU et al., 2019)
<b>Grid-Based Clustering</b>	-	(PARK; LEE, 2004)
<b>Model-Based Clustering</b>	-	(FRALEY; RAFTERY, 2000; HANDCOCK; RAFTERY; TANTRUM, 2007)

**Fonte:** Adaptado de Sun et al. (2017).

Sun et al. (2017) afirmam que, dentre os 3 (três) métodos mais populares, no contexto gerencial e de tomada de decisão, os agrupamentos são baseados em centroide, em conectividade e em densidade. Lopes e Gosling (2021) destacam também 4 (quatro) técnicas diferentes de agrupamento: *K-means*, PAM, DBSCAN e FCM.

O agrupamento dos dados tem sido umas das técnicas mais relevantes na formulação de grupos e no suporte à tomada de decisão gerencial. Contudo, há inúmeros desafios no uso das técnicas de agrupamento tradicionais frente ao novo ambiente organizacional que lida com dados dinâmicos, complexos, em grande escala, representativos, não convexos e de natureza consensual (SUN et al., 2017).

Jain e Dubes (1988) apresentam 9 (nove) desafios fundamentais associados ao



*clustering* que, segundo Jain (2010), ainda permanecem relevantes.

- (i) O que é um agrupamento?
- (ii) Quais características devem ser utilizadas?
- (iii) Os dados devem ser normalizados?
- (iv) Os dados contêm algum outliers?
- (v) Como definimos a similaridade de pares?
- (vi) Quantos *clusters* estão presentes nos dados?
- (vii) Qual método de agrupamento deve ser usado?
- (viii) Os dados têm alguma tendência de agregação?  
Os *clusters* descobertos e a partição são válidos?

### 2.3 APRENDIZADO DE MÁQUINA

Segundo Mohri, Rostamizadeh e Talwalkar (2012) o aprendizado de máquina ou Machine learning (ML) pode ser definido como uma família de métodos computacionais que usam a experiência para melhorar o desempenho ou fazer previsões mais precisas. Sendo que, as informações passadas para o sistema aprendiz são dados em formato eletrônico coletados para análise. De modo mais geral, as técnicas de ML são métodos orientados por dados que combinam conceitos fundamentais em ciência da computação com ideias de estatística, probabilidade e otimização.

Monard e Baranauskas (2003) definem ML como uma área da Inteligência Artificial (IA) que tem por objetivo desenvolver técnicas computacionais de aprendizado e construir sistemas com a capacidade de adquirir conhecimento de forma automatizada. Sistemas dessa natureza, podem tomar decisões tendo por base experiências obtidas na solução bem-sucedida de problemas anteriormente propostos. Tais metodologias, que inicialmente eram utilizadas em trabalhos teóricos, têm sido cada vez mais aplicadas dentro de organizações nas mais diversas áreas. Suas aplicações podem ser encontradas em trabalhos nas áreas de segmentação de clientes, na estimação de riscos, na previsão de demandas, na previsão de vendas, na análise de mercado, nos riscos de fraudes, na análise de crédito e na análise de perfis.

A indução é a forma de inferência lógica que permite obter conclusões genéricas sobre um conjunto particular de exemplos, ou seja, caracteriza-se como um raciocínio que parte do específico para o geral. Essas generalizações devem ser feitas com cautela, porque, se o número de exemplos não forem bem escolhidos ou forem insuficientes, as hipóteses obtidas podem não ter qualquer valor. Portanto, o aprendizado indutivo é



realizado partindo-se de um raciocínio sobre exemplos fornecidos por um processo externo ao sistema de aprendizado e pode ser dividido em supervisionado e não-supervisionado (DUDA; HART; STORK, 2001; MONARD; BARANAUSKAS, 2003).

Monard e Baranauskas (2003) advertem que após a determinação dos agrupamentos, normalmente, é necessária uma análise para determinar o que cada agrupamento significa no contexto do problema que está sendo analisado. Destacam ainda que, embora o ML seja uma ferramenta poderosa para aquisição automatizada de conhecimento, não existe um único algoritmo que apresente o melhor desempenho para todos os problemas. Ou seja, existe a necessidade de compreensão das possibilidades e limitações dos inúmeros algoritmos de ML disponíveis, aliando-os a uma metodologia que permita avaliar os conceitos induzidos por esses algoritmos quando da sua aplicação em determinado problema.

Existe uma grande variedade de problemas de aprendizagem, contudo, segundo Mohri, Rostamizadeh e Talwalkar (2012) esses podem ser agrupados em classes de problemas que diferem: nos tipos de dados de formação disponíveis para o aprendente; na ordem e no método pelo qual os dados de formação são recebidos; e nos dados de teste utilizados para avaliar o algoritmo de aprendizagem, sendo: (i) classificação, (ii) regressão, (iii) ranqueamento, (iv) agrupamento e (v) redução de dimensionalidade. O Quadro 3 apresenta uma síntese desses algoritmos segundo o tipo de aprendizagem e a natureza do problema a ser abordado.

