

**idp**

v.5 n. Único

**97**

# DEBATES EM ECONOMIA APLICADA

## WORKING PAPER

**TRANSFORMAÇÃO DIGITAL: USO DE  
CLUSTERIZAÇÃO NA FORMULAÇÃO DE  
POLÍTICAS DE AUMENTO DA MATURIDADE  
DIGITAL DE EMPRESAS BRASILEIRAS**

**KAREN CRISTINA LEAL DA SILVA**

# TRANSFORMAÇÃO DIGITAL: USO DE CLUSTERIZAÇÃO NA FORMULAÇÃO DE POLÍTICAS DE AUMENTO DA MATURIDADE DIGITAL DE EMPRESAS BRASILEIRAS

**KAREN CRISTINA LEAL DA SILVA<sup>a</sup>**

---

<sup>a</sup>Karen Cristina Leal da Silva é Mestre em Economia pelo Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP). E-mail: [kcilogti@gmail.com](mailto:kcilogti@gmail.com). ORCID: 0009-0002-8004-8639.



## IDP

O IDP é um centro de excelência no ensino, na pesquisa e na extensão nas áreas da Administração Pública, Direito e Economia. O Instituto tem como um de seus objetivos centrais a profusão e difusão do conhecimento de assuntos estratégicos nas áreas em que atua, constituindo-se um think tank independente que visa contribuir para as transformações sociais, políticas e econômicas do Brasil.

### DIREÇÃO E COORDENAÇÃO

#### **Diretor Geral**

Francisco Schertel

#### **Coordenador do Mestrado em Economia**

José Luiz Rossi

### CONSELHO EDITORIAL

#### **Coordenação**

Thiago Caldeira

Emmanuel Brasil

#### **Supervisão e Revisão**

Mathias Tessmann

Lucas Dutra

#### **Projeto Gráfico e Diagramação**

Juliana Vasconcelos

**[www.idp.edu.br](http://www.idp.edu.br)**

Revista Técnica voltada à divulgação de resultados preliminares de estudos e pesquisas aplicados em desenvolvimento por professores, pesquisadores e estudantes de pós-graduação com o objetivo de estimular a produção e a discussão de conhecimentos

## DEBATES EM ECONOMIA APLICADA

técnicos relevantes na área de Economia.

Convidamos a comunidade acadêmica e profissional a enviar comentários e críticas aos autores, visando o aprimoramento dos trabalhos para futura publicação. Por seu propósito se concentrar na recepção de comentários e críticas, a Revista Debates em Economia Aplicada não possui ISSN e não fere o ineditismo dos trabalhos divulgados.

As publicações da Revista estão disponíveis para acesso e download gratuito no formato PDF. Acesse: [www.idp.edu.br](http://www.idp.edu.br)

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do IDP.

Qualquer citação aos trabalhos da Série só é permitida mediante autorização expressa do(s) autor(es).

# debates em economia aplicada

## SUMÁRIO

1	Introdução .....	6
2	Fundamentação Teórica .....	7
2.1	Transformação Digital e Maturidade Digital .....	7
2.2	Clusterização .....	8
3	Metodologia .....	9
3.1	Algoritmos de Clusterização .....	9
3.2	Dados .....	10
4	Resultados .....	11
4.1	Aplicação da Clusterização nos Dados .....	11
4.2	Análise Sob a Perspectiva da Qualidade dos Agrupamentos .....	14
4.3	Análise Sob a Perspectiva das Médias .....	16
5	Considerações Finais .....	21
	Referências .....	22

**RESUMO:** A transformação digital é estratégica para fortalecer a competitividade do setor produtivo brasileiro, mas a heterogeneidade em porte, setor e localização das empresas dificulta políticas públicas padronizadas. Este estudo utiliza técnicas de clusterização para segmentar empresas conforme sua maturidade digital, com base em dados da 6ª edição do Índice de Maturidade Digital (IMD) da Agência Brasileira de Desenvolvimento Industrial (ABDI), aplicado a 2.000 empresas. Foram testados os métodos hierárquico, k-means, k-modes e PAM, com avaliação pelos índices Calinski-Harabasz, Davies-Bouldin, Silhueta e ANOVA. O k-means apresentou os melhores resultados (CH = 630,96; DBI = 1,53; Silhueta = 0,23), demonstrando agrupamentos robustos. A classificação original do IMD mostrou desempenho inferior, mas útil para análises complementares. Porte, setor e região influenciam fortemente os clusters, revelando padrões não capturados previamente. Conclui-se que a clusterização, especialmente via k-means, é uma abordagem robusta para orientar políticas públicas mais precisas, modulares e adaptadas à diversidade do tecido empresarial brasileiro.

**PALAVRAS-CHAVE:** transformação digital, maturidade digital, políticas públicas, clusterização, segmentação empresarial

**ABSTRACT:** Digital transformation is a strategic process to strengthen the competitiveness of Brazil's productive sector, but the heterogeneity in company size, industry, and regional location makes it difficult to implement standardized public policies. This study applies clustering techniques to segment companies according to their digital maturity, based on data from the 6th edition of the Digital Maturity Index (IMD) by the Brazilian Agency for Industrial Development (ABDI), conducted with 2,000 companies. The hierarchical, k-means, k-modes, and PAM methods were tested, with quality assessed using the Calinski-Harabasz, Davies-Bouldin, Silhouette, and ANOVA indices. The k-means algorithm yielded the best results (CH = 630.96; DBI = 1.53; Silhouette = 0.23), indicating robust groupings. The original IMD classification performed less effectively but remains useful for complementary analyses. Company size, industry, and region strongly influence cluster formation, revealing patterns not previously captured. It is concluded that clustering, especially using the k-means algorithm, is a robust approach to guide more precise, modular, and tailored public policies suited to the diversity of the Brazilian business landscape.

**KEYWORDS:** digital transformation, digital maturity, public policies, clustering, business segmentation

**RESUMEN:** La transformación digital es estratégica para fortalecer la competitividad del sector productivo brasileño, pero la heterogeneidad en tamaño, sector y ubicación de las empresas dificulta la formulación de políticas públicas estandarizadas. Este estudio utiliza técnicas de agrupamiento para segmentar empresas según su nivel de madurez digital, con base en los datos de la 6ª edición del Índice de Madurez Digital (IMD) de la Agencia Brasileña de Desarrollo Industrial (ABDI), aplicado a 2.000 empresas. Se probaron los métodos jerárquico, k-means, k-modes y PAM, evaluados mediante los índices de Calinski-Harabasz, Davies-Bouldin, Silueta y ANOVA. El k-means presentó los mejores resultados (CH = 630,96; DBI = 1,53; Silueta = 0,23), demostrando agrupamientos robustos. La clasificación original del IMD mostró un

rendimiento inferior, aunque útil para análisis complementarios. El tamaño, el sector y la región influyen fuertemente en la formación de los clusters, revelando patrones no captados anteriormente. Se concluye que el agrupamiento, especialmente mediante k-means, es un enfoque sólido para orientar políticas públicas más precisas, modulares y adaptadas a la diversidad del tejido empresarial brasileño.

**PALABRAS CLAVE:** transformación digital, madurez digital, políticas públicas, clusterización, segmentación empresarial

**CLASSIFICAÇÃO JEL:** O33; C38; L52

## 1 Introdução

Segundo pesquisa sobre o uso de Tecnologia da Informação e Comunicação - TIC, realizada em 2021 pela FGV<sup>1</sup>, a pandemia de Covid-19 acelerou, em poucos meses, avanços que normalmente levariam anos. Esse movimento acelerou a Transformação Digital, tornando o digital essencial à sobrevivência das empresas e redefinindo mercado e relação com os clientes.

As estimativas do tamanho da economia digital variam conforme os critérios adotados. Dados da União Europeia mostram que, em 2019, a indústria digital representava 8,2% do PIB do bloco (Van Ark et al., 2019). A relevância do setor cresce rapidamente, levando países a estipular metas crescentes de participação no PIB (Caldas, 2022). Segundo o G20 (G20, 2016), a economia digital abrange atividades baseadas em dados e conhecimento digitalizado, uso de redes modernas e TICs como motor de produtividade e crescimento. Apesar do nome, ela não se separa da economia tradicional: o virtual influencia o físico e vice-versa, moldando relações sociais, políticas e econômicas, o que alguns chamam de “transformação figital” (Meira, 2021).<sup>2</sup>

Segundo a Agência Brasileira de Desenvolvimento Industrial – ABDI<sup>3</sup> (ABDI, 2019), a economia digital envolve governo, setor produtivo e sociedade. O governo e o setor produtivo atuam como impulsionadores econômicos, enquanto a sociedade é a principal beneficiária. O estado promove a transformação digital, capacita cidadãos, garante direitos e melhora a gestão pública. Já o setor produtivo, ao digitalizar processos e adaptar-se ao ambiente digital, ganha competitividade e impulsiona o crescimento econômico.

