

Desigualdades Territoriais na Qualidade dos Serviços de Telecomunicações: Uma Análise Interpretável com Clusterização Espacial

Marco Antonio A. Fidos Jr, Renato Machado, Dimas I. Alves, Bartolomeu F. Uchôa Filho e Lucio P. Amaro

Resumo—Este trabalho propõe uma abordagem explicável para análise espacial da qualidade dos serviços de telecomunicações no Brasil. Utilizando dados do sistema RQUAL (Anatel) e indicadores do IBGE, foram aplicadas técnicas de clusterização para identificar perfis municipais e SHAP para ranquear variáveis explicativas. Os resultados revelam padrões de desigualdade técnica e destacam municípios com desempenho atípico. A abordagem contribui para decisões regulatórias mais justas e baseadas em evidência.

Palavras-Chave—Inteligência Artificial Explicável, SHAP, Clusterização, Qualidade em Telecomunicações, Análise Espacial

Abstract—This paper proposes an explainable approach to spatial analysis of telecommunications service quality in Brazil. Using data from Anatel's RQUAL system and socioeconomic indicators from IBGE, clustering techniques and SHAP were applied to identify municipal profiles and explanatory variables. Results reveal inequality patterns and highlight atypical municipalities. This supports evidence-based regulatory decision-making.

Keywords—Explainable Artificial Intelligence, SHAP, Clustering, Telecommunications Quality, Spatial Analysis

I. INTRODUÇÃO

A qualidade dos serviços de telecomunicações é essencial para o desenvolvimento socioeconômico [4], [5], especialmente em países com grande extensão territorial e desigualdades regionais, como o Brasil. A Agência Nacional de Telecomunicações (ANATEL), por meio da plataforma RQUAL [6], disponibiliza indicadores técnicos mensais sobre o desempenho desses serviços em nível municipal.

Apesar da relevância desses dados, sua exploração ainda se limita, em geral, a análises descritivas, com pouca integração a variáveis estruturais (como as do IBGE) e carência de mecanismos explicáveis para identificação de padrões e desigualdades [7], [8].

Neste trabalho, propomos uma abordagem explicável baseada em inteligência artificial para analisar espacialmente a qualidade dos serviços, integrando dados técnicos do RQUAL

Marco Antonio de Almeida Fidos Jr, Renato Machado, Dimas Irion Alves, Bartolomeu Ferreira Uchôa Filho e Lucio Pinheiro Amaro estão com o Centro de Competência em Guerra Eletrônica (CCGE), Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos, SP. E-mails: marco.fidos@ga.ita.br, {rmachado, bartolomeu, dimasirion, lucio}@ita.br. Este trabalho foi financiado, em parte, pelo Instituto Nacional de Ciência e Tecnologia (INCT-Sinais), patrocinado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sob o número de processo 406517/2022-3, pelo Ministério da Ciência, Tecnologia e Inovação (MCTI), pela Agência Nacional de Telecomunicações (ANATEL) e pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), sob o número de processo 20/09838-0 – (BIOS -Brazilian Institute of Data Science).

com indicadores estruturais do IBGE. Utilizamos técnicas de clusterização (KMeans e HDBSCAN) para identificar grupos municipais com características similares, e aplicamos métodos interpretáveis como SHAP para ranquear os fatores que explicam essas formações. O estudo utiliza dados públicos de 2022 com granularidade municipal e busca preencher duas lacunas: (i) a ausência de modelos interpretáveis para explicar variações entre municípios; e (ii) o uso limitado de métodos capazes de evidenciar desigualdades espaciais relevantes para a regulação.

A principal contribuição do trabalho é propor um pipeline interpretável e auditável que preenche uma lacuna regulatória importante: a ausência de mecanismos capazes de transformar grandes volumes de dados técnicos (como os da ANATEL) em evidências estruturadas sobre desigualdades territoriais na qualidade dos serviços [9]. Do ponto de vista científico, o estudo demonstra como a integração entre clusterização espacial e técnicas de interpretabilidade pode gerar diagnósticos baseados em dados, úteis tanto para a formulação de políticas públicas quanto para a priorização de investimentos.

Neste artigo, utilizamos o termo *interpretável* para nos referirmos a técnicas que permitem compreender, de forma transparente e rastreável, a contribuição de variáveis para a formação de agrupamentos territoriais, com destaque para o uso do método SHAP.

