

idp

idm

MESTRADO PROFISSIONAL EM ECONOMIA

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

KAREN CRISTINA LEAL DA SILVA ILOGTI

Brasília-DF, 2025

KAREN CRISTINA LEAL DA SILVA ILOGTI

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

Dissertação apresentada ao Programa de Pós Graduação em Economia, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como requisito parcial para obtenção do grau de Mestre.

Orientador

Professor Doutor Leonardo Monteiro Monastério.

Brasília-DF 2025

KAREN CRISTINA LEAL DA SILVA ILOGTI

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

Dissertação apresentada ao Programa de Pós Graduação em Economia, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como requisito parcial para obtenção do grau de Mestre.

Aprovado em 25 / 02 / 2025

Banca Examinadora

Prof. Dr. Leonardo Monteiro Monastério – Orientador

Prof. Dr. Carlos Eduardo Gasparini

Prof. Dr. Luiz Ricardo Mattos Teixeira Cavalcante

I29t Ilogti, 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 Ilogti. – Brasília: IDP, 2025.

64 f.
Inclui bibliografia.

Dissertação – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa – IDP, Mestrado Profissional em Economia, Brasília, 2025.
Orientador: Prof. Dr. Leonardo Monteiro Monastério.

1. Maturidade digital. 2. Políticas públicas. 3. Segmentação empresarial. I. Título.

CDD: 330

Ficha catalográfica elaborada pela Biblioteca Ministro Moreira Alves
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa

RESUMO

A transformação digital é um processo estratégico essencial para o fortalecimento da competitividade do setor produtivo brasileiro. Contudo, a ampla heterogeneidade das empresas em termos de porte, setor e distribuição regional impõe desafios à formulação de políticas públicas padronizadas. Este trabalho investiga o uso de técnicas de clusterização como alternativa metodológica para segmentar empresas segundo seus níveis de maturidade digital e, assim, subsidiar o desenho de políticas mais direcionadas e eficazes. Utilizando dados da 6ª edição do Índice de Maturidade Digital - IMD da Agência Brasileira de Desenvolvimento Industrial - ABDI, aplicado a 2.000 empresas, foram testados diferentes métodos de clusterização, incluindo hierárquico, k-means, k-modes e Partitioning Around medoids – PAM. A qualidade dos agrupamentos foi avaliada por meio dos índices Calinski-Harabasz - CH, Davies-Bouldin - DBI, Silhueta e ANOVA, permitindo a comparação dos métodos e a identificação da abordagem mais eficiente. Os resultados indicaram que o k-means apresentou o melhor desempenho, com maior separação dos clusters, com um CH de 630,96, menor sobreposição entre grupos segundo o índice DBI de 1,53 e uma estrutura bem definida segundo o método Silhueta, que foi de 0,23. Como segunda melhor opção, a classificação feita pela ABDI apresentou resultados satisfatórios, porém com menor eficiência na separação dos agrupamentos dos clusters. Os achados deste estudo indicam que a clusterização utilizando o algoritmo k-means é a melhor abordagem para segmentação das empresas brasileiras quanto à maturidade digital, permitindo uma melhor definição dos grupos e, consequentemente, facilitando a formulação de políticas públicas mais precisas. A categorização IMD ABDI, embora menos eficiente, pode ser considerada uma alternativa viável para análises complementares. Além disso, observou-se que fatores como porte, setor e região influenciam fortemente a distribuição dos clusters, revelando padrões estruturais que a categorização do IMD ABDI não captura plenamente. Os resultados demonstram que a clusterização é uma ferramenta robusta para segmentação de empresas, permitindo a formulação de políticas públicas modulares, progressivas e mais aderentes à realidade de diferentes perfis empresariais.

Palavras-chave: transformação digital, maturidade digital, políticas públicas, clusterização, segmentação empresarial.

ABSTRACT

Digital transformation is a strategic process essential for strengthening the competitiveness of the Brazilian productive sector. However, the broad heterogeneity of companies in terms of size, sector, and regional distribution poses challenges to the formulation of standardized public policies. This study investigates the use of clustering techniques as a methodological alternative to segment companies according to their levels of digital maturity, thus supporting the design of more targeted and effective policies. Using data from the 6th edition of the Digital Maturity Index (IMD) by the Brazilian Industrial Development Agency (ABDI), applied to 2,000 companies, different clustering methods were tested, including hierarchical, k-means, k-modes, and Partitioning Around Medoids (PAM). The quality of the groupings was evaluated using the Calinski-Harabasz (CH), Davies-Bouldin Index (DBI), Silhouette method, and ANOVA, allowing for a comparison of the methods and the identification of the most efficient approach. The results indicated that k-means performed the best, with greater cluster separation, a CH score of 630.96, lower overlap between groups with a DBI of 1.53, and a well-defined structure according to the Silhouette method, which scored 0.23. As the second-best option, the classification adopted by ABDI showed satisfactory results, although less efficient in separating clusters. The findings of this study indicate that clustering using the k-means algorithm is the most effective approach for segmenting Brazilian companies by digital maturity, allowing for better group definition and, consequently, facilitating the formulation of more precise public policies. Although less efficient, the IMD ABDI categorization can be considered a viable alternative for complementary analyses. Additionally, it was observed that factors such as company size, sector, and region strongly influence cluster distribution, revealing structural patterns that the IMD ABDI categorization does not fully capture. The results demonstrate that clustering is a robust tool for business segmentation, enabling the formulation of modular, progressive, and more context-sensitive public policies tailored to the reality of different business profiles.

Keywords: digital transformation, digital maturity, public policies, clustering, business segmentation.

LISTA DE ABREVIATURAS E SIGLAS

ABDI	Agência Brasileira de Desenvolvimento Industrial
CNAE	Classificação Nacional de Atividades Econômicas
Demais	Empresas de Média e Grande Empresas
DF	Distrito Federal
DNA	deoxyribo nucleic acid
E-Digital	Estratégia Brasileira para a Transformação Digital
EPP	Empresas de Pequeno Porte – EPP
G20	Grupo formado pelos ministros de finanças e chefes dos bancos centrais das 19 maiores economias do mundo mais a União Africana e União Europeia.
IMD	Índice de Maturidade Digital da ABDI
LGPD	Lei Geral de Proteção de Dados
MCTI	Ministério da Ciência, Tecnologia e Inovação
MDIC	Ministério do Desenvolvimento, Indústria, Comércio e Serviços
ME	Microempresas
MEI	Microempreendedores Individuais
MPEs	Micro e Pequenas Empresas
N	Tamanho Populacional
NIB	Nova Indústria Brasil
PAM	Partitioning Around medoids
PIB	Produto Interno Bruto
TIC	Tecnologias de Informação e Comunicação
UF	Unidades da Federação
WEB	World Wide Web

LISTA DE ILUSTRAÇÕES

Gráfico 1

Dispersão dos dados nos algoritmos: HC, Means, Modes e PAM

33

Gráfico 2

Predominância entre os setores econômicos nos algoritmos: HC, Means, Modes e PAM

34

Gráfico 3

Predominância entre as regiões brasileiras nos algoritmos: HC, Means, Modes e PAM

36

Gráfico 4

Predominância do porte das empresas nos algoritmos: HC, Means, Modes e PAM

38

LISTA DE TABELAS

Tabela 1

Definições sobre Transformação Digital18

Tabela 2

Distribuição dos dados entre os algoritmos observados32

Tabela 3

Resultado dos métodos de qualidade39

Tabela 4

Valor médio por dimensão dos clusters do algoritmo k-means41

Tabela 5

Valor médio por dimensão da categorização IMD ABDI49

SUMÁRIO

1. INTRODUÇÃO	13
----------------------------	-----------

2. REVISÃO DA LITERATURA	18
---------------------------------------	-----------

2.1 TRANSFORMAÇÃO DIGITAL.....	18
--------------------------------	----

2.2 MATURIDADE DIGITAL.....	20
-----------------------------	----

2.3 CLUSTERIZAÇÃO.....	22
------------------------	----

3. METODOLOGIA.....	25
----------------------------	-----------

3.1 ALGORITMOS DE CLUSTERIZAÇÃO ADOTADOS.....	25
---	----

3.2 PREPARAÇÃO DOS DADOS.....	27
-------------------------------	----

4. RESULTADOS.....	31
---------------------------	-----------

4.1 APLICAÇÃO DA CLUSTERIZAÇÃO NOS DADOS.....	31
---	----

4.2 ANÁLISE SOB A PERSPECTIVA DA QUALIDADE DOS AGRUPAMENTOS	39
--	----

4.3 ANÁLISE SOB A PERSPECTIVA DAS MÉDIAS	41
--	----

5. CONCLUSÃO	54
---------------------------	-----------

REFERÊNCIAS.....	59
-------------------------	-----------



1

INTRODUÇÃO

Segundo Pesquisa sobre o uso de Tecnologia da Informação e Comunicação - TIC, realizada em 2021 pela FGV¹, no Brasil, o impacto da Covid 19 provocou em meses um avanço no uso de TIC, o que normalmente levaria anos para ocorrer. Os resultados comprovam uma aceleração do processo de Transformação Digital nas empresas e na sociedade. A forma como as empresas atuam em seus mercados e se relacionam com seus clientes foi modificada. Para muitas empresas, o digital passou a ser uma estratégia determinante à sua sobrevivência e a correlação deste tema com a economia digital é inevitável.

As estimativas sobre o tamanho da economia digital variam significativamente de acordo com os critérios utilizados para sua mensuração. Dados da União Europeia indicam que, em 2019, a indústria de produtos digitais representava 8,2% do PIB total do bloco econômico (Van Ark, 2019). Contudo, a relevância da economia digital cresce rapidamente, a ponto de alguns países estabelecerem metas crescentes para sua participação no PIB (Caldas, 2022).

O Fórum Econômico Mundial e o G20 definem a economia digital como uma ampla gama de atividades econômicas que incluem: o uso de informações e conhecimento digitalizados como o principal fator de produção; redes de informações modernas como um importante espaço de atividades; e o uso eficaz das TIC como um importante impulsionador da produtividade, crescimento e otimização estrutural econômica. Embora seja chamada de economia digital, ela não está separada da economia tradicional, que se baseia em bens físicos. Aspectos do “mundo virtual” têm impactos no “mundo físico” e

¹ A pesquisa realizada anualmente pela Fundação Getúlio Vargas - FGV faz parte do Fórum de Informações permanente sobre o Uso de TIC nas empresas. O objetivo é conhecer e estudar a situação das empresas e organizações brasileiras nesta temática.
<https://portal.fgv.br/artigos/panorama-uso-ti-brasil#:~:text=Vamos%20ultrapassar%20200%20milh%C3%B5es%20de,metade%20do%20pico%20de%202013.>

vice-versa, moldando as relações econômicas, políticas e sociais, o que alguns chamam de “transformação digital”² (Meira, 2021).

Segundo a Agência Brasileira de Desenvolvimento Industrial – ABDI³, em sua publicação “E-ABDI: Documento base para a definição das ações da ABDI em transformação digital (2019) a Economia Digital abrange três esferas: governo, setor produtivo e sociedade. O governo e setor produtivo atuam como impulsionadores econômicos, e a sociedade como beneficiária na economia. O governo facilita a transformação digital, capacita a sociedade e garante serviços e direitos, além de melhorar a gestão do gasto público. Para o setor produtivo a digitalização dos processos produtivos e a adaptação ao ambiente digital impulsionam a competitividade e a produtividade das empresas brasileiras, contribuindo para o crescimento econômico do país e para a sociedade.

O governo brasileiro vem estruturando políticas públicas para otimizar esforços na transformação digital e na indústria. Dois exemplos são: i) a Estratégia Brasileira para a Transformação Digital - E-Digital⁴, revisada em 2022, e; ii) a Nova Política Industrial, chamada de Nova Indústria Brasil - NIB⁵, publicada em 2024. A E-Digital visa promover o uso inclusivo de tecnologias digitais, modernizando a infraestrutura digital, estimulando a inovação tecnológica e desenvolvendo habilidades digitais, com parcerias entre governo, setor privado e sociedade civil. A NIB, estruturada em seis missões, busca enfrentar desafios sociais e ambientais por meio de políticas públicas orientadas para a segurança alimentar, saúde, infraestrutura sustentável, transformação digital da indústria, bioeconomia e tecnologias para soberania nacional, promovendo crescimento econômico sustentável e inclusivo. Essas políticas exemplificam como

² 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.

³ A ABDI é um Serviço Social Autônomo criado a partir da Lei N° 11.080, de 30 de dezembro de 2004, para promover a execução de políticas de desenvolvimento industrial, especialmente as que contribuam para a geração de empregos no Brasil, em consonância com as políticas de comércio exterior e de ciência e tecnologia (<https://www.abdi.com.br>).

⁴ A Estratégia de Transformação Digital (E-Digital) é uma política coordenada pelo Ministério da Ciência, Tecnologia e Inovação (MCTI) - <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/estrategia-digital>;

⁵ A Nova Indústria Brasil (NIB) é uma política coordenada pelo Ministério de Desenvolvimento, Indústria, Comércio e Serviços (MDIC) - <https://www.gov.br/mdic/pt-br/composicao/se/cndi/plano-de-acao/nova-industria-brasil-plano-de-acao.pdf>

a transformação digital se configura como um dos principais instrumentos de política pública na era da economia digital e é tema central deste trabalho.