**Quadro 3** – Síntese dos algoritmos de *ML* por paradigma de aprendizagem e natureza do problema.

Tipo de Aprendizagem	Problema(s)	Algoritmo(s)
Supervisionada	Classificação	<i>Logistic Regression</i>
Supervisionada	Classificação	<i>K-Nearest Neighbor</i>
Supervisionada	Classificação	<i>Naive Bayes</i>
Supervisionada	Classificação	<i>Decisionn Trees</i>
Supervisionada	Regressão	<i>Regression Trees</i>
Supervisionada	Regressão	<i>Linear Regression</i>
Supervisionada	Classificação/Regressão	<i>Neural Networks</i>
Supervisionada	Classificação/Regressão	<i>Support Vector Machines</i>
Supervisionada	Classificação/Regressão	<i>Random Forest</i>

Não Supervisionada	Redução de Dimensionalidade	<i>Principal Component Analysis</i>
Não Supervisionada	Deteção de Padrões	<i>Association Rules</i>
Não Supervisionada	Agrupamento	<i>K-means Clustering</i>
Não Supervisionada	Agrupamento	DBSCAN

**Fonte:** Adaptado de ENAP (2020) e de Mohri, Rostamizadeh e Talwalkar (2012).

### 3 PROCEDIMENTOS METODOLÓGICOS

Schwenk (1995) ao tratar dos estudos simulados em gestão aponta pontos positivos e negativos. Os pontos negativos geralmente dizem respeito ao fato de que decisões de estudantes, diferem muito das decisões executivas e de profissionais cujos resultados usualmente apresentam-se generalizados. A artificialidade do contexto simulado com simplicidade das tarefas e, por vezes, baixo nível de motivação dos envolvidos, diferem em muito a tomada de decisões em nível estratégico que é representada por um processo extremamente complexo com muitas variáveis.

No entanto, experimentos se justificam quando utilizados em uma população na qual se deseje produzir resultados como o aprendizado, dentre as principais razões são a acessibilidade e a causalidade (MARCONI; LAKATOS, 2003). Nesse sentido, classifica-se o estudo como descritivo com abordagem quali-quantitativa.

A acessibilidade refere-se as variáveis que, usualmente, não são acessíveis em pesquisas de campo e relacionadas às características dos processos de decisão e às formas como os fatores ambientais e organizacionais afetam os processos de decisão. Posto que, no “mundo real” geralmente é impossível a observação da tomada de decisão e a determinação de quais elementos do processo de decisão ou contexto causaram resultados específicos. Dessa forma, uma vez que possam ser identificados como possíveis agentes causais, é possível examiná-los em contextos controlados para avaliar a relação de causalidade entre os agentes e os resultados pretendidos.

Da extração dos dados e das informações do exercício simulado usado como referência nesse estudo, tem-se que o número de empresas, neste caso 8 (oito), eram constituídas por 4 (quatro) diretorias representando a gestão de pessoas, o comercial, a produção e as finanças. Dos relatórios disponíveis foram selecionadas as variáveis consideradas mais significativas para realização das análises. As variáveis são comuns à todas as empresas, num total de 62 (sessenta e duas) retiradas do Balanço Patrimonial

(BP), da Demonstração do Resultado do Exercício (DRE) e do contexto comercial de atuação apresentadas no Quadro 4.

**Quadro 4** – Variáveis utilizadas no modelo de análise.

Origem	Nº	Variáveis
Empresa	1	Nome de cada uma das 8 (oito) empresas
Tempo	1	Cada período de decisão refere-se a um trimestre no ambiente simulado, num total de 12 (doze) trimestres ou 3 anos de gestão.
Balanço Patrimonial	18	Caixa, Aplicação, Clientes, Estoque de produto acabado, Estoque matéria-prima A, Estoque matéria-prima B, Máquinas, Depreciação Acumulada das Máquinas, Fornecedores, Contas em Atraso, Participações a Pagar, Dividendos a Pagar, Empréstimos e Financiamento a Curto Prazo, Financiamento a Longo Prazo, Capital Social, Lucros Acumulados no Ano, Patrimônio Líquido, Rentabilidade do Ativo
Demonstração do Resultado do Exercício	7	Receita de Vendas, Custo do Produto Vendido (CPV), Despesas com Vendas, Despesas Administrativas, Resultado Financeiro, Outras Receitas e Despesas, Ebitda
Demonstração do Fluxo de Caixa	13	Recebimento de clientes, Pagamento de gastos operacionais, Imposto de renda e participações, Compra de máquinas, Venda de máquinas, Aplicação financeira líquida, Empréstimos e financiamentos obtidos, Amortização (de empréstimos e financiamentos), Juros bancários, Dividendos, Outras Atividades, Variação de Caixa no Período, Saldo inicial
Relatório de Mercado	23	% de participação de mercado em cada uma das 9 regiões, Preço de venda em cada uma das 9 regiões, Propaganda média, Qualidade do produto, Força da marca, Vendas acumuladas, Cotações das ações
Total	62	-