O restante do artigo está organizado da seguinte forma. A Seção II descreve as bases de dados utilizadas no artigo. A Seção III apresenta a metodologia proposta. A Seção IV traz os experimentos realizados bem como os resultados obtidos. A Seção V discute os resultados. Finalmente, a Seção VI conclui o artigo, trazendo alguns aspectos de limitação e possíveis desdobramentos do estudo realizado.

II. BASES DE DADOS E PRÉ-PROCESSAMENTO

O estudo utilizou dados públicos em nível municipal. A base técnica foi o sistema RQUAL [6], da ANATEL, com oito indicadores de qualidade: IND2 (queda de chamadas), IND4 (velocidade contratada), IND5 (latência), IND8 (disponibilidade), IND9 (cumprimento de prazos), INF1 (reparo), INF4-DL e INF4-UP (velocidades médias). Variáveis socioeconômicas e geográficas foram obtidas do IBGE [13], incluindo região geográfica, densidade demográfica (2022), PIB per capita (2021), grau de urbanização (2022) e IDHM (2010). Variáveis identificadoras foram removidas por inconsistência, mantendo-se o código municipal para rastreabilidade. Variáveis com alta taxa de ausência (como INF4-DL e INF4-UP) foram

imputadas por regressão ou KNN, conforme menor MAE. Para ausências moderadas, foi utilizada a mediana global ou KNN. Variáveis demográficas (população urbana/rural) foram reconstruídas a partir de proporções conhecidas. Todos os valores imputados foram truncados por faixas válidas, e a padronização foi feita com *RobustScaler*, escolhido por apresentar menor assimetria média. A Tabela I resume as técnicas de imputação utilizadas. Apenas 0,02% das linhas completas foram modificadas, validando a abordagem como conservadora e coerente.

TABELA I
RESUMO DAS ESTRATÉGIAS DE IMPUTAÇÃO.

Variáveis	Técnica	Limite
INF4-DL/UP, IND9	Regressão ou KNN	Mínimo real
IND2, IND4, INF1	Mediana ou KNN	Mínimo real
pop_urb., pop_rural	Relação proporcional	≥ 0
TxUrb, IDHM	Derivada ou KNN	[0, 1]

O Fator de Inflação da Variância (VIF) foi calculado para avaliar multicolinearidade. Variáveis técnicas apresentaram $VIF < 5$ e foram mantidas. Entre as socioeconômicas, houve colinearidade severa ($VIF > 10^6$), especialmente em atributos populacionais derivados, que foram removidos para preservar a interpretabilidade sem uso de PCA.

III. METODOLOGIA

A metodologia deste trabalho foi concebida para realizar análises quantitativas, explicativas e espaciais [10] com foco na identificação de agrupamentos territoriais e na interpretação dos fatores estruturais associados à qualidade dos serviços de telecomunicações. O pipeline proposto combina técnicas de clusterização multivariada com métodos explicáveis de interpretação de modelos, aplicando-os a dados regulatórios da ANATEL (RQUAL) e variáveis socioeconômicas do IBGE, com granularidade municipal. A Tabela II resume os principais aspectos metodológicos considerados nesta etapa do estudo, incluindo o foco analítico, as técnicas utilizadas e os produtos esperados.

TABELA II
RESUMO METODOLÓGICO.

#	Foco Principal	Técnicas	Produto Esperado
1	Identificar padrões e perfis municipais	KMeans, HDBSCAN, PCA, UMAP	Clusters interpretáveis com padrões comuns [11]
2	Comparar clusters e regiões	Estatísticas descritivas	Comparativo de clusters, estados e grandes regiões
3	Identificar fatores explicativos dos agrupamentos	Decision Tree, Random Forest, XGBoost, Logistic Regression, SHAP	Ranking consensual de variáveis relevantes por subcluster

A. Hipóteses Explicáveis e Organização Temática

Cada frente metodológica foi guiada por hipóteses explicáveis, formuladas a partir de observações empíricas e conhecimento técnico-regulatório, para, não apenas a descoberta de padrões multivariados nos dados, mas também a

interpretação dos fatores determinantes para tais padrões. A Tabela III sistematiza os três eixos analíticos que orientaram a investigação:

TABELA III
TRÊS TEMAS ANALÍTICOS QUE ORGANIZAM AS HIPÓTESES EXPLICÁVEIS.