A rápida evolução tecnológica torna a maturidade digital essencial. Entender seus fatores nas empresas brasileiras ajuda a orientar estratégias e políticas públicas mais eficazes. Além disso, a transformação digital contribui para ganhos de produtividade, que, no médio e longo prazo, tendem a gerar maior bem-estar social. No Brasil, a forte heterogeneidade entre e dentro dos setores industriais reforça a necessidade de políticas segmentadas para promover competitividade, aprendizado e inovação, reduzindo desigualdades estruturais (Cartela e Porcile, 2015).

Este estudo parte do problema de segmentar empresas segundo seus níveis de maturidade digital, com o objetivo de subsidiar políticas públicas mais eficazes. Assim, o estudo partiu da seguinte pergunta de pesquisa: o uso de clusterização na formulação de políticas, aumenta a maturidade digital de empresas brasileiras? Buscando as respostas, foram aplica-

<sup>1</sup>A pesquisa anual da FGV integra o Fórum de Informações sobre o uso de TIC, com o objetivo de analisar a situação das empresas brasileiras nesse tema. Disponível em: <https://shorturl.at/OEt0A>.

<sup>2</sup>Meira (Meira, 2021) usa o termo “figital” para explicar que a economia digital não anula a tradicional (considerando os aspectos físicos) e que, na visão do autor, o termo economia figital seria mais adequado do que economia digital.

<sup>3</sup>A ABDI é um Serviço Social Autônomo criado pela Lei nº 11.080/2004 para executar políticas de desenvolvimento industrial voltadas à geração de empregos, alinhadas ao comércio exterior e à ciência e tecnologia (<https://www.abdi.com.br>).

das diferentes metodologias de clusterização, capazes de identificar padrões entre empresas com características semelhantes, superando classificações fixas ou lineares. Os resultados evidenciam perfis distintos de maturidade digital, permitindo uma tipologia mais precisa das empresas brasileiras. O estudo contribui ao propor uma abordagem empírica baseada em aprendizado não supervisionado e oferece subsídios relevantes para políticas públicas voltadas à transformação digital segmentada e eficiente.

O artigo está estruturado da seguinte forma: a seção 2 aborda a literatura sobre transformação digital e segmentação; a seção 3 apresenta os dados e a metodologia; a seção 4 discute os resultados empíricos; e a seção 5 traz as implicações e recomendações de política pública.

## 2 Fundamentação Teórica

O Brasil tem estruturado políticas públicas para fortalecer a transformação digital. Destacam-se: i) a Estratégia Brasileira para a Transformação Digital - E-Digital, revista em 2022, que busca inclusão tecnológica, inovação e desenvolvimento de habilidades digitais e; ii) a Nova Política Industrial, chamada de Nova Indústria Brasil - NIB, lançada em 2024. Essas iniciativas mostram que a transformação digital é um instrumento central de política pública na era digital.

Existem diversas definições para conceituar a transformação digital. Segundo a ABDI trata-se de estratégia empresarial baseada na integração de tecnologias digitais para modificar operações, criar oportunidades e agregar valor ao cliente [ABDI \(2019\)](#). Para [Vial \(2019\)](#), trata-se de um processo que promove mudanças significativas nas organizações por meio das TICs. [Matarazzo et al. \(2020\)](#) destacam seu impacto nas rotinas e capacidades organizacionais, enquanto [Pagani e Pardo \(2017\)](#) a associam à criação de modelos de negócios inovadores e estratégicos. A incorporação de tecnologias digitais nas empresas exige mudanças profundas na cultura organizacional, nas operações e na gestão, com foco em vantagem competitiva ([Hassani e El Idrissi, 2020](#); [Sampaio, 2018](#)). Para ser efetiva, a transformação digital deve estar alinhada aos objetivos estratégicos da organização ([Hinings et al., 2018](#)).

### 2.1 Transformação Digital e Maturidade Digital

A transformação digital e a maturidade digital são temas recentes nos estudos empíricos, com pouca uniformidade conceitual ou metodológica sobre como as empresas podem acelerar sua digitalização e elevar sua maturidade digital ([Kraus et al., 2022](#)). O conceito abrange desde a adoção de tecnologias em processos analógicos até mudanças estruturais no modelo de negócio ([Kraus et al., 2021](#)).

A conceituação de transformação digital varia entre autores, mas três conceitos fundamentais se inter-relacionam: digitização, digitalização e transformação digital. Segundo

Eller et al. [Eller et al. \(2020\)](#), as empresas passam por três fases rumo à economia digital. A primeira, digitização, é técnica e consiste na conversão de processos analógicos em digitais, gerando dados para os sistemas corporativos ([Verhoef et al., 2021](#)). A segunda, digitalização, é sociotécnica: os gestores interpretam os dados gerados para ajustar processos e explorar novas possibilidades com tecnologias digitais ([Eller et al., 2020](#); [Verhoef et al., 2021](#)). Por fim, a transformação digital representa uma mudança profunda nas bases da competição, integrando tecnologia à estratégia organizacional para criar, entregar e capturar valor de forma inovadora ([Rohn et al., 2021](#); [Verhoef et al., 2021](#)). Como destaca Rogers [Rogers \(2016\)](#), trata-se sobretudo de uma mudança estratégica e não apenas tecnológica.

Segundo o Mapa da Digitalização das MPEs Brasileiras [ABDI \(2023\)](#), as empresas se encontram em diferentes estágios da transformação digital, não sendo adequado tratá-las de forma uniforme. Enquanto algumas estão iniciando a digitização, outras já reformulam seus modelos de negócio com base em tecnologias digitais. Por isso, é essencial identificar o nível de maturidade digital de cada empresa para propor soluções mais adequadas à sua realidade e promover estratégias compatíveis com seu estágio de evolução. Como as empresas se encontram em diferentes estágios desse processo, é inadequado tratá-las de forma uniforme. Quando olhamos para os modelos de maturidade digital, observamos que têm ganhado destaque por combinarem boas práticas na avaliação do comportamento organizacional. Esses modelos baseiam-se no uso das tecnologias digitais para otimizar a transformação digital, identificando fraquezas e potencialidades por meio de métodos qualitativos e quantitativos ([Kirmizi e Kocaoglu, 2022](#)).

Apesar da centralidade da tecnologia, [Dutta et al. \(2021\)](#) destacam que atingir a maturidade digital exige foco em pessoas, práticas e organização do trabalho, pois a tecnologia é meio, não fim. Para [Lobão e Zilli \(2020\)](#), maturidade digital exige alinhar estratégia, cultura, estrutura e equipe às demandas digitais. [Saldanha \(2019\)](#) a define como o objetivo de integrar o “DNA digital” ao negócio. [Gollhardt et al. \(2020\)](#) e [Gökalp & Martinez \(2021\)](#) ressaltam que a transformação digital visa agregar valor ao negócio, melhorando o desempenho organizacional por meio da otimização de processos, aumento da produtividade e criação de novos mercados com base na informação.

## 2.2 Clusterização

Algoritmos de aprendizado de máquina são geralmente classificados em dois tipos: supervisionados, nos quais modelos aprendem a partir de dados rotulados com respostas corretas ([Aggarwal, 2015](#); [Alpaydin, 2016](#); [Grus, 2015](#)), e não supervisionados, onde não há rótulos e o modelo identifica padrões estruturais nos dados por conta própria ([Aggarwal, 2015](#); [Jordan e Mitchell, 2015](#)). A classificação está associada ao aprendizado supervisionado, enquanto a clusterização é típica do aprendizado não supervisionado.

Como o foco deste trabalho é a clusterização, não serão aprofundados os conceitos relacionados à classificação. A clusterização é amplamente utilizada na ciência de dados para

segmentar bases de dados (Deng et al., 2016). Consiste em agrupar objetos semelhantes entre si e distintos de outros grupos, com base em uma métrica de similaridade (Rodrigues et al., 2017).

De Morsier et al. De Morsier et al. (2015) destacam que, sem dados semânticos, agrupar é desafiador. A clusterização revela padrões ocultos (Jayaram Reddy et al., 2018) e é vista como “reconhecimento de padrões” (Kassambara, 2017). Já Omrani, Shafaat e Eemrouznejad Omrani et al. (2018) destacam que a clusterização são conjuntos de objetos similares entre si e distintos dos demais. A clusterização pode ser realizada por diferentes algoritmos, cada um com critérios próprios de formação dos grupos (Liu et al., 2016; Rodrigues et al., 2017), oferecendo insights sobre os dados que métodos supervisionados nem sempre capturam.