**Fonte:** Elaborada pelos autores (2021).

### 3.1 ANÁLISE DOS DADOS

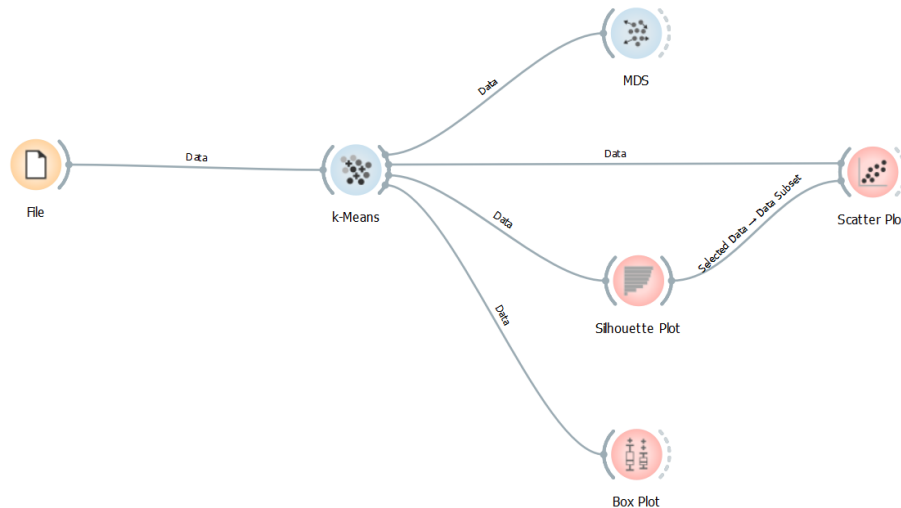
Todos os dados foram tabulados e organizados de forma a se aplicar a técnica de aprendizado não supervisionado e o método *K-means*, um dos mais utilizados para agrupamento e preferido para o reconhecimento de padrões em função da facilidade de implementação, simplicidade, eficiência e sucesso empírico (JAIN, 2010; KAUFMAN; ROUSSEEUW, 2005).

Na prática esse algoritmo encontra localmente soluções ótimas no que diz respeito ao erro de agrupamento. Seu método se baseia na colocação dos centros de agrupamento em posições arbitrárias e prossegue movendo-os a fim de minimizar o erro (LIKAS; VLASSIS; J. VERBEEK, 2003).

Sendo utilizado o software Orange na sua versão 3.28.0 para carregar, processar e visualizar os dados das empresas (DEMŠAR et al., 2013).

Na Figura 1 é apresentado o diagrama esquemático da análise realizada pelo software Orange.

**Figura 1** – Diagrama esquemático da análise realizada no software Orange.



**Fonte:** Elaborada pelos autores (2021).

Observa-se ainda na Figura 1 que cada um dos ícones se vale da nomenclatura utilizada na documentação do software Orange e conhecida como “*widget*”, ou seja, representa um atalho para os métodos disponíveis nesse aplicativo. O *widget* “*File*” lê o arquivo de dados de entrada e envia esse conjunto de dados para o seu canal de saída.

O *widget* “*K-means*” aplica o algoritmo de agrupamento *K-means* aos dados e gera um novo conjunto de dados que tem o índice de agrupamento calculado como atributo de classe.

O *widget* com a sigla “*MDS*” é “*Multidimensional Scaling*” ou escala multidimensional que projeta os itens sobre um plano ajustado a determinadas distâncias entre pontos. Portanto, essa técnica consiste em encontrar uma projeção de pontos de baixa dimensão, neste caso bidimensional, ajustando as distâncias entre os pontos da melhor maneira possível.

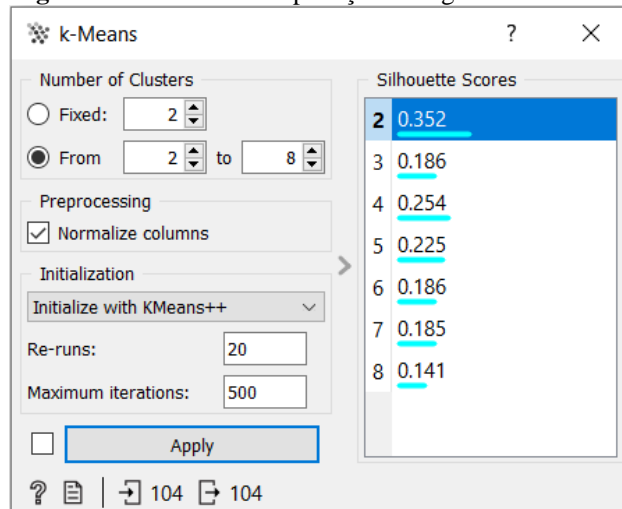
O *widget* “*Silhouette Plot*” constitui uma representação gráfica da consistência dentro de *clusters* de dados, ou seja, permite avaliar visualmente a qualidade de um *cluster*. E os *widgets* “*Box Plot*” e “*Scatter Plot*” são atalhos para gerar visualizações auxiliares na forma de *box plot* (que mostra as distribuições dos valores de um atributo) e de dispersão de dados (visualização bidimensional do gráfico de dispersão para atributos