#	Tema	Descrição
T1	Perfis Territoriais	Agrupar municípios com padrões similares de qualidade técnica e estrutura socioeconômica, identificando também outliers e ilhas [12]
T2	Disparidades Regionais	Analisar a relação entre os agrupamentos encontrados e divisões político-administrativas, evidenciando possíveis desigualdades territoriais
T3	Fatores Explicativos	Ranquear e interpretar variáveis técnicas e estruturais que explicam os agrupamentos, por meio de modelos explicáveis

Essa estrutura metodológica busca garantir transparência, reprodutibilidade e valor regulatório à análise, oferecendo um pipeline auditável e aplicável a outros contextos marcados por desigualdade digital.

IV. EXPERIMENTOS REALIZADOS

Os experimentos foram realizados sobre as bases de dados apresentadas na Seção II-B, com granularidade municipal, considerando prioritariamente o ano de 2022. Este ano foi selecionado como referência principal por ser o único em que todos os oito indicadores técnicos analisados (IND2, IND4, IND5, IND8, IND9, INF1, INF4-DL e INF4-UP) apresentaram ampla cobertura com valores distintos de zero em mais de 93% dos municípios (para seis dos oito indicadores). Nos anos posteriores (2023 a 2025), observou-se queda significativa na disponibilidade de dados para INF4-DL e INF4-UP, além da redução na cobertura dos indicadores INF1 e IND9. Tais limitações reforçam a escolha por utilizar 2022 como base principal da análise, assegurando consistência na entrada de dados. Com o objetivo de sintetizar os indicadores técnicos em uma métrica unificada por município, foram testadas quatro abordagens para a construção do Índice de Qualidade dos Serviços de Telecomunicações (IQST): média simples, média ponderada pela cobertura, z-score médio e análise de componentes principais (PCA). A média ponderada foi selecionada por apresentar distribuição mais centrada (mediana ≈ 61), menor assimetria e maior robustez frente a dados ausentes. As abordagens testadas apresentaram correlações moderadas entre si, sugerindo complementaridade. A média ponderada final variou entre 15,8 e 115,9.

A. Padrões e Perfis Municipais

Adotou-se uma estratégia de clusterização em duas etapas para identificar agrupamentos territoriais com base em indicadores técnicos e socioeconômicos. Inicialmente, o algoritmo KMeans foi utilizado para agrupar os municípios brasileiros em regiões homogêneas, com o objetivo de capturar padrões globais de similaridade. A Tabela IV compara métricas para diferentes valores de k , justificando a escolha de $k = 9$:

Embora $k = 2$ tenha apresentado bons indicadores de separação, optou-se por $k = 9$ por oferecer o melhor equilíbrio entre coesão intracluster, granularidade e interpretabilidade. O

TABELA IV
AVALIAÇÃO DO KMEANS.

k	(1)	(2)	(3)	(4)
2	5070.68	0.873	5375.13	0.507
8	583.09	0.558	12775.31	0.492
9	477.03	0.571	13815.75	0.479
10	417.25	0.550	14126.26	0.497

(1) Inércia em $\times 10^9$, (2) Silhouette Score, (3) Calinski-Harabasz, (4) Davies-Bouldin.

KMeans ($k = 9$) segmentou os 5.570 municípios com base em 14 variáveis técnicas e estruturais normalizadas. Apesar da boa separação entre os clusters principais, uma análise de heterogeneidade intracluster (via desvio padrão médio, amplitude e variância dos 14 indicadores) indicou que os Clusters 1, 2, 6 e 8 apresentavam alta dispersão interna, sugerindo subestruturas relevantes. Para refinar esses agrupamentos, foi aplicado o algoritmo HDBSCAN, que permite identificar subclusters com formatos arbitrários e detectar outliers, sem a necessidade de pré-definir o número de grupos. O HDBSCAN foi calibrado com `min_cluster_size = 5` e `min_samples = 3`, mantendo a mesma base de 14 variáveis normalizadas. Como resultado, foram identificados **6 subclusters válidos** e **4 agrupamentos de outliers**, detalhados na Tabela V. Os rótulos foram preservados (e.g., `1_0`, `1_outlier`) para assegurar rastreabilidade nas etapas posteriores.

TABELA V
SUBCLUSTERS VIA HDBSCAN.