Goldschmidt, Passos e Bezerra Goldschmidt et al. (2015) classificam os métodos de clusterização em três famílias: baseados em distâncias, em distribuição de probabilidades e em densidade. Também os categorizam conforme a forma dos agrupamentos: algoritmos partitivos, que produzem partições fixas, e hierárquicos, que constroem estruturas em árvore. Os métodos utilizados neste estudo, hierárquico, k-means, k-modes e PAM, enquadram-se majoritariamente na família baseada em distâncias, utilizando métricas como distâncias euclidiana ou de Manhattan. K-means, k-modes e PAM são algoritmos partitivos, pois formam k grupos otimizando medidas como média, moda ou medoide. Já o método hierárquico constrói um dendrograma, sem exigir a definição prévia do número de clusters.

## 3 Metodologia

Diante da crescente importância da digitalização para a competitividade dos negócios, este estudo tem como objetivo investigar o uso de técnicas de clusterização como ferramenta de apoio à formulação de políticas públicas voltadas à transformação digital. Parte-se do reconhecimento da heterogeneidade estrutural e comportamental das empresas brasileiras, o que dificulta a adoção de políticas uniformes.

Utilizando dados da 6ª edição do IMD ABDI, desenvolvida pela Agência Brasileira de Desenvolvimento Industrial (ABDI), foram aplicados algoritmos de aprendizado não supervisionado para identificar perfis empresariais com características semelhantes, sem recorrer a classificações rígidas. Para a análise, adota-se o delineamento de pesquisa experimental, que, segundo Gil (2010), é amplamente valorizado na ciência por permitir a definição clara do objeto de estudo, a seleção das variáveis relevantes e o controle e observação sistemática de seus efeitos.

### 3.1 Algoritmos de Clusterização

Este trabalho utiliza quatro algoritmos de clusterização: Hierárquico, k-means, k-modes e Partitioning Around Medoids (PAM), comumente aplicados para identificação de padrões. A

escolha por empregar múltiplos métodos de clusterização se justifica pela necessidade de comparar o desempenho de diferentes abordagens diante da natureza dos dados, majoritariamente numéricos, bem como de testar a robustez dos resultados obtidos por cada técnica. Essa abordagem permite identificar o algoritmo mais adequado para a segmentação das empresas segundo seu nível de maturidade digital.

A escolha do algoritmo depende, em grande parte, do tipo de dado: k-means e PAM são mais adequados para dados numéricos, enquanto k-modes é indicado para dados categóricos. O método hierárquico pode ser aplicado a ambos, dependendo da métrica de distância utilizada. Assim, a adoção de diferentes algoritmos neste estudo acomoda a complexidade dos dados e amplia a capacidade explicativa dos agrupamentos. O algoritmo hierárquico, aqui utilizado em sua versão aglomerativa, constrói uma estrutura em árvore, agrupando objetos com base em suas distâncias e fundindo-os sequencialmente até restar um único cluster (Rencher, 2002; Hair et al., 2010). Sua principal vantagem é não exigir definição prévia do número de grupos, permitindo análise exploratória mais flexível em contextos de alta heterogeneidade, como o das empresas brasileiras.

Os métodos não hierárquicos, como k-means, k-modes e PAM, por sua vez, requerem a definição prévia do número de clusters (Corrar et al., 2009). O k-means, amplamente utilizado por sua simplicidade e eficiência, utiliza a média dos objetos como centro de cada grupo (Jain, Murty e Flynn, 1999; Fontana e Naldi, 2009). O k-modes é uma adaptação voltada a dados categóricos, substituindo a média pela moda como medida de centralidade. Já o PAM utiliza medoides, objetos reais mais centrais no grupo, como representantes dos clusters, o que o torna mais robusto à presença de outliers (Kassambara, 2017; Maechler et al., 2017), alternando fases de construção e otimização para aprimorar a qualidade dos agrupamentos. O algoritmo hierárquico constrói uma estrutura de clusters em forma de árvore (dendrograma), podendo ser aglomerativo (bottom-up) ou divisivo (top-down). No método aglomerativo, os objetos são agrupados com base em suas distâncias, geralmente euclidianas, e a fusão ocorre até restar um único cluster (Rencher, 2002; Hair et al., 2010). A vantagem é que não exige número pré-definido de grupos.

## 3.2 Dados

Os dados utilizados neste estudo são oriundos do IMD ABDI, pesquisa iniciada em 2021 com seis ciclos semestrais até 2023. A amostra contou com 2.000 executivos de empresas brasileiras. Este estudo utiliza apenas os dados da 6ª e última rodada da referida pesquisa (segundo semestre de 2023).

A amostragem utilizada no IMD ABDI foi nacional e setorial, cobrindo empresas de todos os portes (micro, pequenas, médias e grandes) e cinco setores econômicos (agropecuária, comércio, construção civil, indústria e serviços), distribuídas pelas 27 unidades da federação. Para o plano amostral, foram utilizadas duas bases: (i) o cadastro da ABDI, com 1.261.338 empresas ativas; e (ii) a base da Receita Federal, com 9.417.542 registros, usada apenas para excluir Microempreendedores Individuais (MEIs), fora do escopo do estudo. A amostra foi

aleatória simples, proporcional ao setor, porte e unidade federativa, assegurando equilíbrio entre população e amostra, com estrutura voltada à representatividade nacional e à comparação entre segmentos do setor produtivo.

Após a coleta das informações, a ABDI atribuiu uma pontuação às respostas (0 a 100), permitindo a classificação das empresas em quatro faixas de maturidade digital, conforme critérios definidos pela ABDI.

- 0 a 25 – Empresas iniciantes quanto a maturidade digital;
- 26 a 50 – Empresas emergentes quanto a maturidade digital;
- 51 a 75 – Empresas intermediárias quanto a maturidade digital; e
- 76 a 100 – Empresas líderes quanto a maturidade digital.

Para cada uma das cinco dimensões do questionário, a pontuação foi calculada como a média simples de quatro indicadores específicos, atribuindo o mesmo peso a cada item. O índice agregado de maturidade digital foi obtido pela média simples entre as cinco dimensões, com pesos iguais. Esse índice composto foi utilizado apenas nas análises que exigem uma visão geral da maturidade digital, como disposto na Tabela 4, que compara setores e portes empresariais. Nas demais análises, os resultados foram discutidos separadamente por dimensão, preservando a especificidade de cada eixo e permitindo uma análise mais detalhada dos padrões de maturidade digital observados.

$$\text{Score}_{i,f} = \left( \frac{\text{Value}_{i,f} - \text{Wp}_i}{\text{Frontier}_i - \text{Wp}_i} \right) \cdot 100 \quad (1)$$

Sendo:

- *Score*: Pontuação do indicador *i* para a firma *f*. É sempre limitado entre 0 e 100;
- *Value*: Valor da resposta no indicador *i* para a firma *f*;
- *Wp*: Pior valor possível de resposta no indicador *i*, ou menor valor aceitável;
- *Frontier*: Melhor valor possível do indicador *i*. Pode ser determinada com base no valor máximo possível para a resposta, ou o melhor valor observado na amostra.

## 4 Resultados

### 4.1 Aplicação da Clusterização nos Dados

Para a avaliação dos resultados, os algoritmos de clusterização foram executados na linguagem de programação *R*.<sup>4</sup> Os dados são:

- Categóricos: porte da empresa, setor, estado e região;

<sup>4</sup>*R* é uma linguagem de código aberto voltada à análise estatística, criada nos anos 1990 por Ross Ihaka e Robert Gentleman, amplamente usada por estatísticos e cientistas de dados.

- Numéricos: gestão, pessoas, processos, modelo de negócios e vendas.

A definição do número de *clusters* para cada algoritmo levou em consideração o resultado obtido com os seguintes métodos: cotovelo (*Elbow Method*),<sup>5</sup> silhueta (*Silhouette Method*)<sup>6</sup> e Método da Estatística Gap (*GAP Statistic Method*)<sup>7</sup>. O método do cotovelo sugeriu entre  $k = 2$  e/ou  $k = 3$ ; o método da silhueta sugeriu  $k = 2$ ; e o método GAP, executado apenas com as variáveis numéricas, sugeriu  $k = 4$ . De forma a estabelecer uma padronização que possibilitasse comparação entre os algoritmos, foi adotado  $k = 3$ . Dado o resultado positivo do *k-means*, também foi explorado com  $k = 4$  e  $k = 5$ , entretanto o  $k = 3$  nos testes de qualidade foi o que apresentou melhores resultados. A Tabela 1 apresenta a quantidade de empresas alocadas em cada *cluster*, conforme os quatro algoritmos de clusterização aplicados no estudo. Cada modelo foi configurado para gerar três *clusters*, totalizando 2.000 observações por algoritmo.