contínuos) (DEMŠAR et al., 2013; WICKELMAIER, 2003).

#### 4 DISCUSSÃO E ANÁLISE DOS RESULTADOS

Da aplicação do método *K-means* obteve-se as seguintes pontuações de Silhouette para os diferentes números de *clusters* (de 2 a 8 testados) apresentadas na Figura 2.

**Figura 2** – Resultado da aplicação do algoritmo *K-means*.



**Fonte:** Elaborada pelos autores (2021).

Dos resultados obtidos a maior pontuação obtida (0,352) sugere-se a formação de 2 (dois) *clusters*. O *cluster* C1 (identificado em azul) teve sua formação iniciada a partir do período 6 (P06) com a empresa Atemporal. Na sequência, a partir dos P07 e P08, respectivamente, com a empresa DESBRAVE e a TCRB11. Ademais, são apresentadas as classificações de cada empresa em cada período no Quadro 5.

**Quadro 5 – Colocação das empresas em cada período.**

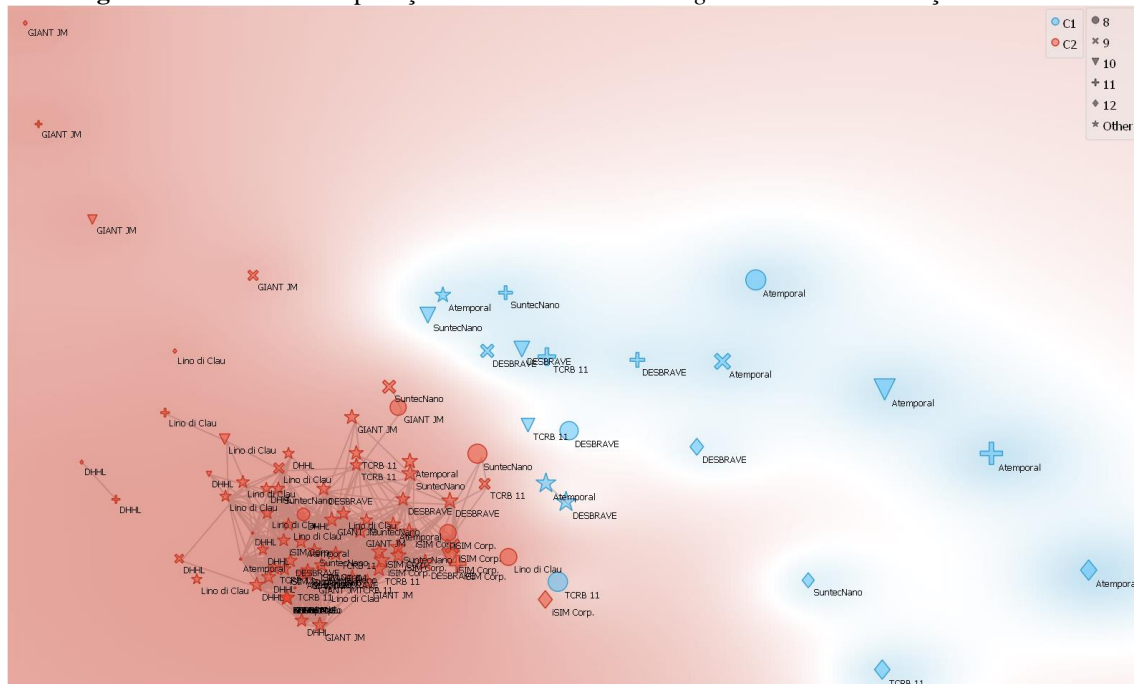
Empresa	P00	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12
iSIM Corp	1	8	4	6	3	1	3	2	4	2	2	2	4
TCRB11	1	2	5	2	2	3	5	6	3*	5	5*	4	2
Lino di Clau	1	4	8	7	7	8	8	8	7	7	6	6	6
Atemporal	1	6	7	5	5	4	1*	3*	1	1	1	1	1
DESBRAVE	1	3	2	4	1	5	2	1*	2	3*	3	3	3
Giant GM	1	1	1	1	6	2	4	5	6	6	7	8	8
SunTecNano	1	5	6	3	4	7	7	4	5	4	4*	5	5
DHHL	1	7	3	8	8	6	6	7	8	8	8	7	7

Fonte: Elaborada pelos autores (2021).