Cluster 1	Cluster 2	Cluster 8
1_0 (62)	2_0 (8)	8_0 (11)
1_1 (39)	2_1 (5)	8_1 (12)
1_outlier (8)	2_outlier (1)	8_outlier (4)
Outlier adicional: 6_outlier (50)		

Nota: Subclusters originados de três agrupamentos principais do KMeans (1, 2, 8). O grupo `6_outlier` agrega municípios dispersos originalmente alocados no cluster 6.

Esta etapa estabelece a base analítica para os módulos seguintes, permitindo a comparação territorial (Módulo 2) e a explicação dos agrupamentos via SHAP (Módulo 3).

B. Coerência Territorial e Diversidade Regional

Esta etapa buscou caracterizar a distribuição geográfica dos subclusters obtidos no Módulo 1 em relação a unidades político-administrativas (Unidades Federativas e Grandes Regiões). Para isso, foram aplicadas estatísticas descritivas — médias, desvios padrão e rankings — sobre dois conjuntos de variáveis: (i) indicadores técnicos de qualidade dos serviços; e (ii) variáveis socioeconômicas estruturais. As operações foram realizadas com funções `groupby()`, preservando rastreabilidade entre os municípios e os agrupamentos. Além disso, foi analisada a distribuição percentual dos subclusters por UF e por região. Essa análise cruzada permitiu observar tanto a predominância de certos perfis de municípios dentro de cada região, quanto a diversidade regional presente em cada subcluster. Os dados tabulados revelam que algumas regiões, como o Nordeste, concentram uma alta proporção de

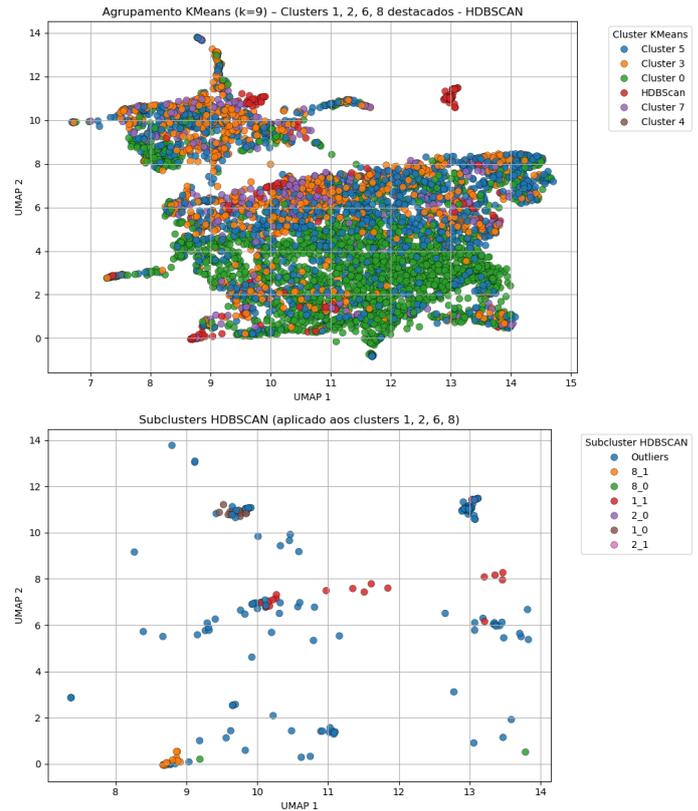


Fig. 1. Acima, Visualização UMAP dos clusters KMeans ($k = 9$), com destaque para os Grupos 1, 2, 6 e 8. Abaixo, Subclusterização via HDBSCAN aplicada sobre os Clusters 1, 2, 6 e 8, revelando estruturas internas e outliers.

municípios em um único subcluster (e.g., 0), sugerindo homogeneidade interna. Em contrapartida, regiões como Centro-Oeste e Sul apresentam maior dispersão entre subclusters, refletindo uma diversidade estrutural e técnica maior entre os municípios[14]. A Tabela VI apresenta a distribuição percentual dos principais subclusters por região, juntamente com as Unidades Federativas com maior representação interna em cada grupo.

Esta etapa contribui para compreender a coerência e a diversidade regional dos agrupamentos identificados, estabelecendo a base para a análise explicável dos fatores estruturais no Módulo 3.

