

Framework de IA Explicável com Indicadores de QoS de Telecomunicações e Socioeconômicos para Classificação e Caracterização de Municípios

Marco Antonio de Almeida Fidos Jr, Felipe Streitenberger Ivo, Paulo Ricardo Branco da Silva, Sarah Negreiros C. Leite, Lucio Pinheiro Amaro, Kai de Andrade Lima e Renato Machado

Resumo— Este artigo propõe um framework de inteligência artificial (IA) explicável estruturado em quatro camadas para a análise de indicadores de Qualidade de Serviço (QoS) de telecomunicações no Brasil. O pipeline metodológico integra: i) segmentação global via K-Means, com definição de $k=5$ por equilibrar separação topológica e granularidade regional; ii) redução de dimensionalidade e agrupamento local via UMAP e HDBSCAN; iii) detecção de anomalias isoladas intracuster com LOF; e iv) explicabilidade baseada em valores SHAP sobre um modelo substituto Random Forest. A aplicação do modelo em 5.570 municípios brasileiros, utilizando dados de QoS dos indicadores regulatórios da ANATEL e socioeconômicos do IBGE, extraiu cinco perfis de desempenho, detectou 204 municípios em subgrupos coesos e 559 anomalias isoladas, com sobreposição nula. A principal contribuição técnica reside na separação explícita de perfis globais, subestruturas locais e anomalias, revelando uma modelagem mais robusta do que aquelas baseadas em clusterização convencionais, complementada pela rastreabilidade individual via SHAP — que identifica, para cada município, quais indicadores mais determinaram sua classificação, fornecendo métricas mais precisas para o monitoramento de QoS e orientando ações regulatórias diferenciadas por perfil territorial.

Palavras-Chave— Inteligência Artificial Explicável; SHAP; K-Means; HDBSCAN; LOF; Qualidade de Serviço (QoS); Análise Espacial; Regulação; ANATEL

Abstract— This paper proposes an explainable artificial intelligence (AI) framework structured in four layers for the analysis of Quality of Service (QoS) indicators in Brazilian telecommunications. The methodological pipeline integrates: i) global segmentation via K-Means, with $k = 5$ defined to balance topological separation and regional granularity; ii) dimensionality reduction and local clustering via UMAP and HDBSCAN; iii) intracuster isolated anomaly detection with LOF; and iv) explainability based on SHAP values over a Random Forest surrogate model. Applied to 5,570 Brazilian municipalities using QoS data from ANATEL's regulatory indicators and socioeconomic data from IBGE, the model extracted five performance profiles, detected 204 municipalities in cohesive subgroups and 559 isolated anomalies, with zero overlap. The main technical contribution lies in the explicit separation of global profiles, local substructures, and anomalies, yielding a more robust modeling approach than conventional clustering methods, complemented by individual traceability via SHAP — which identifies, for each municipality, which

indicators most determined its classification, providing more precise metrics for QoS monitoring and guiding differentiated regulatory actions by territorial profile.

Keywords— Explainable Artificial Intelligence; SHAP; K-Means; HDBSCAN; LOF; Quality of Service (QoS); Spatial Analysis; Regulation; ANATEL

I. INTRODUÇÃO

Sistemas de monitoramento de Qualidade de Serviço (QoS) em redes de telecomunicações geram volumes massivos de dados, cuja análise é ...fundamental para ações em relação a otimização de infraestrutura e fiscalização de operadoras [1], [2]. No Brasil, o Regulamento de Qualidade dos Serviços de Telecomunicações (RQUAL) da Agência Nacional de Telecomunicações (ANATEL) [3] estabelece indicadores de QoS a nível municipal, conforme resultados publicados no Painel de dados da Agência. Contudo, a exploração desses dados frequentemente se limita a análises estatísticas descritivas. Assim, existe a lacuna de soluções computacionais capazes de processar alto volume de dados, integrando parâmetros de rede a variáveis espaciais por meio de mecanismos de Inteligência Artificial Explicável (XAI) [4], [5].

Em [6], propôs-se um *pipeline* interpretável que integrou dados do RQUAL com variáveis socioeconômicas e estruturais do Instituto Brasileiro de Geografia e Estatística (IBGE), utilizando os métodos k-Means e HDBSCAN para segmentação e o método SHAP para ranqueamento explicativo. Os resultados obtidos em [6] demonstraram a viabilidade da abordagem, mas apresentou três limitações algorítmicas, a saber: i) não detecção de anomalias *isoladas*; ii) SHAP calculado sobre classificadores independentes sem validação baseada em modelo substituto unificado; e iii) impossibilidade de rastrear a explicação preditiva para um município específico.