**Tabela 1:** Distribuição dos dados entre os algoritmos observados

Número de <i>Clusters</i>	Hierárquico (HC)	k-Means	k-Modes	PAM
1	817	837	452	804
2	596	369	603	371
3	587	794	945	825
Total	2000	2000	2000	2000

Fonte: Dados elaborados pela autora.

A comparação mostra que o *Cluster 3* é consistente nos métodos *Means*, *Modes* e PAM, enquanto o *Cluster 2*, de baixa representatividade, indica um nicho específico. No HC, o *Cluster 1* predomina; já nos métodos centrados (*Means* e PAM), o *Cluster 3* domina, sugerindo grupos mais homogêneos. O *Means* gera um *Cluster 2* muito pequeno, e o *Modes*, por usar a moda, mostra maior flexibilidade na identificação. Assim sendo, o *Cluster 3* pode ser composto por empresas que compartilham características muito comuns e amplamente representadas, como porte ou setor predominante. O *Cluster 2*, menor e mais específico, pode representar nichos distintos ou *outliers* nos dados. A dispersão dos dados para cada algoritmo de clusterização pode ser observada na Figura 1:

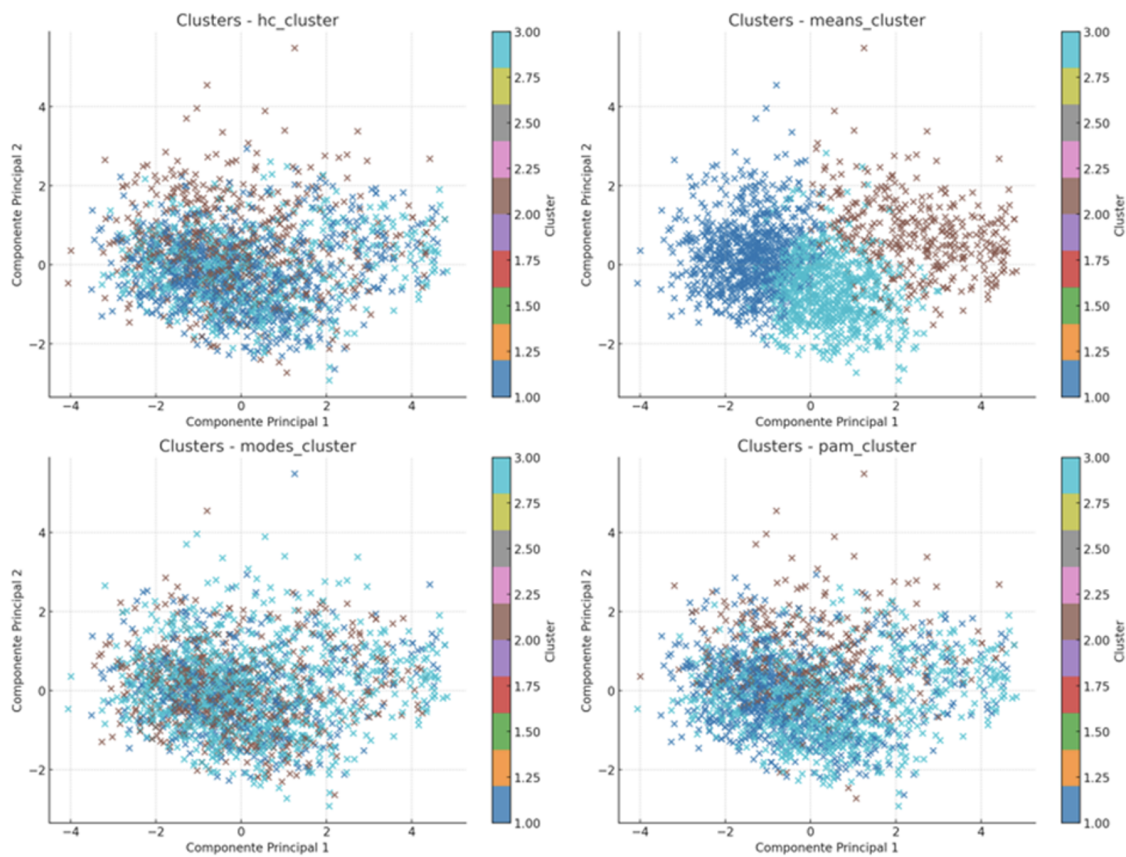
Cada ponto representa uma empresa por *cluster*. No HC, há boa distribuição com sobreposição; o *Means* forma *clusters* compactos ao minimizar a variância; o *Modes* é mais disperso e flexível; e o PAM, semelhante ao *Modes*, é mais robusto a *outliers* por usar medoides. Em geral, os algoritmos *Modes* e HC mostram maior sobreposição, enquanto *Means* e PAM formam *clusters* mais distintos e visualmente coesos. Segue a Figura 2, considerando a predominância entre os setores econômicos nos algoritmos:

<sup>5</sup>O método do Cotovelo tem origem na análise da soma dos erros quadráticos dentro dos *clusters*. Seu conceito é frequentemente associado a trabalhos iniciais de Thorndike (1953) e Kodinariya e Makwana (2013).

<sup>6</sup>O método Silhueta, proposto por Rousseeuw (1987), avalia o quão bem cada ponto está agrupado, comparando sua proximidade com pontos do próprio e de outros *clusters*.

<sup>7</sup>O método *Gap Statistic*, proposto por Tibshirani, Walther e Hastie (2001), estima o número ideal de *clusters* comparando a dispersão observada com a de uma distribuição aleatória.

Figura 1: Dispersão dos dados para cada algoritmo de clusterização

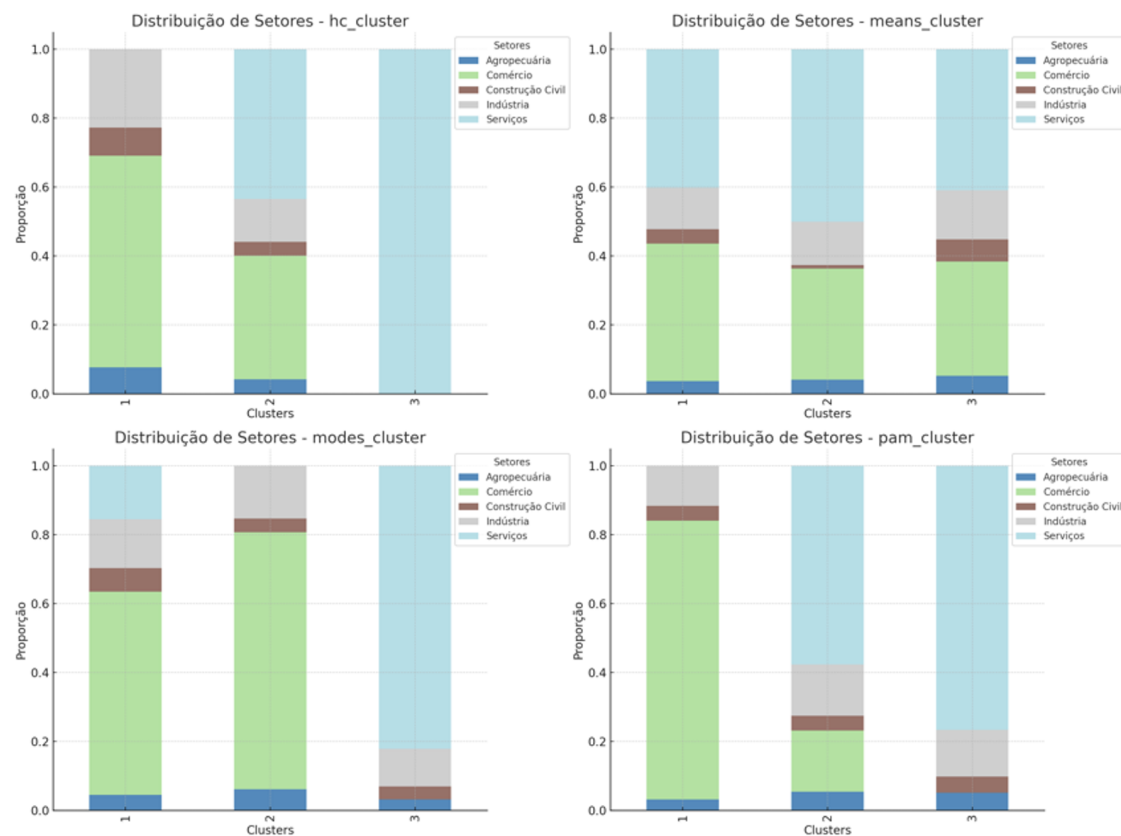


Fonte: Dados elaborados pela autora.

