

Uso da internet em domicílios brasileiros: uma análise de perfil baseada em técnicas de agrupamento

Helton Messias Adigneri^[1], Bruno Samways dos Santos^{[2]*}, Rafael Henrique Palma Lima^[3], Gislaíne Camila Lapasini Leal^[4]

[1] adigneri@hotmail.com, [2] brunosantos@utfpr.edu.br, [3] rafaelhlma@utfpr.edu.br, [4] gclleal@uem.br. Programa de Pós-Graduação em Engenharia de Produção, Universidade Estadual de Maringá (UEM), Maringá, Paraná, Brasil

* autor correspondente

Resumo

Este estudo teve como objetivo desenvolver e analisar perfis de indivíduos brasileiros que utilizam a internet em domicílio, empregando técnicas de mineração de dados para subsidiar processos de tomada de decisão em organizações públicas ou privadas interessadas em compreender o uso e as habilidades relacionadas aos recursos de TIC domiciliares. Para tanto, foram selecionadas duas pesquisas realizadas pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br) sobre TIC domiciliares, abrangendo os anos de 2018 e 2025. Esses dados permitiram a criação de perfis de usuários que descrevem as características dos indivíduos estudados ao longo desses cinco anos. O algoritmo k-modes foi aplicado para agrupar os conjuntos de dados totalmente categóricos. Como resultado, foram desenvolvidos perfis de usuários para 2018 e 2025. A análise dos perfis gerados indicou uma melhora na estrutura socioeconômica desses indivíduos durante o período. Observou-se também migração para áreas urbanas, bem como uma maior representatividade numérica de indivíduos autodeclarados pardos e do sexo feminino dentro dos clusters. No geral, houve um declínio no uso de computadores para acesso à internet, acompanhado por um aumento no uso de dispositivos móveis e uma presença notável de smart TVs. O acesso à internet em casa e o uso da internet móvel também aumentaram. Além disso, o uso de serviços digitais específicos e a demanda por entretenimento online cresceram. No entanto, durante o mesmo período, o interesse em cursos online e buscas acadêmicas diminuiu. Também foi constatado que, entre alguns grupos com condições socioeconômicas mais baixas, o acesso à internet e aos serviços online permanece limitado. As características identificadas nos perfis de usuários domiciliares de TIC no Brasil correspondem àquelas relatadas em estudos realizados em países da América Latina e do Caribe.

Palavras-chave: *k-modes*; mineração de dados; perfis de usuários; tecnologia da informação e comunicação; uso domiciliar da internet.

Internet use in Brazilian households: a profiles analysis based on clustering techniques

Abstract

This study aimed to develop and analyze profiles of Brazilian individuals who use the internet at home, employing data mining techniques to support decision-making processes in public or private organizations interested in understanding the use and skills related to household ICT resources. To this end, two surveys conducted by the Regional Center for Studies on the Development of the Information Society (Cetic.br) on household ICT were selected, covering the years 2018 and 2025. These data enabled the creation of user profiles that describe the characteristics of individuals studied over these five years. The k-modes algorithm was applied to cluster the fully categorical datasets. As a result, user profiles were developed for 2018 and 2025. Analysis of the generated profiles indicated an improvement in the socioeconomic structure of these individuals during the period. Migration towards urban areas was also observed, as well as a higher numerical representation of self-identified mixed-race and female individuals within the clusters. Overall, there was a decline in computer use for

internet access, accompanied by an increase in mobile device use and a notable presence of smart TVs. Internet access at home and mobile internet usage also increased. Furthermore, the use of specific digital services and the demand for online entertainment grew. However, during the same period, interest in online courses and academic searches declined. It was also found that, among some groups with lower socioeconomic conditions, access to the internet and online services remains limited. The characteristics identified in the household ICT user profiles in Brazil correspond to those reported in studies conducted in Latin American and Caribbean countries.

Keywords: data mining; household internet use; information and communication technology; k-modes; user profiles.

1 Introdução

As Tecnologias da Informação e Comunicação (TIC) vêm ganhando destaque ao possibilitar o desenvolvimento de soluções que aumentam a eficiência das atividades realizadas por seus usuários, tanto no setor público quanto no privado, promovendo o fortalecimento comunitário e um cenário propício à inovação. Além disso, as TIC têm impulsionado transformações em cidades antigas e em novas localidades, contribuindo para que se tornem mais inteligentes (Yang; Lam, 2021).

As cidades inteligentes apresentam potencial para enfrentar diversos desafios por meio do uso de informações e tecnologias que compõem as TIC em processos de inovação social, visando alcançar o desenvolvimento sustentável pautado na sustentabilidade, equidade e acessibilidade (Lee *et al.*, 2022).

As TIC exercem um papel central na sociedade contemporânea. Fundamentadas nos princípios da informação, comunicação e tecnologia, impulsionam o crescimento e o desenvolvimento social, tornando-se parte integrante da vida cotidiana (Sharma, 2021). Na educação, transformam o ambiente e os métodos de aprendizagem; no setor financeiro, possibilitam serviços estratégicos e inovadores; nas indústrias, simplificam e automatizam processos produtivos; e no comércio, favorecem a expansão em escala global.

Atualmente, atividades como renovação de documentos, compras e agendamentos médicos podem ser realizadas digitalmente. Além das aplicações práticas, as TIC são amplamente utilizadas para entretenimento, como jogos, streaming de vídeos, música e acesso a notícias. No entanto, apesar de seus inúmeros benefícios, também apresentam desafios e impactos negativos que precisam ser considerados (Rudrawar, 2014).

As TIC englobam diferentes tecnologias, entre elas a internet (Mwiinga, 2023). Compreender seu uso e as competências necessárias para o manuseio de recursos tecnológicos baseados na rede por indivíduos em domicílios é fundamental para analisar sua evolução nesse contexto. Essa compreensão é essencial para avaliar os meios de acesso e os comportamentos adotados por esses indivíduos, aspectos que contribuem diretamente para o desenvolvimento social.

Nesse sentido, organizações demonstram interesse em obter o perfil de indivíduos que utilizam TIC em residências, sobretudo em relação ao uso e às habilidades relacionadas à internet, com vistas à tomada de decisão. Uma das formas de obter tais informações consiste na aplicação de ferramentas e técnicas de mineração de dados em levantamentos como os conduzidos pelo Cetic.br no Brasil.

Segundo Datta e Davim (2021), a mineração de dados constitui atividade essencial na análise de dados e integra o processo de descoberta de conhecimento em bancos de dados, conhecido como KDD (Knowledge Discovery in Databases). De forma geral, a mineração de dados identifica padrões ocultos em grandes volumes de informações, enquanto a análise subsequente permite gerar percepções ou testar hipóteses e modelos.

Para alcançar esse objetivo, a mineração de dados recorre a técnicas de aprendizado de máquina, como o agrupamento, que busca identificar grupos homogêneos de registros a partir das variáveis fornecidas.

De acordo com Rençber, Zorkirişçi e Aytekin (2023), a literatura apresenta ampla produção sobre alfabetização digital e condições de uso das TIC por alunos, professores e funcionários. No entanto, ainda há pouca exploração acerca do uso de dispositivos domésticos de TIC, tema de grande relevância. Essa lacuna justifica o objetivo deste estudo, que consistiu em construir perfis de

indivíduos brasileiros usuários de internet em domicílios, a partir de técnicas de mineração de dados, seguidos de análises descritivas em diferentes dimensões.

A justificativa para esta pesquisa baseia-se na crescente dependência dos indivíduos em relação às TIC para a realização de atividades cotidianas, o que lhes confere importância crescente em diferentes setores, como o educacional, o de defesa e o tecnológico (Singh *et al.*, 2020).

Conforme Cogo, Dutra-Brignol e Fragoso (2015), compreender hábitos tecnológicos, padrões de acesso à internet, recursos tecnológicos, habilidades e frequência de uso das TIC em uma população contribui para que governos e instituições elaborem planos estratégicos voltados à inclusão digital, ao planejamento de políticas públicas, a projetos de impacto social e ao desenvolvimento sustentável, além de apoiar iniciativas de tomada de decisão em organizações públicas e privadas.

Como contribuição, este estudo oferece informações sobre agrupamentos naturais de usuários das TIC em domicílios, especialmente no que se refere ao uso da internet, permitindo que autoridades e tomadores de decisão compreendam melhor as necessidades e lacunas de acesso e uso. Nesse sentido, a pesquisa pode servir de referência para o desenvolvimento de uma infraestrutura digital mais inclusiva e acessível, beneficiando não apenas instituições públicas, mas também indústrias de produtos e serviços tecnológicos.

Esses dados podem subsidiar a criação de novos produtos e serviços dependentes da internet, adequados à realidade brasileira.

O restante deste artigo está organizado em distintas seções. Na Seção 2, o referencial teórico apresenta as principais teorias, conceitos e estudos que fundamentam a pesquisa, com destaque para os conceitos de TIC, internet e o método de agrupamento utilizado. Em seguida, a Seção 3 descreve os trabalhos relacionados, a fim de contextualizar o estudo, identificar lacunas, justificar sua relevância e comparar diferentes abordagens. A Seção 4 detalha o método de pesquisa empregado. Na Seção 5, apresentam-se os resultados e discussões, incluindo a análise e interpretação dos dados obtidos por meio dos agrupamentos. Por fim, a Seção 6 expõe as conclusões da pesquisa.

2 Referencial teórico

As Tecnologias da Informação e Comunicação (TIC) constituem um termo que se refere a um conjunto de tecnologias voltadas à coleta, armazenamento, processamento e transferência de informações, englobando hardwares e softwares, sistemas de telecomunicações, meios de compartilhamento de informações, redes de comunicação e sistemas de computação (Wicht; Reder; Lechner, 2021). As TIC têm impulsionado a modernização de setores como educação, negócios e governança, por meio de formas mais eficientes de comunicação e de sistemas de aprendizagem (Al-Rahmi *et al.*, 2020). Essa definição evidencia a capacidade das TIC de moldar a vida cotidiana, reorganizando processos de trabalho, estruturando ambientes de ensino e aprendizagem e promovendo o desenvolvimento social.

Han, Pei e Tong (2022) explicam que, com a informatização dos processos humanos, gera-se diariamente uma grande quantidade de dados oriundos de diferentes dispositivos. Esses conjuntos de dados podem ser explorados pela mineração de dados, definida como o processo de descoberta de padrões relevantes, enquanto a aprendizagem de máquina corresponde a um conjunto de técnicas baseadas em algoritmos que permitem ao computador aprender a partir desses dados. No presente estudo, o conjunto utilizado corresponde às bases do cetic.br, e a etapa de mineração de dados consistiu na aplicação de um algoritmo de aprendizagem não supervisionada denominado *k-modes*. O modelo é detalhado a seguir, e a justificativa de seu uso encontra-se descrita na Seção 4, que apresenta as etapas e ferramentas da pesquisa.

Segundo Aggarwal e Reddy (2014), o algoritmo *k-modes* implementa um método de agrupamento voltado exclusivamente a dados categóricos, isto é, dados que representam categorias. Determinar semelhanças ou diferenças entre categorias é mais complexo do que em modelos estatísticos contínuos, pois tais categorias contêm valores aleatórios sem ordem intrínseca, ao contrário dos dados numéricos. Embora dois valores categóricos possam ser iguais ou diferentes, não há comparação direta entre eles, como ocorre com números, que admitem relações de ordem, como “maior que” ou “menor que”.