Este artigo propõe um framework de IA explicável estruturado em quatro camadas para a análise de indicadores de QoS de telecomunicações móveis no Brasil, estendendo a abordagem apresentada em [6]. Como resultado de pesquisa, revela-se que a separação explícita de perfis globais, subestruturas locais e anomalias, permite uma modelagem muito mais precisa do que abordagens baseadas apenas em clusterização global, fornecendo métricas mais precisas para o monitoramento de QoS e suporte à gestão regulatória.

Marco A. A. Fidos Jr, Felipe S. Ivo, Paulo R. B. da Silva, Sarah N. C. Leite, Lucio P. Amaro e Renato Machado estão com o Instituto Tecnológico de Aeronáutica (ITA), Centro de Competência de Guerra Eletrônica (CCGE), São José dos Campos-SP, e-mails: marco.fidos@ga.ita.br, {fivo, pbranco, sarahnc, lucio, rmachado}@ita.br. Kai de Andrade Lima está com a Agência Nacional de Telecomunicações (ANATEL), e-mail: kai.lima@anatel.gov.br.

O restante deste artigo está organizado da seguinte forma. A Seção II detalha a base de dados e o pré-processamento. A Seção III apresenta o *framework* metodológico e as formulações em quatro camadas. A Seção IV discute os resultados de segmentação, anomalias e explicabilidade. Por fim, a Seção V apresenta as conclusões do estudo.

II. DADOS E PRÉ-PROCESSAMENTO

A base de dados primária provém do sistema RQUAL [3] da ANATEL, totalizando 5,96 milhões de registros operacionais (2022–2025). Adotou-se 2022 como ano-base por apresentar cobertura de dados superior a 93% dos municípios — a maior completude entre os anos disponíveis, o que minimiza a necessidade de imputação e assegura maior confiabilidade da segmentação. Variáveis socioeconômicas e estruturais foram extraídas do IBGE [8], compreendendo: PIB municipal (2021), Censo Demográfico (2022), taxa de urbanização (2022), IDHM (2010) e geometria espacial. O IDHM segue divulgação decenal; apesar do Censo 2022, sua base municipal atualizada ainda não constava nas fontes oficiais até o fechamento deste artigo, permanecendo 2010 a edição mais recente disponível. A integração dos conjuntos de dados foi realizada de forma consistente utilizando o código municipal (`cod_mun`) de 7 dígitos como chave primária.

Pela sua disponibilidade, foram utilizados cinco indicadores de QoS, do serviço móvel: Queda de chamadas (IND2), Conexão de dados móveis (IND3), Cumprimento de velocidade (IND4), Latência bidirecional (IND5), Disponibilidade (IND8), e dois informativos, definidos pelo RQUAL como secundários aos INDs: cobertura (INF1) e *throughput* de *upload* (INF4-UP). Para a definição do vetor de *features* ótimo, 78 variáveis brutas, x^j , (dados RQUAL + dados IBGE) foram submetidas a um *pipeline* de redução de dimensionalidade estruturado em seis etapas: i) higienização por dominância de valor único (remoção de variáveis com $\geq 99\%$ das observações concentradas em um mesmo valor), critério invariante à escala e distinto de um filtro de variância bruta; ii) imputação espacial baseada na mediana estadual (UF), preservando a heterogeneidade regional do Brasil na correção de dados faltantes dos indicadores RQUAL e do IBGE; iii) Winsorização (1%–99%), mitigando o efeito de valores extremos nos indicadores socioeconômicos e de QoS com assimetria estatística elevada ($|\text{skew}| > 2$); iv) padronização via *z-score* foi aplicada a cada um dos 20 indicadores do vetor de *features*, subtraindo a média e dividindo pelo desvio padrão de cada variável, garantindo que indicadores de escalas distintas — como PIB municipal e taxa de urbanização — contribuam de forma equilibrada para a segmentação dos municípios; v) eliminação de colinearidade por correlação de Pearson ($|\rho| \geq 0,80$), assegurando a retenção prioritária dos parâmetros RQUAL em detrimento de variáveis socioeconômicas redundantes; e vi) O VIF de cada variável foi calculado como o inverso de um menos o coeficiente de determinação (R^2) obtido ao regredi-la sobre as demais, permitindo identificar e remover redundâncias entre os indicadores RQUAL e socioeconômicos do IBGE, com limiar rigoroso de $\text{VIF} \leq 5$.