Num contexto de transformação digital, em que se observa uma alteração no perfil de adoção de TIC pelas empresas, é importante que sejam priorizadas ações voltadas à pesquisa, ao desenvolvimento, à geração de inovações e à produção de tecnologias que estimulem e modernizem todo o tecido produtivo brasileiro.

A Transformação Digital consiste no estabelecimento de uma estratégia de jornada das empresas em que a integração de tecnologias digitais resulte em mudanças na forma como as empresas operam, exploram novas oportunidades de negócios e como elas agregam valor aos seus clientes (ABDI, 2019).

Segundo Vial (2019), a transformação digital é um processo que visa melhorar determinada organização através da utilização de tecnologias digitais de informação e comunicação, tendo como intuito possibilitar mudanças significativas em suas propriedades. Para Matarazzo *et al.* (2020), a transformação digital muda o processo de criação de valor para o consumidor, afetando as rotinas operacionais e capacidades organizacionais. A Transformação Digital introduz um modelo de negócios inovador ao adotar uma nova abordagem estratégica para gerar e aproveitar valor (Pagani; Pardo, 2017).

A incorporação das tecnologias digitais em todas as áreas da empresa (Hassani; Idrissi, 2020) demanda transformações significativas na cultura, nas operações, na estratégia e nos princípios de gestão, visando alcançar vantagem competitiva (Sampaio, 2018). Essas mudanças estruturais, que colocam a tecnologia como elemento central, embora não exclusivo, requerem alinhamento do objetivo da transformação digital com toda a organização (Hinings *et al.*, 2018).

Entretanto, o processo de transformação digital nas empresas não é tarefa simples, já que envolve vários desafios, como acompanhar a evolução das tecnologias que ocorrem de forma muito rápida no mercado. Nesse sentido, aumentar a maturidade digital é algo a ser perseguido em uma jornada de transformação digital pelas empresas, uma consequência de se trilhar um caminho que busca desenvolver a empresa para atuar digitalmente. O processo de transformação digital das empresas precisa estar alinhando à sua estratégia de negócios, dada a relevância do tema. Mas como? Entender os fatores que

determinam o aumento ou não da maturidade digital em empresas brasileiras pode apoiar na escolha de estratégias mais precisas por parte das empresas, por exemplo, como também na construção de políticas públicas mais adequadas.

Sob outra perspectiva, a transformação digital impulsiona melhorias na produtividade. Países com níveis mais altos de produtividade tendem, no médio e longo prazo, a proporcionar maior bem-estar para suas populações.

Um fato conhecido é a ampla heterogeneidade das empresas no Brasil. Segundo Cartela e Porcile (2015), existe uma forte heterogeneidade intersetorial dentro da indústria de transformação. Esse estudo também identificou forte heterogeneidade intrassetorial em empresas que apresentam processos produtivos semelhantes, em termos de intensidade tecnológica. Com isso, a análise destaca a importância de políticas específicas por grupos de empresas para aumentar a eficácia e fomentar a competitividade, aprendizado e inovação, contribuindo para a redução da heterogeneidade estrutural.

Sendo assim, a ABDI, no cumprimento de sua missão, vem monitorando o nível de maturidade digital das empresas brasileiras, visando propor políticas mais adequadas para esse grupo de empresas. No período de 2021 a 2023, a ABDI realizou uma pesquisa com aplicações semestrais a fim de medir o índice de maturidade digital do setor produtivo brasileiro. A pesquisa contou com a participação de 2.000 empresas e a coleta dos dados se deu por meio de um questionário que contemplava cinco dimensões de análise, tais quais: Processos, Gestão, Pessoas, Modelo de Negócios e Vendas e Consumidores. O resultado da sexta onda dessa pesquisa será a base empírica dessa dissertação.

Diante da crescente importância da transformação digital aos negócios, esta dissertação busca criar uma classificação das empresas que seja útil para a elaboração e implementação de políticas públicas, considerando que as empresas brasileiras são bastante heterogêneas em diversas dimensões e as políticas para a busca da maturidade digital não costumam levar esse aspecto em consideração.



2

REVISÃO DA LITERATURA

2.1 TRANSFORMAÇÃO DIGITAL

O tema da transformação digital ainda é bastante recente no desenvolvimento de estudos científicos empíricos capazes de demonstrar uniformidade nos entendimentos conceituais e nas abordagens sobre como as empresas podem acelerar um processo de digitalização, ou seja, aumentar sua maturidade digital, em suas estruturas (Kraus *et al.*, 2022). De fato, a conceituação de transformação digital pode envolver desde a simples aplicação de novas tecnologias a processos analógicos até mudanças mais profundas em modelo de negócio (Kraus *et al.*, 2021). Alguns conceitos são apresentados na Tabela 1 abaixo:

Tabela 1 – Definições sobre Transformação Digital	
Autor	Definição de Transformação Digital
Liu, Chen, and Chou (2011), p.1728	Transformação Digital é uma transformação organizacional que integra tecnologias digitais e processos de negócios em uma economia digital.
Matt <i>et al.</i> (2015), p. 339	Estratégias de transformação digital assumem uma perspectiva diferente e buscam diferentes objetivos. Vindo de uma perspectiva centrada nos negócios, essas estratégias concentram-se na transformação de produtos, processos e aspectos organizacionais devido às novas tecnologias.
Parviainen <i>et al.</i> (2017), p. 64	transformação digital é definida como mudanças nos métodos de trabalho, funções e ofertas comerciais causadas pela adoção de tecnologias digitais em uma organização ou no ambiente operacional da organização.
Bondar, Hsu, Pfouga, e Stjepandić (2017), p. 33	A Transformação Digital é uma rede consistente de todos os setores econômicos e uma adaptação dos atores às novas circunstâncias da economia digital.
Schallmo <i>et al.</i> (2019); Schallmo, Williams,	O <i>framework</i> da Transformação Digital inclui a interconexão de atores como empresas e clientes em todos os segmentos da cadeia de valor, e a aplicação de novas tecnologias. Como tal, a Transformação Digital

Boardman <i>et al.</i> (2019), p. 4	requer habilidades que envolvem a extração e troca de dados, bem como a análise e conversão desses dados em informações acionáveis. Essas informações devem ser usadas para calcular e avaliar opções, a fim de permitir decisões e/ou iniciar atividades. Para aumentar o desempenho e o alcance de uma empresa, a Transformação Digital envolve empresas, modelos de negócios, processos, relacionamentos, produtos etc.
Hinings, Gegenhuber, e Greenwood (2018), p. 53	A Transformação Digital é o efeito combinado de várias inovações digitais que trazem novos atores (e constelações de atores), estruturas, práticas, valores e crenças que alteram, ameaçam, substituem ou complementam as regras existentes dentro de organizações, ecossistemas, indústrias ou áreas específicas.
Heavin e Power (2018), p. 40	Embora a transformação digital tenha seus desafios, pesquisas existentes indicam que o fenômeno digital é uma oportunidade para inovar e redefinir como as organizações fazem negócios. Os dois principais aspectos da transformação digital são definidos em termos de (1) tecnologia e (2) cliente ou usuário.
Vial (2019), p. 121	A transformação digital é um processo que visa aprimorar uma entidade ao desencadear mudanças significativas em suas propriedades por meio de combinações de tecnologias de informação, computação, comunicação e conectividade.
Warner e Wager, 2019, p. 344	A transformação digital é um processo contínuo de renovação estratégica que utiliza avanços em tecnologias digitais para construir capacidades que atualizam ou substituem o modelo de negócio, abordagem colaborativa e cultura de uma organização.

Fonte: Adaptado de Kraus *et al.*, 2022.

Como pode ser percebido, a conceituação de transformação digital é sensivelmente diferente de acordo com a abordagem de cada autor ao tema. É notória a presença de três conceitos que se inter-relacionam permanentemente nas abordagens teóricas. Trata-se dos conceitos de digitização, digitalização e transformação digital.

De acordo com Eller *et al.* (2020) de uma forma geral, as empresas se movem em direção a nova economia digital passando por três fases, cada uma delas com seus próprios desafios, recursos, estrutura e métricas. A primeira fase é chamada de “digitização” e pode ser entendida como um processo técnico que envolve transformar processos analógicos em digitais, ao mesmo tempo em que passa a criar dados para sistemas de informações e para a empresa como um todo (Verhoef *et al.*, 2021).

A segunda fase é descrita por Eller *et al.* (2020) como “digitalização”. Trata-se de um processo sociotécnico no qual as informações geradas pelos sistemas implementados na fase anterior são lidas pelos gestores organizacionais e dado início a um novo período de ajustes nas empresas, que passam a alterar os seus processos existentes a partir da leitura que os dados permitem ter sobre métricas e objetivos empresariais. Nessa fase as empresas começam a experimentar novas possibilidades como artefatos digitais e tecnologias como habilitadoras de novas possibilidades (Verhoef *et al.*, 2021).

Por fim, a terceira fase compreende a transformação digital enquanto fenômeno que altera as bases de competição da empresa através de um processo integrado entre a tecnologia e a estratégia organizacional para repensar a forma como a empresa cria valor, entrega valor para o cliente e captura valor (Rohn *et al.*, 2021). A transformação digital é o terceiro e último estágio de metamorfose empresarial e tem relação com repensar a empresa para competir em uma nova economia de mercado com um novo modelo de negócio trazendo a tecnologia para o centro da estratégia (Verhoef *et al.*, 2021). No entanto é importante entender que a transformação digital é um processo fundamentalmente sobre estratégia e não sobre tecnologia (Rogers, 2016).

Desta forma, as empresas que estão trilhando esse caminho não podem ser tratadas da mesma forma, uma vez que naturalmente haverá empresas menos avançadas e ainda iniciando atividades de sua digitização, enquanto outras já estarão em processos mais avançados repensando modelos de negócio e uso de artefatos digitais como estratégia de diferenciação e criação de vantagens competitivas.

Dada essa variação de estágios quanto ao processo de transformação das empresas, sugere-se uma abordagem para entender em que nível de maturidade digital as empresas estão e assim propor soluções levando em consideração o momento de cada uma.

2.2 MATURIDADE DIGITAL

Os modelos de maturidade digital estão se tornando cada vez mais populares, pois combinam as melhores práticas e recursos disponíveis para avaliar o comportamento das organizações. Eles se baseiam nos níveis de habilidade em tecnologia digital das

organizações para otimizar a transformação digital. Esses modelos visam identificar áreas de melhoria por meio da digitalização, revelando tanto fraquezas quanto pontos fortes. As ações resultantes são comunicadas de forma clara e documentadas minuciosamente, utilizando métodos qualitativos e quantitativos (Kirmizi; Kocaoglu, 2022).

Embora as tecnologias digitais sejam essenciais para a transformação digital das organizações, Dutta *et al.* (2021) ressaltam que, para atingir o nível de maturidade digital desejado, as organizações devem dar prioridade à organização do trabalho, às pessoas e às práticas, considerando-as como elementos-chave que sustentam a tecnologia na execução de procedimentos para alcançar um desempenho superior, ou seja, a tecnologia é meio e não o fim em si mesma.

Para Lobão e Zilli (2020) maturidade digital não acontece de um dia para o outro. Na opinião dos autores, a maturidade digital extrapola a simples implementação de novas ferramentas tecnológicas. Ela consegue alinhar a estratégica, a força de trabalho, a cultura e a estrutura da empresa para atender às expectativas digitais de clientes, funcionários e parceiros. A maturidade digital é, portanto, um processo contínuo de adaptação a um cenário que não para de mudar. Nesse sentido, maturidade é algo a ser perseguido em uma jornada de transformação digital, uma consequência de se trilhar um caminho que busca desenvolver a empresa para atuar digitalmente. Essa jornada precisa conseguir alinhar a estratégia de negócios da empresa com sua estratégia digital. Para Saldanha (2019) a maturidade digital é um resultado a ser perseguido pelas organizações que querem ter o “DNA digital” em seus negócios.

Considerando que a maturidade digital permite que as organizações avancem em direção à transformação digital, Gollhardt *et al.* (2020) e Gökalp e Martinez (2021) ressaltam que o objetivo da transformação digital é agregar valor ao negócio por meio de uma mudança focada no aprimoramento do desempenho organizacional. Isso é alcançado pela otimização dos processos, aumento da produtividade e criação de novos segmentos de mercado, através de um contínuo processamento de informações.

2.3 CLUSTERIZAÇÃO

Os algoritmos de aprendizado de máquina são geralmente classificados em dois cenários principais: (i) cenários supervisionados, que são direcionados por um objetivo específico. Esses cenários envolvem um conjunto de dados rotulados com as respostas corretas, onde o objetivo é aprender a mapear entradas e saídas para identificar os valores corretos fornecidos pelo supervisor (Aggarwal, 2015; Alpaydin, 2016; Grus, 2015); e ii) cenários não supervisionados, que, nesses casos, não há um conjunto de dados específico para "ensinar" o modelo sobre agrupamentos adequados. Em vez disso, o modelo faz suposições sobre as propriedades estruturais dos dados sem a orientação de um supervisor, utilizando apenas os dados de entrada (Aggarwal, 2015; Alpaydin, 2016; Jordan *et al.*, 2015). Esses cenários são aplicados para resolver problemas de aprendizagem, conforme discutido por Jordan e Mitchell (2015). Os dados supervisionados são chamados de classificação e os dados não supervisionado de clusterização. A classificação e a clusterização são dois métodos fundamentais na área de aprendizado de máquina e análise de dados. Ambos possuem características distintas e aplicações específicas.