De acordo com Han, Pei e Tong (2022), o algoritmo *k-modes* é uma variação do tradicional *k-means*, substituindo a distância euclidiana, geralmente utilizada, pela dissimilaridade de Hamming, e as médias pelo cálculo das modas na função objetivo. Seja X, Y dois conjuntos de dados categóricos, com m variáveis categóricas, a dissimilaridade pode ser medida pelo total de diferenças entre duas instâncias, conforme Equações (1) e (2) (Huang, 1998):

$$d_1(X, Y) = \sum_{j=1}^m \delta(x_j, y_j) \quad 1)$$

onde:

$$\delta(x_j, y_j) = \begin{cases} 0, & \text{se}(x_j = y_j) \\ 1, & \text{se}(x_j \neq y_j) \end{cases} \quad 2)$$

Inicialmente, selecionam-se k modas, uma para cada agrupamento. Na etapa seguinte, cada instância é alocada ao agrupamento cuja moda é a mais próxima dela. Em seguida, a moda de cada agrupamento é atualizada, à semelhança do que ocorre no *k-means* (no qual se atualiza a média dos centróides). Para um agrupamento i e uma dimensão j , a moda é atualizada para o valor mais frequente dessa dimensão entre todos os objetos atribuídos ao agrupamento. Caso dois ou mais valores ocorram com a mesma frequência, um deles pode ser escolhido aleatoriamente. Essas etapas se repetem até a minimização da função de custo representada pela Equação (3) (Huang, 1998):

$$P(W, Q) = \sum_{l=1}^k \sum_{i=1}^n \sum_{j=1}^m w_{i,l} \delta(x_{i,j}, q_{l,j}) \quad 3)$$

em que k representa o número de grupos a serem criados, n o número de instâncias, W é uma matriz de partição $n \times k$, na qual $w_{i,l}$ indica a alocação (ou não) da instância i ao grupo k , e $Q = \{Q_1, Q_2, \dots, Q_k\}$ representa o conjunto de instâncias no mesmo domínio (Huang, 1998).

Os algoritmos *k-means* e *k-modes* apresentam diversas similaridades em termos de execução. Contudo, alguns pontos importantes merecem destaques, conforme mostrado no Quadro 1.

Quadro 1 – Comparações entre os algoritmos *k-means* e *k-modes*

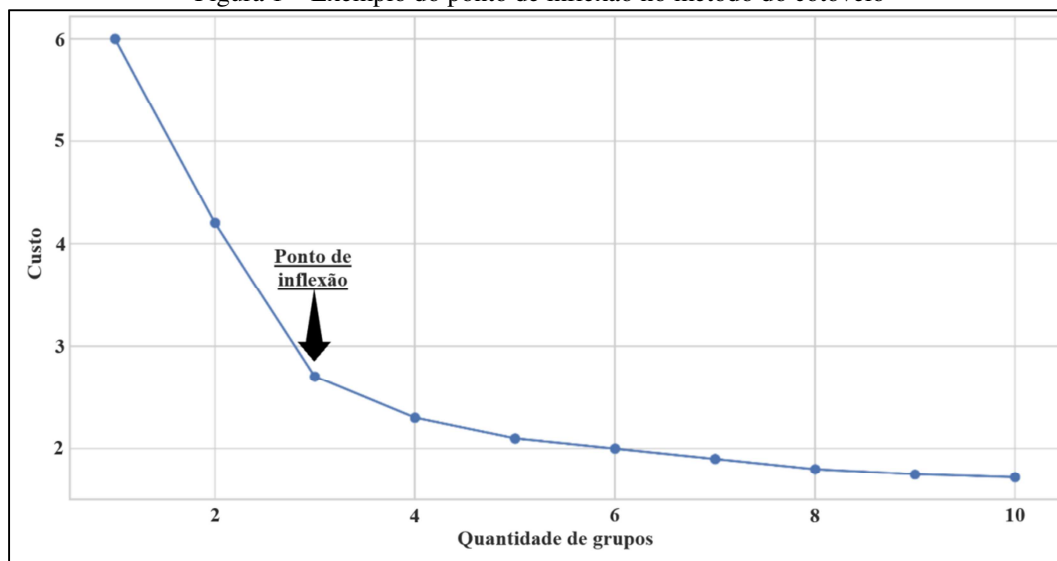
Aspecto	<i>k-means</i>	<i>k-modes</i>
Inicialização	Pontos aleatórios para centróides	Escolha aleatória de instâncias
Tipos de dados	Númerico contínuo	Categóricos
Métrica de distância	Euclidiana	Hamming
Centro do grupo	Média das variáveis	Moda das variáveis

Fonte: adaptado de Yuan, Yang e Yuan (2020) e Ahmad e Dey (2007)

A partir da análise apresentada no Quadro 1, observa-se que, para variáveis categóricas — como as utilizadas nesta pesquisa — recomenda-se o uso do algoritmo *k-modes*, uma vez que a definição da distância, do centro do grupo e da inicialização se torna mais adequada às características do conjunto de dados com esse tipo de variável (Ahmad; Dey, 2007).

Uma das limitações do método *k-modes* é a determinação da quantidade ideal de agrupamentos. Liu e Deng (2021) explicam que, conforme Thorndike (1953), o método do cotovelo foi desenvolvido para avaliar a consistência da análise de agrupamentos, auxiliando na definição do número adequado de grupos. Essa técnica funciona como uma heurística baseada na observação da curva de variância intra-agrupamentos: o primeiro ponto de inflexão mais significativo indica a melhor quantidade de agrupamentos a ser adotada (Han; Pei; Tong, 2022). A Figura 1 ilustra um exemplo de gráfico que representa a curva de uma função de custo em relação ao número total de grupos, permitindo identificar o ponto de inflexão.

Figura 1 – Exemplo do ponto de inflexão no método do cotovelo



Fonte: Murato, Santos e Lima (2024)

Ao observar a Figura 1, nota-se que o ponto de inflexão ocorre no trecho em que há mudança na inclinação da curva, indicando uma redução da queda do custo a cada acréscimo de novo grupo. Assim, as quedas mais significativas no custo são verificadas até o número de três grupos.

3 Trabalhos relacionados

A partir de pesquisas realizadas em bases de dados como Scopus, Google Scholar e Scielo, foram identificados alguns trabalhos similares, descritos na presente seção.

Spizzirri *et al.* (2017) realizaram um estudo, por meio de pesquisa estruturada, sobre o uso da internet por adolescentes brasileiros. O perfil identificado mostrou que esses jovens tinham entre 12 e 17 anos e acessavam a rede, em média, de duas a três horas por dia, principalmente em suas residências, utilizando computadores pessoais. Verificou-se que 89,6% possuíam computador com acesso à internet e que 40,6% utilizavam a rede para o envio de mensagens instantâneas. O estudo também destacou que tanto a democratização do acesso à internet quanto as condições socioeconômicas influenciavam diretamente esse padrão de uso.

Outra pesquisa, realizada por Santos e Gomes (2023), apresentou o perfil das pessoas idosas no Brasil em relação ao uso da internet e às habilidades digitais relacionadas às TIC. Foram utilizadas fontes secundárias da pesquisa TIC Domicílios, conduzida pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), em junho de 2022. A partir da análise dos dados em software específico, concluiu-se que as mulheres brancas são as que mais utilizam a internet em benefício próprio. Os homens, por sua vez, apresentaram maior alfabetização digital e literacia computacional. Já os piores indicadores relacionaram-se à falta de interesse, às habilidades precárias, à baixa alfabetização e à renda reduzida, sobretudo entre idosos de raça preta.

O estudo de Cortés *et al.* (2020) foi realizado no Chile, a partir de uma amostra probabilística estratificada em três estágios de seleção aleatória: bloco, domicílio e indivíduo, envolvendo pessoas de 18 a 69 anos. Por meio da análise de agrupamentos, foram identificados quatro grupos de perfis de indivíduos, com base na intensidade e na variabilidade do uso da internet e de suas tecnologias. Concluiu-se que o menor uso e a menor diversidade no uso da internet estavam associados ao nível reduzido de escolaridade, à idade mais avançada, às condições socioeconômicas mais baixas e ao sexo feminino. Além disso, a pesquisa demonstrou que indivíduos que utilizam a internet de forma mais intensa e diversificada tendem a se preocupar mais com os riscos associados ao uso de mídias sociais, além de associarem o uso frequente da internet tanto ao bem-estar quanto a problemas psicológicos.

Ainda no contexto da América Latina, Flores-Cueto, Hernández e Garay-Argandoña (2020) conduziram uma revisão de literatura sobre o acesso à internet e a exclusão digital no Peru, no âmbito das TIC. O estudo resultou em um perfil geral dos usuários de internet no país, no qual 80,5% dos internautas são adolescentes e jovens adultos, que acessam principalmente por meio de dispositivos móveis, como celulares. Verificou-se também que uma pessoa com ensino superior completo tem pelo menos o dobro de chances de utilizar a internet em comparação àquela que não concluiu o ensino fundamental. Além disso, observou-se que, independentemente do recurso de TIC utilizado, os mais jovens tendem a utilizar a rede com maior frequência e de forma mais diversificada.

Com base nos estudos apresentados, é possível afirmar que fatores como escolaridade, idade, gênero e condições socioeconômicas influenciam fortemente o acesso e o uso das TIC.

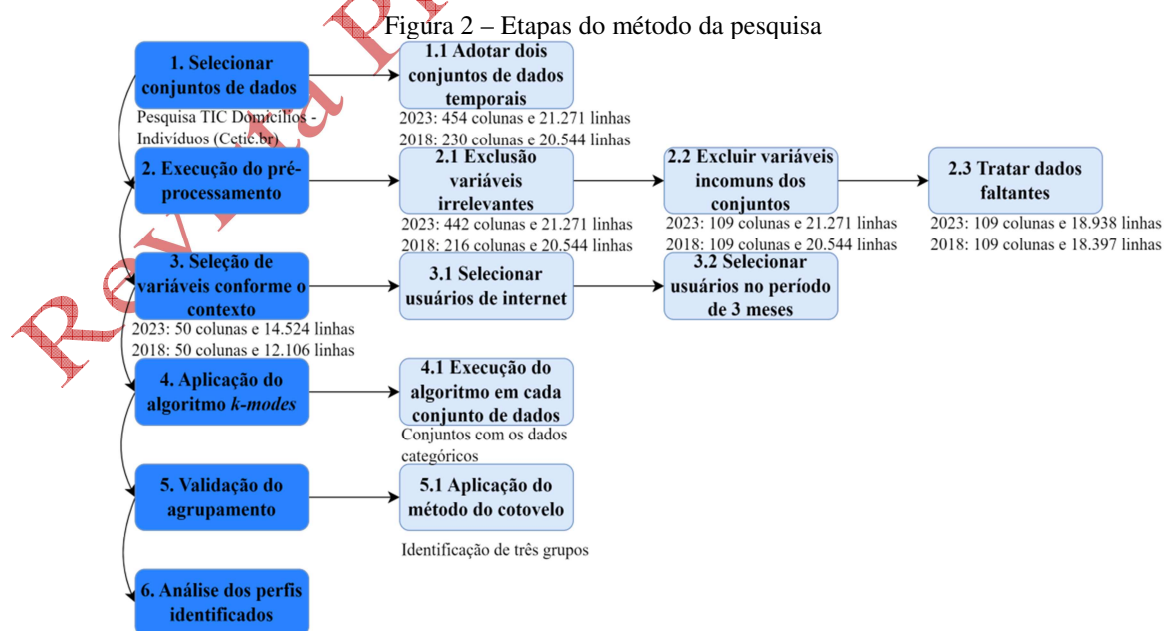
No Chile, a baixa escolaridade, a idade avançada, a vulnerabilidade social e o sexo feminino estão associados a um uso mais restrito da internet. No Peru, adolescentes e jovens lideram o acesso, sobretudo por meio de dispositivos móveis, e a escolaridade destaca-se como variável determinante, evidenciando o impacto das desigualdades estruturais na inclusão digital.

