IA EXPLICÁVEL E GENERATIVA NO SMART K-MEANS: SHAP E LLM PARA CARACTERIZAÇÃO SEMÂNTICA DE INDICADORES DE EXCLUSÃO SOCIAL

Daniel Cirino Martins[1] - it.danielcirino@gmail.com
Matheus Pereira Libório[1] - matheus.liborio@unimontes.br
Marcos Flávio Silveira Vasconcelos D'Angelo[1] - marcos.dangelo@unimontes.br
Angélica Cidália Gouveia dos Santos[1] - angelicacidalia@gmail.com
[1]Programa de Pós-Graduação em Modelagem Computacional e Sistemas, PPGMCS,
Universidade Estadual de Montes Claros, UNIMONTES - Montes Claros, MG, Brasil

Abstract. *Indicadores compostos produzem* scores *que ordenam, mas não classificam, as unidades de decisão. O Smart K-Means responde a essa lacuna combinando o índice de Entropia de Shannon com o K-Means, descartando iterativamente os indicadores de maior entropia até obter agrupamentos com qualidade controlada pela Silhueta. Este artigo estende o método com duas camadas de Inteligência Artificial. A primeira, de explicação pós-agrupamento baseada em IA Explicável (XAI), usa um Random Forest sobre os rótulos dos grupos, triangulado com valores SHAP. A segunda, de IA Generativa, emprega um Modelo de Linguagem de Grande Escala (LLM) que gera automaticamente rótulos e descrições semânticas para cada grupo. O método é aplicado à exclusão social de Maringá e Londrina (PR), municípios de escala contrastante (483 e 770 setores censitários). Ambos convergem para três grupos com Silhueta acima do limiar (0,568 e 0,630) e fidelidade do modelo explicativo superior a 0,98, mas com indicadores discriminantes distintos, evidenciando adaptação ao contexto local. Em Maringá, a sequência de descarte do estudo original é reproduzida.*

Keywords: *Exclusão Social, Agrupamento Estatístico, Inteligência Artificial Explicável (XAI), SHAP, Modelos de Linguagem de Grande Escala (LLM)*

1. INTRODUÇÃO

Métodos consolidados de indicadores compostos sintetizam fenômenos sociais multidimensionais em *scores* unidimensionais que ordenam as unidades de decisão, mas não as classificam em grupos interpretáveis (Mazziotta & Pareto, 2016). Essa lacuna compromete a comunicação dos resultados e o planejamento territorial.

Buscando superar essa lacuna, o Smart K-Means (Martins et al., 2024) parte do uso do k-means para representar a exclusão social (Libório et al., 2022) e o estende à seleção de

indicadores. Nesse processo, o índice de Entropia de Shannon ordena o descarte iterativo, e o coeficiente de Silhueta controla a qualidade do agrupamento. Os resultados obtidos na aplicação à exclusão social de Maringá (PR), relatados por Martins et al. (2024), identificaram três grupos interpretáveis com Silhueta média de 0,51. Nenhuma configuração sugerida pelos métodos tradicionais atingiu o limiar.

O estudo original reconhece três limitações (Martins et al., 2024): (i) ausência de evidência de robustez para outros dados de entrada; (ii) falta de comparação com critérios alternativos de descarte; e (iii) ausência de informações sobre as características dos grupos, o que exige esforço adicional do tomador de decisão.

Este artigo estende o Smart K-Means com duas camadas de Inteligência Artificial. A Seção 2 traz o referencial teórico; a Seção 3, a base de dados e as etapas do método; a Seção 4, os resultados em Maringá e Londrina; e a Seção 5 conclui, respondendo às três limitações do estudo original.

2. REFERENCIAL TEÓRICO

O K-Means minimiza a inércia intragrupo (Kassambara, 2017), mas indicadores pouco informativos degradam a qualidade do agrupamento. O índice de Entropia de Shannon (Shannon, 1948) é empregado em sua forma operacional. Para cada indicador, computa-se $H = -\sum_i p_i \log_2 p_i$ sobre a distribuição de frequências p_i de seus valores distintos. O Smart K-Means descarta iterativamente o indicador de maior índice, retendo aqueles cujos valores se concentram em poucos níveis distintos. A entropia estabelece a ordem do descarte, e o coeficiente de Silhueta (Kassambara, 2017), o ponto de parada (Martins et al., 2024).