Como o foco da dissertação será a clusterização, não será aprofundado os conceitos e tipos de classificações.

Deng *et al.* (2016) ressaltam que a clusterização é uma técnica fundamental na ciência de dados, sendo amplamente empregada para segmentar bases de dados. Rodrigues *et al.* (2017) definem a clusterização, ou o problema da clusterização, como o processo de agrupar dados em conjuntos, onde os objetos dentro de um mesmo grupo tendem a ser mais semelhantes entre si do que em relação aos objetos de outros grupos. Essa definição pressupõe a existência de uma medida de qualidade que avalia a similaridade ou dissimilaridade entre os dados.

De acordo com De Morsier *et al.* (2015), o desafio de separar conjuntos de dados em diferentes grupos é amplamente estudado em várias áreas onde a informação semântica dos dados não está disponível. Dessa forma, a clusterização é classificada como uma técnica de aprendizagem não supervisionada, que oferece diferentes perspectivas sobre a estrutura inerente dos dados, organizando-os em múltiplos grupos, que podem ser sobrepostos ou separados (Jayaram Reddy *et al.*, 2018).

Kassambara (2017) complementa essa visão, observando que na literatura a clusterização é frequentemente denominada "reconhecimento de padrões" ou "aprendizagem de máquina não supervisionada". O termo "não supervisionada" refere-se à ausência de diretrizes prévias sobre quais variáveis ou amostras pertencem a cada *cluster*, enquanto "aprendizagem" diz respeito ao processo pelo qual o algoritmo "aprende" a agrupar os dados.

O resultado de uma clusterização é chamado de *cluster*, sendo uma coleção de objetos similares, distintos dos objetos de outros *clusters* (Omrani; Shafaat; Emrouznejad, 2018).

Liu *et al.* (2016) afirmam que a clusterização pode ser realizada por meio de diversos algoritmos, cada um com sua própria noção do que constitui um *cluster* e de como identificá-los de maneira eficiente. Segundo Rodriguez *et al.* (2017), essa técnica de análise de dados está intimamente associada à criação de modelos que simplificam as propriedades de um conjunto de dados, permitindo uma compreensão intuitiva dos aspectos mais relevantes. Embora os métodos de clusterização sejam geralmente mais complexos do que as abordagens supervisionadas, eles oferecem uma compreensão mais profunda de dados complexos.

Goldschmidt, Passos e Bezerra (2015) dividem os métodos de clusterização em três famílias básicas: baseados em distâncias, baseados em distribuição de probabilidades e os algoritmos de agrupamento baseados em densidade. Os autores apresentam também outro modelo de categorizar que é baseado na forma do agrupamento gerado. Nesse contexto, segundo os autores, existem os algoritmos de agrupamento que geram partição (Algoritmos Partitivos) e os que geram hierarquia (Algoritmos Hierárquicos).



De forma a organizar o raciocínio e a condução da análise das informações levantadas, a metodologia científica a ser adotada é o método de pesquisa experimental. Para Gil (2010) a pesquisa experimental constitui o delineamento mais prestigiado dos meios científicos. Consiste essencialmente em determinar um objeto de estudo, selecionar as variáveis capazes de influenciá-lo e definir as formas de controle e de observação dos efeitos que a variável produto no objeto.

3.1 ALGORITMOS DE CLUSTERIZAÇÃO ADOTADOS

Este trabalho focará em 4 algoritmos de clusterização que são: Hierárquico, *k-means*, *k-modes* e *Partitioning Around Medoids* (PAM).

Os algoritmos de clusterização são utilizados com muita frequência em aplicações que necessitem de busca por padrões. O algoritmo hierárquico é uma abordagem que constrói uma hierarquia de *clusters*. Ela pode ser aglomerativa (*bottom-up*) ou divisiva (*top-down*). De acordo com Rencher (2002), técnicas hierárquicas envolvem um processo de múltiplos passos sequenciais. No método aglomerativo, conforme Hair et al. (2010), as amostras são agrupadas com base em suas distâncias, geralmente euclidianas; os *clusters* formados continuam sendo agrupados conforme sua proximidade até que todos se unam em um único grupo. Por outro lado, os métodos divisíveis seguem o caminho oposto, começando com um único *cluster* que é dividido sucessivamente com base em sua distância interna. Uma vez que um objeto é alocado em um grupo, ele não pode ser realocado, assegurando a hierarquia do processo. Esses métodos geralmente são representados por dendrogramas (Rencher, 2002). É importante destacar que a clusterização hierárquica não requer que o número de *clusters* seja pré-definido e o resultado pode ser representado em um dendrograma, que mostra a fusão ou divisão dos *clusters*.

Já os métodos não hierárquicos alocam objetos a *clusters* com base em um número pré-definido de grupos (Corrar et al., 2009). Exemplos incluem o *k-means*, *k-modes* e o *Partitioning Around Medoids* (PAM). O *k-modes*, uma variação do *k-means*, é usado para

dados categóricos e usa a "moda" (valor mais frequente) para definir os centros dos *clusters*, em vez da média.

O algoritmo de clusterização *k-means*, também conhecido como k-médias, é amplamente utilizado devido à sua simplicidade e eficiência. Segundo Jain *et al.* (1999), o *k-means* é popular principalmente por sua facilidade de implementação e sua baixa complexidade computacional. Fontana e Naldi (2009) destacam que o *k-means* utiliza o conceito de centroides como protótipos representativos dos *clusters*. Esses centroides são calculados como a média de todos os objetos pertencentes ao grupo, representando, assim, o centro do *cluster* de maneira eficaz.

O PAM é um algoritmo de clusterização que utiliza medoides, representações dos *clusters* similares aos centroides do *k-means*, mas que sempre correspondem a pontos reais dos dados. Esses medoides são os pontos mais centralmente localizados em cada agrupamento, minimizando a distância entre eles e os demais pontos do *cluster* (Kassambara, 2017; Maechler *et al.*, 2017). Inicialmente, na fase de construção, os medoides são escolhidos para os *k clusters* desejados e, em seguida, cada ponto não medoide é associado ao medoide mais próximo (Kassambara, 2017). Após isso, a fase de intercâmbio ocorre, onde medoides e não medoides são trocados para otimizar a função objetivo, melhorando a qualidade dos agrupamentos (Kassambara, 2017). O PAM é mais robusto que o *k-means* contra *outliers* e ruídos nos dados (Kassambara, 2017; Maechler *et al.*, 2017).

Os métodos de clusterização podem ser aplicados a diferentes tipos de dados. O *k-means* e o PAM são mais adequados para dados numéricos, pois dependem de medidas de distância como a Euclidiana. O *k-modes*, por outro lado, é especificamente projetado para dados categóricos, pois utiliza a moda como centro dos *clusters*. Já o algoritmo hierárquico pode ser adaptado tanto para dados numéricos quanto categóricos, dependendo da métrica de similaridade escolhida. Assim, a seleção do método apropriado depende diretamente da natureza dos dados em análise.

Este trabalho utilizará os dados categóricos, colhidos na mensuração do Índice de Maturidade Digital - IMD da ABDI, que são: estado, região, porte e setor das empresas participantes. Os dados numéricos utilizados são o resultado do índice agregado de maturidade digital da ABDI, que é computado com a média da

pontuação das cinco dimensões aplicadas no questionário da pesquisa, que são: gestão, pessoas, processos, modelo de negócios e vendas e consumidor.

3.2 PREPARAÇÃO DOS DADOS

Como mencionado, os dados a serem utilizados são oriundos da mensuração do Índice de Maturidade Digital - IMD da ABDI. Trata-se de pesquisa iniciada em 2021, com o objetivo de monitorar o nível de maturidade digital das empresas do setor produtivo brasileiro no período de 36 meses, com 6 ciclos de aplicação, a fim de subsidiar formulação de políticas públicas relacionadas à transformação digital. O primeiro ciclo aconteceu no primeiro semestre de 2021, e o sexto e último no segundo semestre de 2023. Para a pesquisa, foram definidos os setores: agropecuária, comércio, construção civil, indústria e serviços. Em painéis semestrais, foram ouvidos 2.000 (dois mil) executivos de empresas (matriz ou filiais) do setor produtivo brasileiro.

O questionário utilizado foi composto por 2 blocos, sendo o primeiro relativo ao perfil da empresa e do respondente e o segundo contendo 21 questões quantitativas, de múltipla escolha, pertinentes à mensuração da maturidade digital das empresas, distribuídas em 5 dimensões, com 4 questões específicas em cada uma delas. Foi adotada a anonimização do respondente para atendimento às normas da Lei Geral de Proteção de Dados - LGPD. O questionário foi testado em aplicação piloto.

As 5 dimensões que compõe o índice de maturidade digital da ABDI, relacionadas aos esforços de transformação tecnológica das empresas são: gestão, pessoas, processos, modelo de negócios, vendas e consumidores.

A amostra possui representação nacional e setorial, sendo consultadas Microempresas - ME, Empresas de Pequeno Porte - EPP, e Média e Grande Empresas - Demais. O plano amostral foi estruturado de acordo com a base de dados disponibilizada pela ABDI com tamanho populacional de N=1.261.338 (setor e unidade da federação - UF) e pela base de informações da Receita Federal (2020) com tamanho populacional de N=9.417.542 (excluídas os Microempreendedores Individuais - MEIs). Como a coleta foi feita por amostragem aleatória simples, onde a probabilidade de participar da pesquisa é a mesma para todas as empresas que constam no banco de dados enviado pela

ABDI, o plano amostral foi proporcional e relacionou o setor da empresa com o porte em relação aos estados, objetivando o equilíbrio entre população e amostra.

Foi realizado um cálculo de proporção do quantitativo de empresas por setor em suas respectivas UFs em relação à proporção do porte das empresas fornecidos pela Receita Federal, proporcionando o equilíbrio entre setor e porte por UFs no plano amostral.

Importante destacar que neste trabalho foram utilizados apenas os dados referentes a última aplicação da pesquisa (6ª aplicação realizada no segundo semestre de 2023) e que as entrevistas se deram por meio de aplicativo, presenciais e por telefone, nas 27 (vinte e sete) unidades da federação, incluindo o Distrito Federal.

Após a aplicação da pesquisa, para cada resposta do questionário, foi atribuído um valor numérico que possibilitou estabelecer faixas e/ou categorias para IMD. Assim sendo, ABDI criou 4 (quatro) categorizações conforme as faixas de pontuação estabelecidas. A pontuação vai de 0 a 100 e a classificação ficou da seguinte forma:

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

Para cada dimensão do questionário, a pontuação foi calculada como uma média simples de todos os indicadores, isto é, cada indicador tem o mesmo peso no cálculo da pontuação. O índice agregado de maturidade digital é computado como a média da pontuação em cada dimensão.

$$score_{i,f} = \left(\frac{value_{i,f} - wp_i}{frontier_i - wp_i} \right) \cdot 100$$

Sendo:

- a) Score: Pontuação do indicador i para a firma f.**
- b) Value: Valor da resposta no indicador i para a firma f.**
- c) Wp: Pior valor possível de resposta no indicador i, ou menor resposta aceitável.**

- d) **Frontier:** Melhor cenário possível do indicador i. Pode ser determinada com base no valor máximo possível para a resposta, ou o melhor valor observado na amostra.
- e) **Score** é sempre limitado entre 0 e 100.

Para fins de comparação com os algoritmos de clusterização adotados, foi atribuído um valor a cada classificação dada pela ABDI, onde foi considerado:

- a) **1 para empresas consideradas iniciantes;**
- b) **2 para empresas consideradas emergentes;**
- c) **3 para empresas consideradas intermediárias; e**
- d) **4 para empresas consideradas líderes.**



4

RESULTADOS

4.1 APLICAÇÃO DA CLUSTERIZAÇÃO NOS DADOS

Considerando que a base de dados continha dados categóricos e dados numéricos, foram testados 4 (quatro) algoritmos de clusterização que são: Hierárquico (HC), *k-mens* (*Means*), *k-modes* (*Modes*) e *Partitioning Around Medoids* (PAM). Para a avaliação dos resultados, os algoritmos de clusterização foram executados na linguagem de programação “R⁶”. Os dados são:

- a) **Categóricos: porte da empresa, setor, estado e região;**
- b) **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*)⁷, silhueta (*Silhouette Method*)⁸ e Método da Estatística Gap (*GAP Statistic Method*)⁹. 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. Na tabela abaixo é possível observar a distribuição dos dados entre os *clusters*:

⁶ A linguagem de programação R é uma linguagem de código aberto, focada em análise estatística, mineração de dados e gráficos estatísticos. Ela foi criada por Ross Ihaka e Robert Gentleman no início da década de 1990, sendo amplamente utilizada por estatísticos, cientistas de dados, analistas e pesquisadores em diversas áreas.

⁷ 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 & Makwana (2013).

⁸ O método Silhueta mede o quão bem cada ponto está agrupado, comparando a proximidade entre pontos do mesmo cluster e de clusters vizinhos. Esse método foi introduzido por Peter J. Rousseeuw (1987) e se tornou um padrão para avaliação de clusters.