No Brasil, observa-se um cenário convergente ao de outros países latino-americanos, em que os adolescentes assumem papel central no uso da internet, predominantemente por meio de dispositivos pessoais. Em contraste, entre os idosos, o uso das TIC permanece limitado e condicionado por desigualdades estruturais relacionadas a sexo, raça e renda. Assim, o Brasil compartilha com seus vizinhos latino-americanos o desafio da exclusão digital vinculada a determinantes socioeconômicos e educacionais.

4 Método da pesquisa

Inspirando-se nas contribuições de Saccol (2010), esta pesquisa adotou o paradigma positivista, considerando que o fenômeno investigado pode ser compreendido a partir de dados mensuráveis e objetivos. Alinhada a essa perspectiva, a investigação utilizou uma abordagem quantitativa, de natureza aplicada, voltada à análise de dados para identificar padrões e perfis relacionados ao uso das TIC.

O trabalho está constituído por etapas que abrangem desde a concepção e a seleção das bases de dados utilizadas na pesquisa até a análise dos resultados. Essas etapas foram adaptadas a partir dos fluxos de trabalho do processo KDD descritos por Datta e Davim (2021). A Figura 2 apresenta a sequência metodológica adotada.



4.1 Seleção das bases de dados

A primeira etapa desta pesquisa correspondeu à seleção de duas bases de dados para o desenvolvimento do estudo. Optou-se por bases públicas contendo informações sobre pesquisas relacionadas às TIC nos lares brasileiros (TIC Domicílios), conduzidas pelo cetic.br. Tais bases são de acesso aberto, licenciadas sob a CC BY 4.0, o que possibilita seu uso livre mediante atribuição apropriada. Assim, os conjuntos de dados podem ser consultados por meio do Núcleo de Informação e Coordenação do Ponto BR (NIC.BR, 2019; NIC.BR, 2024).

Desde 2005, o Comitê Gestor da Internet no Brasil (CGI.br), por intermédio do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), realiza anualmente a pesquisa intitulada *TIC Domicílios*. Esse estudo avalia o acesso e o uso das TIC nas residências brasileiras, utilizando indicadores relacionados às habilidades digitais e ao uso das tecnologias pela população. A pesquisa possui duas unidades de análise: os domicílios particulares permanentes e a população residente com 10 anos ou mais (CGI.br, 2024).

A investigação brasileira adota padrões metodológicos e indicadores internacionalmente estabelecidos, de modo a assegurar a comparabilidade dos resultados com outras pesquisas. Ela segue as diretrizes da iniciativa multissetorial *Partnership on Measuring ICT for Development*, coordenada pela International Telecommunication Union (ITU). Sua principal base de referência é o *Manual for Measuring ICT Access and Use by Households and Individuals*, utilizado por diversos países para orientar a coleta e a disseminação de estatísticas sobre TIC. Nesse manual, a ITU ressalta a relevância das pesquisas relacionadas às TIC Domicílios e destaca organizações que analisam tais informações para elaborar planos estratégicos de impacto global, como a Organização das Nações Unidas (ONU), que desde 2000 estuda questões relacionadas à produção, uso e descarte das TIC (ITU, 2014).

Foram selecionadas as bases de dados referentes aos anos de 2018 e 2023, com o objetivo de gerar perfis de usuários de internet em residências. A escolha desses dois períodos possibilitou a comparação entre perfis com intervalo de cinco anos.

A definição desse intervalo temporal, a partir da última pesquisa divulgada, foi inspirada na recomendação de De Solla Price (1963) *apud* Salvati (2021), segundo a qual um quinquênio constitui um período adequado para a busca de referências bibliográficas voltadas à compreensão de determinado tema.

A base de 2023 continha originalmente 454 colunas e 21.171 linhas, enquanto a de 2018 possuía 230 colunas e 20.544 linhas.

4.2 Execução do pré-processamento

Nesta etapa, os dois conjuntos de dados originais foram transformados em estruturas organizadas, com informações relevantes e apropriadas para a aplicação do agrupamento. Inicialmente, as colunas estavam nomeadas de forma codificada, conforme o dicionário de dados, e receberam nomes mais intuitivos, a fim de facilitar a compreensão.

Na sequência, foram excluídas as variáveis que não estavam presentes simultaneamente nos dois bancos de dados, aquelas utilizadas para controle estatístico, como identificação do domicílio e códigos censitários, bem como as de mesmo valor semântico (por exemplo, idade e faixa etária), mantendo-se, nesses casos, apenas uma das variáveis.

O processo resultou em dois conjuntos: o de 2023, com 442 variáveis e 21.271 instâncias, e o de 2018, com 216 variáveis e 20.544 instâncias. Para facilitar a interpretação dos resultados, mantiveram-se apenas as variáveis comuns, totalizando 109 em ambos os conjuntos, preservando-se também o número de instâncias da etapa anterior.

Na verificação de dados faltantes, as respostas classificadas como “não sabe” foram removidas, uma vez que o interesse era apenas em perfis de indivíduos que demonstrassem convicção em suas respostas. Não foram identificadas respostas em branco ou com caracteres inválidos. Ao final dessa etapa, os conjuntos de dados resultantes apresentaram 109 variáveis cada, com 18.938 instâncias no ano de 2023 e 18.397 no ano de 2018.

4.3 Seleção de variáveis conforme o contexto

Na sequência, foi aplicado um filtro nos conjuntos de dados com a finalidade de selecionar apenas indivíduos que efetivamente utilizaram a internet, uma vez que o objetivo consistiu em obter perfis de usuários frequentes, reduzindo, assim, a quantidade de registros disponíveis. As pesquisas realizadas em 2018 e 2023 contemplaram indivíduos que utilizaram a internet por, no mínimo, três meses consecutivos anteriores à coleta de dados.

Para alcançar os objetivos desta pesquisa, variáveis sociodemográficas dos usuários, como faixa etária e sexo, não foram encaminhadas ao processo de agrupamento. Essa abordagem visa evitar possíveis vieses que tais variáveis poderiam introduzir no processo, interferindo na formação dos perfis criados pelo algoritmo, que se baseia exclusivamente no uso da internet.

As variáveis sociodemográficas pertencem a um grupo mais generalista, de acordo com a classificação de bases segmentadas, enquanto o uso da internet se enquadra em uma classe denominada “específica ao produto” (Wedel; Kamakura, 2000). Assim, optou-se por utilizar variáveis orientadas à segmentação específica do estudo para a realização do agrupamento.

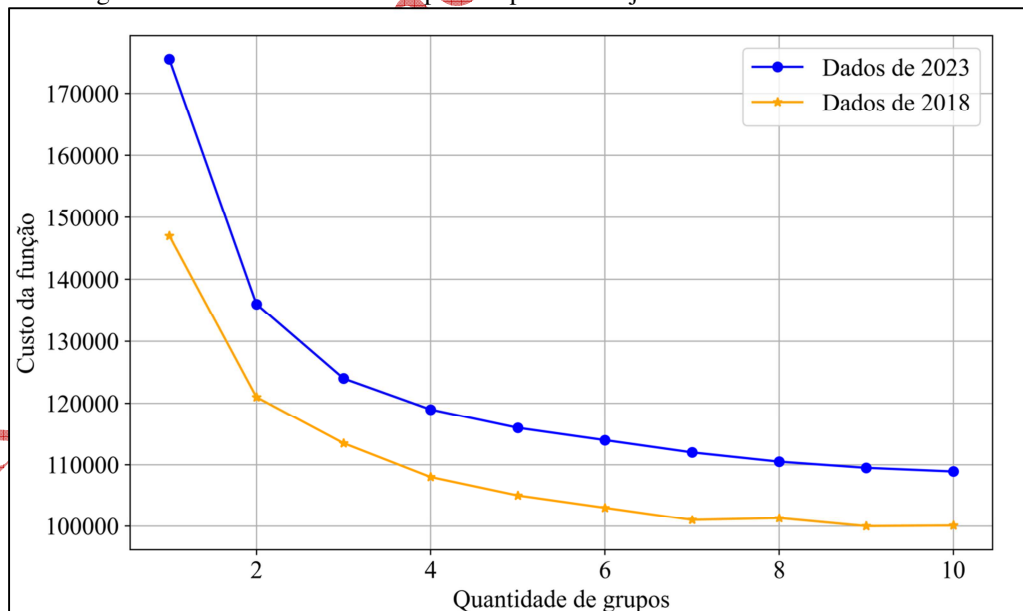
Após o filtro, os conjuntos resultaram em 50 variáveis cada, com 14.524 instâncias referentes a 2023 e 12.206 instâncias relativas a 2018.

4.4 Aplicação do algoritmo *k-modes* e validação do agrupamento

Nesta etapa, verificaram-se as características dos dados presentes nos dois conjuntos, a fim de selecionar a técnica de agrupamento mais adequada. Ao analisar os dados fornecidos pela pesquisa do cetic.br, observou-se a presença exclusiva de variáveis categóricas. Dessa forma, optou-se pelo algoritmo *k-modes*, que, mesmo em sua formulação clássica, permanece amplamente utilizado em tarefas de aprendizado não supervisionado voltadas ao processamento de dados categóricos (Gavva; Karthik; Punna, 2024).

O método do cotovelo foi aplicado às bases de 2018 e 2023, resultando nos gráficos apresentados na Figura 3.

Figura 3 – Método do cotovelo aplicado para os conjuntos de dados de 2018 e 2023.



Fonte: dados da pesquisa

A Figura 3 evidencia a possibilidade de dois ou três agrupamentos para 2023 e de dois ou quatro agrupamentos para 2018. Entretanto, após análise empírica, concluiu-se que três agrupamentos se

mostraram ideais para a base de 2023. Tal escolha permitiu a obtenção de grupos mais consistentes, com maior diferenciação em relação às características dos perfis originados.

4.5 Análise dos perfis identificados na etapa de agrupamento

Nesta etapa final, as dimensões extraídas de cada perfil (agrupamento) foram utilizadas como objeto de análise. Essas dimensões referem-se às características, comportamentos e padrões de uso que distinguem os usuários da internet.