Complementarmente, o Random Forest (Breiman, 2001), treinado com os rótulos dos grupos como variável-alvo, opera como modelo explicativo pós-agrupamento (*post hoc surrogate*). Ele explica a estrutura já encontrada, sem pretensão de predizer novos dados (Molnar, 2022). Sua importância por decréscimo de impureza (*Mean Decrease in Impurity* — MDI) resume em um valor por indicador a contribuição para a separação entre grupos; os valores SHAP (Lundberg & Lee, 2017) a decompõem por observação e por classe. Construídas de formas independentes, as duas medidas admitem triangulação.

Modelos de linguagem de grande escala (LLM) generalizam a tarefas não vistas a partir de instruções em linguagem natural (Brown et al., 2020). Aqui o LLM não participa do agrupamento nem da atribuição de importância, recebendo apenas as estatísticas já computadas de cada grupo e o dicionário de variáveis, e produzindo rótulos e descrições. Essa é uma camada aditiva e verificável, cujas afirmações são rastreáveis às estatísticas fornecidas.

Dado esse contexto, o que distingue peso de importância? O *peso* em um índice composto expressa quanto a dimensão constitui o conceito medido; a *importância* de um modelo explicativo, quanto o indicador separa os grupos formados. São, portanto, perguntas distintas, e aqui Random Forest e SHAP operam apenas como camada de explicação, não como esquema de ponderação.

3. MATERIAIS E MÉTODOS

3.1 Bases de dados e critérios de seleção dos municípios

O método é aplicado aos mesmos 15 indicadores de exclusão social, organizados nas dimensões demográfica (VD), econômica (VE), educacional (VED), habitacional (VH) e de entorno urbano (VA), extraídos do Censo Demográfico de 2010 por setor censitário, segundo o *framework* de Martinuci & Libório (2022) adotado em Martins et al. (2024). Maringá (357.077 habitantes; 483 setores censitários) é a base do estudo original. Londrina (506.701 habitantes; 770 setores) é o segundo município mais populoso do Paraná e foi selecionada por contrastar em escala e perfil socioeconômico, mantendo indicadores, fonte e período. Isso assegura comparabilidade.

3.2 Método Smart K-means

O Smart K-means é executado em cinco etapas, descritas a seguir.

Etapa 1. Para cada K em $[3, 7]$, o algoritmo parte do conjunto completo e descarta iterativamente o indicador de maior entropia, reexecutando o K-Means, até a Silhueta média superar o limiar de aceitação de 0,50 ou restarem dois indicadores. Entre os K que atingem o limiar, seleciona-se o que retém mais indicadores; empates são resolvidos pela maior Silhueta, com parâmetros idênticos nos dois municípios.

Etapa 2. Os rótulos da Etapa 1 são a variável-alvo de um Random Forest treinado sobre os indicadores retidos. A acurácia *out-of-bag* (OOB) mede a *fidelidade do modelo explicativo à partição* — o quanto os indicadores retidos reproduzem os grupos —, não validação preditiva sobre novos dados. A contribuição de cada indicador é quantificada pelo MDI e pela média dos valores absolutos de SHAP por grupo, computada sobre todos os setores em Maringá e sobre amostra de 500 em Londrina. Adota-se o MDI por serem as variáveis homogêneas — contínuas, em escala $[0, 1]$ —, condição sob a qual seu viés contra atributos de alta cardinalidade não se manifesta (Strobl et al., 2007).

Etapa 3. Independentemente do modelo explicativo, a média de cada indicador retido por grupo é padronizada entre os grupos (*z-score*) e ordenada em postos de 1 a K. O Índice Global do grupo g é $IG_g = (\overline{r}_g - 1)/(K - 1)$, em que $\overline{r}_g$ é a média de seus postos. Essa camada pressupõe polaridade alinhada a montante — maior valor corresponde, uniformemente, a melhor condição —, responsabilidade da preparação dos dados. Por preservar a direção de cada indicador, essa camada permite auditar esse pressuposto.

Etapa 4. Para cada grupo, selecionam-se os seis indicadores de maior valor SHAP daquele grupo; os *z-scores* informam direção e magnitude do desvio ante a média municipal. O conjunto é submetido a um Modelo de Linguagem de Grande Escala (Brown et al., 2020) — *deepseek-v4-flash* — com *prompt* padronizado (contexto metodológico, dicionário de indicadores e instrução de geração). A geração inicial é automática, a partir exclusivamente dos dados estatísticos. O texto final é revisado pelo analista antes da publicação.