Nos gráficos, o HC isola serviços no *Cluster* 3, enquanto comércio e indústria aparecem nos *Clusters* 1 e 2. O *Means* mostra maior equilíbrio entre setores. O *Modes* tende a isolar serviços e comércio, com menor flexibilidade. O PAM distribui comércio e serviços entre os *clusters*, sendo mais robusto a *outliers*. De forma geral, HC, *Modes* e PAM tendem a isolar setores homogêneos como o de serviços, enquanto *Means* e PAM formam *clusters* mais balanceados, diluindo a segmentação setorial mais rígida. Segue a Figura 3, considerando a predominância das regiões brasileiras para cada algoritmo de clusterização:

A análise mostra predominância do Sudeste nos *clusters* do HC, com Sul e Nordeste como coadjuvantes. O *Means* redistribuiu melhor no *Cluster* 2. No *Modes*, o Sudeste domina e o Nordeste não aparece no *Cluster* 2. Já o PAM mantém o Sudeste no *Cluster* 1, mas inclui mais o Nordeste, Norte e Centro-Oeste nos demais, revelando maior abrangência regional. Em geral, todos os algoritmos refletem o peso do Sudeste, mas o PAM e o *Modes* mostram maior inclusão de regiões menos representadas. O HC enfatiza a concentração industrial, enquanto o *Means* e o PAM promovem distribuição regional mais equilibrada, e o *Modes* forma agrupamentos mais geograficamente diversos. Segue a Figura 4, considerando a predominância do porte das empresas para cada algoritmo de clusterização:

**Figura 2:** Predominância entre os setores econômicos para cada algoritmo de clusterização



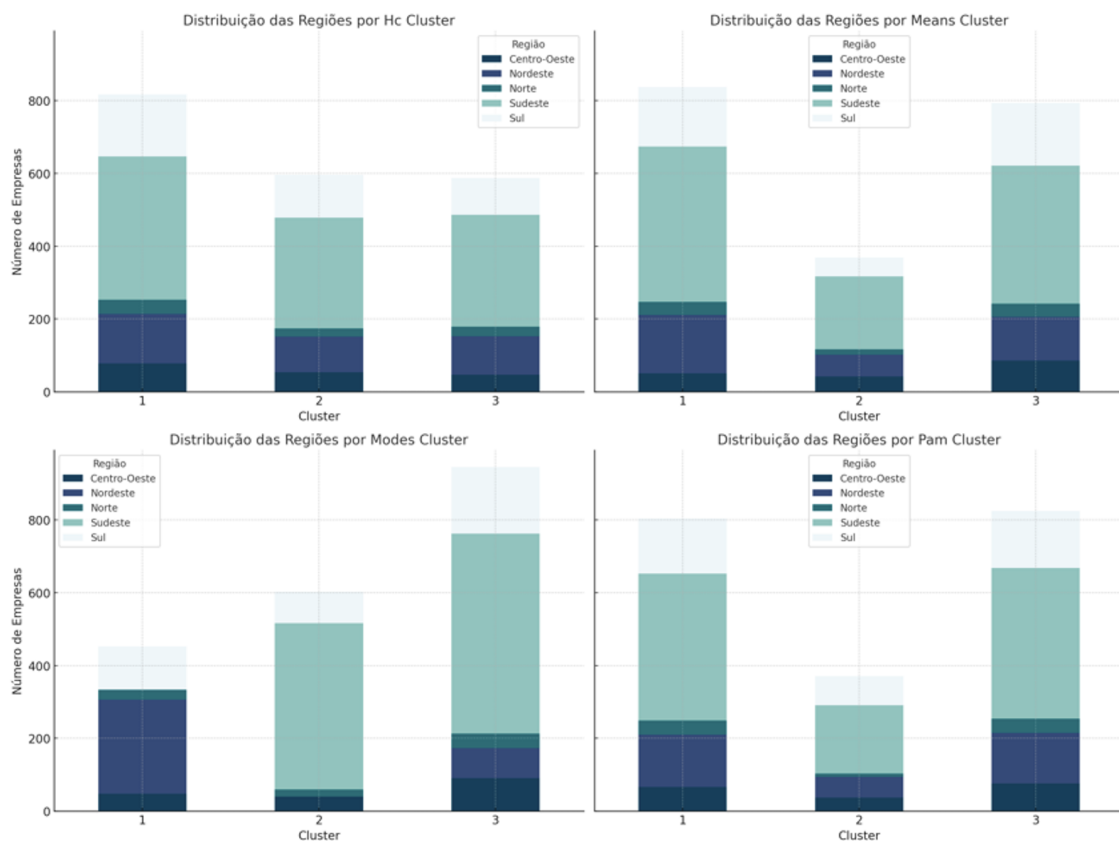
Fonte: Dados elaborados pela autora.

Nos gráficos do HC, os *clusters* 1 e 3 são compostos exclusivamente por microempresas, indicando menor capacidade financeira e ampla dispersão. O *cluster* 2 reúne pequenas, médias e grandes empresas, sugerindo um perfil de crescimento. No método *Means*, há maior equilíbrio entre os portes, com microempresas predominando no *Cluster* 1 e distribuição mais uniforme no *cluster* 2. O *Cluster* 3 é similar ao 1. O método *Modes* mostra bom balanceamento entre os portes nos *clusters*. Já o PAM segue padrão próximo ao HC, com maior variabilidade entre grupos. Em todos os algoritmos, microempresas dominam os *clusters*, refletindo sua ampla presença no mercado e na amostra. Pequenas empresas são mais equilibradas entre os grupos, enquanto médias e grandes têm baixa representatividade.

## 4.2 Análise Sob a Perspectiva da Qualidade dos Agrupamentos

Como forma de avaliar a qualidade dos *clusters* observados, foram utilizados quatro métodos de avaliação, que são: Índice de Calinski-Harabasz - CH (*Calinski-Harabasz Index*); Índice de Davies-Bouldin - DBI (*Davies-Bouldin Index*); Valor-p do ANOVA (*ANOVA p-value*); e Índice da Silhueta (*Silhouette Score*). Os métodos de qualidade de clusterização são métricas

**Figura 3:** Predominância entre as regiões brasileiras para cada algoritmo de clusterização

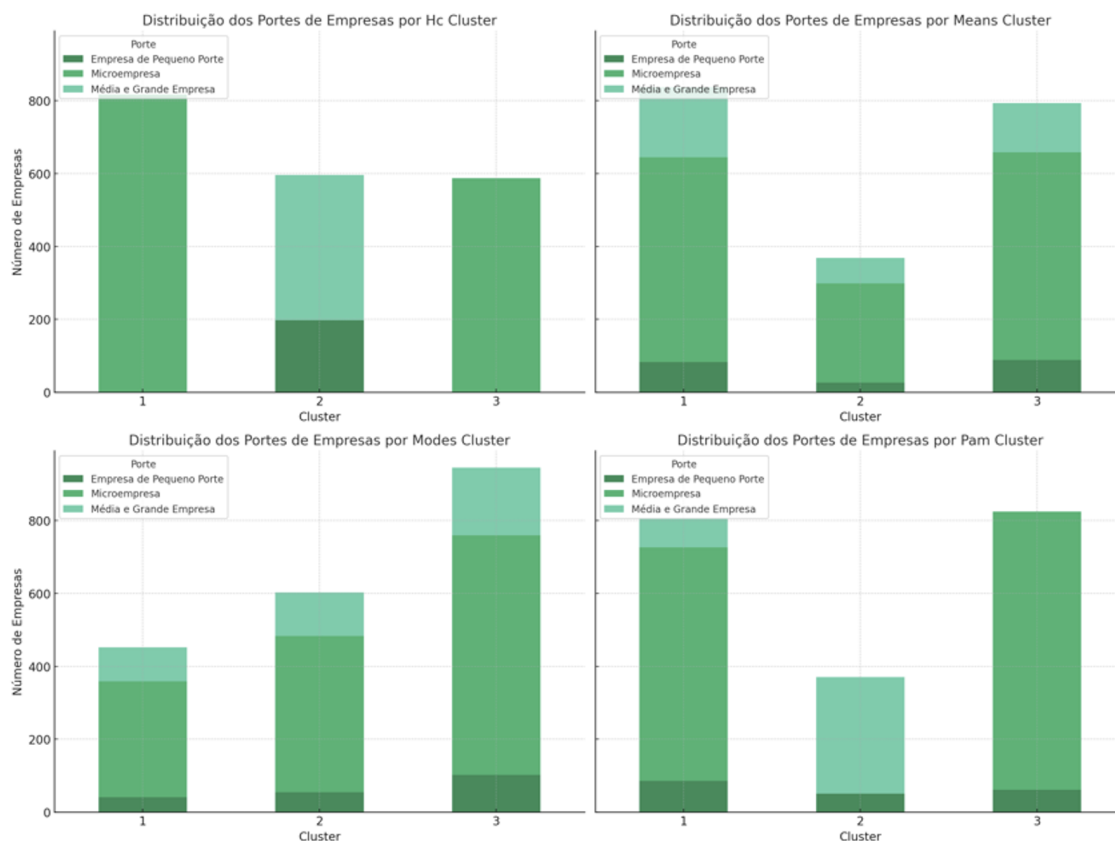


Fonte: Dados elaborados pela autora.