⁹ O método Gap Statistic foi proposto por Tibshirani, Walther & Hastie (2001) como uma alternativa estatisticamente fundamentada para determinar k . Ele compara a dispersão dos clusters em relação a uma distribuição aleatória de referência.

Tabela 2 – Distribuição dos dados entre os algoritmos observados

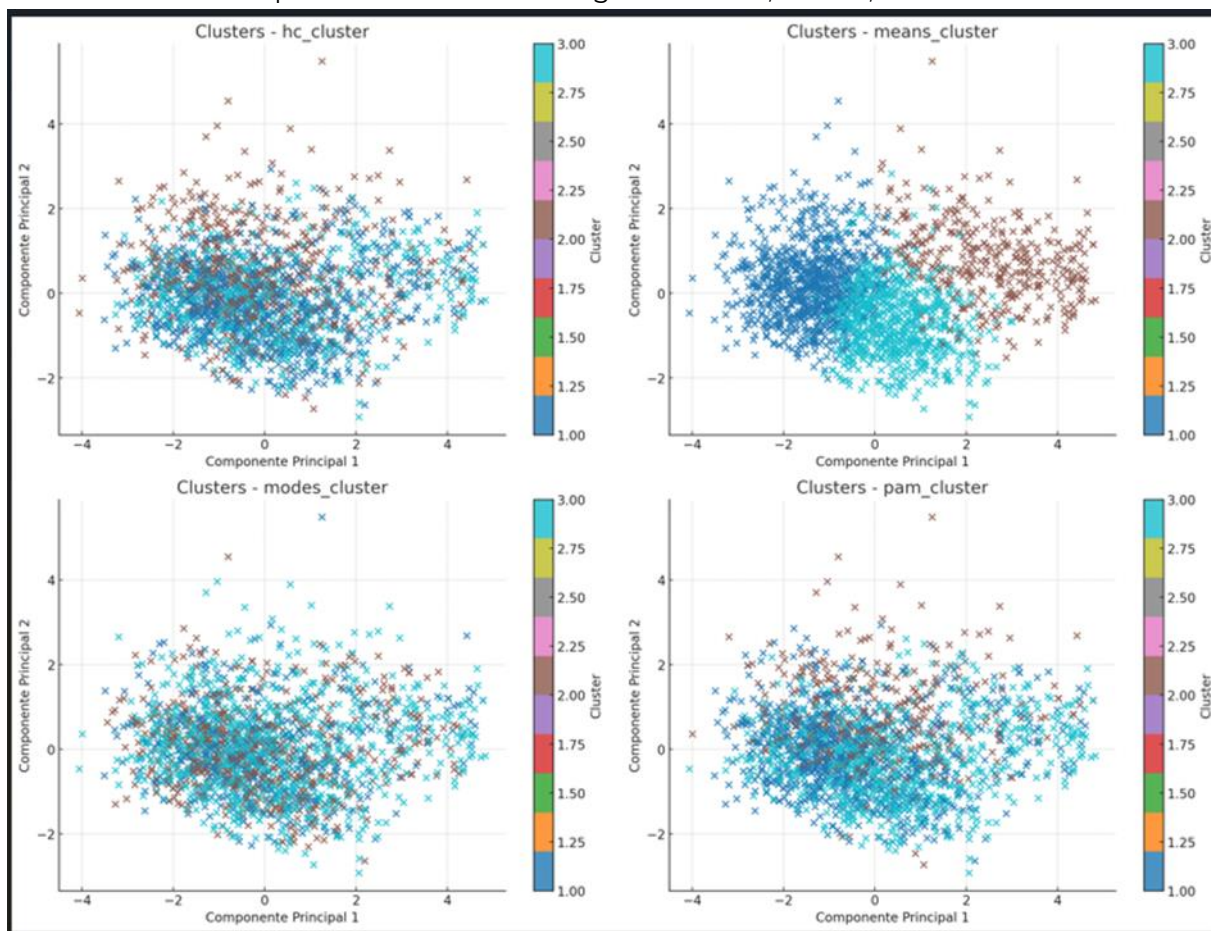
Número de Clusters	HC	Means	Modes	PAM
1	817	837	452	804
2	596	369	603	371
3	587	794	945	825
Total	2000	2000	2000	2000

Fonte: Dados elaborados pela autora.

Comparando os algoritmos, é percebido que o *cluster 3* é um grande grupo coeso entre os algoritmos *Means*, *Modes* e PAM e que o *cluster 2* apresenta menor representatividade em quase todos os métodos, sugerindo que as empresas deste grupo são mais específicas e menos representativas da base total. O *cluster 1* é o maior no HC, contudo nos métodos baseados em medidas centrais, como o *Means* e o PAM, o *cluster 3* assume a liderança. Isso porque o HC captura grupos com características amplas, enquanto os outros métodos favorecem agrupamentos mais homogêneos. No *Means* o *cluster 2* é extremamente pequeno, sugerindo que o método *Means* pode estar agrupando um subconjunto muito específico de empresas. Por outro lado, no *Modes*, o *cluster 2* é mais representativo, sugerindo que o método que utiliza a moda é mais flexível na identificação desse grupo.

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. Abaixo, segue o gráfico de dispersão dos algoritmos:

Gráfico 1 – Dispersão dos dados nos algoritmos: HC, Means, Modes e PAM



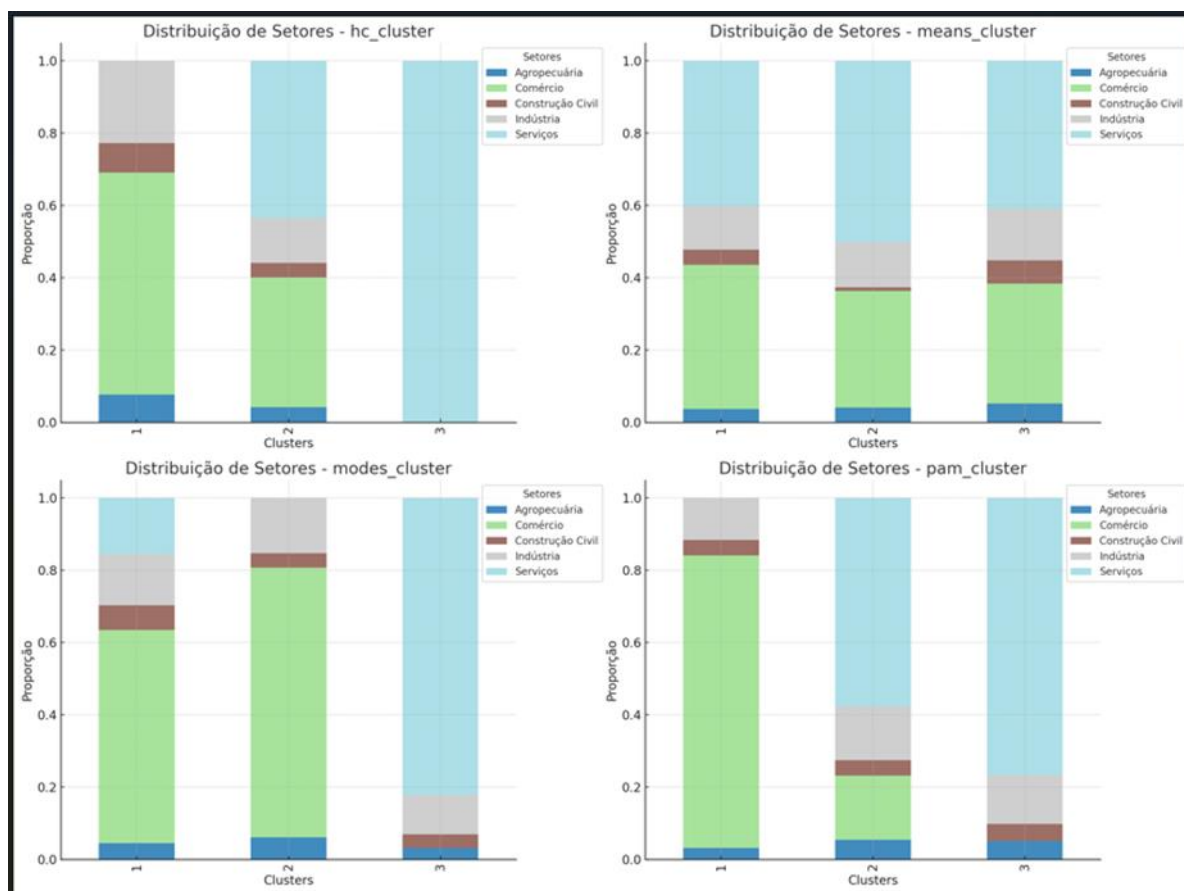
Fonte: Dados elaborados pela autora.

Cada ponto representa uma empresa, e as cores indicam o *cluster* ao qual cada empresa pertence. No Algoritmo HC os *clusters* estão bem distribuídos no espaço de projeção e apresentam alguma sobreposição em certas regiões. Isso indica que a hierarquia pode estar agrupando alguns pontos mais próximos de forma um pouco arbitrária, pois esse algoritmo tende a formar *clusters* com base em proximidade relativa, mas não otimiza a homogeneidade interna como métodos baseados em centroides. No gráfico referente ao algoritmo *Means* observa-se que os *clusters* formam grupos mais compactos e bem definidos. Isso é característico do *Means*, que minimiza a variação dentro dos *clusters*. O algoritmo *Modes* apresenta um gráfico com *clusters* mais flexíveis, com grupos mais espalhados. Há menor compactação em comparação ao *Means*, o que pode refletir uma abordagem menos restritiva. No gráfico do algoritmo PAM os *clusters* estão bem definidos, entretanto apresentam maior robustez contra *outliers* em comparação ao *Modes*. A disposição dos *clusters* é semelhante ao *Modes*, mas existem diferenças nos limites dos agrupamentos. Como o PAM usa *medoids* em vez de centroides, ele é

menos sensível a valores extremos. Em geral, considerando os gráficos de dispersão acima, observa-se uma sobreposição nos algoritmos *Modes* e o HC, possivelmente devido à flexibilidade em seus critérios de agrupamento. Já o *Means* e o PAM apresentam *clusters* mais compactos e visualmente distintos, refletindo seu foco em otimizar a variância interna.

Abaixo, segue gráfico, considerando a predominância entre os setores econômicos nos algoritmos:

Gráfico 2 – Predominância entre os setores econômicos nos algoritmos: HC, Means, Modes e PAM



Fonte: Dados elaborados pela autora.

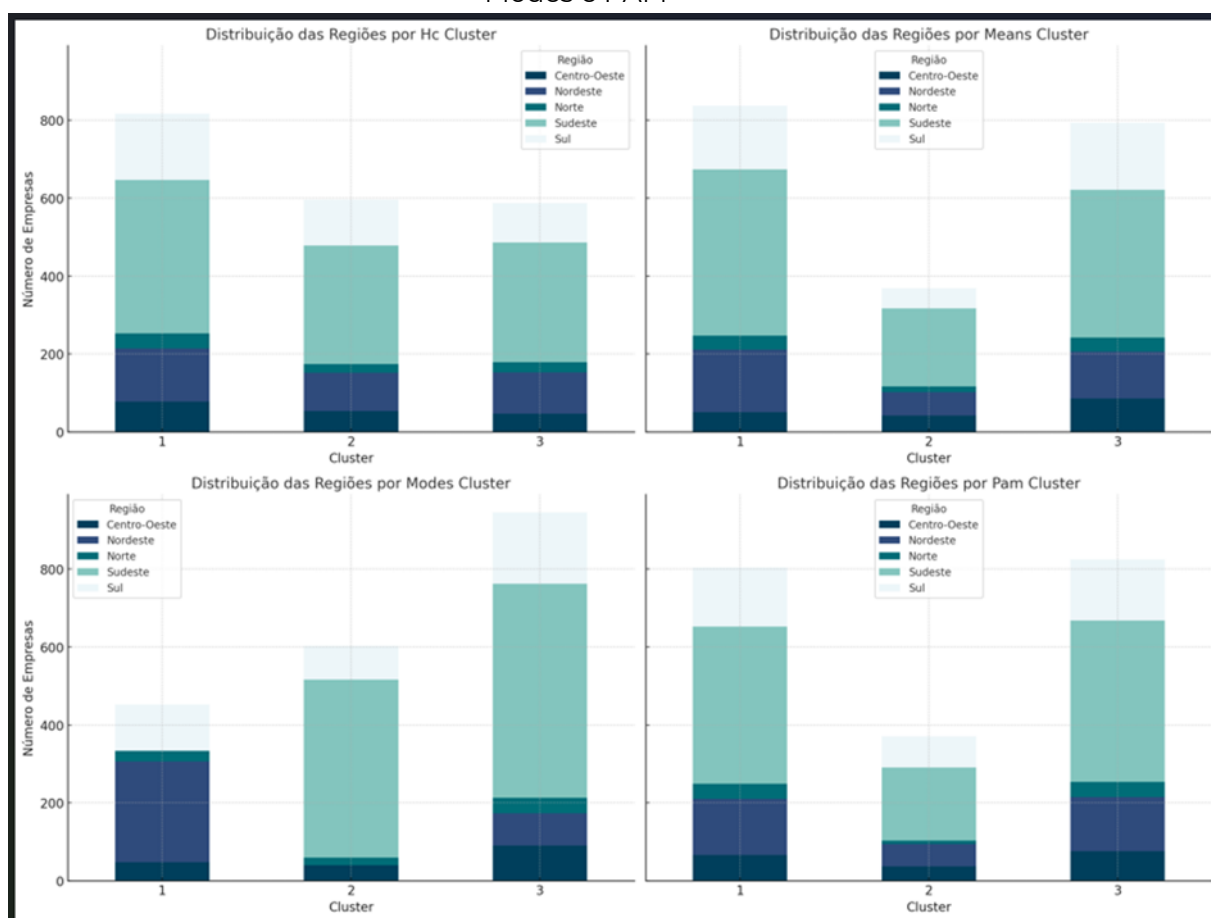
Ao analisar os gráficos, percebe-se que no algoritmo HC existe uma predominância do setor de comércio com uma presença relevante da indústria no *cluster* 1. O setor de serviço predomina seguido de comércio no *cluster* 2 e que há exclusividade do setor de serviços no *cluster* 3. Assim sendo, o algoritmo HC tende a isolar completamente empresas do setor de serviços em um único *cluster*, enquanto os outros setores são mais distribuídos nos demais *clusters*. No algoritmo *Means*, os setores de comércio e de serviços dividem a