Etapa 5. Os rótulos de grupo são unidos por código à malha de setores censitários de 2010 — mesma referência temporal dos dados, assegurando cobertura integral —, produzindo um mapa coroplético. Essa camada de comunicação para o planejamento intraurbano não altera o agrupamento.

Na sequência, são examinados os resultados obtidos para os municípios de Maringá e Londrina.

4. RESULTADOS

4.1 Agrupamento

A Tabela 1 mostra que ambos os municípios convergem para três grupos com Silhueta acima do limiar de aceitação, apesar do contraste de escala. As demais configurações de $[3, 7]$ produzem Silhueta inferior. Maringá seleciona o mesmo $K = 3$ do estudo original (Martins et al., 2024) e reproduz sua sequência de descarte, inclusive nos dois primeiros indicadores eliminados (VA4 e VE3), diferindo apenas pela exclusão adicional de VD4. O resultado obtido apresenta Silhueta maior (0,568 contra 0,51), porém o número de indicadores retidos passa de oito para sete. Esses resultados replicam os de Martins et al. (2026), base para as camadas de IA.

Tabela 1: Resultados do agrupamento em Maringá e Londrina.

	Maringá	**Londrina**
Setores censitários (N)	483	770
Grupos (K)	3	3
Silhueta média	0,568	0,630
Indicadores retidos	7 de 15	9 de 15
Avaliações até convergência	9	7
Distribuição dos setores	413 / 38 / 32	624 / 120 / 26
Fidelidade do modelo explicativo (OOB)	0,983	0,995

4.2 Explicação pós-agrupamento

A Tabela 2 apresenta os indicadores de maior contribuição segundo as duas medidas. Em Maringá, MDI e SHAP convergem nos três indicadores mais contributivos, na mesma ordem. Em Londrina, convergem no par dominante — VH3 e VA4, responsáveis por 89,9% da importância agregada — e divergem a partir do terceiro colocado (o MDI aponta VED2; o SHAP, VH4). Nenhum dos dois indicadores dominantes de Londrina coincide com os de Maringá. O método não impõe padrão fixo de importância. A explicação se ajusta ao perfil socioespacial de cada município, ainda que ambos convirjam para o mesmo K.

Tabela 2: Indicadores de maior contribuição para a separação entre grupos (MDI e SHAP).

	Maringá			**Londrina**		
Posto	Indicador	MDI	|SHAP|	Indicador	MDI	|SHAP|
1º	VH4	0,398	0,076	VH3	0,742	0,165
2º	VE2	0,251	0,047	VA4	0,156	0,031
3º	VH2	0,218	0,041	VED2	0,023	0,006

A fidelidade, embora elevada, não é uniforme. As divergências *out-of-bag* (4 de 770 setores em Londrina; 8 de 483 em Maringá) concentram-se entre grupos adjacentes, e o menor *recall* ocorre no grupo de menor suporte — G3 de Maringá ($n = 32$), 0,906 ante o agregado de 0,983 —, ressalva para sua caracterização.

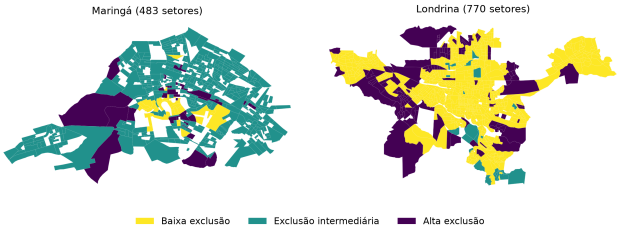

Figura 1: Distribuição espacial dos grupos por nível de exclusão (Seção 4.3); paleta segura para daltonismo, mais escuro = maior exclusão. Agrupamento independente por cidade. Junção com a malha censitária de 2010 do IBGE, cobertura integral.

4.3 Caracterização semântica

A geração inicial é automática (Etapa 4). A seguir, apresenta-se a versão revisada pelo analista. Nos dois municípios, o Índice Global ordena os grupos consistentemente com os rótulos, do menor ao maior nível de exclusão: G3, G1 e G2.