que avaliam o quão bem os dados foram agrupados em *clusters* distintos. Para fins de análise, foi incluído o resultado de agrupamento feito pela ABDI, denominado IMD ABDI. Segue a Tabela 2, considerando o resultado dos métodos de qualidade obtidos para cada uma dessas métricas nos cinco algoritmos analisados, incluindo nessa análise o IMD ABDI:

Com base nos resultados, O método Means foi o mais eficaz, com CH = 630,96, DBI = 1,53 e Silhueta = 0,23, indicando *clusters* bem definidos. O p-valor (6,46e-82) confirma a significância estatística da segmentação. Como segunda opção viável, o IMD ABDI apresentou DBI de 1,70, Silhueta de 0,11 e p-valor 0, indicando boa significância, embora seu CH de 325,70 seja inferior ao do Means. Já os métodos Modes e HC tiveram os piores resultados, com Silhueta negativa e DBI elevados (17,69 e 9,55), sugerindo *clusters* sobrepostos e mal definidos. Assim, o algoritmo Means é o mais recomendado, com melhor separação e significância; o IMD ABDI é uma alternativa razoável, enquanto Modes e HC devem ser evitados.

**Figura 4:** Predominância do porte das empresas para cada algoritmo de clusterização



Fonte: Dados elaborados pela autora.

### 4.3 Análise Sob a Perspectiva das Médias

Como o algoritmo k-means apresentou o melhor desempenho geral nas métricas de avaliação e a categorização IMD ABDI foi identificada como a segunda melhor alternativa, optou-se por aprofundar a análise desses dois agrupamentos. Para isso, foram examinadas as médias por dimensão e o perfil das empresas próximas a essas médias, com o objetivo de compreender os traços característicos dos grupos identificados e validar a coerência interna dos *clusters*. As médias por dimensão indicam a distribuição central das empresas em cada eixo avaliado, permitindo identificar padrões comportamentais relevantes. Para enriquecer essa análise, selecionaram-se as empresas cujos valores em cada dimensão estavam próximos da média de seu respectivo *cluster*, adotando-se uma margem de variação de  $\pm 0,2$ . Segue a Tabela 3, considerando o valor médio por dimensão dos *clusters* do algoritmo k-means:

O Cluster 1 apresenta os menores valores médios, refletindo fragilidades estruturais e operacionais. Com média de 4,93 em gestão, indica baixa maturidade gerencial e vulnerabilidade a crises. A média de 5,79 em pessoas revela dificuldades em gestão de equipes, capacitação e clima organizacional. Em processos, a média de 6,58 sugere baixa automação e

**Tabela 2:** Resultado dos métodos de qualidade

Métodos de qualidade	Hierárquico (HC)	k-Means	k-Modes	PAM	IMD ABDI
CH (↑ Melhor)	221,80	630,96	183,51	250,37	325,70
DBI (↓ Melhor)	9,55	1,53	17,69	4,88	1,70
Valor-p (↓ Melhor)	3,82e-04	6,46e-82	7,79e-03	8,81e-12	0,00
Índice da Silhueta (↑ Melhor)	-0,01	0,23	-0,01	0,00	0,11

Fonte: Dados elaborados pela autora.

**Tabela 3:** Valor médio por dimensão dos *clusters* do algoritmo k-means

Clusters	Dimensões				
	Gestão	Pessoas	Processos	Modelo de Negócios	Vendas e Consumidores
1	4,93	5,79	6,59	5,27	5,07
2	8,33	9,14	6,54	7,77	9,08
3	7,60	8,43	6,46	5,95	4,55

Fonte: Dados elaborados pela autora.

padronização. Já em modelo de negócios (5,27), evidencia limitações para inovar, diversificar receitas e competir no meio digital.

No caso da dimensão Vendas e Consumidores, com média 5,07, o Cluster 1 indica dificuldades em atrair e reter clientes, reflexo de marketing ineficaz, baixa presença digital e falta de estratégias claras. Isso compromete o crescimento e a sustentabilidade dessas empresas no mercado.

Considerando o resultado das médias por dimensão do Cluster 1, observa-se que esse grupo é composto por empresas com menor maturidade digital, que apresentam desempenho inferior em todas as dimensões avaliadas. Diante desse diagnóstico, os elementos a seguir oferecem subsídios relevantes para o delineamento de políticas públicas voltadas a empresas em estágios iniciais de digitalização. Essas orientações podem contribuir para o desenvolvimento de estratégias mais adequadas às necessidades específicas desse perfil empresarial:

- Capacitação gerencial: A baixa média em gestão aponta para a necessidade de formar lideranças, qualificar gestores e modernizar a governança por meio de mentorias, consultorias e cursos gratuitos.
- Desenvolvimento de pessoas: A média de 5,79 sugere falhas na gestão de equipes e retenção de talentos. Políticas devem promover RH estratégico, capacitação e melhoria do ambiente organizacional.
- Digitalização de processos: Apesar da média ligeiramente melhor em processos, há baixa automação. Políticas podem apoiar mapeamento, reestruturação e uso de tecnologias simples e eficazes.
- Inovação nos modelos de negócio: A média de 5,27 indica estruturas ultrapassadas. Sugere-se incentivo a laboratórios de inovação, remodelagem de ofertas e parcerias com startups e universidades.
- Melhoria nas vendas: Com média de 5,07, evidenciam-se dificuldades em presença digi-

tal e fidelização. Ações em marketing, *marketplaces* e canais digitais são recomendadas.

- Abordagem gradual: A política deve ser segmentada, com soluções simples no início, evoluindo conforme a capacidade interna das empresas avança.

O Cluster 2 reúne empresas com maior maturidade digital, com as maiores médias em todas as dimensões. A média de 8,33 em gestão reflete liderança estruturada, planejamento estratégico e uso de dados, favorecendo inovação, parcerias e atração de investimentos. Apesar do bom desempenho, a média de 6,54 em processos aponta necessidade de avanços em padronização e automação para escalar operações e manter a competitividade.

Com 7,77 de média, a dimensão modelo de negócios do cluster 2 sugere que as empresas próximas a essa média possuem um modelo de negócios bem definido e inovador. Elas conseguem adaptar suas estratégias às mudanças do mercado e exploram diferentes fontes de receita, garantindo maior estabilidade financeira e crescimento sustentável.

Com média de 9,08 em vendas e consumidores, o cluster 2 indica domínio de estratégias comerciais e digitais, com bom uso de canais online, marketing estruturado e fidelização de clientes. Essas empresas mostram alta maturidade digital, com gestão sólida, inovação e uso avançado de ferramentas como *e-commerce* e automação. Assim, políticas públicas devem focar em aceleração da inovação, expansão internacional e sofisticação tecnológica, e não mais em fundamentos estruturais. Essas orientações podem contribuir para o desenvolvimento de estratégias mais adequadas às necessidades específicas desse perfil empresarial:

- Inovação contínua e Pesquisa e Desenvolvimento: Empresas do Cluster 2 precisam de políticas que incentivem inovação incremental e disruptiva, com estímulo a parcerias com universidades, testes de tecnologias emergentes (IA, *blockchain*, quântica) e acesso a recursos não reembolsáveis.
- Internacionalização: Com gestão e presença digital consolidadas, devem ser apoiadas na inserção em cadeias globais, missões comerciais, *soft landing* e adequação regulatória internacional.
- Automação e integração: A política deve fomentar automação e integração de sistemas, promovendo eficiência operacional e adoção de tecnologias da Indústria 4.0.
- Governança de dados e transformação digital: Apoio a soluções avançadas como *machine learning* e plataformas digitais próprias, via consultorias, laboratórios digitais e infraestrutura tecnológica.
- Ecossistemas de inovação: Inserção em *hubs*, redes e *clusters*, atuando como empresas âncora para fortalecer cadeias produtivas e disseminar boas práticas.
- Financiamento diferenciado: Necessitam de linhas robustas de crédito para inovação, fundos de investimento público-privado e mecanismos para mitigar riscos tecnológicos.

O Cluster 3 tem desempenho intermediário. Com média de 7,60 em gestão, mostra boa maturidade organizacional, mas ainda enfrenta desafios em inovação e adaptação. A média de 8,43 em pessoas indica valorização da equipe e investimento em desenvolvimento, com espaço para melhorar retenção de talentos e cultura organizacional. A média de 6,46 na dimensão processos desse cluster sugere que a eficiência operacional pode ser um ponto de atenção. Embora possuam processos estruturados, essas empresas podem enfrentar

dificuldades na automação e integração de setores.