O processamento convergiu para um espaço de características (*features*) reduzido a 20 *features* por município, distribuídas em quatro grupos: i) sete indicadores de QoS (RQUAL); ii) três variáveis econômicas (PIB agropecuário, industrial e *per capita*); iii) seis variáveis demográficas e espaciais (área, densidade, população rural, IDHM, latitude e longitude); e iv) quatro variáveis indicadoras (*dummies*) regionais.

III. METODOLOGIA

Este artigo propõe um *framework* multicamada (4 camadas) para tratamento dos dados de QoS de telecomunicações (ANATEL) e socioeconômicos e estruturais (IBGE). Cada camada captura uma forma distinta de estrutura nos dados: perfis globais (Camada 1), subestruturas locais coesas (Camada 2) e anomalias isoladas (Camadas 3 e 4), conforme ilustrado na Fig. 1.

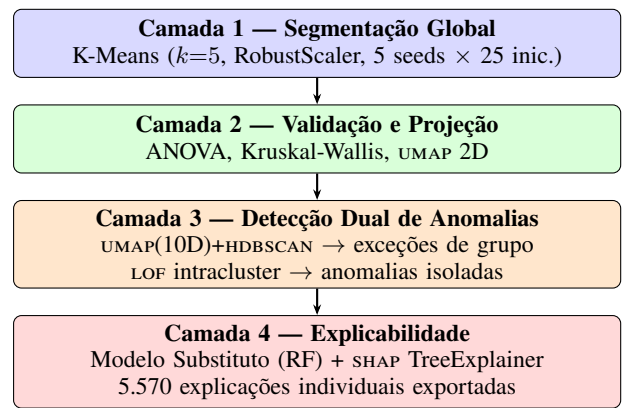


Fig. 1. Framework em quatro camadas. Cada camada captura uma forma distinta de estrutura nos dados: perfis globais, subestruturas locais coesas e anomalias isoladas.

O Algoritmo 1 traz a implementação do método proposto, descrevendo o *pipeline* completo com todos os hiperparâmetros utilizados, garantindo reprodutibilidade integral dos resultados. O código usado para gerar os resultados deste artigo pode ser acessado na íntegra em [7]. Na Camada 1: Segmentação Global via K-Means utilizou o algoritmo K-Means precedido por RobustScaler (mitigando o impacto de 2,47% de outliers). Avaliou-se $k = 2$ a $k = 12$ com cinco sementes aleatórias e 25 inicializações por rodada, definindo o vetor ótimo via rank ponderado (Calinski-Harabász $\times 2$, Davies-Bouldin $\times 1,5$, Silhueta $\times 1,2$, Inércia $\times 1$). Os pesos foram calibrados empiricamente; a seleção de $k = 5$ manteve-se estável sob pesos iguais e sob inversão de prioridade entre os índices, indicando baixa sensibilidade à ponderação adotada. O Silhouette Score de cada município mede a qualidade do seu agrupamento comparando a distância média aos demais municípios do seu cluster (a_i) com a distância média ao cluster vizinho mais próximo (b_i), variando entre -1 (má atribuição) e 1 (atribuição ideal), sendo utilizado como um dos critérios para a seleção de $k=5$. A estabilidade da solução $k = 5$ foi confirmada pela convergência entre média e mediana da silhueta entre as cinco sementes (0,6203 vs.

0,6203, $\Delta < 0,01\%$), indicando que a segmentação é estável e robusta entre inicializações distintas.