A partir dos dados do Quadro 5 percebe-se que existe uma diferenciação que não se relaciona com a obtenção das primeiras colocações (ranking).

Na Figura 3 tem-se uma representação bidimensional dos 2 (dois) *clusters* formados através da escala multidimensional (MDS), o tamanho dos pontos representa a rentabilidade do ativo de cada uma das empresas.

**Figura 3 – Resultado da aplicação do método de *clustering K-means* e visualização via MDS.**

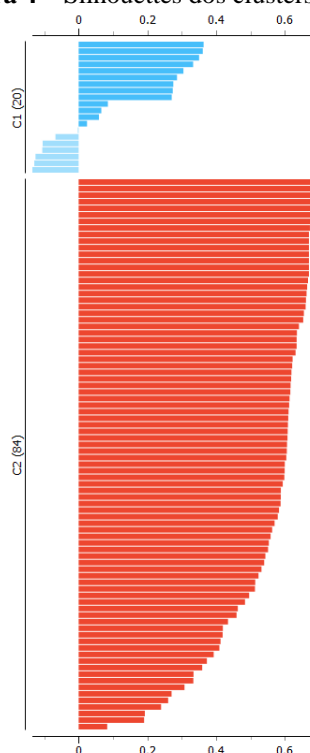


Fonte: Elaborada pelos autores (2021).

Uma medida da acurácia do modelo pode ser obtida através do gráfico Silhouette. A pontuação da silhueta é uma medida da semelhança de um elemento ao seu próprio *cluster* em comparação com outros *clusters*. Os valores próximos de 1 (um) indicam que a instância (ou ponto do conjunto) de dados está próxima do centro e aquelas que possuem valores próximos de 0 (zero) estão na fronteira entre 2 (dois) *clusters* (DEMŠAR et al., 2013; ROUSSEEUW, 1987).

A Figura 4 apresenta o gráfico Silhouette das instâncias (ou pontos) pertencentes a cada um dos *clusters* obtidos na análise.

**Figura 4** – Silhouettes dos clusters obtidos.



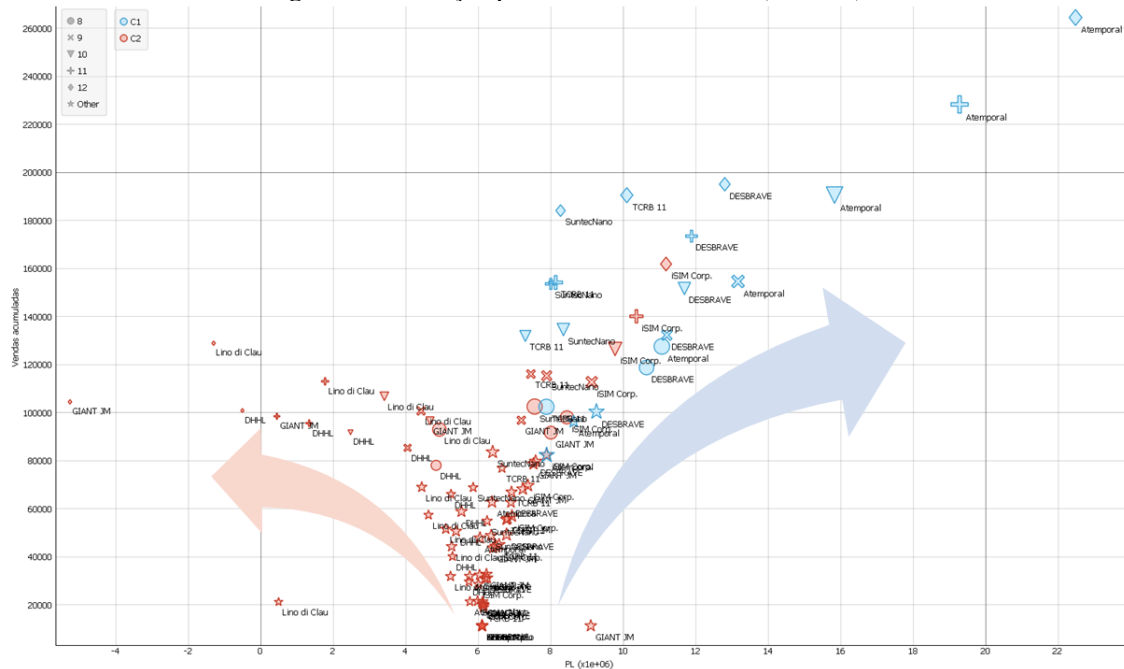
**Fonte:** Elaborada pelos autores (2021).