Considerando que a proposta do estudo foi analisar os perfis criados para identificar a evolução do uso da internet domiciliar, bem como as habilidades associadas, o objetivo geral foi desdobrado em objetivos específicos, distribuídos em seis dimensões:

- **Informações pessoais:** analisa dados como sexo e idade dos indivíduos, composta pelas variáveis sexo, faixa etária, renda pessoal, raça, área urbana ou rural, renda familiar, condição da atividade (PEA), região geográfica, nível de instrução e classe social;
- **Meio de acesso e frequência:** examina os dispositivos utilizados para acessar a internet e a frequência desse uso, constituída pelas variáveis frequência no uso de internet e verificação do uso da internet por meio de computador, notebook, *tablet*, videogame, TV e outros aparelhos;
- **Locais de acesso:** verifica os locais de onde os indivíduos acessaram a rede mundial de computadores, formada por variáveis que analisam o uso da internet na residência do indivíduo, no trabalho, em instituições de ensino, na casa de outras pessoas, em locais públicos (gratuitos e não gratuitos), em deslocamento ou em outros locais, além do local de maior frequência de uso;
- **Habilidades com recursos de comunicação e informação:** avalia como os usuários empregam a internet para comunicação e obtenção de informações, composta por variáveis que verificam o uso de e-mail, envio de mensagens instantâneas (com ou sem áudio e vídeo), utilização de listas de discussão ou fóruns, uso de *microblogs*, busca por informações sobre produtos ou serviços, saúde, viagens, emprego, informações em sites de enciclopédias, em portais governamentais, uso de serviços públicos online e realização de transações financeiras;
- **Entretenimento:** analisa a utilização da internet para lazer e acesso a recursos culturais, formada por variáveis que verificam o uso da internet para jogos, músicas online e conteúdos em vídeo, bem como variáveis que avaliam o acesso a jornais, revistas e notícias, além do uso para visitar exposições e museus virtuais;
- **Habilidades com recursos de interesses individuais:** examina o uso da internet para finalidades pessoais, como estudo, busca de emprego, lazer e outras atividades de interesse próprio, composta por variáveis que constatarem atividades como realização de pesquisas ou tarefas escolares, participação em cursos online, busca de informações sobre graduações, pós-graduações ou cursos de extensão, busca de informações para estudo autônomo, uso de serviços de armazenamento, realização de atividades relacionadas ao trabalho, compartilhamento de conteúdos publicados, criação ou atualização de sites, postagem de conteúdos autorais, realização de downloads de filmes, músicas, jogos, softwares, e-books e séries, além da realização de compras online.

Dessa forma, cada objetivo específico resultou em agrupamentos de indivíduos, compondo as seis dimensões de análise.

No que se refere à consistência dos agrupamentos entre os anos de 2018 e 2023, verificou-se a porcentagem mais significativa de indivíduos em cada grupo, considerando as variáveis que os compõem. Observou-se que, em algumas dimensões, essa representatividade percentual sofreu mudanças ao longo do tempo.

As dimensões habilidades com recursos de comunicação e informação, entretenimento e habilidades para acessar serviços com recursos pessoais mostraram-se mais estáveis. Por outro lado, as

dimensões meio de acesso e frequência, locais de acesso e informações pessoais (ainda que esta última não tenha passado pelo processo de agrupamento) apresentaram maior variação na representatividade percentual das variáveis entre os grupos.

5 Resultados e discussões

Para a análise dos perfis, os dados foram organizados em tabelas que apresentam a predominância intra-agrupamento sobre as dimensões estudadas.

Inicialmente, foram examinados os protótipos finais (centroídes) obtidos nos agrupamentos realizados. O agrupamento 1 correspondeu a usuários com uso restrito da internet; o agrupamento 2, a usuários com uso moderado ou intermediário, pois apresentaram características médias em relação aos demais agrupamentos; e o agrupamento 3, a usuários com uso intenso e diversificado. Essas interpretações foram válidas para ambos os anos da pesquisa.

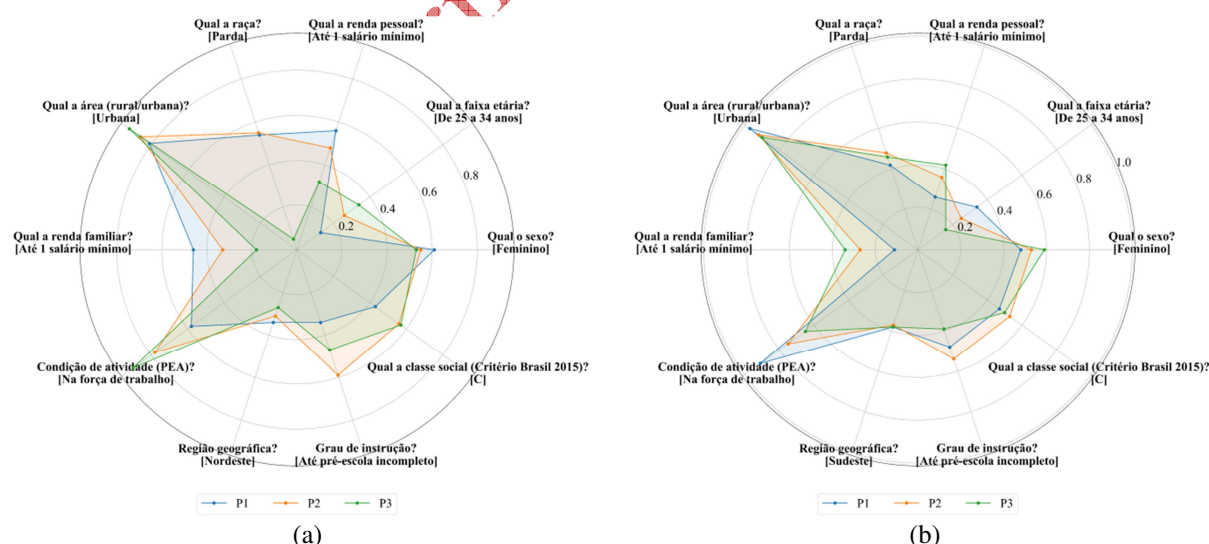
Outra análise introdutória consistiu na avaliação da moda de cada agrupamento, evidenciando os comportamentos majoritários. Em 2023, o Grupo 1 utilizava a internet em casa para envio de mensagens instantâneas; o Grupo 2 acessava a internet principalmente pela TV em seus lares; e o Grupo 3 fazia uso predominante de notebooks, seguido por *smart TV*. Já em 2018, o Grupo 1 acessava a internet em notebooks, sobretudo em casa e, em seguida, no trabalho; o Grupo 2 utilizava a internet em casa, em dispositivos de terceiros e durante deslocamentos; e o Grupo 3 se caracterizou pelo uso frequente da internet residencial para envio de mensagens instantâneas.

Essas análises permitem identificar inicialmente algumas características dos usuários de TIC Domicílio. Nas seções seguintes, os agrupamentos e dados da pesquisa foram representados em gráficos de radar, cujos rótulos no eixo categórico indicam a pergunta e, entre colchetes, a resposta mais recorrente.

5.1 Avaliação dos perfis correspondentes aos anos de 2023 e 2018

As Figuras 4a e 4b apresentam as avaliações dos perfis referentes às pesquisas de 2023 e 2018, respectivamente. Ressalta-se que os dados relativos à dimensão informações pessoais não foram processados por agrupamento, sendo analisados apenas na etapa pós-agrupamento.

Figura 4 – Gráficos da dimensão informações pessoais. (a) Ano de 2023. (b) Ano de 2018



Fonte: dados da pesquisa

A partir da Figura 4a, foi possível caracterizar três perfis distintos:

- **Perfil 1:** composto principalmente por mulheres, com baixa renda familiar e pessoal, grande participação no mercado de trabalho e maior proporção de indivíduos residentes em áreas rurais;
- **Perfil 2:** grupo intermediário, com predominância de indivíduos autodeclarados parados e com nenhuma ou muito baixo grau de instrução;
- **Perfil 3:** formado majoritariamente por jovens adultos, com baixa presença de pessoas pardas, pertencentes à classe trabalhadora e com aparente melhor estrutura econômica.

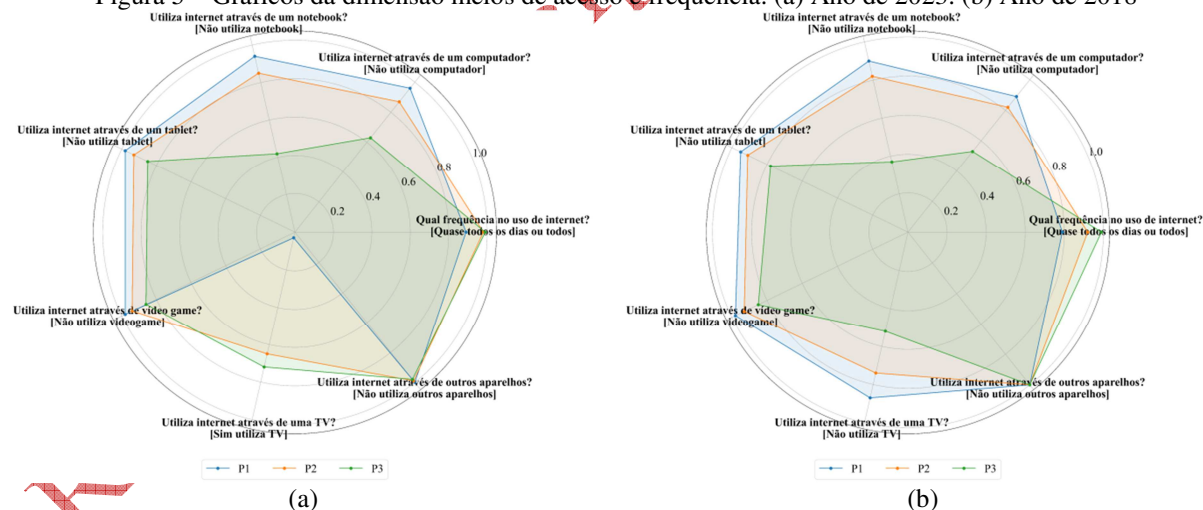
De forma semelhante, a partir da Figura 4b, observaram-se as principais características dos perfis de 2018:

- **Perfil 1:** composto em sua maioria por homens jovens, residentes em áreas urbanas, que trabalham para se sustentar;
- **Perfil 2:** grupo composto de forma quase equilibrada entre homens e mulheres, com melhores condições financeiras que o Perfil 1, maior concentração de indivíduos pardos, predominância de pessoas com escolaridade até a pré-escola incompleta ou sem formação educacional, pertencentes à classe social C;
- **Perfil 3:** predominantemente formado por mulheres, com baixa renda individual e familiar.

A Figura 5a mostra que, em 2023, os perfis apresentaram as seguintes características:

- **Perfil 1:** indivíduos que acessam a internet com menor frequência e não utilizam múltiplos dispositivos;
- **Perfil 2:** indivíduos que acessam a internet frequentemente, com pouca variação de dispositivos, mas com uso mais frequente de TV;
- **Perfil 3:** usuários de acesso intensivo e diário, com maior frequência de uso da TV e do notebook em comparação aos outros perfis.