Algoritmo 1: Pipeline do Framework Explicável

```

Entrada:  $\mathbf{X} \in \mathbb{R}^{5570 \times 20}$  (indicadores RQUAL + IBGE)
Saída: Rótulos  $C$ , subclusters  $H$ , anomalias  $L$ , valores SHAP  $S$ 

// Camada 1 — Segmentação global
1:  $\mathbf{X}_s \leftarrow \text{RobustScaler}(\mathbf{X})$ 
2:  $C \leftarrow \text{K-Means}(\mathbf{X}_s, k=5, \text{seeds}=\{42, 7, 123, 2025, 99\}, n_{\text{init}}=25)$ 
// Camada 2 — Validação e projeção
3: Validar  $C$  via ANOVA e Kruskal-Wallis ( $p < 0,001$ )
4:  $\mathbf{U}_{2D} \leftarrow \text{UMAP}(\mathbf{X}_s, n_{\text{comp}}=2, n_{\text{neigh}}=15, \text{min\_dist}=0,1, \text{seed}=42)$ 
// Camada 3 — Detecção dual de anomalias
5:  $\mathbf{U}_{10D} \leftarrow \text{UMAP}(\mathbf{X}_s, n_{\text{comp}}=10, n_{\text{neigh}}=15, \text{min\_dist}=0,0, \text{seed}=42)$ 
6:  $H \leftarrow \text{HDBSCAN}(\mathbf{U}_{10D}, \text{min\_size}=30, \text{min\_samp}=10, \text{method}=\text{eom})$ 
7: para  $c \in C$  faça
     $mcs_c \leftarrow \max(10, \lfloor 0,05 \times |c| \rfloor)$ 
     $H_c \leftarrow \text{HDBSCAN}(\mathbf{X}_s[c], mcs_c, \text{min\_samp}=5, \text{eom})$ 
     $L_c \leftarrow \text{LOF}(\mathbf{X}_s[c], n_{\text{neigh}}=20, \text{contam}=0,10)$ 
8: fim
9:  $L \leftarrow \{i : \text{score\_lof}(i) > 1,5\}$ 
// Camada 4 — Explicabilidade via modelo substituto
10:  $\text{RF} \leftarrow \text{RandomForest}(C, n_{\text{est}}=300, \text{depth}=\infty, \text{min\_leaf}=5, \text{seed}=42)$ 
11: Validar RF via StratifiedKFold( $k=5, \text{seed}=42$ )
12:  $S \leftarrow \text{SHAP\_TreeExplainer}(\text{RF}, \mathbf{X}_s)$ 
13: retorne  $C, H, L, S$ 
    
```

Embora $k = 2$ maximize a silhueta (0,831), resulta em macro-grupos genéricos, perdendo as nuances operacionais do Nordeste, Norte e capitais. Portanto, $k = 5$ equilibra separação topológica e granularidade. A solução $k = 5$ produziu cinco perfis globais, i.e., Clusters C10, C11, C12, C13 e C14, resumidos na Tabela I. A segmentação reflete a estrutura regional do Brasil: dos municípios do C10, 67% pertencem à região Sudeste; do C12, 84% ao Nordeste; do C13, 68% ao Norte; e o C14 distribui-se por todas as regiões — composição regional interna, distinta da participação total já mostrada na Tabela I. O radar da Fig. 2 evidencia a separação entre perfis. Todos os indicadores apresentam $p < 0,001$ em ANOVA e Kruskal-Wallis, confirmando que a separação não é artefato algorítmico.

TABELA I. Cinco perfis municipais — K-Means $k = 5$

Cluster	Perfil	n (%)
C10	Urbano-Avançado (Sul/Sudeste)	216 (3,9%)
C11	Intermediário (padrão nacional)	3.054 (54,8%)
C12	Nordeste (vulnerabilidade operacional)	2.057 (36,9%)
C13	Norte/Amazônico extensivo	216 (3,9%)
C14	Capitais/Destaques (benchmark)	27 (0,5%)

Na Camada 3, realizou-se a detecção dual de anomalias através da aplicação do HDBSCAN direto no espaço de 20 dimensões (20D), que apresentou taxa de ruído de 96,3% (maldição da dimensionalidade). A redução prévia via UMAP(10D) ($\text{min_dist}=0,0$ para maior compactação dos clusters) integrada ao HDBSCAN ($\text{min_cluster_size}=30, \text{min_samples}=10, \text{method}=\text{eom}$) reduziu essa taxa para 27,6%, extraindo 204 municípios em subclusters coesos (exceções de grupo).