C. Ranking Explicável de Variáveis

Este módulo teve como objetivo identificar os fatores técnicos e estruturais que melhor explicam a formação dos subclusters identificados nos módulos anteriores. Para isso, foram utilizados classificadores supervisionados combinados com técnicas de interpretabilidade, permitindo ranquear variáveis segundo sua importância preditiva. O rótulo-alvo foi o `subcluster_hdbscan`, e todas as variáveis numéricas disponíveis — técnicas, socioeconômicas e demográficas — foram avaliadas. Foram aplicados quatro modelos com diferentes pressupostos e níveis de interpretabilidade: `DecisionTreeClassifier` (regras diretas), `RandomForestClassifier` (média de ganhos de impureza), `XGBoostClassifier` (interações

TABELA VI
DISTRIBUIÇÃO DOS SUBCLUSTERS POR REGIÃO E UFS MAIS REPRESENTATIVAS.

SC	CO	NE	N	SE	S
0	8.4%	85.5%	50.9%	32.4%	3.3%
	GO (19.1)	PB (95.1)	AM (66.1)	MG (55.1)	PR (4.5)
	MT (10.6)	CE (90.8)	PA (72.2)	SP (19.4)	RS (2.8)
3	19.0%	1.8%	6.3%	13.2%	41.0%
	MT (21.3)	AL (2.9)	AM (1.6)	SP (23.4)	SC (47.5)
	GO (20.7)	BA (1.4)	RO (13.5)	RS (15.2)	RS (40.6)
5	27.3%	10.6%	33.2%	43.2%	39.0%
	GO (47.2)	AL (29.4)	RO (76.9)	SP (45.6)	PR (50.1)
	MT (29.1)	BA (8.9)	PA (17.4)	ES (48.7)	RS (31.4)
	MS (32.9)	RN (14.4)	RR (33.3)	RJ (50.0)	SC (35.6)

Nota: SC = Subcluster; valores indicam proporção média de municípios por região e as UFs com maiores participações internas por subcluster (entre parênteses).

complexas) e `LogisticRegression` (coeficientes normalizados). A diversidade de modelos visou garantir robustez metodológica e complementaridade interpretativa. Todos os classificadores foram treinados com divisão estratificada de 70/30 (`train_test_split`) e avaliados por métricas como `F1-score`, `accuracy`, `precision` e `confusion matrix`. Apenas modelos com desempenho estável ($F1 \geq 0,65$) foram considerados para análise explicável. Para os modelos baseados em árvores, aplicou-se SHAP (SHapley Additive Explanations), permitindo gerar explicações tanto globais quanto locais sobre a contribuição marginal de cada variável para a classificação dos subclusters. No caso da regressão logística, SHAP foi utilizado de forma ilustrativa, sendo os coeficientes padronizados a principal fonte interpretativa. A Tabela VII resume as estratégias explicativas adotadas.

TABELA VII
MODELOS SUPERVISIONADOS E ESTRATÉGIAS EXPLICATIVAS.

Modelo	Fonte	Interp.	SHAP	Complemento
DT	SP	Regra if/then	N	Caminho explícito
RF	GM	Global (med.)	S	Explicação global/local
XGB	GV	Não linear	S	Interações + robustez
LR	CN	Efeitos marginais	P	Visual coeficiente

Siglas: DT = Decision Tree, RF = Random Forest, XGB = XGBoost, LR = Logistic Regression. SP = Splits, GM = Ganhos Médios, GV = Ganho por Variável, CN = Coef. Normalizados. Interp. = Interpretação, Marg. = Marginais, Med. = Médias, P = Parcial, S = Sim, N = Não.

Para consolidar os resultados, adotou-se uma estratégia de decisão por consenso. Foram consideradas variáveis explicativas consensuais aquelas que apareceram entre as cinco mais relevantes em pelo menos três dos quatro modelos. Em seguida, essas variáveis foram organizadas por subcluster, compondo uma base interpretativa robusta para os agrupamentos. Essa abordagem multitécnica, explicável e auditável permitiu não apenas compreender os determinantes estruturais dos subclusters, mas também oferecer subsídios concretos para diagnósticos territoriais e formulação de políticas públicas mais precisas e baseadas em evidência.

V. DISCUSSÃO DOS RESULTADOS

A. Padrões e Perfis Municipais

A análise dos nove clusters obtidos via KMeans revelou diferenças estruturais relevantes entre os grupos. A Tabela VIII resume os rankings médios dos clusters em indicadores técnicos e socioeconômicos. O Cluster 2 obteve os melhores escores em ambas as dimensões, enquanto os Clusters 0 e 8 apresentaram os piores desempenhos combinados. A aplicação do HDBSCAN aprofundou a caracterização, identificando subclusters com padrões distintos dentro dos clusters mais heterogêneos. Subclusters como 2_0, 2_1 e 1_0 destacaram-se por altos desempenhos técnico e estrutural, enquanto 8_0, 8_1 e 8_outlier concentraram os piores indicadores. Grupos intermediários, como 1_outlier e 2_outlier, apresentaram boas condições socioeconômicas, mas qualidade técnica abaixo da média, sugerindo gargalos de infraestrutura em territórios com maior potencial de conectividade.