Figura 5 – Gráficos da dimensão meios de acesso e frequência. (a) Ano de 2023. (b) Ano de 2018



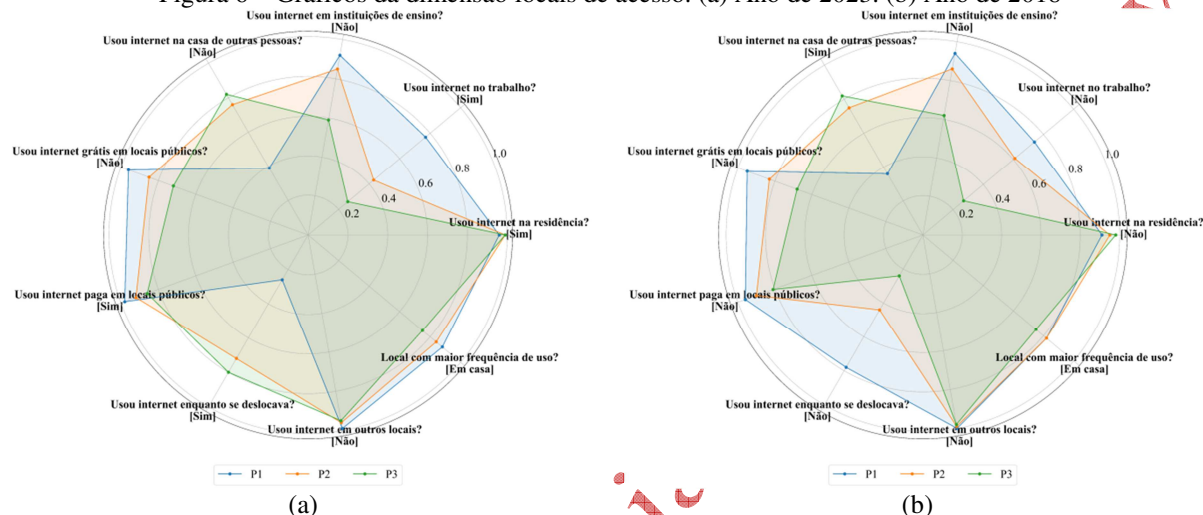
Na Figura 5b, identificaram-se as seguintes características para os perfis de 2018:

- **Perfil 1:** usuários com alta frequência de acesso, principalmente em notebooks e computadores, e menor utilização de *tablets* e videogames;
- **Perfil 2:** indivíduos com frequência de acesso intermediária;
- **Perfil 3:** pessoas com menor frequência de uso e que recorrem a poucos dispositivos.

A Figura 6a revela que, em 2023, os perfis foram assim caracterizados:

- **Perfil 1:** indivíduos que acessaram principalmente de suas residências, com pouca utilização em locais públicos, sendo o acesso nesses casos majoritariamente pago. Acesso limitado em deslocamentos e inexistente em instituições de ensino;
- **Perfil 2:** grupo intermediário, com quase metade dos indivíduos sem acesso no trabalho, mas com uso frequente em casa;
- **Perfil 3:** indivíduos que acessaram intensivamente durante deslocamentos, com destaque para o uso em locais públicos gratuitos. Foi o grupo com menor acesso no ambiente de trabalho.

Figura 6 – Gráficos da dimensão locais de acesso. (a) Ano de 2023. (b) Ano de 2018



Fonte: dados da pesquisa

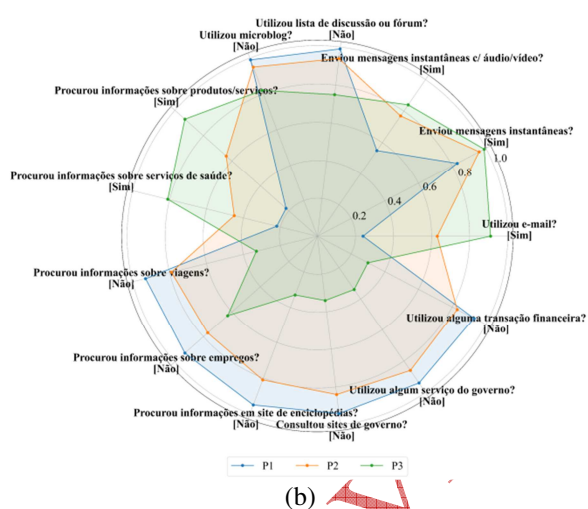
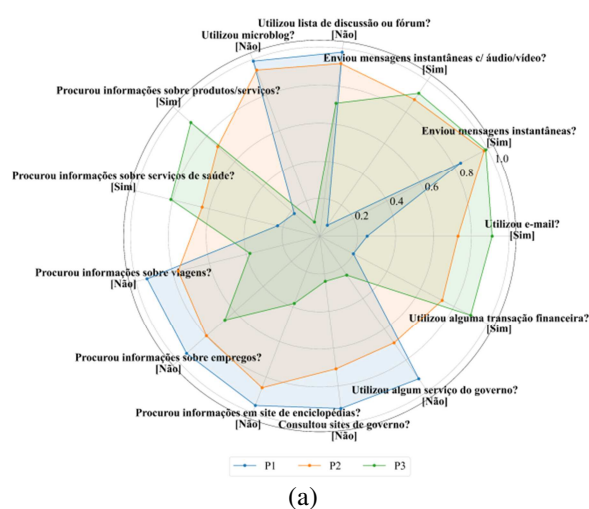
Por sua vez, a Figura 6b mostra os perfis de 2018:

- **Perfil 1:** baixo uso em residências, com predomínio do acesso na casa de terceiros, no trabalho, em meios de transporte e em locais públicos gratuitos;
- **Perfil 2:** uso intermediário no trabalho, associado ao acesso em residências próprias e de terceiros;
- **Perfil 3:** uso preferencial em residências (assim como o Perfil 2), com pouca utilização no trabalho, em instituições de ensino, na casa de terceiros ou em locais públicos (gratuitos ou pagos).

A partir da Figura 7a, as características dos perfis de 2023 foram sintetizadas em:

- **Perfil 1:** indivíduos que quase não realizam buscas por informações sobre áreas específicas, como empregos ou viagens, utilizam poucos recursos multimídia e ferramentas digitais, como serviços bancários, e em sua maioria não utilizam e-mail;
- **Perfil 2:** indivíduos com uso elevado de recursos multimídia, ainda que não em grande escala;
- **Perfil 3:** usuários com alto consumo de recursos tecnológicos, que buscam diversos tipos de informações em diferentes meios, além de utilizar serviços financeiros, bancários e outros disponibilizados pelo governo.

Figura 7 – Gráficos da dimensão habilidades com recursos de comunicação e informação. (a) Ano de 2023. (b) Ano de 2018



Fonte: dados da pesquisa

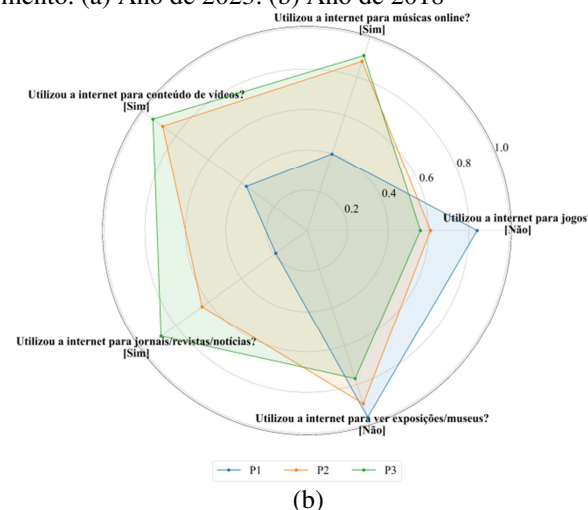
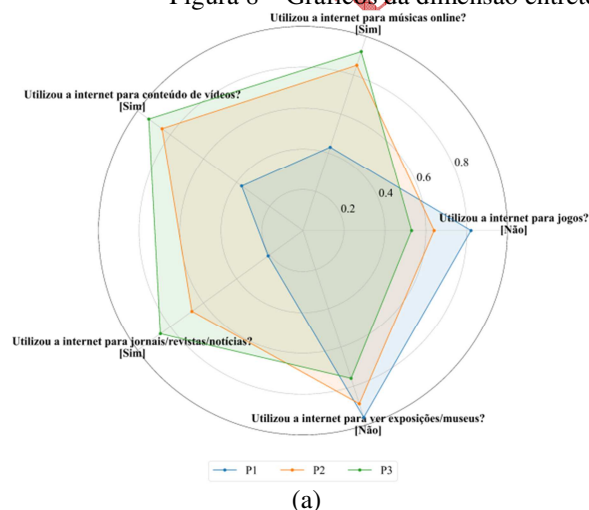
Observando a Figura 7b, obtiveram-se as principais características dos perfis de 2018:

- **Perfil 1:** grupo que mais utilizou e-mail e aplicativos de mensagens instantâneas, além de realizar buscas por serviços de saúde, informações sobre produtos, transações financeiras, sites governamentais, empregos e viagens;
- **Perfil 2:** pessoas que utilizaram recursos de comunicação e informação de forma moderada;
- **Perfil 3:** indivíduos que menos recorreram ao e-mail e a mensagens instantâneas, não utilizaram massivamente microblogs ou listas de discussão e foram os que menos buscaram serviços de saúde, informações sobre viagens, empregos, governo, entre outros.

A Figura 8a sintetiza os perfis de 2023 na dimensão entretenimento:

- **Perfil 1:** indivíduos sem interesse em jogos online, notícias, exposições ou museus virtuais;
- **Perfil 2:** pessoas que acessam de forma moderada serviços de música, vídeo e notícias;
- **Perfil 3:** indivíduos que mais consomem música, vídeos e notícias, com maior interesse em exposições ou museus virtuais e também em jogos.

Figura 8 – Gráficos da dimensão entretenimento. (a) Ano de 2023. (b) Ano de 2018



Fonte: dados da pesquisa

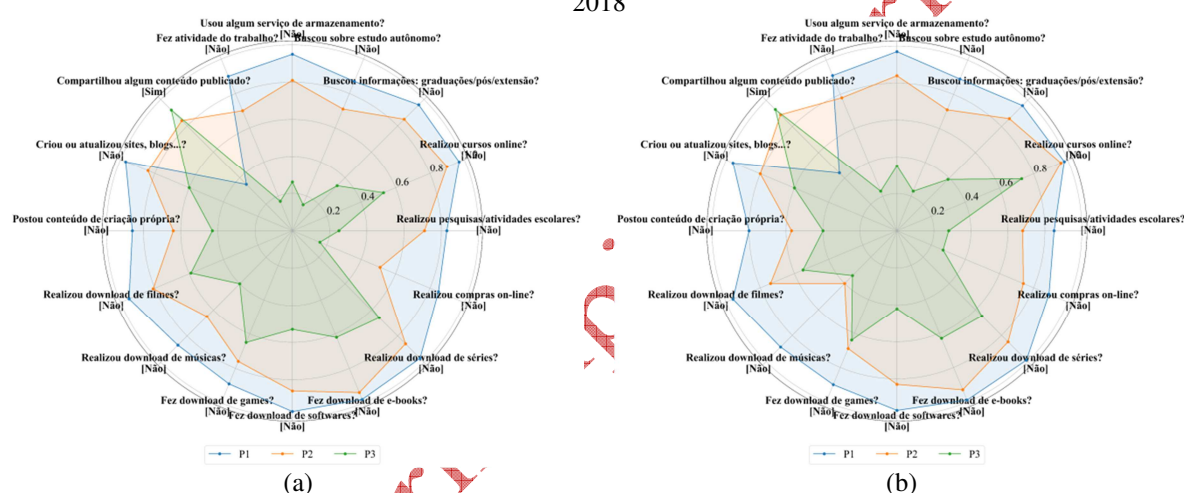
A Figura 8b apresenta os perfis de 2018:

- **Perfil 1:** usuários que mais acessaram jogos, música, vídeos e meios de informação digital, como revistas;
- **Perfil 2:** grupo com uso moderado de entretenimento digital;
- **Perfil 3:** grupo de menor representatividade, com pouco ou nenhum uso da internet para jogos, música, vídeos ou mídias digitais, e baixo interesse em exposições ou museus virtuais.