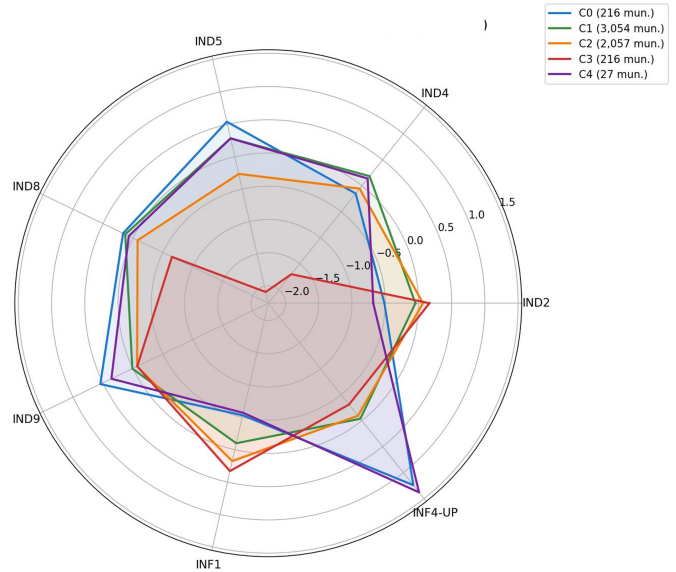


Fig. 2. Radar dos perfis médios por cluster (indicadores RQUAL, z-scores). C13 colapsa em IND4 ($-1,70\sigma$) e IND5 ($-2,09\sigma$); C10 e C14 dominam em INF4-UP e IND8.

Em paralelo, o LOF intracluster foi configurado com $n_{\text{neighbors}}=20$ e $\text{contamination}=0,10$, parâmetro que já fixa internamente $\sim 10\%$ de cada cluster como candidato a anomalia; o $\text{score} > 1,5$ atua como filtro adicional, e o resultado (559 municípios, 10,04%) é consistente por construção com esse parâmetro, não uma validação independente. A estabilidade foi verificada entre score 1,4–1,6; a sensibilidade a outros valores de contamination é direção futura. O procedimento detectou 559 anomalias em todos os cinco clusters, incluindo C11 e C12, onde o HDBSCAN não identificou subestruturas. A sobreposição $\text{HDBSCAN} \cap \text{LOF}$ resultou em zero municípios, conforme ilustrado na Fig. 3, resultado consistente com a natureza complementar — em parte esperada pela própria definição dos resultado consistente com a complementaridade — em parte esperada pela própria definição dos métodos — entre os dois critérios: subclusters HDBSCAN têm score LOF baixo por construção, enquanto outliers LOF são isolados por definição. A divergência nas distribuições de scores LOF foi confirmada pelo teste U de Mann-Whitney ($p < 0,001$). A sobreposição nula e a complementaridade mantiveram-se consistentes para limiares de LOF entre 1,4 e 1,6 e variações de $\pm 20\%$ em min_cluster_size . Por fim, na Camada 4 aplicou-se Explicabilidade via Modelo Substituto e SHAP. Como o K-Means não possui função de decisão analítica interpretável, implementou-se um modelo substituto utilizando Random Forest (RF) para reproduzir os rótulos do K-Means ($n_{\text{estimators}}=300, \text{max_depth}=\infty, \text{min_samples_leaf}=5, \text{random_state}=42$; acurácia CV-5: $95,0\% \pm 0,6\%$, Fig. IV).

Em seguida, utilizou-se o SHAP TreeExplainer para extrair os valores de contribuição exatos para os 5.570 municípios em 16 *features* contínuas — excluídas as quatro variáveis indicadoras (*dummies*) regionais, que integram o vetor de treinamento do

modelo substituto mas não produzem contribuições marginais interpretáveis no contexto municipal. Como a segmentação correlaciona-se fortemente com a região, as *dummies* regionais podem absorver poder preditivo que, de outra forma, seria atribuído a *lat*, *lon* e *idhm*; quantificar essa sobreposição é direção futura.