Ranking Técnico e Socioeconômico dos Subclusters							
Subcluster	Qtd	IQST_rank	IDH_rank	PIB_rank	Urb_rank	Dem_rank	Area_rank
0	2_0	4	1	4	4	4	14
1	1_0	12	2	2	7	2	9
2	4	2	3	10	1	3	6
3	2_outlier	7	4	11	2	14	8
4	2_1	3	5	1	3	1	15
5	7	353	6	6	9	8	7
6	3	912	7	5	10	7	12
7	6_outlier	50	8	7	5	6	5
8	5	1562	9	8	11	9	11
9	1_1	17	10	3	8	5	10
10	1_outlier	80	11	9	6	10	4
11	0	2541	12	12	14	12	10
12	8_1	19	13	13	13	11	13
13	8_0	3	14	15	15	13	15
14	8_outlier	5	15	14	12	15	14

Fig. 2. Ranking técnico e socioeconômico médio por subcluster.

B. Comparações Territoriais entre Subclusters

A distribuição dos subclusters por região e por UF indicou padrões de coerência e diversidade territorial. O Subcluster 0, de baixo desempenho, concentra-se fortemente no Nordeste (85,5%), enquanto o Sul apresenta maior heterogeneidade, com predominância de Subclusters 3 e 5. As “ilhas percentuais” — subclusters com presença inferior a 5% em determinadas regiões ou UFs — revelam municípios com perfil técnico atípico dentro de seu contexto geográfico. Essas exceções, como a presença do Subcluster 3 no Nordeste ou do Subcluster 0 no Sul, representam potenciais alvos de políticas públicas diferenciadas [15].

TABELA VIII
DISTRIBUIÇÃO PERCENTUAL DE SUBCLUSTERS POR REGIÃO (VALORES * <5%).

Região	SC_0	SC_3	SC_5	SC_7
Centro-Oeste	13.9%	23.1%	39.2%	14.1%
Nordeste	85.5%	1.8%*	10.6%	1.1%*
Norte	50.9%	6.3%	33.2%	2.4%*
Sudeste	32.4%	13.2%	43.2%	5.6%
Sul	3.3%*	41.0%	39.0%	13.7%

Nota: Valores com * indicam “ilhas percentuais” (presença inferior a 5%).

VI. CONCLUSÕES

Este trabalho apresentou uma abordagem explicável de IA aplicada à análise territorial da qualidade dos serviços de telecomunicações no Brasil. A partir da integração de dados técnicos da ANATEL (RQUAL) com indicadores estruturais do IBGE, foi possível identificar agrupamentos municipais com padrões distintos de desempenho técnico e contexto socioeconômico, por meio de técnicas de clusterização (KMeans e HDBSCAN). A análise da coerência regional dos subclusters revelou tanto a existência de regiões homogêneas em termos de qualidade quanto a presença de “ilhas percentuais” — municípios com perfis divergentes dos padrões regionais, que podem demandar atenção regulatória específica. A etapa explicável, baseada em classificadores supervisionados e SHAP, permitiu ranquear variáveis técnicas e estruturais segundo sua contribuição para a formação dos subclusters, promovendo maior transparência e auditabilidade nos achados. Os resultados reforçam o potencial da análise espacial explicável como ferramenta para o direcionamento de políticas públicas, priorização de investimentos e tomada de decisão regulatória orientada por evidências, sensível às desigualdades regionais e à diversidade territorial brasileira.

A. Limitações

Apesar dos resultados promissores, este estudo possui algumas limitações que devem ser consideradas. Primeiramente, a análise concentrou-se em dados agregados de 2022, não explorando ainda o potencial de séries temporais para identificar tendências ou variações sazonais na qualidade dos serviços.

Além disso, a abordagem proposta não contempla, nesta etapa, mecanismos formais de auditoria de justiça algorítmica ou inferência causal, o que limita a capacidade de estabelecer relações de causa e efeito entre variáveis estruturais e qualidade técnica.