predominância no *cluster* 1. O setor de serviços predomina no *cluster* 2 com menor presença dos outros setores. O setor de serviços mantém predominância no *cluster* 3, seguido do setor de comércio. Assim, observa-se que o *Means* forma *clusters* mais equilibrados, sem isolar completamente nenhum setor. Isso reflete a otimização baseada em centroides, que força uma segmentação mais uniforme. No algoritmo *Modes*, o comércio possui predominância no *cluster* 1 seguido da participação da indústria. O setor de comércio também domina o *cluster* 2 com presença tímida dos outros setores. Já o setor de serviços predomina no *cluster* 3, sendo o único setor significativo presente. Observa-se que esse algoritmo isola o setor de serviços no *cluster* 3 de forma semelhante ao HC e também forma um *cluster* quase exclusivo do setor de comércio no *cluster* 2, indicando maior rigidez em capturar padrões dominantes. Por fim, no algoritmo PAM evidencia-se um forte domínio do setor de comércio no *cluster* 1, seguido do setor de indústria. No *cluster* 2 o setor de serviços predomina seguido pelo setor de comércio e o setor de serviços possui maior presença no *cluster* 3. O PAM reflete maior robustez ao lidar com valores extremos, destacando o setor de comércio no *cluster* 1 e isolando o setor de serviços no *cluster* 3, que possui *clusters* menos homogêneos que o *Modes*.

Um fato que salta aos olhos é que os algoritmos diferem na forma como lidam com a predominância dos setores. Os algoritmos HC e os baseados em moda (*Modes* e PAM) são mais propensos a isolar grupos homogêneos como serviços. O algoritmo *Means* e PAM criam *clusters* mais equilibrados que podem diluir a especificidade setorial.

Abaixo, segue gráfico, considerando a predominância entre as regiões brasileiras nos algoritmos:

Gráfico 3 – Predominância entre as regiões brasileiras nos algoritmos: HC, Means, Modes e PAM



Fonte: Dados elaborados pela autora.

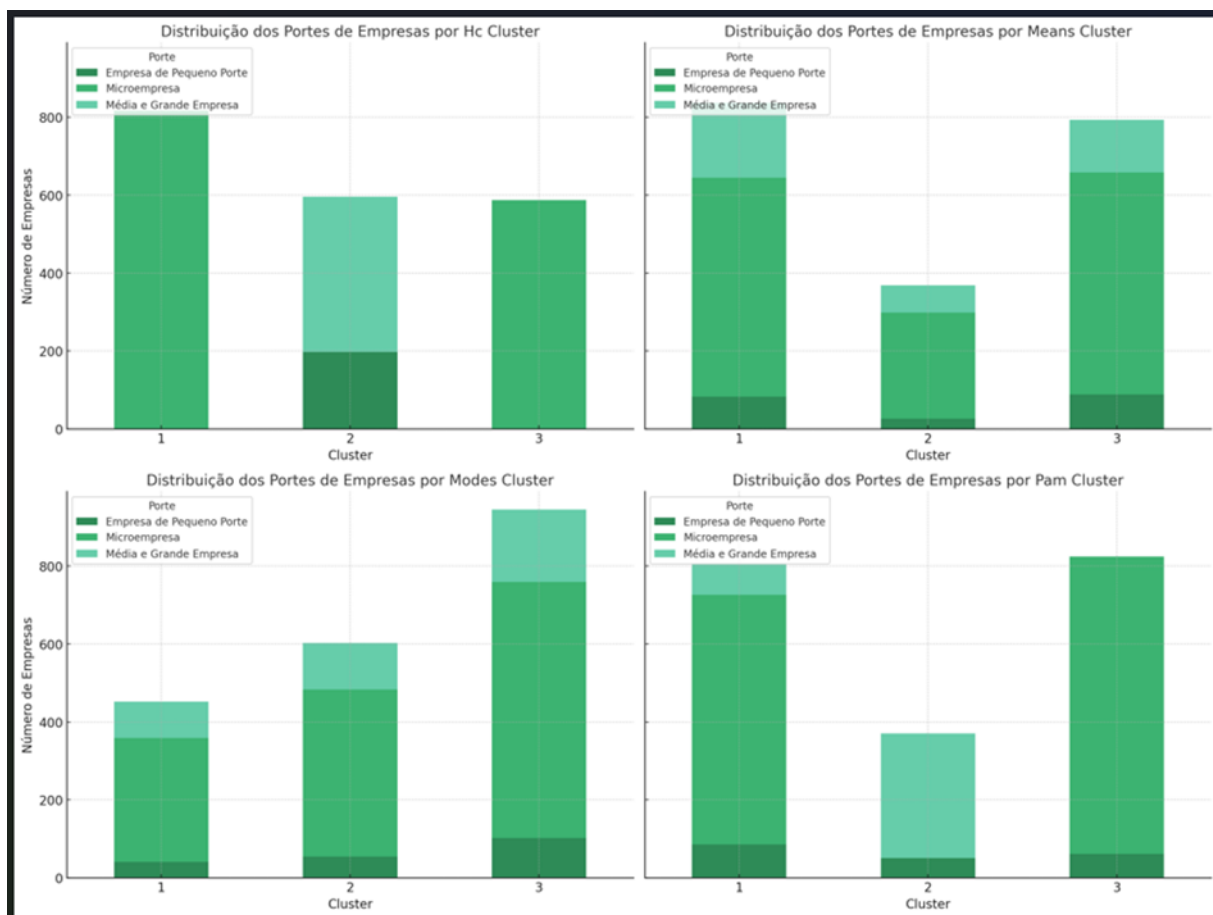
Ao analisar os gráficos, percebe-se que no algoritmo HC a região Sudeste predomina de forma significativa no *cluster* 1, seguido pelo Sul e Nordeste. Isso sugere que esse *cluster* agrupa empresas altamente concentradas nas regiões mais industrializadas. No *cluster* 2 o Sudeste ainda é majoritário e o Sul e Nordeste têm uma contribuição mais equilibrada. No *cluster* 3 a participação do Nordeste aumenta e o Sudeste ainda predomina. O HC reflete tanto a concentração econômica do Sudeste quanto a diversidade econômica de outras regiões. Já no algoritmo *Means* o *cluster* 1 ficou muito parecido com o *cluster* 1 do algoritmo HC, onde há predominância do Sudeste, com forte presença também do Sul e Nordeste. Este *cluster* pode representar empresas de maior porte ou complexidade. No *cluster* 2 há reduzida participação do Sudeste, com o Sul e Nordeste ganhando destaque. No *cluster* 3 a região Sudeste ainda é predominante. Como um algoritmo particional, o *Means* busca minimizar a variação dentro dos *clusters*, criando uma separação mais uniforme. Isso explica a redistribuição mais balanceada em relação ao HC. No algoritmo *Modes*,

o *cluster* 1 é menor, porém existe uma concentração majoritária de empresas da região Centro-Oeste. Os *clusters* 2 e 3 possuem predominância da região Sudeste, porém no *cluster* 2 não se observa empresas do Nordeste. O *Modes* cria agrupamentos com base em características modais, resultando em *clusters* com maior diversidade geográfica em comparação aos outros algoritmos. No algoritmo PAM, a região Sudeste continua predominante no *cluster* 1, seguindo o padrão do HC e *Means*, apresentando um leve aumento na participação do Nordeste em comparação aos outros algoritmos. No *cluster* 2, a região Sul ganha mais representatividade, entretanto as regiões Norte e o Centro-Oeste aparecem com maior relevância do que nos outros algoritmos. Como o PAM é mais robusto a *outliers*, e isso parece beneficiar regiões menores como Norte e Centro-Oeste na amostra, o resultado em *clusters* incluem empresas dessas regiões de forma mais consistente.

Comparando todos os algoritmos é observado que a região Sudeste lidera em todos os algoritmos e *clusters*, refletindo a sua posição econômica e o número de empresas da base de dados. As regiões Sul e Sudeste apresentam contribuições secundárias e de peso significativos, dependendo do algoritmo. As regiões Norte e Centro-Oeste têm menor representatividade, entretanto aparecem de forma mais evidente nos algoritmos PAM e *Modes*. Assim sendo o algoritmo HC tende a refletir a concentração econômica do Sudeste. Os algoritmos *Means* e PAM promovem uma redistribuição mais uniforme, capturando melhor as características regionais e o algoritmo *Modes* destaca *clusters* mais diversificados, incluindo regiões menos representadas como o Norte e o Centro-Oeste.

Abaixo, segue gráfico, considerando a predominância entre o porte das empresas nos algoritmos:

Gráfico 4 – Predominância do porte das empresas nos algoritmos: HC, Means, Modes e PAM



Fonte: Dados elaborados pela autora.

Ao analisar os gráficos, percebe-se que os *clusters* 1 e 3, do algoritmo HC, foram formados em sua totalidade por microempresas, indicando que eles representam empresas com menor capacidade financeiro, menor número de funcionários e mais distribuídas graficamente. O *cluster* 2 do algoritmo HC mesclou médias e grandes empresas com empresas de pequeno porte. Neste caso demonstra um perfil intermediário, grupando empresas com características de crescimento e expansão. No algoritmo Means existe um equilíbrio na distribuição dos portes por *clusters*. O *cluster* 1, similar ao algoritmo HC, possui predominância de microempresas. O *cluster* 2, dada a proporcionalidade com relação ao quantitativo de empresas da amostra foi que apresentou maior equilíbrio na distribuição de empresas por porte dentro do *cluster*. O *cluster* 3 ficou muito parecido com o *cluster* 1. O algoritmo Modes destaca-se o balanceamento entre os *clusters*. O algoritmo PAM tende a seguir um padrão similar ao HC, apresentando maior variabilidade entre *clusters*.

Microempresas dominam em quase todos os *clusters* e algoritmos, refletindo sua maior representatividade geral no mercado e na amostra. As pequenas empresas aparecem em menor proporção e são distribuídas de forma mais equilibrada entre os *clusters* e as médias e grandes empresas têm participações residuais, indicando que são menos representadas na amostra.

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*)¹⁰; Índice de Davies-Bouldin - DBI (*Davies-Bouldin Index*)¹¹; Valor-p do ANOVA (*ANOVA p-value*)¹²; e Índice da Silhueta (*Silhouette Score*). Os métodos de qualidade de clusterização são métricas 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.

Tabela 3 – Resultado dos métodos de qualidade					
Métodos de qualidade	HC	Means	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

¹⁰ O índice de Calinski-Harabasz (ou critério de Calinski-Harabasz) mede a qualidade da clusterização avaliando a dispersão entre e dentro dos *clusters*. Um valor maior indica que os *clusters* são mais bem separados e mais compactos. O conceito foi proposto por Caliński, T. e Harabasz, J. em 1974.

¹¹ O índice de Davies-Bouldin mede a qualidade da clusterização baseado na similaridade entre os *clusters*, ou seja, compara a compactação de um *cluster* com sua separação em relação aos outros. Um valor menor indica melhores *clusters*. O conceito foi proposto por Davies, D. L., e Bouldin, D. W. em 1979.

¹² O p-valor do ANOVA é usado para avaliar se os *clusters* são estatisticamente diferentes entre si, comparando a variância intra e inter-cluster. Um p-valor pequeno (< 0.05) indica que os *clusters* são significativamente diferentes conforme Fisher, R. A. (1925).

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.

Com base nos resultados, o método *Means* foi o melhor método de clusterização. Ele apresentou o maior índice de CH, no valor de 630,96, indicando uma separação forte e bem definida dos *clusters*. Além disso, teve o menor índice DBI, no valor de 1,53, sugerindo que os *clusters* estão bem separados e compactos. O p-valor baixo, de 6,46e-82, reforça a significância estatística desse agrupamento, demonstrando que há diferenças claras entre os *clusters*. Finalmente, o Índice de Silhueta no valor de 0,23, o maior entre os métodos avaliados, sugere que a estrutura dos *clusters* está relativamente bem definida.

Como uma segunda opção viável ficou o método de categorização da ABDI, o IMD ABDI, que apresentou um índice de DBI no valor de 1,70, um índice de Silhueta positivo, no valor de 0,11 e um p-valor de 0, indicando alta significância estatística. No entanto, seu índice de CH ficou em 325,70, que sugere que a categorização pode não ser tão forte quanto ao *Means*.