Verifica-se no *cluster* C1 a presença de algumas instâncias com valores abaixo de 0 (zero) indicando que essas não se ajustaram bem ao *cluster* sugerido. Contudo, este foi o agrupamento ( $K = 2$ ) com o melhor ajuste conforme já apresentado na Figura 3. As instâncias (ou pontos) com este parâmetro abaixo de 0 (zero) são os seguintes: Desbrave (P07), Desbrave (P09), Atemporal (P07), TCRB11 (P10), Atemporal (P06), TCRB11 (P08) e SunTecNano (P10). Estas 7 (sete) instâncias estão marcadas com \* no Quadro 4.

Dentre os elementos que podem ter levado a essa diferenciação temos a evolução do patrimônio líquido (PL) das empresas apresentado na Figura 5.



**Figura 5** – Evolução patrimonial dos *clusters* (C1 e C2).

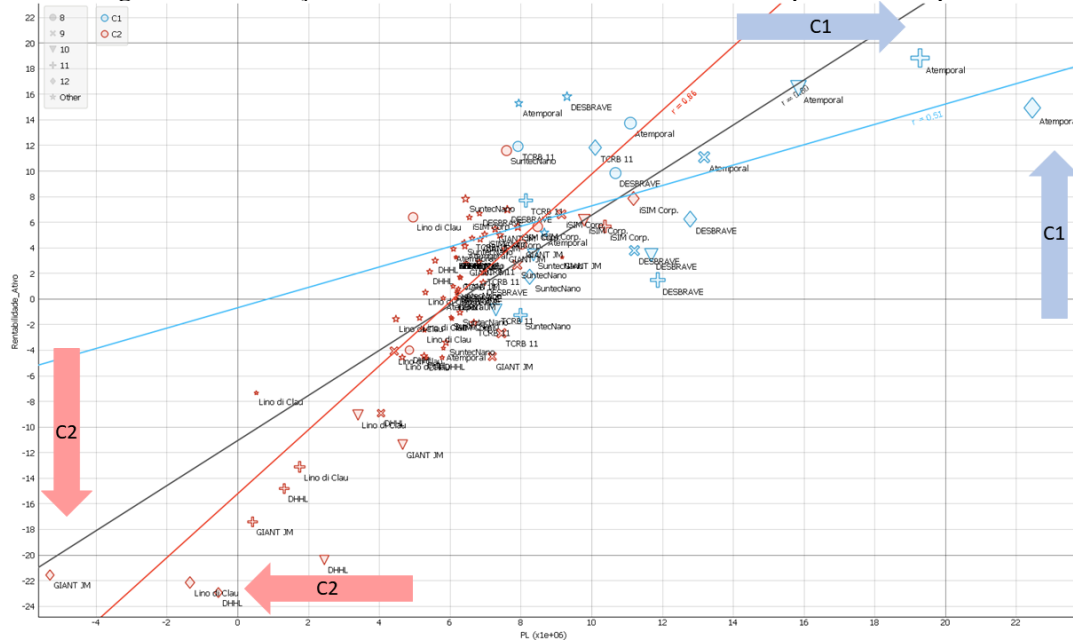


**Fonte:** Elaborada pelos autores (2021).

Na Figura 6 observam-se os sentidos opostos do patrimônio das empresas dos *clusters* (C1 e C2). Para o *cluster* C1 (em azul) há uma tendência de crescimento e tamanho dos pontos, indicando a rentabilidade do ativo seguindo a mesma tendência de crescimento. E para o grupo C2 (em vermelho) tem-se um comportamento inverso, ou seja, o PL decrescendo e a rentabilidade do ativo também cada vez menor. Observa-se como exceção em C2 a empresa iSIM Corp. apresentando um comportamento mais próximo às empresas de C1 mas sem diferenciar-se efetivamente do C2.

Além disso, avaliou-se a existência de correlação entre as variáveis Rentabilidade do Ativo e Patrimônio Líquido (PL) para todo o conjunto de dados e para cada um dos *clusters* individualmente. O resultado é apresentado na Figura 6.

**Figura 6** – Correlação entre as variáveis rentabilidade do ativo e patrimônio líquido.



Fonte: Elaborada pelos autores (2021).

O resultado da correlação de Pearson entre as variáveis PL e Rentabilidade do Ativo foi +0,801 para todos os dados. Ao aplicar essa correlação para cada um dos *clusters* obteve-se em C1 (+0,51) e em C2 (+0,86). Apesar do maior valor para o grupo C2, observa-se que em alguns períodos, essa correlação ocorre em sentido inverso, ou seja, na medida em que o tempo avança (a forma dos pontos indica o período) a rentabilidade e o PL diminuíram. Como exceção a esse comportamento dentro de C2 identificamos novamente a iSIM Corp.

## 5 CONCLUSÕES

Para Ketchen Jr e Shook (1996) e Sun et al. (2017) a análise de *clusters* constitui uma ferramenta útil na avaliação das relações entre estratégia, ambiente, tomada de decisão e desempenho. E em especial, quando o número de variáveis e dados a serem analisados apresentam grande volume. Contudo, é necessária cautela na interpretação e aprofundamento na análise dos resultados obtidos.