Na dimensão modelo de negócios, o *cluster 3* apresentou uma média de 5,95 indicando que essas empresas próximas a média podem ter dificuldade em inovar e expandir seus negócios, necessitando de ajustes para aumentar sua competitividade. Com média de 4,55 em vendas e consumidores, o *cluster 3* revela dificuldades em atrair e fidelizar clientes, possivelmente por estratégias de marketing frágeis ou baixa presença digital. Embora apresente práticas estruturadas, ainda enfrenta desafios para consolidar sua competitividade e maturidade digital. Essas orientações podem contribuir para o desenvolvimento de estratégias mais adequadas às necessidades específicas desse perfil empresarial:

- Aceleração digital estratégica: com boa estrutura e gestão, essas empresas precisam de apoio para superar barreiras tecnológicas, por meio de consultorias, soluções escaláveis e formação em liderança digital.
- Fortalecimento da presença digital: a baixa média em vendas indica necessidade de ações em marketing digital, estruturação de canais online e uso de ferramentas para entender o consumidor.
- Inovação nos modelos de negócio: com média de 5,95, essas empresas ainda operam de forma tradicional. Políticas podem incentivar mentorias, imersões digitais e redesenho estratégico.
- Eficiência operacional: com média de 6,46 em processos, há espaço para automação. Diagnósticos, tecnologias acessíveis e metodologias ágeis podem otimizar operações.
- Integração em redes de inovação: políticas podem conectá-las a *Hubs*, arranjos produtivos locais e empresas mais maduras, promovendo cooperação e aprendizado.
- Financiamento e acesso a tecnologia: políticas específicas de crédito, vouchers de inovação e parcerias com fornecedores podem facilitar a modernização tecnológica.

Com o objetivo de aprofundar a análise do agrupamento realizado pela categorização institucional da ABDI, também foram avaliadas as médias por dimensão em cada uma das categorias do IMD ABDI. Essa etapa permite verificar se os grupos definidos previamente refletem padrões consistentes de maturidade digital entre as empresas classificadas. Abaixo, segue a Tabela 4, considerando o valor médio por dimensão da categorização IMD ABDI, apresentando os resultados médios obtidos em cada uma das cinco dimensões, considerando as quatro categorias propostas:

**Tabela 4:** Valor médio por dimensão da categorização IMD ABDI

Categorias	Dimensões				
	Gestão	Pessoas	Processos	Modelo de Negócios	Vendas e Consumidores
Iniciantes	4,16	5,05	6,10	4,73	4,21
Emergentes	6,23	7,08	6,54	5,65	4,95
Intermediários	8,31	9,11	6,71	7,11	7,29
Líderes	9,94	10,77	6,52	8,79	9,92

Fonte: Dados elaborados pela autora.

Conforme descrito na Subseção 3.2, o índice de maturidade digital é a média das dimensões avaliadas. A categoria empresas iniciantes, reúne empresas com os menores

níveis de maturidade. Com média de 4,16 em gestão e 5,05 em pessoas, indicam fraca capacidade estratégica e baixa qualificação da equipe. A média de 6,10 em processos revela estrutura operacional ainda pouco automatizada e desorganizada.

Empresas próximas à média dessa categoria enfrentam desafios em inovação, com média de 4,73 em modelo de negócios, e sérias dificuldades em vendas e relacionamento com clientes (média de 4,21). A baixa presença digital e ausência de estratégias de marketing impactam negativamente sua competitividade. Trata-se de empresas ainda dependentes de práticas tradicionais, que necessitam de mudanças estruturais para avançar na digitalização.

A categoria de empresas emergentes, mostra maturidade digital um pouco superior aos iniciantes, mas ainda com desafios. Com médias de 6,23 em gestão e 7,08 em pessoas, indicam estrutura mais definida e foco em capacitação, mas com dificuldades estratégicas. A média de 6,54 em processos sinaliza início da digitalização, e 5,65 em modelo de negócios mostra transição ainda incipiente. A baixa média de 4,95 em vendas revela fragilidade no uso de canais digitais. São empresas em fase inicial de transição digital.

As empresas intermediárias, apresentam maturidade digital mais avançada, com gestão estruturada (8,31) e alto investimento em pessoas (9,11), indicando valorização de equipes qualificadas. Em processos (6,71), já adotam metodologias ágeis, embora com potencial de aprimorar automação e integração. O modelo de negócios (7,11) e vendas e consumidores (7,29) sugerem capacidade de inovação e uso de estratégias digitais como *CRM* e *e-commerce*. Esse grupo aplica bem a transformação digital, mas ainda pode expandir sua atuação para atingir maturidade mais elevada.

O último grupo de empresas, líderes, reúne as empresas mais avançadas digitalmente, com as maiores médias em todas as dimensões. Destacam-se em gestão (9,94) e pessoas (10,77), com estruturas eficientes, equipes qualificadas e metodologias inovadoras. Em processos (6,52), apesar da média próxima às intermediárias, adotam tecnologias avançadas como IA e automação. Apresentam modelo de negócios inovador (8,79) e excelência em vendas e consumidores (9,92), com domínio dos canais digitais e uso intensivo de dados. São empresas que lideram a transformação digital e definem padrões competitivos no mercado.

Diante do exposto, ressalta-se que a metodologia de clusterização identifica grupos reais de empresas com padrões semelhantes, sem impor categorias a priori. Isso contrasta com modelos como o IMD ABDI, que normalmente aplicam escalas fixas e cortes definidos previamente, como “iniciante”, “intermediário” e “avançado”.

A análise por *cluster* permite capturar nuances e combinações de características que modelos fixos podem não perceber. Por exemplo, o *cluster 3* tem boas práticas em gestão e pessoas e péssimo desempenho em vendas, um perfil híbrido que talvez fosse rotulado como “intermediário” no IMD, sem captar suas demandas específicas por apoio em marketing e canais digitais.

## 5 Considerações Finais

Diante da relevância crescente da transformação digital para os negócios, este estudo teve como objetivo investigar o uso de técnicas de clusterização como ferramenta de apoio à formulação de políticas públicas voltadas à digitalização, considerando a heterogeneidade das empresas brasileiras. Com base nos dados da 6ª edição da pesquisa do IMD ABDI, foram aplicados algoritmos de aprendizado não supervisionado para identificar perfis empresariais semelhantes, sem recorrer a classificações rígidas.

Os resultados deste estudo indicam que o algoritmo *k-means* foi o mais eficiente na segmentação das empresas brasileiras segundo sua maturidade digital. Apresentou os melhores índices de avaliação (CH = 630,96; DBI = 1,53; Silhueta = 0,23), demonstrando capacidade superior de formar agrupamentos robustos, bem definidos e com baixa sobreposição entre *clusters*. Em comparação, a categorização do IMD ABDI, embora razoável, mostrou-se menos precisa, enquanto os métodos HC e k-modes apresentaram baixo desempenho. A escolha do *k-means*, portanto, se justifica pela clareza na discriminação de perfis distintos de maturidade digital, viabilizando análises mais sólidas e confiáveis.

A partir da segmentação gerada, foram identificados três perfis de empresas: um grupo com baixa maturidade digital, composto majoritariamente por microempresas de serviços e comércio; um segundo grupo mais avançado, com forte estrutura de gestão e presença digital consolidada; e um grupo intermediário, com estrutura razoável, mas carência em capacidades digitais. Esses perfis fornecem subsídios importantes para o desenho de políticas públicas mais segmentadas, que considerem os diferentes estágios de maturidade digital. Além disso, a clusterização permite traçar trilhas evolutivas específicas, orientar o apoio técnico e financeiro de forma mais eficiente e fomentar parcerias estratégicas entre empresas de diferentes *clusters*.

As evidências apresentadas neste estudo são relevantes tanto para formuladores de políticas públicas quanto para a literatura sobre transformação digital e métodos de segmentação empresarial. Para o poder público, os achados oferecem uma abordagem metodológica robusta e adaptável para diagnosticar e apoiar a digitalização do setor produtivo. Este trabalho busca contribuir à literatura ao demonstrar o valor do uso de algoritmos de clusterização, especialmente o *k-means*, como ferramenta complementar ou alternativa a classificações rígidas, fortalecendo a análise empírica em estudos sobre maturidade digital e políticas de inovação.

Em relação à literatura existente, os resultados convergem com estudos que apontam a eficácia do *k-means* na identificação de perfis empresariais, mas, em contrapartida, trazem uma contribuição inovadora ao aplicá-lo de forma comparativa ao IMD ABDI. Assim, reforçam achados já estabelecidos, ao mesmo tempo em que oferecem evidências novas e contextualizadas para o caso brasileiro.