Por fim, algumas variáveis utilizadas apresentaram lacunas significativas de cobertura, como as velocidades INF4-DL e INF4-UP, exigindo técnicas de imputação que, embora controladas, podem introduzir viés residual. Ainda assim, os critérios de validação adotados indicaram preservação da coerência semântica dos dados.

B. Trabalhos futuros

Como desdobramento deste estudo, propõem-se várias direções para aprofundamento e ampliação da abordagem. A primeira envolve a incorporação de técnicas de auditoria de justiça algorítmica, como Equalized Odds e Disparate Impact, permitindo avaliar se os agrupamentos gerados reproduzem ou mitigam desigualdades entre grupos sociais ou territoriais.

Outra frente relevante consiste na aplicação de métodos de inferência causal, como DoubleML ou Causal Forest, para estimar os efeitos estruturais de variáveis socioeconômicas sobre a qualidade técnica dos serviços, superando limitações correlacionais.

Pretende-se ainda incorporar mecanismos de validação de robustez e sensibilidade, como perturbações sistemáticas nos dados, validação cruzada estratificada e simulações de cenários, de modo a aumentar a confiabilidade dos achados.

Por fim, destaca-se o potencial de aplicar essa abordagem a dados mais recentes, mesmo com cobertura parcial, explorando técnicas semi-supervisionadas e modelagem temporal. Além disso, sua adaptação para outros setores regulados, como saúde, educação ou saneamento, poderá ampliar o impacto social e institucional da análise espacial explicável.

AGRADECIMENTOS

Este trabalho foi financiado, em parte, pelo Instituto Nacional de Ciência e Tecnologia (INCT-Sinais), patrocinado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sob o número de processo 406517/2022-3, pelo Ministério da Ciência, Tecnologia e Inovação (MCTI), pela Agência Nacional de Telecomunicações (ANATEL) e pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), sob o número de processo 20/09838-0 – (BIOS -Brazilian Institute of Data Science).

REFERÊNCIAS

- [1] MacQueen, J. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14, 1967, pp. 281–297.
- [2] Campello, R. J. G. B., Moulavi, D., and Sander, J. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2013, pp. 160–172. Springer.
- [3] Lundberg, S. M. and Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 4765–4774.
- [4] Castells, M. *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford University Press, 2002.
- [5] Meena, M. E.; Geng, J. Dynamic competition in telecommunications: A systematic literature review. *SAGE Open*, vol. 12, no. 2, p. 1-10, 2022. Disponível em: DOI: 10.1177/21582440221094609. Acesso em: 20 abr. 2025.
- [6] Anatel. *Sistema de Acompanhamento da Qualidade - RQUAL*. Brasília, 2023. Disponível em: <https://www.gov.br/anatel/pt-br/dados/qualidade>. Acesso em: 20 abr. 2025.
- [7] Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Lulu.com, 2020. <https://christophm.github.io/interpretable-ml-book/>.
- [8] Ribeiro, M. T., Singh, S., and Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.
- [9] Anatel. *Resolução nº 717, de 23 de dezembro de 2019*. Diário Oficial da União, Brasília, 2019. Disponível em: <https://informacoes.anatel.gov.br/legislacao/resolucoes/2019/1371-resolucao-717>. Acesso em: 20 abr. 2025.
- [10] Cliff, A. D.; Ord, J. K. *Spatial Autocorrelation*. Pion, London, 1973.
- [11] McInnes, L., Healy, J., and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [12] Anselin, L. Local indicators of spatial association—LISA. *Geographical analysis*, vol. 27, no. 2, 1995, pp. 93–115.
- [13] IBGE – Instituto Brasileiro de Geografia e Estatística. *Censo Demográfico 2022: Resultados do universo por município*. Rio de Janeiro: IBGE, 2023. <https://censo2022.ibge.gov.br/resultados.html>. Acesso em: 20 abr. 2025.
- [14] Barredo, J. I., Benitez, M., Pesaresi, M., and Bosque, J. Spatial analysis of telecommunication and socio-economic disparities in European regions. *Telecommunications Policy*, vol. 33, no. 9, 2009, pp. 493–507.
- [15] Oughton, E. J.; Comini, N.; Foster, V.; Hall, J. W. Policy choices can help keep 4G and 5G universal broadband affordable. *Telecommunications Policy*, vol. 46, no. 3, p. 102291, 2022.