No estudo foi possível aplicar a análise de *clusters* e verificar que as diferenças nas decisões dos gestores das empresas industriais permitiu agrupá-las em 2 (dois) grupos distintos (C1 e C2). Dos resultados resultados obtidos, percebe-se uma identificação com os modelos teóricos de perfis de decisão de Eisenhardt e Zbaracki (1992) que destacam a importância das ações dos gestores na alocação dos recursos disponíveis, ajustando-as à execução da estratégia proposta. Observa-se ainda que os *players* (gestores) com o tempo

que dispunham para as decisões, foram levados a dois tipos distintos de comportamentos, sendo: o de assumirem riscos ou o de esquivarem-se deles.

Em tempo, como propostas de aprofundamento e amadurecimento desse estudo sugere-se:

- (i) a inclusão de dados de mais usuários do sistema de simulação;
- (ii) aplicação de métodos de validação interna e externa conforme proposto por Liu et al. (2010) e Sun et al. (2017);
- (iii) a avaliação das escalas das variáveis, de *outliers* e a aplicação de outras técnicas de agrupamento de modo a obter mais elementos na fundamentação e plausibilidade do modelo proposto conforme recomendam Lopes e Gosling (2021);
- (iv) a utilização de métodos dinâmicos de *clustering*, tais como *streaming data clustering* (AGGARWAL et al., 2003; GUHA et al., 2003), *clustering* evolutivo (CHAKRABARTI; KUMAR; TOMKINS, 2006), e *clustering* incremental (BAGIROV; UGON; WEBB, 2011; CHARIKAR et al., 1997) para avaliar como esse processo de agrupamento ocorre ao longo do tempo;
- (v) o aprofundamento da interpretação dos perfis de decisão segundo modelos teóricos para inferências mais consistentes.

## REFERÊNCIAS

- AMARAL, B. F. DO. Classificação semissupervisionada de séries temporais extraídas de imagens de satélite. São Carlos: Universidade de São Paulo, 18 nov. 2016.
- ANDREWS, K. R. The Concept of Corporate Strategy. [s.l: s.n.].
- ANSOFF, H. I. Corporate Strategy. New York: McGraw Hill, 1965.
- BURNS, T.; STALKER, G. M. The Management of Innovation. London: Oxford University Press, 1961.
- CHAPELLE, O.; SCHOLKOPF, B.; ZIEN, A. Semi-Supervised Learning. [s.l.] The MIT Press, 2006.
- DEMŠAR, J. et al. Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research, v. 14, p. 2349–2353, 2013.
- DERPANIS, K. G. Mean Shift Clustering Lecture Notes. [s.l: s.n.].
- DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern Classification. 2nd. ed. New York: Wiley, 2001.
- EISENHARDT, K. M.; ZBARACKI, M. J. Strategic decision making. Strategic Management Journal, v. 13, n. S2, p. 17–37, 1992.
- ELBANNA, S. Strategic decision-making: Process perspectives. International Journal of Management Reviews, v. 8, n. 1, p. 1–20, mar. 2006.
- ESCOLA NACIONAL DE ADMINISTRAÇÃO PÚBLICA - ENAP. Análise de Dados em Linguagem R Brasília ENAP, , 2020.
- ESTER, M. et al. Density-based spatial clustering of applications with noise. Knowledge Discovery and Data Mining. Anais...1996
- FRALEY, C.; RAFTERY, A. E. Model-Based Clustering Analysis and Density Estimation. n. 0, p. 68–70, 2000.
- HAMBRICK, D. C. ; MASON, P. A. . Upper Echelons: The Organization as a Reflection of Its Top Managers. Academy of Management Review, v. 9, n. 2, p. 193–206, abr. 1984.
- HAN, J.; KAMBER, M.; PEI, J. Data Mining: Concepts and Techniques. 3ed. ed. [s.l.] Elsevier, 2011.
- HANDCOCK, M. S.; RAFTERY, A. E.; TANTRUM, J. M. Model-based clustering for social networks. Journal of the Royal Statistical Society. Series A: Statistics in Society, v. 170, n. 2, p. 301–354, 2007.
- HOFER, C. W.; SCHENDEL, D. Strategy formulation: Analytical concepts. [s.l.] West Publ, 1980.

HOSKISSON, R. E. et al. Theory and research in strategic management: Swings of a pendulum. *Journal of Management*, v. 25, n. 3, p. 417–456, 1 jun. 1999.

JAIN, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, v. 31, n. 8, p. 651–666, 2010.

JAIN, A. K.; DUBES, R. C. *Algorithms for Clustering Data*. 1st. ed. New Jersey: Prentice Hall, 1988.