Por outro lado, os algoritmos de clusterização *Modes* e HC tiveram os piores desempenhos na avaliação. Ambos apresentaram os Índices de Silhueta negativos, o que sugere que muitos pontos podem estar classificados incorretamente nos *clusters*. Além disso, os índices de DBI desses algoritmos foram muito altos, 17,69 e 9,55, respectivamente, indicando que os *clusters* gerados estão muito sobrepostos e mal definidos.

Dessa forma, diante do resultado da avaliação de qualidade dos agrupamentos, recomenda-se a utilização do algoritmo *Means* como o melhor método de clusterização, pois apresentou a melhor separação, compactação e significância estatística. Se for necessária uma segunda abordagem, o IMD ABDI pode ser considerado, embora tenha uma performance um pouco inferior. Os algoritmos *Modes*, e HC devem ser evitados, pois apresentam uma definição fraca de *clusters*.

4.3 ANÁLISE SOB A PERSPECTIVA DAS MÉDIAS

Considerando que o resultado da qualidade dos agrupamentos foi o *Means* como primeira opção e o IMD da ABDI como segunda, optou-se também em avaliar o resultado da média dos *clusters* por dimensão (gestão, pessoas, processos, modelo de negócios e vendas e consumidores), desses dois agrupamentos e o perfil das empresas próximas a essa média como forma de enriquecer a análise.

As médias indicam a distribuição central das empresas dentro de cada uma das dimensões por cluster dos algoritmos *Means* e da categorização IMD ABDI. Para compreender melhor as características das empresas que estão dentro dessa faixa, identificou-se aquelas cujos valores estão próximos à média, dentro de uma margem de variação de ± 0.2 .

Considerando o algoritmo *Means*, na tabela abaixo é possível observar o valor médio por dimensão em cada *cluster*:

Tabela 4 – 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.

O *cluster* 1 apresenta os menores valores médios em quase todos os eixos, sugerindo que essas empresas possuem mais dificuldades estruturais e operacionais. As empresas desse *cluster* apresentam o menor desempenho em gestão, com uma média de 4,93. Isso sugere que as empresas próximas a essa média possuem uma estrutura gerencial pouco desenvolvida, com baixa formalização de processos administrativos e tomada de decisão menos estratégica. Esse cenário pode estar relacionado à falta de capacitação dos gestores ou à ausência de metodologias de gestão eficazes. Empresas com baixa maturidade gerencial tendem a ter dificuldades na alocação de

recursos, na delegação de tarefas e no acompanhamento de indicadores de desempenho, tornando-se mais vulneráveis a crises e mudanças de mercado.

Na dimensão pessoas, o *cluster* 1 apresentou uma média de 5,79 indicando possível dificuldade dessas empresas na gestão de equipe e na retenção de talentos. Esse resultado sugere que os funcionários podem não estar suficientemente engajados, e as práticas de desenvolvimento de equipe podem ser limitadas. Empresas com baixa pontuação nesse eixo geralmente têm alta rotatividade de funcionários, menos investimento em capacitação e menor nível de satisfação interna (clima organizacional). A falta de um ambiente de trabalho estruturado pode impactar diretamente a produtividade e a inovação dentro da empresa. Funcionários sem treinamento adequado e sem uma cultura organizacional forte tendem a apresentar menor eficiência e comprometimento, o que reflete no desempenho geral do negócio.

As empresas do *cluster* 1 apresentam uma média razoável em processos, no valor de 6,58, indicando que, apesar das dificuldades em gestão e pessoas, há uma certa organização operacional. No entanto, essa pontuação ainda não demonstra um nível elevado de maturidade digital, sugerindo que as empresas podem operar com processos pouco otimizados, com falta de padronização e baixa automação. A presença de processos medianos pode indicar que essas empresas possuem algum nível de controle operacional com baixa eficiência. Elas podem enfrentar gargalos produtivos, dificuldades na integração de setores e problemas na implementação de tecnologias que melhorem a eficiência.

A média do *cluster* 1 no eixo modelo de negócios foi de 5,27, o que pode demonstrar que as empresas próximas a média não apresentam um modelo de negócios bem estruturado. Isso pode significar dificuldades em definir propostas de valor claras, segmentar corretamente o público-alvo e adaptar-se às mudanças do mercado. Muitas dessas empresas podem estar presas a modelos ultrapassados e não conseguir diversificar suas fontes de receita, tendo mais dificuldade com modelagens digitais dos negócios. A dificuldade em estruturar um modelo de negócios pensando no mercado digital pode reduzir a competitividade dessas empresas. Sem uma estratégia bem definida, elas podem perder oportunidades de mercado e ter dificuldades em captar investimentos ou escalar suas operações.

Ao analisar a média da dimensão vendas e consumidores do *cluster 1*, que foi de 5,07, observa-se que as empresas desse cluster podem enfrentar dificuldades significativas em atrair e fidelizar clientes. Isso pode ser reflexo de estratégias de *marketing* ineficazes, baixa presença digital e dificuldades em entender as necessidades do consumidor. Empresas com essa pontuação podem depender excessivamente de clientes recorrentes, sem um fluxo constante de novos consumidores. A ausência de estratégias sólidas de captação e retenção de clientes pode gerar impactos negativos na sustentabilidade do negócio. Sem um plano claro de vendas e relacionamento com o consumidor, ampliado para o mercado digital, essas empresas podem ter dificuldades em crescer e se consolidar no mercado.

Considerando o resultado das médias por dimensão do *cluster 1*, pode-se entender que se trata de agrupamento de empresas com menor maturidade digital, precisando evoluir em todas as dimensões. Diante disso, políticas públicas mais aderentes a essa realidade devem considerar as seguintes diretrizes:

- a) Foco em capacitação gerencial e técnica: a baixa média na dimensão gestão evidencia a necessidade de programas voltados à formação de lideranças, qualificação de gestores e introdução de metodologias modernas de gestão. A política pode incluir mentorias, consultorias técnicas subsidiadas, cursos online gratuitos e diagnósticos personalizados para melhoria da governança organizacional.**
- b) Desenvolvimento de capital humano: a média de 5,79 na dimensão pessoas sugere fragilidade na gestão de equipes, baixa retenção de talentos e clima organizacional comprometido. Políticas devem incentivar programas de implementação de departamentos de recursos humanos mais estratégicos, iniciativas de educação corporativa, subsídios à contratação e capacitação de colaboradores, além de ações voltadas à melhoria do ambiente de trabalho e engajamento dos times.**
- c) Apoio à estruturação e digitalização de processos: apesar de uma média um pouco mais elevada na dimensão processos, ainda há sinais de baixa automação e padronização. A política pública pode oferecer acesso a tecnologias básicas, apoio para mapeamento e reestruturação de processos, e iniciativas para transformação digital gradual, com foco em pequenas**

soluções de alto impacto, como por exemplo a implementação e automação do processo de vendas, de gestão de estoques etc.

- d) Inovação no modelo de negócios:** com uma média de 5,27, há indícios de modelos de negócios defasados ou pouco adaptados ao mercado digital. O Estado pode incentivar laboratórios de inovação para Micro e Pequenas Empresas (MPEs), programas de modelagem de negócios, e apoio para reestruturação de ofertas, canais e estratégias comerciais. Parcerias com startups, universidades e *Hubs* de inovação também podem ser estratégias eficazes.
- e) Fortalecimento da presença no mercado e nas vendas:** a média de 5,07 em vendas e consumidores sugere que muitas dessas empresas têm baixa presença digital e dificuldade de captação e fidelização de clientes. A política pública pode fomentar ações de marketing digital, formação em vendas, apoio para entrada em *marketplaces*, e desenvolvimento de canais digitais de relacionamento com o consumidor.
- f) Abordagem segmentada e progressiva:** a política deve reconhecer que essas empresas partem de um nível inicial de maturidade e, portanto, precisam de intervenções graduais, customizadas e com baixo grau de complexidade tecnológica inicialmente, evoluindo conforme os ganhos em capacidade interna.

Ao analisar o *cluster* 2 percebe-se que apresenta os maiores valores médios em todas as dimensões, sugerindo um grupo de empresas mais estruturadas. As empresas do *cluster* 2 se destacam na dimensão gestão, com uma média de 8,33, indicando forte maturidade gerencial. Essas empresas provavelmente possuem liderança bem definida, processos administrativos organizados e uma cultura de gestão baseada em dados e métricas de desempenho. Elas demonstram capacidade de planejamento estratégico, eficiência na alocação de recursos e maior resiliência diante de desafios do mercado. A presença de uma gestão sólida permite que essas empresas cresçam de forma sustentável e inovem constantemente. Além disso, um bom nível de governança corporativa facilita a captação de investimentos e o desenvolvimento de parcerias estratégicas que favorecem a transformação digital.

O *cluster* 2 apresenta um alto desempenho na dimensão pessoas, com uma média de 9,14, sugerindo que essas empresas

próximas a média investem fortemente na qualificação e no bem-estar dos funcionários. Provavelmente, possuem políticas estruturadas de recrutamento, treinamento contínuo e retenção de talentos, criando um ambiente de trabalho positivo e produtivo. Empresas com alta pontuação nessa dimensão geralmente têm colaboradores mais engajados e motivados, resultando em maior produtividade e inovação. Esse fator pode ser um diferencial competitivo importante, garantindo que a empresa mantenha um time qualificado e alinhado com os objetivos organizacionais, favorecendo um processo de transformação digital nos negócios.

Apesar do forte desempenho nas dimensões gestão e pessoas, a dimensão processos no *cluster 2* apresentou uma média de 6,54, indicando que ainda há espaço para melhorias na padronização e automação operacional. Essas empresas podem estar em uma fase de transição, buscando adotar novas tecnologias e metodologias para aprimorar sua eficiência. A otimização dos processos internos é essencial para que essas empresas mantenham sua competitividade e consigam escalar suas operações.

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.

A média elevada de 9,08 na dimensão vendas e consumidores do *cluster 2* demonstra que, possivelmente, essas empresas dominam estratégias de vendas e relacionamento com clientes. Provavelmente, utilizam canais digitais de forma eficiente, possuem estratégias bem estruturadas de *marketing* e trabalham com atendimento ao cliente de alta qualidade. Esse desempenho sugere que essas empresas conseguem crescer rapidamente e fidelizar consumidores, garantindo previsibilidade na receita e maior competitividade no mercado.

Considerando o resultado das médias por dimensão do *cluster 2*, pode-se entender que se trata de agrupamento de empresas mais maduras digitalmente, mostrando que essas empresas têm forte capacidade de inovação, gestão estruturada e estratégias eficazes para lidar com o consumidor digital. Esse grupo provavelmente já utiliza ferramentas digitais, tais como: automação de processos, e-commerce

e *marketing* digital, colocando-as à frente dos outros dois *clusters*. Esse diagnóstico impõe uma abordagem distinta de política pública, voltada não mais à base estrutural, ou seja, para empresas em fases iniciais de seu processo de transformação digital, mas à aceleração da inovação, expansão internacional e sofisticação tecnológica. Assim sendo, políticas públicas mais aderentes a essa realidade devem considerar as seguintes diretrizes:

- a) **Fomento à inovação contínua e a pesquisa e desenvolvimento:** empresas já estruturadas, como as do *Cluster 2*, necessitam de políticas que promovam inovação incremental e disruptiva. É fundamental criar instrumentos que estimulem projetos de pesquisa aplicada, parcerias com universidades e centros tecnológicos, bem como acesso a recursos não-reembolsáveis para testes de soluções emergentes, considerando tecnologias como inteligência artificial, *blockchain*, tecnologias quânticas etc.
- b) **Internacionalização e competitividade global:** com gestão sólida e presença digital consolidada, essas empresas estão mais aptas à expansão. Políticas públicas podem oferecer apoio para inserção em cadeias globais de valor, missões internacionais, *soft landing* em mercados estratégicos e apoio à adequação regulatória internacional.
- c) **Fomento à automação avançada e integração de processos:** apesar do bom desempenho geral, a dimensão processos revela oportunidades de melhoria. A política pública pode fomentar projetos de automação e integração de sistemas com foco em eficiência operacional e escalabilidade, além de apoiar a adoção de tecnologias de indústria 4.0.
- d) **Apoio à governança de dados e transformação digital de ponta:** com alto nível de gestão e cultura organizacional, essas empresas estão prontas para estratégias mais sofisticadas, como *data driven*, *machine learning*, plataformas digitais próprias, entre outros. Políticas públicas podem subsidiar consultorias especializadas, laboratórios de experimentação digital e acesso facilitado a infraestrutura tecnológica.
- e) **Articulação em ecossistemas de inovação:** essas empresas devem ser inseridas ou fortalecidas em ambientes colaborativos, como *Hubs*, *clusters* produtivos e redes de inovação. A política pode facilitar sua atuação como âncoras tecnológicas para outras empresas menores (como as do

Cluster 1), promovendo encadeamento produtivo, mentorias e difusão de boas práticas.