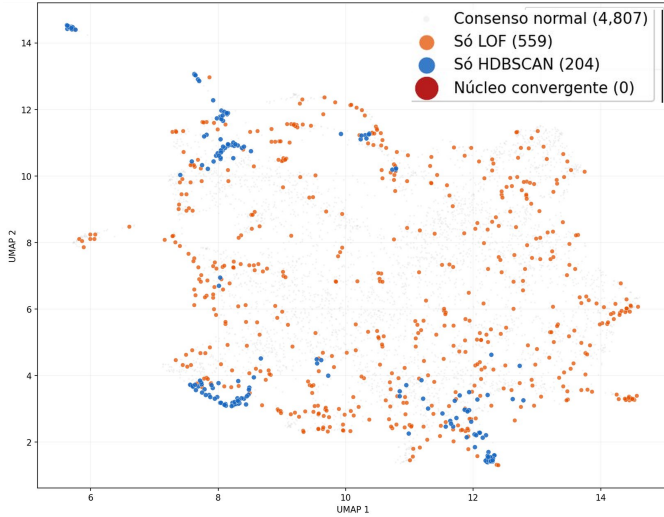


Fig. 3. Sobreposição HDBSCAN x LOF (UMAP 2D). Laranja = só LOF (559); Azul = só HDBSCAN (204); Núcleo convergente = 0. A ausência de sobreposição é consistente com a distinção formal entre os dois tipos de excepcionalidade ($p < 0,001$, Mann-Whitney U).

Diferentemente de [6], que usava SHAP sobre quatro classificadores independentes sem validação unificada, este trabalho usa um único modelo substituto validado (CV-5, 95,0%), garantindo escala comparável e rastreabilidade total [9].

Cabe ressaltar que os valores SHAP explicam o modelo substituto, não a decisão do K-Means diretamente; os $\sim 5\%$ de municípios cuja classificação pelo RF diverge do K-Means (fronteira C11–C12, Fig. IV) têm explicações calculadas sobre um rótulo divergente do original, o que deve ser considerado na leitura de casos de fronteira.

Os principais drivers SHAP mantiveram a mesma ordenação relativa dos três primeiros atributos por cluster sob variações moderadas (*area_km2* em C13, *pib_industria* em C10/C14, *idhm+lat* em C11). As matrizes de explicação em nível municipal são exportadas em *shap_explicacoes_municipios.csv*.

IV. RESULTADOS E DISCUSSÃO

Os cinco perfis globais resumidos na Tabela I e visualizados na Fig. 2 capturam diferenças reais do território ($p < 0,001$, afastando a hipótese de agrupamento espúrio), oferecendo um subsídio objetivo para definição de medidas regulatórias, tais como a priorização de fiscalização e a atuação diferenciada por contexto territorial: C10 (Sudeste) e C11 (padrão nacional) demandam, respectivamente, ações de **Manutenção** e **Monitoramento** contínuo; C12 (Nordeste) e C13 (Norte),

com vulnerabilidades em atendimento como Cumprimento de velocidade (IND4) e Latência bidirecional (IND5), configuram os perfis **Prioritário** e **Crítico**; e C14 serve como **Benchmark** regulatório para avaliação comparativa dos demais perfis. A validação estatística confirmou que os cinco perfis são separáveis em todos os sete indicadores RQUAL ($p < 0,001$), afastando a hipótese de agrupamento espúrio. A projeção UMAP em 2D confirmou que os cinco perfis formam regiões espacialmente coesas e não sobrepostas — condição necessária para que a detecção de anomalias da Camada 3 opere sobre estruturas genuínas.

A Camada 3 revelou que municípios do mesmo perfil global podem esconder comportamentos radicalmente distintos, ilustrados na Fig. 3. Em C13 (Norte/Amazônico), um subcluster de municípios do agronegócio (MT/MS, $n = 19$) apresenta IND5 substancialmente superior ao restante do cluster — saindo de $-2,19\sigma$ para $-0,59\sigma$ — demonstrando que restrições geográficas não determinam uniformemente a qualidade de serviço. Em C10, um subcluster de gargalo (SP/RJ, $n = 23$) combina infraestrutura de ponta (INF4-UP $+1,36\sigma$) com Cumprimento de velocidade (IND4) negativo, revelando que o problema nesses municípios é capacidade operacional sob alta demanda, não infraestrutura. O LOF identificou 559 anomalias isoladas que demandam atenção regulatória individualizada — entre os mais críticos, Tunas/RS ($-4,31\sigma$ em IND5) e Silvanópolis/TO ($-4,11\sigma$ em IND8), desempenhos completamente ocultos nas médias do perfil. A sobreposição nula ($\text{HDBSCAN} \cap \text{LOF} = 0$, $p < 0,001$, Mann-Whitney U) é compatível com a complementaridade entre os dois critérios — não uma validação empírica independente de ortogonalidade: HDBSCAN identifica *onde há grupos dentro do grupo*; LOF identifica *quem destoa de todos ao redor*.