## Referências

- ABDI, Agência Brasileira de Desenvolvimento Industrial** (2019). “E-ABDI: Documento base para a definição das ações da ABDI em transformação digital.”
- ABDI, Agência Brasileira de Desenvolvimento Industrial** (2023). “Maturidade Digital das MPEs Brasileiras.”
- Aggarwal, Charu** (2015). *Data Mining: The Textbook*, 14, Springer International Publishing Switzerland, New York, USA, , 1st edition.
- Alpaydin, Ethem** (2016). *Machine Learning: The New AI*, MIT Press, Cambridge, MA.
- Caldas, Sérgio** (2022). “China quer impulsionar fatia da economia digital no PIB para 10% até 2025,” CNN Brasil.
- Cartela, Eva Yamila Da Silva e José Gabriel Porcile** (2015). “Heterogeneidade estrutural na produtividade das firmas brasileiras,” CEPAL e IPEA.
- Corrar, Luiz J. et al.** (2009). *Análise multivariada: para os cursos de administração, ciências contábeis e economia*, Atlas.
- De Morsier, Frank et al.** (2015). “Cluster validity measure and merging system for hierarchical clustering considering outliers,” *Pattern Recognition*, 48(4): 1478–1489.
- Deng, Zhenyun et al.** (2016). “Efficient kNN classification algorithm for big data,” *Neurocomputing*, 195 143–148.
- Dutta, Gautam et al.** (2021). “Digitalization priorities of quality control processes for SMEs: A conceptual study in perspective of Industry 4.0 adoption,” *Journal of Intelligent Manufacturing*, 32(6): 1679–1698.
- Eller, Robert et al.** (2020). “Antecedents, consequences, and challenges of small and medium-sized enterprise digitalization,” *Journal of Business Research*, 112 119–127.
- Fontana, André e Murilo C. Naldi** (2009). “Estudo de Comparação de Métodos para Estimação de Números de Grupos em Problemas de Agrupamento de Dados,” Universidade de São Paulo, ISSN 0103-2569.
- G20** (2016). “G20 Digital Economy Development and Cooperative Index.”
- Gil, Antonio Carlos** (2010). *Como elaborar projetos de pesquisa*, São Paulo, , 5th edition.
- Goldschmidt, Ronaldo et al.** (2015). *Data mining: conceitos, técnicas, algoritmos, orientações e aplicações*, Elsevier, Rio de Janeiro.
- Gollhardt, Torsten et al.** (2020). “Development of a Digital Transformation Maturity Model for IT Companies,” in *2020 IEEE 22nd Conference on Business Informatics (CBI)*, 1 94–103.

- Grus, Joel** (2015). *Data Science from Scratch*, 1, O'Reilly Media, Cambridge, Mass., , 1st edition.
- Gökalp, Ebru e Veronica Martinez** (2021). “Digital transformation capability maturity model enabling the assessment of industrial manufacturers,” *Computers in Industry*, 132 p. 103522.
- Hair, Joseph F. et al.** (2010). *Multivariate Data Analysis*, Pearson, 7th edition.
- Hassani, Reza e Younès El Idrissi** (2020). “A framework to succeed IT project management in an era of digital transformation (Open Access),” *International Journal of Advanced Trends in Computer Science and Engineering*, 9(1): 630–636.
- Hinings, Bob et al.** (2018). “Digital innovation and transformation: An institutional perspective,” *Information and Organization*, 28(1): 52–61.
- Jain, A. K., M. N. Murty, e P. J. Flynn** (1999). “Data clustering: a review,” *ACM Computing Surveys*, 31(3): 264–323.
- Jayaram Reddy, Avulapalli et al.** (2018). “Performance Analysis of Clustering Algorithm in Data Mining in R Language,” in *Communications in Computer and Information Science*, 837 364–372.
- Jordan, Michael I. e Tom M. Mitchell** (2015). “Machine learning: Trends, perspectives, and prospects,” *Science*, 349(6245): 255–260.
- Kassambara, Alboukadel** (2017). *Multivariate Analysis 1: Practical Guide To Cluster Analysis in R: Unsupervised machine learning*, STHDA, 1st edition.
- Kirmizi, Mehmet e Batuhan Kocaoglu** (2022). “Digital transformation maturity model development framework based on design science: case studies in manufacturing industry,” *Journal of Manufacturing Technology Management*, 33(7): 1319–1346.
- Kodinariya, Trupti M. e Prashant R. Makwana** (2013). “Review on determining number of Cluster in K-Means Clustering,” *International Journal of Advance Research in Computer Science and Management Studies*, 1(6): 90–95.
- Kraus, Sascha et al.** (2022). “Digital transformation in business and management research: An overview of the current status quo,” *International Journal of Information Management*, 63.
- Kraus, Sascha, Paul Jones, Norbert Kailer, Alexandra Weinmann, Nuria Chaparro-Banegas, e Norat Roig-Tierno** (2021). “Digital Transformation: An Overview of the Current State of the Art of Research,” *SAGE Open*, 11(3): 1–15.
- Liu, Qihong et al.** (2016). “Distributed k-means algorithm for sensor networks based on multi-agent consensus theory,” in *Proceedings of the IEEE International Conference on Industrial Technology*, 2016-May 2114–2119.

- Lobão, Luis e Carlos Zilli** (2020). *A jornada da transformação digital: um guia prático: casos, fundamento e ferramentas*, Editora Lamonica, São Paulo.
- Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, e Kurt Hornik** (2017). *cluster: Cluster analysis basics and extensions*, R package version 2.0.6.
- Matarazzo, Michela et al.** (2020). “Digital transformation and customer value creation in Made in Italy SMEs: A dynamic capabilities perspective,” *Journal of Business Research*, 123 642–656.
- Meira, Silvio** (2021). “Direções, desafios e dimensões para uma estratégia de Brasil digital.”
- Omrani, Hashem et al.** (2018). “An integrated fuzzy clustering cooperative game data envelopment analysis model with application in hospital efficiency,” *Elsevier Ltd*, 114.
- Pagani, Margherita e Catherine Pardo** (2017). “The impact of digital technology on relationships in a business network,” *Industrial Marketing Management*, 67 185–192.
- Rencher, Alvin C.** (2002). *Methods of Multivariate Analysis*, John Wiley & Sons, 2nd edition.
- Rodrigues, Eugénio, David Sousa-Rodrigues, Mafalda Teixeira de Sampayo, Adélio R. Gaspar, Álvaro Gomes, e Carlos H. Antunes** (2017). “Clustering of architectural floor plans: A comparison of shape representations,” *Automation in Construction*, 80 48–65.
- Rogers, David L.** (2016). *The Digital Transformation Playbook: rethink your business for the digital age*, Columbia University Press.
- Rohn, Daniel et al.** (2021). “Digital platform-based business models – An exploration of critical success factors,” *Journal of Engineering and Technology Management - JET-M*.
- Rousseeuw, Peter J.** (1987). “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, 20 53–65.
- Saldanha, Tony** (2019). *Why Digital Transformations Fail: The Surprising Disciplines of How to Take Off and Stay Ahead*, Berrett-Koehler Publishers, Oakland, CA.
- Sampaio, Rafael** (2018). *Vantagem digital: Um guia prático para a transformação digital*, Alta Books, Rio de Janeiro.
- Thorndike, Robert L.** (1953). “Who belongs in the family?” *Psychometrika*, 18(4): 267–276.
- Tibshirani, Robert, Guenther Walther, e Trevor Hastie** (2001). “Estimating the number of clusters in a dataset via the Gap statistic,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2): 411–423.
- Van Ark, Bart et al.** (2019). “Productivity & innovation competencies in the midst of the digital transformation age: A EU-US comparison.”

**Verhoef, Peter C. et al.** (2021). "Digital transformation: A multidisciplinary reflection and research agenda," *Journal of Business Research*, 122 889–901.





**Vial, Gregory** (2019). "Understanding digital transformation: A review and a research agenda," *The Journal of Strategic Information Systems*, 28(2): 118–144.



The background features a dark blue color scheme with various data visualization elements. On the left, there is a bar chart with five bars of increasing height, with values 138, 178, 175, 172, and 190. A line graph with two upward-trending lines is overlaid on the bars. At the top center, a circular progress indicator shows 68%. To the right, another circular progress indicator shows 75%. Below these, there are several abstract geometric shapes, including a triangle and a sphere-like structure composed of interconnected nodes and lines. The overall aesthetic is modern and technical.

# idp

SGAS Quadra 607 - Módulo 49  
Via L2 Sul, Brasília-DF  
CEP: 70200-670

  /sejaidp  
 (61) 3535-6565  
 idp.edu.br