A Figura 9a mostra os perfis de 2023:

- **Perfil 1:** indivíduos com menor interesse ou habilidade em realizar atividades pessoais pela internet, como atividades escolares ou busca por cursos. Também apresentam baixo interesse em realizar downloads de conteúdos;
- **Perfil 2:** Grupo com uso moderado da internet para recursos pessoais;
- **Perfil 3:** Indivíduos que mais compartilham conteúdos disponíveis na internet, com alto interesse em realizar downloads de diversos itens e em efetuar compras online.

Figura 9 – Gráfico da dimensão habilidades para acessar serviços de caráter pessoal. (a) Ano de 2023. (b) Ano de 2018



Fonte: dados da pesquisa

Analogamente, a Figura 9b evidencia os perfis de 2018:

- **Perfil 1:** grupo que mais realizou pesquisas escolares e profissionais, atividades de estudo, compartilhamento de conteúdos e downloads de mídias diversas. Foi também o grupo que mais demonstrou interesse em compras online;
- **Perfil 2:** grupo de uso moderado da internet para serviços de caráter pessoal;
- **Perfil 3:** grupo com menor frequência de uso, menor interesse ou habilidade para downloads e compras online, além de baixa participação no compartilhamento de conteúdo.

Após essa análise detalhada, realizou-se o levantamento das características mais relevantes de cada dimensão, permitindo, assim, a identificação das principais características dos perfis de usuários de TIC Domiciliares nos anos estudados. Dessa forma, obteve-se um perfil resultante para 2023 e outro para 2018, conforme apresentado no Quadro 2.

Quadro 2 – Principais atributos representativos dos perfis resultantes

Dimensão	Perfil 2023	Perfil 2018
Informações pessoais	Mulheres com rendimento de até um salário mínimo, renda familiar de um salário mínimo, residentes na região Nordeste	Homens com rendimento de até um salário mínimo, renda familiar de um salário mínimo, residentes na região Sudeste

Meio de acesso e frequência	Acessam a internet quase diariamente e utilizam a TV como principal meio de acesso	Acessam a internet quase diariamente e utilizam outros aparelhos como meio de acesso
Locais de acesso	Utilizam a internet em suas residências e durante deslocamentos	Utilizam a internet nas casas de outras pessoas
Habilidades com recursos de comunicação e informação	Utilizam e-mail, enviam mensagens instantâneas, inclusive com áudio e vídeo, buscam informações sobre produtos e serviços, incluindo serviços de saúde, e realizam transações financeiras	Utilizam e-mail, enviam mensagens instantâneas, inclusive com áudio e vídeo, e buscam informações sobre produtos e serviços, incluindo serviços de saúde
Entretenimento	Utilizam a internet para ouvir músicas online, acessar conteúdos em vídeo, jornais, revistas e notícias	Utilizaram a internet para ouvir músicas online, acessar conteúdos em vídeo, jornais, revistas e notícias
Habilidades para acessar serviços com recursos pessoais	Compartilham conteúdos publicados na internet	Compartilham conteúdos publicados na internet

Fonte: dados da pesquisa

Os resultados permitiram traçar uma visão geral dos perfis de indivíduos usuários de TIC domiciliares nos anos de 2023 e 2018, possibilitando comparações entre dois contextos temporais. Para embasar as análises e identificar tendências no uso da internet residencial, foi utilizada como referência a pesquisa conduzida em 2021 pelo Grupo Banco Mundial (*World Bank Group – WBG*) e com o apoio do Programa das Nações Unidas para o Desenvolvimento (*United Nations Development Programme – UNDP*, publicada em 2022, que reuniu dados sobre emprego, renda, segurança alimentar, saúde, educação, serviços financeiros, conectividade de banda larga e impactos da pandemia de COVID-19 (WBG, 2022).

5.2 Dimensão sobre informações pessoais

De modo geral, observou-se alteração nos perfis quanto à renda familiar e pessoal, sendo o aumento da escolaridade mais perceptível no perfil intermediário (P2). Quanto à localização geográfica, em 2018 a maioria dos respondentes estava no Nordeste, enquanto em 2023 concentrou-se no Sudeste, com aumento da proporção de moradores em áreas urbanas.

O percentual de mulheres manteve-se em destaque, embora com leve queda em todos os agrupamentos. A autodeclaração de raça parda permaneceu predominante.

O perfil socioeconômico mais representativo para o ano de 2023 foi o agrupamento formado predominantemente por mulheres de cor parda, pertencentes à classe social C, com renda individual de até um salário mínimo e renda familiar de um salário mínimo, residentes na região Nordeste. Já para o ano de 2018, o agrupamento mais representativo era composto por homens, metade dos quais pertencentes à classe C, com renda individual de até um salário mínimo, renda familiar de um salário mínimo e residentes na região Sudeste.

Percebe-se que os entrevistados de 2018 apresentavam melhores condições socioeconômicas, embora grande parte pertencesse à classe social C, o que reforça a predominância de indivíduos das classes mais desfavorecidas. Nas descrições das políticas públicas nacionais atuais, observa-se que o foco está direcionado para grupos em condições socioeconômicas desfavoráveis, como aponta o estudo de Franco (2021).

Esses resultados sugerem mudanças estruturais na composição dos grupos ao longo do tempo, indicando melhoria nos indicadores socioeconômicos e redistribuição geográfica dos respondentes. Os achados dialogam com a pesquisa conduzida pelo WBG (2022), segundo a qual domicílios com indivíduos de nível superior têm quase o dobro de probabilidade de possuir conexões fixas de internet em comparação àqueles com apenas ensino fundamental, resultado associado à maior renda e necessidade de conectividade intensificada pela pandemia de COVID-19. O estudo também aponta

que, em média, 74% das famílias urbanas da região possuem acesso à internet fixa, contra 42% em áreas rurais.

Esses resultados estão em consonância com o relatório do WBG (2022), segundo o qual as famílias na América Latina e no Caribe enfrentam dificuldades de acesso à internet devido a fatores socioeconômicos desfavoráveis, sendo as tarifas elevadas um dos grandes obstáculos ao acesso.

5.3 Dimensão sobre acesso e frequência

A análise dos anos de 2018 e 2023 indica que o acesso diário à internet se tornou mais abrangente, sobretudo no perfil P1, que concentra a população com maiores dificuldades financeiras. O uso de computadores e notebooks caiu, com exclusão superior a 85% em todos os perfis. *Tablets* e videogames mantiveram participação reduzida no acesso.

As *smart TVs* consolidaram-se como meio de acesso, enquanto o uso de outros dispositivos permaneceu praticamente inalterado.

Keshav (2005) já havia identificado a tendência de substituição de notebooks e desktops por dispositivos móveis, impulsionada por vantagens técnicas e não técnicas. Desde a década passada, destaca-se que o *smartphone* pode ser considerado um novo tipo de computador, por ser geralmente mais barato, mais conveniente devido à sua portabilidade e útil no contexto da geolocalização (Allen; Graupera; Lundrigan, 2010).

Oliveira e Barroco (2023) também afirmam que, atualmente, o *smartphone* é uma ferramenta central para a vida cotidiana devido à sua tecnologia, portabilidade e operacionalidade, apresentando múltiplas funções que conferem versatilidade às atividades diárias. Essa tendência também foi observada em outros países da América Latina e do Caribe, onde a internet móvel por *smartphones* é predominante. Em média, 95% dos domicílios urbanos da região possuem ao menos um membro com acesso a um *smartphone* (WBG, 2022).

A análise dos perfis criados por meio do agrupamento demonstrou que, em 2023, houve um aumento no uso de outros dispositivos, em comparação aos perfis de 2018, não especificados na pesquisa, para o acesso à internet. Provavelmente, os *smartphones* estão entre eles, uma vez que, conforme apontado pelo relatório WBG (2022), a maioria das famílias na América Latina e no Caribe acessa a internet por meio desses aparelhos.

5.4 Dimensão sobre locais de acesso

Verificou-se que o acesso domiciliar tornou-se o principal meio de conexão, em detrimento de locais como instituições de ensino. No ambiente de trabalho, o uso aumentou no perfil P1 e diminuiu nos perfis de maior renda (P2 e P3).

O acesso gratuito em locais públicos reduziu, enquanto aumentaram o acesso móvel, o acesso público pago e o uso durante deslocamentos. A expansão das redes 4G e 5G em municípios brasileiros explica, em parte, a ampliação desse acesso (Bertollo, 2024).

O uso da internet em residências de terceiros diminuiu, o que reforça a expansão da internet fixa e móvel nos domicílios próprios. Outra explicação está no período da pandemia de COVID-19, quando o isolamento social ampliou a demanda por tráfego de internet nos lares (Feldmann *et al.*, 2020).

5.5 Dimensão sobre habilidades com recursos de comunicação e informação

Verificou-se que o e-mail é mais utilizado entre os grupos digital e financeiramente mais estruturados ("P2" e "P3"), que também preferem o envio de áudios e vídeos em mensagens instantâneas, enquanto o perfil "P1" tende a preferir mensagens em formato de texto. O uso de fóruns e microblogs caiu drasticamente no geral, embora tenha apresentado crescimento no perfil "P3".

A busca por informações sobre produtos, saúde e serviços governamentais aumentou, enquanto as pesquisas relacionadas a empregos e viagens sofreram redução. A utilização de serviços financeiros e governamentais cresceu, em especial no perfil "P3", o mais estruturado socioeconomicamente.

De modo geral, observa-se que o acesso a serviços digitais avançou de forma mais expressiva entre indivíduos dos perfis mais bem estruturados tecnologicamente e financeiramente ("P3"). Em

contrapartida, os pertencentes ao perfil "P1" ainda enfrentam barreiras no acesso a esses serviços no Brasil. Boerkamp *et al.* (2024) destacam que fatores como custos de serviços e dispositivos, habilidades digitais limitadas e baixa frequência de uso são elementos que dificultam a apropriação da internet entre famílias em situação de vulnerabilidade.

O WBG (2022) também apontam que essas barreiras estão presentes em outros países da América Latina e do Caribe, especialmente no que diz respeito ao uso de mensagens instantâneas. Nesses contextos, indivíduos com apenas ensino fundamental utilizam a internet predominantemente para comunicação por mensagens instantâneas. A utilização desse recurso foi o principal caso de uso da internet em países como Nicarágua e República Dominicana, além de figurar entre os mais relevantes em Dominica e Brasil.

A análise dos perfis criados evidenciou que, em todos os agrupamentos, houve envio de mensagens instantâneas por grande número de indivíduos, tanto em 2018 quanto em 2023.

5.6 Dimensão sobre entretenimento

Nesta dimensão, concluiu-se que o consumo de músicas e vídeos online é predominante nos perfis "P2" e "P3", havendo crescimento discreto também no perfil "P1". A utilização da internet para jogos apresentou queda entre "P2" e "P3".

O acesso a notícias online aumentou nos perfis "P1" e "P2", enquanto no perfil "P3" observou-se diversificação nos meios de consumo de notícias. O interesse por exposições e museus virtuais permanece baixo, ainda que com discreto crescimento no perfil "P1".