TABELA II. Visão consolidada: segmentação, SHAP e anomalias

Cluster	n	SHAP dom.	LOF	Direcion.
C10 Urb.-Av.	216	<i>pib_industria</i>	22	Manutenção
C11 Interm.	3.054	<i>idhm + lat</i>	306	Monitoramento
C12 NE Vuln.	2.057	<i>idhm + lat + lon</i>	206	Prioritário
C13 N/Amaz.	216	<i>area_km2</i>	22	Crítico
C14 Capitais	27	<i>pib_ind + dens.</i>	3	Benchmark

A Tabela II sintetiza os resultados consolidados por cluster. A Fig. 4 confirma que o modelo substituto reproduz com precisão os cinco clusters, com os poucos erros concentrados na fronteira entre C11 e C12 — os perfis com características mais semelhantes entre si.

A Fig. 5 revela três achados regulatórios: i) em C13, a extensão territorial (*area_km2*) domina, indicando barreira geográfica, não técnica [10]; ii) em C12, localização e indicadores socioeconômicos têm pesos comparáveis, apontando para foco em atendimento e resolução de reclamações, não expansão de cobertura; e iii) em C10/C14, o PIB industrial domina ambos, confirmando desafio de capacidade de atendimento sob alta demanda, não infraestrutura.

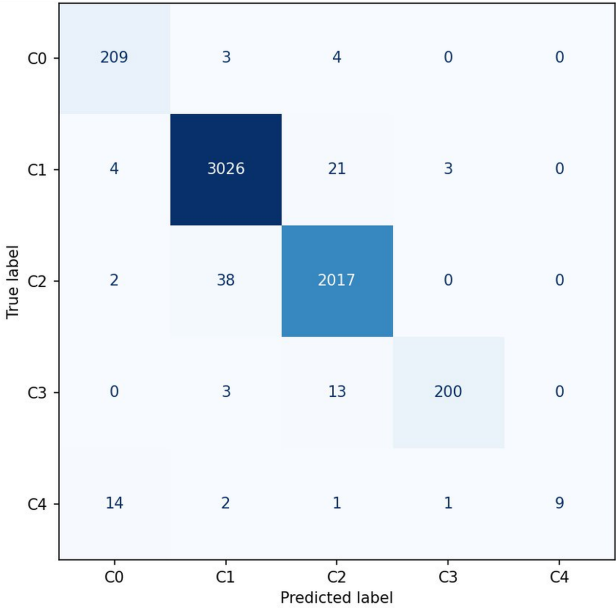


Fig. 4. Matriz de confusão: modelo substituto RF vs. K-Means original. A diagonal dominante valida que os valores SHAP explicam os clusters reais (acurácia CV-5: 95,0% ± 0,6%).

V. CONCLUSÃO

Este artigo demonstrou que a segmentação global pode ocultar aspectos relevantes à atuação regulatória. O caso mais direto é o subcluster do agronegócio em C13: a segmentação global o classifica no pior perfil do Brasil, mas a análise de subestruturas revela desempenho em latência bidirecional (IND5) próximo da média nacional — invisível sem a separação proposta. Da mesma forma, os 559 municípios anômalos jamais seriam detectados por análise de grupo, pois seu desvio se manifesta apenas no nível individual. As implicações regulatórias diferem por perfil: barreiras territoriais no Norte, atendimento e resolução no Nordeste, e capacidade operacional sob alta demanda em C10 e C14 — padrões invisíveis à segmentação global. O framework, disponível em [7], foi desenvolvido sobre indicadores RQUAL de banda larga móvel, mas sua arquitetura é diretamente aplicável a outros serviços ANATEL (TV por assinatura, telefonia fixa), desde que estruturados no mesmo formato de *features*.

As análises limitaram-se aos dados RQUAL de 2022 e ao IDHM (2010, Seção II), a edição mais recente disponível. Direções futuras incluem extensão para 2023–2025, atualização das variáveis do IBGE, verificação de equidade entre grupos populacionais, análise de causalidade, sensibilidade do LOF ao contamination, importância relativa das *dummies* regionais, e uso de índices compostos para segmentação.

AGRADECIMENTOS

Este trabalho foi financiado em partes pelo Instituto Nacional de Ciência e Tecnologia (INCT-Sinais), CNPq (processo 406517/2022-3), ANATEL e FAPESP (processo 20/09838-0 — BIOS).