- f) Instrumentos diferenciados de financiamento: empresas maduras não precisam apenas de capital inicial, e sim de linhas de crédito mais robustas para inovação, investimento de risco, também chamados de *venture capital*, fundos de co-investimento público-privado e mecanismos de mitigação de risco tecnológico.**

Comparativamente aos *clusters* 1 e 2, o *cluster* 3 apresenta um desempenho intermediário, sugerindo que as empresas próximas à média estão bem estruturadas, e ainda enfrentam desafios em algumas dimensões. A média gerencial das empresas do *cluster* 3 foi de 7,60, indicando uma possível boa maturidade organizacional, com espaço para melhorias. Essas empresas podem ter processos bem definidos, entretanto podem enfrentar dificuldades na implementação de mudanças estratégicas e inovação. A capacidade de adaptação e modernização pode ser um fator crítico para que essas empresas alcancem um desempenho ainda melhor nessa dimensão.

A média de 8,43 da dimensão pessoas no *cluster* 3 pode indicar que essas empresas valorizam a equipe e investem no desenvolvimento profissional, e podem melhorar em alguns aspectos, como 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.

A pontuação mais baixa no *cluster* 3 foi na dimensão vendas e consumidores, no valor de 4,55. Observa-se que com essa média esse grupo de empresas podem ter dificuldades na captação e fidelização de clientes. Isso pode indicar falta de estratégias de *marketing* eficazes ou baixa presença digital. Assim sendo, o *cluster* 3 apresentou um resultado intermediário do ponto de vista de maturidade digital, com boas práticas, sem consolidar sua presença digital. Esse grupo de empresas enfrentam entraves para consolidar sua competitividade

digital e alcançar maior capacidade de inovação. Com base nisso, as políticas públicas devem considerar as seguintes diretrizes:

- a) **Aceleração da transformação digital com foco estratégico:** essas empresas já têm um nível razoável de estrutura e gestão, portanto, a política pública deve atuar para reduzir barreiras à modernização tecnológica, oferecendo consultorias especializadas em transformação digital, acesso a soluções escaláveis e formação de lideranças digitais.
- b) **Fortalecimento de *marketing* e presença digital:** a pontuação mais baixa em vendas e consumidores indica a necessidade de concentrar ações voltadas ao mercado. A política pública pode incluir programas de capacitação em *marketing* digital, subsídios para estruturação de canais digitais, como por exemplo o *e-commerce*, redes sociais, plataformas de venda, e acesso a ferramentas de análise de comportamento do consumidor.
- c) **Estímulo à inovação no modelo de negócios:** a média de 5,95 em modelo de negócios mostra que muitas dessas empresas ainda operam com modelos tradicionais. Iniciativas públicas podem apoiar a modelagem de negócios inovadores, com mentorias especializadas, ações de imersão em ecossistemas digitais, e ferramentas de redesenho estratégico que permitam maior adaptação às exigências do mercado atual.
- d) **Melhoria da eficiência operacional:** a média de 6,46 em processos revela que, embora existam estruturas mínimas, falta automação e integração. A política pode incentivar diagnósticos de eficiência operacional, adesão a tecnologias acessíveis de automação, e implantação de metodologias ágeis e *lean* para otimizar fluxos internos e reduzir desperdícios.
- e) **Integração em redes de cooperação e inovação:** esse grupo se beneficia fortemente ao ser conectado com outros atores do ecossistema. A política pública pode favorecer sua participação em programas de inovação aberta, encadeamento com empresas líderes do *Cluster 2*, por exemplo, e inserção em arranjos produtivos locais ou Hubs regionais, promovendo intercâmbio de conhecimento, recursos e experiências.
- f) **Acesso orientado a financiamento e tecnologia:** muitas dessas empresas estão na fase de escalar soluções. Políticas podem criar linhas de crédito específicas para modernização

tecnológica, voucher de inovação e acesso facilitado a tecnologias digitais por meio de parcerias com fornecedores, reduzindo o custo de transição digital.

Exemplos bem-sucedidos de políticas no Brasil que apoiam as empresas a melhorarem seus níveis de maturidade digital são o Brasil Mais Produtivo¹³ do governo federal e o Jornada Digital da ABDI¹⁴.

Considerando a categorização IMD ABDI, na tabela abaixo é possível observar o valor médio por dimensão:

Tabela 5 – Valor médio por dimensão da categorização IMD ABDI					
Categorias	Dimensões				
	Gestão	Pessoas	Processos	Modelo de Negócios	Vendas e Consumidores
1 - Iniciantes	4,16	5,05	6,10	4,73	4,21
2 - Emergentes	6,23	7,08	6,54	5,65	4,95
3 - Intermediários	8,31	9,11	6,71	7,11	7,29
4 - Líderes	9,94	10,77	6,52	8,79	9,92

Fonte: Dados elaborados pela autora.

Conforme mencionado no item 3.2 – preparação dos dados, o índice agregado de maturidade digital da ABDI é computado com a média da pontuação em cada dimensão. Observando as médias, podemos considerar que o agrupamento 1 - iniciantes é formado por empresas com o menor nível de maturidade digital, apresentando dificuldades em todos os eixos analisados. O desempenho gerencial é fraco, com média de 4,16 indicando possível baixa adoção de metodologias estruturadas e pouca capacidade de planejamento estratégico. A média de 5,05 na dimensão pessoas sugere que essas

¹³ O Programa Brasil Mais Produtivo está à Política Industrial Nova Indústria Brasil (NIB), especialmente à Missão 4 que trata da transformação digital da indústria para ampliar a produtividade. Informações disponíveis no link: <https://brasilmaisprodutivo.mdic.gov.br/>

¹⁴ O projeto Jornada Digital da ABDI está voltado para MPES e busca apoiar as empresas brasileiras em seus processos de aumento da maturidade digital dos negócios. Informações disponíveis no link: <https://jornadadigital.abdi.com.br/>

empresas não investem em capacitação de funcionários, tornando a equipe pouco qualificada para lidar com ferramentas digitais e metodologias ágeis. Além disso, a estrutura operacional ainda é rudimentar, com média de 6,10 em processos, o que demonstra baixa automação e organização interna.

Essas empresas em torno da média também enfrentam problemas para inovar, como evidenciado pelo resultado obtido na dimensão modelo de negócios, que foi de 4,73, mostrando resistência à adaptação ao mercado digital. Na dimensão vendas e consumidores, a média foi de 4,21 que é a mais baixa entre todas as categorias, indicando que essas empresas têm dificuldades em conquistar novos clientes e manter um bom relacionamento com a base existente. A falta de estratégias digitais e o baixo investimento em *marketing* e canais *online* podem impactar na competitividade dessas empresas. Em resumo, este agrupamento representa empresas atrasadas na digitalização, que ainda dependem de práticas tradicionais e precisam de mudanças estruturais para se tornarem mais competitivas.

O agrupamento 2 - emergentes apresenta um nível um pouco mais avançado do que o agrupamento 1 - iniciantes, e enfrenta muitos desafios. A média de 6,23 em gestão pode indicar que essas empresas possuem uma estrutura organizacional mais definida, que podem ter dificuldades na tomada de decisões estratégicas e na implementação de boas práticas gerenciais. A dimensão pessoas apresentou um resultado médio de 7,08, o que pode demonstrar que há um esforço maior em treinar e qualificar a equipe, e o engajamento e a retenção de talentos podem ser problemáticos. A média de 6,54 na dimensão processos sugere que algumas dessas empresas já começaram a adotar ferramentas e metodologias digitais para otimizar suas operações, sendo que a digitalização ainda não é totalmente eficiente. O modelo de negócios dessas empresas também está em fase de transição, com uma média de 5,65, indicando que algumas empresas podem ter começado a explorar novas formas de atuação no mercado digital, mas sem uma estratégia consolidada. No entanto, a dimensão vendas e consumidores apresentou uma média baixa, de 4,95, o que demonstra que ainda há dificuldades em atrair e fidelizar clientes por meio de estratégias digitais eficazes. Muitas dessas empresas podem estar começando a investir em *marketing* digital, sem um retorno expressivo. Em suma, este agrupamento representa empresas em estágios também iniciais de transformação digital, um pouco mais

avanzadas que o agrupamento 1 – iniciantes e precisam evoluir na estruturação de processos e no relacionamento digital com os clientes.

As empresas próximas a média do agrupamento 3 - intermediárias possuem um nível mais avançado de maturidade digital, com gestão bem estruturada, apresentando um resultado médio de 8,31 e uma abordagem mais estratégica para a transformação digital. Na dimensão pessoas, a média foi de 9,11, indicando que pode existir um investimento sólido na qualificação e retenção de talentos, sugerindo que essas empresas valorizam equipes altamente capacitadas para lidar com desafios do mercado digital. A média de 6,71 em processos revela que essas empresas já adotam metodologias ágeis e ferramentas digitais para aumentar a eficiência operacional, embora ainda possam melhorar a automação e a integração dos fluxos de trabalho. Na dimensão modelo de negócios, a média apresentada foi de 7,11, demonstrando que algumas empresas conseguem inovar e adaptar-se às novas tendências do mercado digital. Além disso, a média de 7,29 em vendas e consumidores sugere que essas empresas já exploram estratégias digitais para atrair e fidelizar clientes, utilizando ferramentas como CRM, automação de *marketing* e *e-commerce*. Essa aglomeração de empresas já compreendem e aplicam a transformação digital, entretanto podem expandir suas operações para alcançar um nível de maturidade digital mais alto.

A última categorização, denominada pela ABDI como empresas líderes é composta pelas empresas mais avançadas digitalmente, com os maiores valores médios em todos os eixos. A média de 9,94 em gestão indica uma estrutura altamente eficiente e baseada em dados, permitindo decisões estratégicas ágeis e bem fundamentadas. Na dimensão pessoas, a média de 10,77 sugere que essas empresas possuem equipes extremamente qualificadas e engajadas, muitas vezes adotando metodologias de trabalho inovadoras, como *squads*, *design thinking* e gestão descentralizada. Na parte dos processos, a média apresentada foi de 6,52. Embora o valor seja semelhante ao agrupamento de empresas intermediárias, a diferença está na adoção de tecnologias avançadas, como inteligência artificial e automação, garantindo maior eficiência operacional. Essas empresas também se destacam pelo modelo de negócios, com média de 8,79, o que demonstra uma abordagem flexível e inovadora, com estratégias bem estruturadas para capturar novas oportunidades de mercado. Além disso, a média de 9,92 em vendas e consumidores confirma que essas empresas dominam os canais digitais, utilizando estratégias avançadas

de *marketing*, atendimento automatizado e inteligência de dados para personalizar a experiência do cliente. Este agrupamento representa empresas líderes na transformação digital, que se adaptaram ao mercado digital e impulsionam inovações e estabelecem novos padrões competitivos.

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

CONCLUSÃO

Diante da crescente importância da transformação digital para os negócios, este trabalho teve como objetivo central investigar o uso de técnicas de clusterização como instrumento de apoio à formulação de políticas públicas orientadas à transformação digital, considerando a elevada heterogeneidade das empresas brasileiras. A partir de dados empíricos oriundos da 6ª edição da pesquisa do IMD da ABDI, foi possível aplicar algoritmos de aprendizado não supervisionado para identificar padrões e perfis de empresas que compartilham características semelhantes, independentemente de classificações fixas ou lineares.

Do ponto de vista metodológico, a estratégia adotada combinou quatro categorias: porte, setor, estado e região com variáveis numéricas, que foram as cinco dimensões do IMD ABDI: gestão, pessoas, processos, modelo de negócios e vendas e consumidores, permitindo uma análise integrada de natureza estrutural e comportamental das empresas.

A análise da composição dos *clusters* revelou importantes diferenças estruturais e territoriais entre os grupos formados, especialmente no que diz respeito ao porte das empresas, setor de atuação e localização regional. Observou-se a predominância de microempresas na maioria dos *clusters*, refletindo sua representatividade nacional, com variações relevantes entre os algoritmos. O setor de serviços, em particular, foi frequentemente isolado em agrupamentos próprios, especialmente nos métodos HC e por moda, indicando padrões homogêneos que podem ser associados ao baixo grau de digitalização. Já as empresas do setor industrial apareceram mais diluídas, enquanto o comércio demonstrou forte presença em todos os grupos, com comportamentos diversos quanto à maturidade digital.

Geograficamente, a região Sudeste concentrou a maior parte das empresas em todos os *clusters*, reflexo de sua densidade econômica e concentração de infraestrutura. No entanto, algoritmos mais robustos como o PAM revelaram a presença de perfis específicos nas regiões Norte e Centro-Oeste, sugerindo que empresas localizadas

em regiões menos desenvolvidas também formam agrupamentos próprios com desafios e oportunidades específicos. Isso reforça a importância de incorporar a dimensão territorial nas políticas públicas digitais, uma vez que o acesso à infraestrutura tecnológica, à mão de obra qualificada e às redes de apoio varia substancialmente entre as regiões brasileiras.

A análise de qualidade revelou padrões distintos na segmentação das empresas brasileiras quanto à maturidade digital. Dentre os métodos testados, o *Means* demonstrou ser a abordagem mais eficaz, proporcionando *clusters* bem definidos e estatisticamente robustos. Essa técnica obteve um CH de 630,96, indicando alta coesão interna e separação clara entre os grupos. Além disso, o Índice de DBI apresentou um valor de 1,53, sugerindo baixa sobreposição entre os *clusters*, e o Índice de Silhueta registrou 0,23, reforçando a qualidade da segmentação obtida.