Esses resultados evidenciam que o consumo de entretenimento digital está em ascensão, embora com transformações nos meios de acesso. O estudo do WBG (2022) identificaram tendência semelhante em países latino-americanos e caribenhos, com crescimento do uso de jogos online, especialmente em Costa Rica, México e Uruguai. A partir dos perfis resultantes deste estudo, observa-se que os jogos online não eram tão populares em 2018, mas apresentaram crescimento significativo entre os usuários de internet em 2023, sendo o grupo com melhores condições financeiras o principal consumidor desse tipo de entretenimento. Ressalta-se que o relatório WBG (2022) não abordou outros tipos de entretenimento analisados nesta pesquisa.

Um estudo mais antigo elaborado por Grazzi e Vergara (2014) identificou que países como Brasil, Chile e Costa Rica usam muito a internet para o entretenimento, e a renda das pessoas possui uma relação positiva significativa neste aspecto, ou seja, quanto maior a renda, maior o uso para o entretenimento, diferentemente de países como El Salvador, Honduras, México e Paraguai. Nesta mesma pesquisa, especificamente para o Brasil, encontrou-se uma relação negativa para a dimensão do entretenimento e variáveis demográficas como idade, ser estudante e estar empregado, indicando que a questão da falta de tempo devido ao trabalho, estudos e responsabilidades não permitem o uso da internet para entretenimento.

5.7 Dimensão sobre habilidades para acessar serviços com recursos pessoais

Verificou-se uma redução no interesse por cursos online e pesquisas acadêmicas, sobretudo nos perfis "P1" e "P2". Em contrapartida, houve aumento no compartilhamento de conteúdos nesses perfis. Nos três perfis, observou-se redução no uso de serviços de armazenamento em nuvem, enquanto as compras online cresceram significativamente entre os grupos mais favorecidos ("P2" e "P3").

Também foi registrada redução no download de filmes, músicas e softwares, possivelmente em decorrência da ampliação do acesso a serviços de streaming. Esse fenômeno remodelou a forma como as pessoas consomem mídia (Falkowski-Gilski; Uhl, 2020).

Os dados indicam mudanças relevantes no comportamento digital: aumento nas compras online, menor uso de armazenamento físico e maior consumo de serviços digitais, governamentais e bancários. A pandemia de COVID-19 impulsionou o comércio eletrônico, modificando hábitos de consumo (Afonso; Carneiro; Azevedo, 2024). Paralelamente, a consolidação das ferramentas baseadas em nuvem, que reduzem custos e oferecem maior flexibilidade, contribuiu para a diminuição no uso de armazenamento físico (Akindote *et al.*, 2023). Com o aumento do tráfego de dados na internet, durante o período da pandemia de COVID-19, especialmente nas redes que prestam serviços a usuários

residenciais, observou-se um crescimento no uso de diversos serviços online por esses indivíduos (Feldmann *et al.*, 2020).

O WBG (2022) destaca que famílias da América Latina e do Caribe cujo chefe possui ensino superior tendem a utilizar a internet principalmente para fins educacionais e de trabalho remoto. O Chile é um exemplo expressivo, sendo o único país em que o trabalho remoto aparece entre os três principais usos da internet, em contraste com nações caribenhas, onde sequer figura entre os cinco principais usos. Essa diferença reflete estímulos governamentais para o uso de recursos digitais durante a pandemia.

No presente estudo, a redução no interesse por cursos online e pesquisas acadêmicas nos perfis "P1" e "P2" em 2023 pode estar associada ao término do período pandêmico e ao retorno das atividades presenciais.

De modo geral, observou-se que:

- Houve melhorias socioeconômicas significativas apenas entre indivíduos de perfil intermediário;
- Pessoas pardas e mulheres continuam em maior número entre os usuários de internet;
- O acesso à internet aumentou, sobretudo entre os mais carentes, com substituição progressiva de computadores por dispositivos móveis e *smart TVs*;
- O acesso domiciliar tornou-se quase universal, enquanto entre os mais pobres cresceu o acesso em ambientes de trabalho;
- Os perfis mais estruturados tecnologicamente ampliaram o uso de serviços digitais, enquanto os mais vulneráveis ainda enfrentam dificuldades;
- O consumo de entretenimento digital e notícias se intensificou, mas o acesso a museus virtuais permaneceu limitado;
- Houve aumento expressivo de compras online nas classes mais favorecidas e redução na busca por cursos e conteúdos acadêmicos nos grupos mais pobres;
- Em todos os perfis houve queda no uso de armazenamento em nuvem e downloads de conteúdo.

Ao final da apresentação dos resultados, em que foram analisadas detalhadamente as variáveis de cada dimensão, de modo a caracterizá-las, verifica-se que este estudo se mostrou coerente com o relatório do WBG (2022), o qual destaca que o uso da internet varia consideravelmente de acordo com o nível educacional e o gênero, além de apontar que a lacuna digital tende a acentuar as desigualdades socioeconômicas.

O trabalho de Franco (2021) descreve as diversas políticas públicas desenvolvidas no Brasil para ampliar o acesso à internet e promover a inclusão digital. Foram identificadas dez iniciativas do Governo Federal, entre 2010 e 2019, voltadas à ampliação do acesso à internet no país. Além disso, constatou-se que a inclusão digital deve ser observada sob duas perspectivas: o acesso à infraestrutura e a educação para o uso das TIC, com ênfase no letramento digital. De modo geral, essas iniciativas tiveram como objetivo ampliar o acesso à banda larga, promover a inclusão digital, reduzir desigualdades e fortalecer o desenvolvimento socioeconômico e tecnológico do país.

A análise dos perfis dos agrupamentos realizados neste estudo demonstrou que houve avanços no acesso à internet no quinquênio analisado, sobretudo entre os grupos mais pobres. Ainda assim, os fatores socioeconômicos continuam a interferir no acesso reduzido à internet e aos serviços digitais disponíveis. Por exemplo, os grupos com maior acesso a jogos online e à realização de compras pela internet estão entre os mais bem estruturados socioeconomicamente.

Embora tenha sido constatada uma melhora nas condições socioeconômicas dos grupos intermediários, ainda se faz necessária a elaboração e a implementação efetiva de políticas públicas que garantam a todos o acesso e o conhecimento necessários para a plena utilização das TIC domiciliares.

6 Conclusão

Ao concluir esta pesquisa, constatou-se que as técnicas de mineração de dados apresentam aplicação relevante na criação de perfis a partir de bases de dados, permitindo a análise posterior de suas características, especialmente com o uso do algoritmo *k-modes* em bases de dados totalmente categóricas.

Quanto ao objetivo principal do estudo que era a construção de agrupamentos de perfis de indivíduos usuários de TIC, sobretudo de internet em nível domiciliar, este foi alcançado tanto para os dados referentes a 2023 quanto para os de 2018. Esse resultado possibilitou verificar mudanças ao longo do tempo e elaborar um perfil geral para o quinquênio analisado, oferecendo uma visão abrangente dos indivíduos nesse período.

A escolha do parâmetro *k* do algoritmo *k-modes*, que corresponde ao número de agrupamentos ideais para validar a qualidade do processo, mostrou-se adequada nesta pesquisa. A definição em três agrupamentos permitiu identificar perfis com características socioeconômicas distintas, facilitando a interpretação e a análise dos resultados, bem como a compreensão dos hábitos e habilidades dos indivíduos usuários de internet em relação às suas condições socioeconômicas.

Observou-se que os perfis de 2023 e 2018 apresentaram diferenças significativas. Com base nelas, foi possível propor um perfil resultante que representasse os indivíduos em nível domiciliar usuários de TIC, especialmente de internet, durante o quinquênio considerado. Esse perfil apresentou convergência com características de usuários de internet em domicílios nos países da América Latina e do Caribe.

No que se refere à dimensão “habilidades para acessar serviços com recursos pessoais”, verificou-se redução expressiva no interesse por cursos on-line, assim como nas buscas relacionadas a áreas acadêmicas, entre 2018 e 2023. Como discutido, essa mudança pode estar relacionada ao fim da pandemia de COVID-19, período no qual houve incentivo ao encerramento de atividades de estudo e trabalho domiciliar. Esse aspecto poderá ser aprofundado em estudos futuros.

Este trabalho configurou-se como um estudo preliminar, cujo propósito foi a construção de perfis de indivíduos usuários de TIC em nível domiciliar, sobretudo de internet, e a análise de suas características e mudanças ao longo de cinco anos. O algoritmo *k-modes* foi aplicado com êxito, confirmando sua adequação como método tradicional para processos de agrupamento nesse tipo de estudo.

Para pesquisas futuras, sugere-se a aplicação de modelos de inteligência artificial explicável (como o *SHapley Additive exPlanations* – SHAP) e de técnicas de comparação entre grupos (por exemplo, Análise de Variância – ANOVA ou Kruskal-Wallis). A incorporação desses métodos poderá evidenciar quais variáveis da pesquisa realizada pelo cetec.br exercem maior influência na formação dos perfis de usuários de TIC em nível domiciliar, trazendo novas percepções e confirmando resultados já discutidos nesta investigação.

Financiamento

Esta pesquisa não recebeu financiamento.

Conflito de interesses

Os autores declaram não haver conflito de interesses.

Nota (se aplicável; não deve ser numerada)

Este artigo é fruto parcial do desenvolvimento de dissertação do curso de Mestrado em Engenharia de Produção da Universidade Estadual de Maringá (UEM).

Contribuições ao artigo

ADIGNERI, H. M.: concepção ou desenho do estudo/pesquisa; análise e/ou interpretação dos dados; elaboração e redação do manuscrito. **SANTOS, B. S.:** supervisão geral e coordenação do projeto ou estudo; coleta, análise e/ou interpretação dos dados. **LIMA, R. H. P.:** coleta, análise e/ou interpretação dos dados; revisão final com participação crítica e intelectual no manuscrito. **LEAL, G.**

C. L.: revisão final com participação crítica e intelectual no manuscrito. Todos os autores participaram da escrita, discussão, leitura e aprovação da versão final do artigo.