KAUFMAN, L.; ROUSSEEUW, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey: John Wiley & Sons, Inc., 2005.

KETCHEN JR., D. J.; SHOOK, C. L. THE APPLICATION OF CLUSTER ANALYSIS IN STRATEGIC MANAGEMENT RESEARCH: AN ANALYSIS AND CRITIQUE. *Strategic Management Journal*, v. 17, n. 6, p. 441–458, jun. 1996.

KOŚCIELNIAK, H.; PUTO, A. BIG DATA in Decision Making Processes of Enterprises. *Procedia Computer Science*, v. 65, n. Iccmit, p. 1052–1058, 2015.

LIKAS, A.; VLASSIS, N.; J. VERBEEK, J. The global K-means clustering algorithm. *Pattern Recognition*, v. 36, n. 2, p. 451–461, 2003.

LOPES, H. E. G.; GOSLING, M. DE S. Cluster Analysis in Practice: Dealing with Outliers in Managerial Research. *Revista de Administração Contemporânea*, v. 25, n. 1, p. 181–196, 2021.

MARCH, J. G.; SIMON, H. A. *Organizations*. 1st. ed. [s.l: s.n.].

MARCONI, M. DE A.; LAKATOS, E. M. *Fundamentos de Metodologia Científica*. 5. ed. São Paulo: Atlas, 2003.

MEYER, A. D.; TSUI, A. S.; HININGS, C. R. Configurational Approaches to Organizational Analysis. *Academy of Management Journal*, v. 36, n. 6, p. 1175–1195, 1993.

MILLER, D.; FRIESEN, P. H.; MINTZBERG, H. *Organizations: A quantum view*. [s.l.] Prentice Hall, 1984.

MINTZBERG, H. *The Structuring of Organizations*. [s.l.] Prentice Hall, 1979.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. *Foundations of Machine Learning*. Cambridge, Massachusetts: MIT Press, 2012.

MONARD, M. C.; BARANAUSKAS, J. A. Capítulo 4: Conceitos sobre Aprendizado de Máquina. In: *Sistemas inteligentes: Fundamentos e Aplicações*. São Paulo: Manole, 2003. p. 39–56.

MURPHY, K. P. *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts: MIT Press, 2012.

PAPADAKIS, V. M.; BARWISE, P. How much do CEOs and top managers matter in strategic decision-making? *British Journal of Management*, v. 13, n. 1, p. 83–95, 2002.

PARK, N. H.; LEE, W. S. Statistical grid-based clustering over data streams. *ACM SIGMOD Record*, v. 33, n. 1, p. 32–37, mar. 2004.

REGER, R. K.; HUFF, A. S. Strategic groups: A cognitive perspective. *Strategic Management Journal*, v. 14, n. 2, p. 103–123, 1993.

RIBEIRO, I. et al. RELAÇÕES TEÓRICAS E CONCEITUAIS EM TOMADA DE DECISÃO ESTRATÉGICA. *Revista Eletrônica de Estratégia & Negócios*, v. 9, n. 2, p. 58, 1 set. 2016.

SCHWENK, C. R. Strategic Decision Making. *Journal of Management*, v. 21, n. 3, p. 471–493, 1995.

SERRA, B. P. DE C.; SERRA, F. R.; TOMEI, P. A Pesquisa em Tomada de Decisão Estratégica no Alto Escalão: evolução e base intelectual do tema. *Revista de Ciências da Administração*, p. 11, 16 dez. 2014.

SUMMER, C. E. et al. Doctoral Education in the Field of Business Policy and Strategy. *Journal of Management*, v. 16, n. 2, p. 361–398, 1990.

SUN, L. et al. Cluster Analysis in Data-Driven Management and Decisions. *Journal of Management Science and Engineering*, v. 2, n. 4, p. 227–251, dez. 2017.

THOMPSON, J. D. *Organizations in Action: Social Science Bases of Administrative Theory*. New York: McGraw Hill, 1967.

VENKATRAMAN, N.; RAMANUJAM, V. Measurement of Business Performance in Strategy Research: A Comparison of Approaches. *Academy of Management Review*, v. 11, n. 4, p. 801–814, out. 1986.

WEBER, M. *The Theory of Social and Economic Organization*. [s.l.] The Free Press, 1947.

WICKELMAIER, F. *An Introduction to MDS*. [s.l.] Aalborg Universitetsforlag, 2003.

WU, K.-L.; YANG, M.-S. Mean shift-based clustering. *Pattern Recognition*, v. 40, n. 11, p. 3035–3052, nov. 2007.

YU, D. et al. Density Peaks Clustering Based on Weighted Local Density Sequence and Nearest Neighbor Assignment. *IEEE Access*, v. 7, p. 34301–34317, 2019.

ZHU, X.; GOLDBERG, A. B. *Introduction to Semi-Supervised Learning*. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, v. 3, n. 1, p. 1–130, jan. 2009.