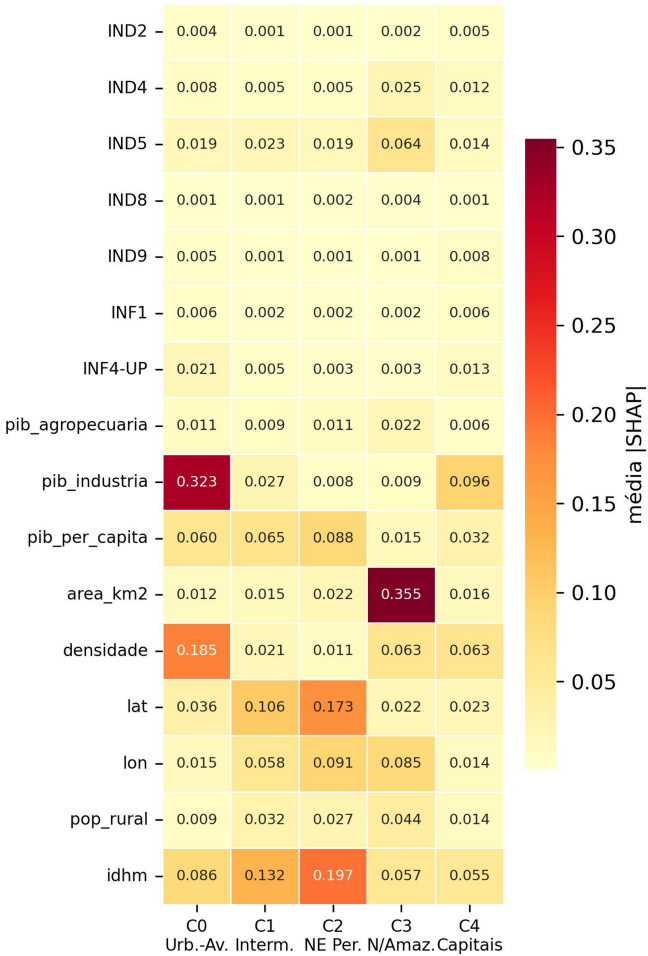


Fig. 5. Importância SHAP média por cluster. Células mais escuras = maior peso explicativo. *area_km2* domina C13 (0,355); *pib_industria* domina C10 (0,323); *idhm*+*lat*+*lon* dominam C12 (0,197). Indicadores RQUAL apresentam peso residual em todos os clusters.

REFERÊNCIAS

[1] S. O. Folorunsho, O. A. Adenekan, C. Ezeigweneme, I. C. Somadina e P. A. Okeleke, “Optimizing network performance and QoS with AI-driven solutions for future telecom,” *Int. J. Front. Eng. Technol. Res.*, vol. 7, n. 1, pp. 73–92, 2024. DOI: 10.53294/ijfetr.2024.7.1.0041

[2] M. E. Meena e J. Geng, “Dynamic competition in telecommunications,” *SAGE Open*, vol. 12, n. 2, pp. 1–10, 2022.

[3] ANATEL, “Sistema RQUAL,” Brasília, 2023. gov.br/anatel/pt-br/dados/qualidade

[4] C. Molnar, *Interpretable Machine Learning*, 2.^a ed., 2022. christophm.github.io/interpretable-ml-book

[5] M. T. Ribeiro, S. Singh e C. Guestrin, “Why Should I Trust You?,” *ACM SIGKDD*, 2016, pp. 1135–1144.

[6] M. A. de A. Fidos Jr, R. Machado, D. I. Alves e B. F. Uchôa-Filho, “Qualidade de Serviços em Telecomunicações e Desigualdades Territoriais: Uma Abordagem Explicável com Clusterização Espacial”, *XLIII SBrt*, Natal, RN, 2025. DOI: 10.14209/sbrt.2025.1571157091

[7] M. A. de A. Fidos Jr, “Framework de Explicabilidade”, GitHub, 2026. github.com/mfidosjr/Framework-Explicabilidade-Indicadores

[8] IBGE, “Censo Demográfico 2022”, Rio de Janeiro: IBGE, 2023. censo2022.ibge.gov.br/resultados.html

[9] S. M. Lundberg e S.-I. Lee, “A unified approach to interpreting model predictions,” *NeurIPS*, vol. 30, pp. 4765–4774, 2017.

[10] E. J. Oughton *et al.*, “Policy choices can help keep 4G and 5G affordable,” *Telecom. Policy*, vol. 46, n. 3, p. 102291, 2022.