Referências

- AFONSO, A. P.; CARNEIRO, J.; AZEVEDO, A. I. The impact of COVID-19 on e-commerce: a systematic review of the literature on the purchasing behavior of online retail consumers. **Journal of Marketing Research and Case Studies**, v. 2024, e403212, 2024. DOI: <https://doi.org/10.5171/2024.403212>.
- AGGARWAL, C. C.; REDDY, C. K. **Data clustering**: algorithms and applications. Boca Raton: CRC Press, 2014.
- AHMAD, A.; DEY, L. A *k*-mean clustering algorithm for mixed numeric and categorical data. **Data & Knowledge Engineering**, v. 63, n. 2, p. 503-527, 2007. DOI: <https://doi.org/10.1016/j.datak.2007.03.016>.
- AKINDOTE, O. J.; ADEGBITE, A. O.; DAWODU, S. O.; OMOTOSHO, A.; ANYANWU, A. Innovation in data storage technologies: from cloud computing to edge computing. **Computer Science & IT Research Journal**, v. 4, n. 3, p. 273-299, 2023. DOI: <https://doi.org/10.51594/csitrj.v4i3.661>.
- AL-RAHMI, W.M.; ALZHRANI, A. I.; YAHAYA, N.; ALALWAN, N.; KAMIN, Y. B. Digital communication: information and communication technology (ICT) usage for education sustainability. **Sustainability**, v. 12, n. 12, e5052, 2020. DOI: <https://doi.org/10.3390/su12125052>.
- ALLEN, S.; GRAUPERA, V.; LUNDRIGAN, L. The smartphone is the new PC. In: ALLEN, S.; GRAUPERA, V.; LUNDRIGAN, L. (Eds.). **Pro smartphone cross-platform development**. Berkeley: Apress, 2010, p. 1-14. DOI: https://doi.org/10.1007/978-1-4302-2869-1_1
- BERTOLLO, M. Internet e telefonia móvel no Brasil: qualidade e densidade das redes no território. **Confins**, v. 64, 2024. DOI: <https://doi.org/10.4000/12f3k>.
- BOERKAMP, L. G. P.; VAN DEURSEN, A. J. A. M.; VAN LAAR, E.; VAN DER ZEEUW, A.; VAN DER GRAAF, S. Exploring barriers to and outcomes of internet appropriation among households living in poverty: a systematic literature review. **SAGE Open**, v. 14, n. 1, 2024. DOI: <https://doi.org/10.1177/21582440241233047>.
- CGLBR – COMITÊ GESTOR DA INTERNET NO BRASIL. **TIC Domicílios 2023**: pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros. São Paulo: Comitê Gestor da Internet no Brasil, 2024. Disponível em: https://cetec.br/media/docs/publicacoes/2/2024/104102822/tic_domicilios_2023_livro_eletronico.pdf. Acesso em: 22 set. 2025.
- COGÓ, D.; DUTRA-BRIGNOL, L. FRAGOSO, S. Práticas cotidianas de acesso às TIC: outro modo de compreender a inclusão digital. **Palavra Chave**, v. 18, n. 1, p. 156-183, 2015. DOI: <https://doi.org/10.5294/pacla.2015.18.1.7>.
- CORTÉS, F.; TEZANOS-PINTO, P.; HELSPER, E.; LAY, S.; MANZI, J.; NOVOA, C. ¿Se ha reducido la brecha digital en Chile? Diferencias entre acceso, uso y factores asociados al empleo de Internet. **MIDevidencias**. Disponível em: <https://mideuc.cl/wp-content/uploads/2022/11/MIDevidencias-N22.pdf>. Acesso em: 23 set. 2025.

DATTA, S.; DAVIM, J. P. **Machine learning in industry**. Cham: Springer, 2021. DOI: <https://doi.org/10.1007/978-3-030-75847-9>.

FALKOWSKI-GILSKI, P.; UHL, T. Current trends in consumption of multimedia content using online streaming platforms: a user-centric survey. **Computer Science Review**, v. 37, e100268, 2020. DOI: <https://doi.org/10.1016/j.cosrev.2020.100268>.

FELDMANN, A.; GASSER, O.; LICHTBLAU, F.; PUJOL, E.; POESE, I.; DIETZEL, C.; WAGNER, D.; WICHTLHUBER, M.; TAPIADOR, J.; VALLINA-RODRIGUEZ, N.; HOHLFELD, O.; SMARAGDAKIS, G. The lockdown effect: implications of the COVID-19 pandemic on internet traffic. In: ACM INTERNET MEASUREMENT CONFERENCE (IMC '20), 2020, Virtual Event. **Proceedings [...]**. ACM, p. 1-18, 2020. DOI: <https://doi.org/10.1145/3419394.3423658>.

FLORES-CUETO, J. J.; HERNÁNDEZ, R. M.; GARAY-ARGANDOÑA, R. Tecnologías de información: acceso a internet y brecha digital en Perú. **Revista Venezolana de Gerencia (RVG)**, v. 25, n. 90, p. 504-527, 2020. Disponível em: <https://produccioncientificaluz.org/index.php/rvg/article/view/32396>. Acesso em: 23 set. 2025.

FRANCO, A. H. C. Políticas públicas de informação: um olhar para o acesso à internet e para a inclusão digital no cenário brasileiro. **Em Questão**, v. 27, n. 4, p. 61-83, 2021. DOI: <https://doi.org/10.19132/1808-5245274.61-83>.

GAVVA, S. T.; KARTHIK, C. S.; PUNNA, S. Clustering categorical data: soft rounding k-modes. **Information and Computation**, v. 296, e105115, 2024. DOI: <https://doi.org/10.1016/j.ic.2023.105115>.

GRAZZI, M.; VERGARA, S. Internet in Latin America: who uses it? ... and for what? **Economics of Innovation and New Technology**, v. 23, n. 4, p. 327-352, 2014. DOI: <http://dx.doi.org/10.1080/10438599.2013.854513>.

HAN, J.; PEI, J.; TONG, H. **Data mining concepts and techniques**. 4. ed. Cambridge: Morgan Kaufmann, 2022.

HUANG, Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. **Data Mining and Knowledge Discovery**, v. 2, n. 3, p. 283-304, 1998. DOI: <https://doi.org/10.1023/A:1009769707641>.

ITU – INTERNATIONAL TELECOMMUNICATION UNION. **Manual for measuring ICT access and use by households and individuals**, 2014. Disponível em: https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ITCMEAS-2014-PDF-E.pdf. Acesso em: 23 set. 2025.

KESHAV, S. Why cell phones will dominate the future internet. **ACM SIGCOMM Computer Communication Review**, v. 35, n. 2, p. 83-86, 2005. DOI: <https://doi.org/10.1145/1064413.1064425>.

LEE, J.; BABCOCK, J.; PHAM, T. S.; BUI, T. H.; KANG, M. Smart city as a social transition towards inclusive development through technology: a tale of four smart cities. **International Journal of Urban Sciences**, v. 27, n. 1, p. 75-100, 2022. DOI: <https://doi.org/10.1080/12265934.2022.2074076>.

LIU, F.; DENG, Y. Determine the number of unknown targets in open world based on elbow method. **IEEE Transactions on Fuzzy Systems**, v. 29, n. 5, p. 986-995, 2021. DOI: <http://doi.org/10.1109/TFUZZ.2020.2966182>.

MURATO, D. M.; SANTOS, B. S.; LIMA, R. H. P. Clustering and analysis of tweets related to Petrobras. **Cadernos do IME - Série Informática**, v. 49, p. 113-131, 2024. DOI: <https://doi.org/10.12957/cadinf.2024.82401>.

MWIINGA, P. **Information and communications technology**. Zenodo, v. 1, 2023. DOI: <https://www.doi.org/10.5281/zenodo.10406885>.

NIC.BR – Núcleo de Informação e Coordenação do Ponto BR. **Pesquisa sobre o uso das tecnologias de informação e comunicação**: pesquisa TIC Domicílios, ano 2018. Comitê Gestor da Internet no Brasil, 2019. Disponível em: <http://cetic.br/pt/arquivos/domicilios/2018/domicilios>. Acesso em: 23 set. 2025.

NIC.BR – Núcleo de Informação e Coordenação do Ponto BR. **Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros**: pesquisa TIC Domicílios, ano 2023. Comitê Gestor da Internet no Brasil, 2024. Disponível em: <https://cetic.br/pt/arquivos/domicilios/2023/domicilios>. Acesso em: 23 set. 2025.

OLIVEIRA, F. A. F.; BARROCO, S. M. S. Revolução tecnológica e smartphone: considerações sobre a constituição do sujeito contemporâneo. **Psicologia em Estudo**, v. 28, 2023. DOI: <https://doi.org/10.4025/psicoestud.v28i0.51648>.

RENÇBER, Ö. F.; ZORKIRIŞCI, E. D.; AYTEKIN, M. Clustering of households based on their use of information technologies: a review of Türkiye. **Journal of Political Administrative and Local Studies**, v. 6, n. 2, p. 75-86, 2023. Disponível em: <https://dergipark.org.tr/en/pub/jpaljournal/issue/80403/1356848>. Acesso em: mar. de 2025.

RUDRAWAR, S. S. The use of ICT in our everyday life can have both positive and negative impacts. **Vidyawarta: Interdisciplinary Multilingual Referred Journal**, v. 3, n. 7, p. 122-125, jul. 2014.

SACCOL, A. Z. Um retorno ao básico: compreendendo os paradigmas de pesquisa e sua aplicação na pesquisa em administração. **Revista de Administração da UFSM**, v. 2, n. 2, p. 250-269, 2010. DOI: <https://doi.org/10.5902/198346591555>.

SALVIATI, R. C. M. **Fatores críticos de sucesso na captação de recursos junto a agências internacionais de fomento à pesquisa e desenvolvimento: um estudo na área de saúde no Brasil**. 2021. Tese (Doutorado em Tecnologia e Sociedade) – Universidade Tecnológica Federal do Paraná, Curitiba, 2021. Disponível em: <http://repositorio.utfpr.edu.br/jspui/handle/1/26810>. Acesso em: 10 maio 2025.

SANTOS, M. R.; GOMES, M. M. F. Perfil das pessoas idosas no Brasil: um estudo sobre acesso à internet e habilidades digitais por gerações. **Caderno Pedagógico**, v. 20, n. 7, 2023. DOI: <https://doi.org/10.54033/cadpedv20n7-003>.

SHARMA, M. Influence of ICT and its dynamic change in daily life of human being. **The Journal of Contemporary Issues in Business and Government**, v. 27, n. 3, p. 635-639, abr. 2021. Disponível em: <https://cibgp.com/index.php/1323-6903/article/view/1643>. Acesso em: 23 set. 2025.

SINGH, J.; SHIEKH, A. S.; KOUR, M.; KUMAR, P. Student learning and the role of information and communication technology (ICT) in 21st century: a review. **International Journal on Integrated Education**, v. 3, n. 9, p. 181-185, 2020. DOI: <https://doi.org/10.31149/ijie.v3i9.617>.

SPIZZIRRI, R. C. P.; WAGNER, A.; MOSMANN, C. P.; ARMANI, A. B. Adolescência conectada: mapeando o uso da internet em jovens internautas. **Psicologia Argumento**, v. 30, n. 69, p. 327-335, 2017.

THORNDIKE, R. L. Who belongs in the family? **Psychometrika**, v. 18, p. 267-276, 1953. DOI: <https://doi.org/10.1007/BF02289263>.

WBG – WORLD BANK GROUP, UNDP – UNITED NATIONS DEVELOPMENT PROGRAMME. **Internet access and use in Latin America and the Caribbean: From the LAC high frequency phone surveys 2021, 2022**. Disponível em: <https://www.undp.org/sites/g/files/zskgke326/files/2022-09/undp-rblac-Digital-EN.pdf>. Acesso em: 10 mar. 2025.

WEDEL, M.; KAMAJURA, W. A. **Market segmentation**. Boston: Springer, 2000. DOI: <https://doi.org/10.1007/978-1-4615-4651-1>.

WICHT, A.; REDER, S.; LECHNER, C. M. Sources of individual differences in adults' ICT skills: a large-scale empirical test of a new guiding framework. **PLOS ONE**, v. 16, n. 4, e0249574, 2021. DOI: <https://doi.org/10.1371/journal.pone.0249574>.

YANG, W.; LAM, P. T. I. An evaluation of ICT benefits enhancing walkability in a smart city. **Landscape and Urban Planning**, v. 215, e104227, 2021. DOI: <https://doi.org/10.1016/j.landurbplan.2021.104227>.

YUAN, F.; YANG, Y.; YUAN, T. A dissimilarity measure for mixed nominal and ordinal attribute data in k-Modes algorithm. **Applied Intelligence**, v. 50, p. 1498-1509, 2020. DOI: <https://doi.org/10.1007/s10489-019-01583-5>.