Como alternativa viável, a categorização proposta pelo IMD ABDI também demonstrou coerência na separação dos grupos, ainda que de maneira menos eficiente. Seus CH de 325,70 e DBI de 1,70 indicam uma segmentação funcional, com menor precisão na diferenciação entre empresas. O Índice de Silhueta de 0,11 sugere que a estrutura dos grupos, embora perceptível, apresenta mais ambiguidades do que o algoritmo gerado pelo algoritmo *Means*.

Os algoritmos HC e *Modes*, por outro lado, não conseguiram estruturar *clusters* com separação clara. Ambos apresentaram Índices de Silhueta negativos, o que sugere que muitas empresas foram agrupadas de forma inconsistente. O HC obteve um DBI de 9,55, enquanto o *Modes* atingiu um valor ainda mais elevado de 17,69, revelando forte sobreposição entre os *clusters*. Esses resultados indicam que esses métodos não capturaram adequadamente os padrões subjacentes na distribuição dos dados.

Além das métricas de qualidade, a distribuição dos *clusters* entre os diferentes métodos de agrupamento fornece insights importantes sobre as características das empresas em cada segmento. O *cluster* dominante identificado pelo *Means* reflete a concentração de micro e pequenas empresas, predominantemente dos setores de comércio e serviços. As regiões Sudeste e Sul lideram em representatividade, sugerindo que a maturidade digital nesses locais está mais avançada do que em outras partes do país.

A análise dos perfis médios dos *clusters* do algoritmo *Means* oportunizou a construção de uma tipologia útil para orientar o desenho de políticas segmentadas:

- a) O *Cluster 1*, formado por empresas com baixa maturidade digital, exige políticas estruturantes, com foco em capacitação básica, acesso a tecnologias essenciais, fortalecimento da gestão e apoio à estruturação organizacional. Em geral, são micro ou pequenas empresas, dos setores de serviços e comércio, localizadas em regiões com menor densidade tecnológica.**
- b) O *Cluster 2*, composto por empresas maduras digitalmente, apresenta alto desempenho em todas as dimensões analisadas. São empresas com gestão profissionalizada, presença digital consolidada e alto grau de inovação, que já operam com automação e ferramentas analíticas. Para este grupo, políticas devem promover a sofisticação tecnológica, internacionalização, integração em ecossistemas de inovação e o uso de tecnologias de fronteira.**
- c) O *Cluster 3* representa empresas em estágio intermediário, com estrutura gerencial sólida, apresentando dificuldades operacionais e frágil presença comercial digital. Essas empresas têm alto potencial de avanço e demandam políticas de aceleração digital, com foco em redesenho do modelo de negócios, integração de processos e fortalecimento de canais digitais de relacionamento com o mercado.**

Em comparação com a categorização IMD ABDI, observa-se que os modelos baseados em faixas fixas (Iniciantes, Emergentes, Intermediários e Líderes) não conseguem capturar a complexidade e os padrões híbridos identificados pela clusterização. A segmentação tradicional presume linearidade entre as dimensões de maturidade, enquanto a clusterização permite identificar empresas com comportamentos mistos e trajetórias assimétricas, como, por exemplo, organizações com boa gestão, contudo sem estratégia digital consolidada. Essa capacidade analítica é essencial para desenhar intervenções mais precisas, eficazes e adaptadas à realidade de cada grupo.

Além disso, os resultados reforçam que a clusterização pode servir como base para estratégias modulares de política pública, permitindo a construção de trilhas evolutivas distintas para empresas

em diferentes níveis de maturidade digital. Essas trilhas podem articular diferentes tipos de apoio técnico e financeiro, com níveis graduais de exigência e complexidade, aumentando as chances de sucesso na transformação digital. Também se abrem oportunidades para o uso estratégico desses perfis em programas de encadeamento produtivo, nos quais empresas do *Cluster 2* possam atuar como âncoras de inovação para empresas dos *Clusters 1 e 3*, promovendo sinergias sistêmicas em nível regional e setorial. Essa lógica colaborativa, baseada em redes de aprendizagem e complementaridades, pode acelerar a digitalização de cadeias produtivas inteiras, reduzindo desigualdades e ampliando o impacto das políticas públicas.

Em síntese, os achados demonstram que o uso de modelos de clusterização, com destaque para o algoritmo *Means*, é metodologicamente robusto e estrategicamente valioso para segmentação de empresas conforme sua maturidade digital. Tal abordagem permite desenhar políticas públicas mais segmentadas, eficazes e com melhor custo-benefício, adaptadas às necessidades de diferentes perfis empresariais, setores produtivos e realidades territoriais.

Em última instância, este trabalho comprova que a integração entre técnicas quantitativas de aprendizado não supervisionado e análise qualitativa de perfis organizacionais representa um caminho promissor para a modernização da formulação de políticas públicas no Brasil. A clusterização, ao capturar a complexidade e diversidade das empresas brasileiras, fortalece a capacidade do Estado em promover a transformação digital de forma justa, segmentada e eficaz, contribuindo para o aumento da competitividade nacional em um cenário econômico cada vez mais orientado pela tecnologia.



REFERÊNCIAS

REFERÊNCIAS

REFERÊNCIAS

ABDI, Agência Brasileira de Desenvolvimento Industrial. **E-ABDI: Documento base para a definição das ações da ABDI em transformação digital.** Brasília, 2019. Disponível em <https://api.abdi.com.br/storage/files/boas-praticas/gEJOuQOeYJUaFPHIRW2d/E-ABDI.pdf>

AGGARWAL, C. **Data Mining: The Textbook.** 1. ed. New York, USA: Springer International Publishing Switzerland, 2015. v. 14

CALIŃSKI, T., & Harabasz, J. 1974. **"A dendrite method for cluster analysis."** *Communications in Statistics*, 3(1), 1-27

CALDAS, Sérgio. **China quer impulsionar fatia da economia digital no PIB para 10% até 2025.** CNN Brasil, 2022. Disponível em: <https://www.cnnbrasil.com.br/business/china-quer-impulsionar-fatia-da-economia-digital-no-pib-para-10-ate-2025/>.

CARTELA, E.Y.S; & PORCILE, G. **Heterogeneidade estrutural na produtividade das firmas brasileiras.** CEPAL e IPEA, 2015. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/11645>.

CORRAR L. J., PAULO E., DIAS FILHO J. M. **Análise multivariada: para os cursos de administração, ciências contábeis e economia.** Atlas, 2009.

DAVIES, D. L., & Bouldin, D. W. 1979. **"A cluster separation measure."** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224-227

DE MORSIER, F. et al. **Cluster validity measure and merging system for hierarchical clustering considering outliers.** *Pattern Recognition*, v. 48, n. 4, p. 1478–1489, 2015.

DENG, Z. et al. **Efficient kNN classification algorithm for big data.** *Neurocomputing*, v. 195, p. 143–148, 2016.

Digital (E-Digital). Ciclo 2022- 2026. Brasília, 2022. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosestrategiadigital/e-digital_ciclo_2022-2026.pdf

DUTTA, G., Kumar, R., Sindhwani, R., & Singh, R., Jr. **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, 2021. Disponível em: <https://doi.org/10.1007/s10845-021-01783-2>

ELLER, R., Alford, P., Kallmünzer, A., & Peters, M. **Antecedents, consequences, and challenges of small and medium-sized enterprise digitalization.** Journal of Business Research, 112, 119–127, 2020. Disponível em: <https://doi.org/10.1016/j.jbusres.2020.03.004>

FISHER, R. A. 1925. **"Statistical methods for research workers."** *Oliver and Boyd, Edinburgh*

FONTANA, A., Naldi, M. C. **Estudo de Comparação de Métodos para Estimção de Números de Grupos em Problemas de Agrupamento de Dados.** 2009. Universidade de São Paulo. ISSN – 0103-2569.

G20. **G20 Digital Economy Development and Cooperative Index.** 2016. Disponível em: <http://www.mofa.go.jp/files/000185874.pdf>

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa.** 5 ed. São Paulo, 2010.

GÖKALP, E., & Martinez, V. **Digital transformation capability maturity model enabling the assessment of industrial manufacturers.** Computers in Industry, 132, 103522, 2021. Disponível em: <https://doi.org/10.1016/j.compind.2021.103522>

GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. **Data mining: conceitos, técnicas, algoritmos, orientações e aplicações.** Rio de Janeiro: Elsevier, 2015.

GOLLHARDT, T., Halsbenning, S., Hermann, A., Karsakova, A., & Becker, J. **Development of a Digital Transformation Maturity Model for IT Companies.** 2020 IEEE 22nd Conference on Business Informatics (CBI), 1, 94–103, 2020. Disponível em: <https://doi.org/10.1109/CBI49978.2020.00018>

GRUS, J. **Data Science from Scratch.** 1. ed. Cambridge, Mass.: O'Reilly Media, 2015. v. 1

HAIR J. F., BLACK W. C., BABIN B. J., ANDERSON R. E. **Multivariate Data Analysis,** 7ª ed. Pearson, 2010.

HASSANI, R.; El Idrissi, Y. **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, v. 9, n. 1, p. 630-636, 2020.

HININGS, B.; Gegenhuber, T.; Greenwood, R. **Digital innovation and transformation: An institutional perspective**. Information and Organization, v. 28, n. 1, p. 52-61, 2018.

JAYARAM REDDY, A. et al. **Performance Analysis of Clustering Algorithm in Data Mining in R Language**. Communications in Computer and Information Science, v. 837, n. November, p. 364–372, 2018.

KASSAMBARA, A. **Multivariate Analysis 1: Practical Guide To Cluster Analysis in R: Unsupervised machine learning**. 1. ed. [s.l.] STHDA, 2017.

KRAUS, S., Durst, S., Ferreira, J. J., Veiga, P., Kailer, N., & Weinmann, A. **Digital transformation in business and management research: An overview of the current status quo**. International Journal of Information Management, 63, 2022. Disponível em: <https://doi.org/10.1016/j.ijinfomgt.2021.102466>

KRAUS, S., Jones, P., Kailer, N., Weinmann, A., Chaparro-Banegas, N., & Roig-Tierno, N. **Digital Transformation: An Overview of the Current State of the Art of Research**.

SAGE Open, 11(3), 2021. Disponível em: <https://doi.org/10.1177/21582440211047576>

LIU, Q. et al. **Distributed k-means algorithm for sensor networks based on multi-agent consensus theory**. Proceedings of the IEEE International Conference on Industrial Technology, v. 2016- May, p. 2114–2119, 2016.

LOBÃO, L. Zilli, C. **A jornada da transformação digital: um guia prático: cases, fundamento e ferramentas**. São Paulo: Editora Lamonica, 2020.

MATARAZZO, M.; Penco, L.; Profumo, G.; Quaglia, R. **Digital transformation and customer value creation in Made in Italy SMEs: A dynamic capabilities perspective**. Journal of Business Research, v. 123, p. 642-656, 2020.

MDIC, Ministério do Desenvolvimento, Indústria, Comércio e Serviços. **Nova indústria Brasil – Nova indústria Brasil – forte, transformadora e sustentável: Plano de Ação para a neindustrialização 2024-2026.** Brasília, 2024. Disponível em: <https://www.gov.br/mdic/pt-br/composicao/se/cndi/plano-de-acao/nova-industria-brasil-plano-de-acao.pdf>

MEIRA, Silvio. **Direções, desafios e dimensões para uma estratégia de Brasil digital.** Recife: tds.company, 2021. Disponível em: <https://biblioteca.tds.company/ebookbrasil-digital>

OMRANI, H.; SHAFAT, K.; & EMROUZNEJAD, A. **An integrated fuzzy clustering cooperative game data envelopment analysis model with application in hospital efficiency.** [s.l.] Elsevier Ltd, 2018. v. 114

PAGANI, M.; Pardo, C. **The impact of digital technology on relationships in a business network.** Industrial Marketing Management, v. 67, p. 185-192, 2017.

RENCHER, A. C. **Methods of Multivariate Analysis.** 2ª ed., John Wiley & Sons, 2002.

ROGERS, D. **The Digital Transformation Playbook: rethink your business for the digital age.** Columbia University Press, 2016.

ROHN, D., Bican, P. M., Brem, A., Kraus, S., & Clauss, T. **Digital platform-based business models – An exploration of critical success factors.** Journal of Engineering and Technology Management - JET-M, 2021. Disponível em: <https://doi.org/10.1016/j.jengtecman.2021.101625>

ROUSSEAU, P. J. 1987. **“Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.”** *Journal of Computational and Applied Mathematics*, 20, 53-65

SAMPAIO, R. **Vantagem digital: Um guia prático para a transformação digital.** Rio de Janeiro: Alta Books, 2018.

THORNDIKE, R. L. (1953). **“Who belongs in the family?”** *Psychometrika*, 18(4), 267–276

VAN ARK, Bart et al. **Productivity & innovation competencies in the midst of the digital transformation age: A EU-US comparison.** *European Economy-Discussion Papers*, n. 119, 2019. Disponível em:

https://ec.europa.eu/info/sites/default/files/economy-finance/dp119_en.pdf.

economy-

VERHOEF, P. C.; Broekhuizen, T.; Bart, Y.; Bhattacharya, A.; Qi Dong, J.; Fabian, N.; & Haenlein, M. **Digital transformation: A multidisciplinary reflection and research agenda**. Journal of Business Research, 122 (September 2019), 889–901, 2021. Disponível em: <https://doi.org/10.1016/j.jbusres.2019.09.022>



idp

A ESCOLHA QUE
TRANSFORMA
O SEU CONHECIMENTO