Maringá. *G3 — Baixa exclusão* ($n = 32$; 6,6%): alta renda, moradias amplas de baixa densidade e infraestrutura plena. *G1 — Exclusão intermediária* ($n = 413$; 85,5%): a massa urbana, com infraestrutura básica universalizada, sem concentração de alta renda. *G2 — Alta exclusão* ($n = 38$; 7,9%): maior adensamento domiciliar, acesso mais precário à água encanada — o indicador de maior |SHAP| neste grupo (0,652) — e maior analfabetismo.

Londrina. *G3 — Baixa exclusão* ($n = 624$; 81,0%): grupo majoritário, com saneamento quase universal e o melhor desempenho geral. *G1 — Exclusão intermediária* ($n = 26$; 3,4%): pequeno e heterogêneo (Silhueta 0,24), combina a maior presença de domicílios com quatro ou mais banheiros à maior vulnerabilidade de entorno — polarização que o rótulo intermediário sintetiza. *G2 — Alta exclusão* ($n = 120$; 15,6%): déficit de saneamento — o esgoto é o indicador de maior |SHAP| neste grupo (0,753) — e maior analfabetismo infantil.

A revisão humana da caracterização, ao confrontar a direção de cada indicador (camada descritiva) com a semântica declarada no dicionário de variáveis, evidenciou sentidos armazenados que contradizem o esperado. O caso mais nítido: o grupo de menor exclusão de cada município exibe, sob leitura literal, quase 100% de domicílios "sem banheiro exclusivo". Trata-se de inconsistência de polaridade dos dados de entrada, não de erro do agrupamento. A partição, as medidas de importância e a matriz de confusão independem do sentido dos eixos, e a ordenação pelo Índice Global permanece inalterada. A caracterização atua como verificação de qualidade de dados, tornando legíveis as inconsistências que exigem avaliação, cuja correção cabe à preparação dos dados a montante.

4.4 Espacialização dos grupos

A leitura espacial, apresentada na Fig. 1, revela padrões contíguos e coerentes com a estrutura urbana. Em Maringá, o grupo majoritário recobre quase toda a malha. Em Londrina, forma mancha central contínua, com os demais grupos em bolsões periféricos.

5. CONCLUSÕES

Os resultados obtidos em Maringá e Londrina respondem às três limitações identificadas por Martins et al. (2024). A limitação (i), robustez do método para outros dados de entrada, é

superada pela convergência consistente para três grupos em municípios de escalas distintas, com Silhueta acima do limiar e fidelidade explicativa superior a 0,98 (Tabela 1). A limitação (ii), vantagem do critério de entropia sobre alternativas, é abordada pelos resultados complementares de Martins et al. (2026): nos dois municípios aqui analisados, a entropia supera carga fatorial, intercorrelação e variância em Silhueta. A limitação (iii), ausência de caracterização dos grupos, é solucionada pela camada de IA Explicável — triangulação entre MDI e SHAP — e pela geração automática de rótulos e descrições semânticas via LLM (Seção 4.3).

A integração das camadas de explicação e geração semântica amplia o escopo do método: além de formar grupos com qualidade mensurada, produz interpretações auditáveis e identifica inconsistências de polaridade nos dados de entrada. O caso dos domicílios "sem banheiro exclusivo" (Seção 4.3) ilustra a caracterização semântica atuando como mecanismo de verificação de qualidade dos indicadores, sem afetar a partição nem as medidas de importância.

O método consolida-se, assim, como abordagem integrada para representação, explicação e comunicação de fenômenos multidimensionais no planejamento urbano.

Como trabalho futuro, destacam-se a automação da verificação de polaridade a partir do catálogo de variáveis, a avaliação da camada explicativa sob multicolinearidade elevada e a extensão do método a séries temporais, para acompanhar a evolução da exclusão social.

Agradecimentos

Os autores agradecem à FAPEMIG (grants APQ-00528-24 e APQ-00691-23).

REFERÊNCIAS

Breiman, L. (2001), "Random Forests", *Machine Learning*, vol. 45, n. 1, pp. 5–32.

Brown, T. B. et al. (2020), "Language Models are Few-Shot Learners", *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901.

Kassambara, A. (2017), *Practical Guide to Cluster Analysis in R*. STHDA, 2ª ed.

Libório, M. P.; Martinuci, O. S.; Machado, A. M. C.; Lyrio, R. M.; Bernardes, P. (2022), "Time–Space Analysis of Multidimensional Phenomena: A Composite Indicator of Social Exclusion Through k-Means", *Social Indicators Research*, vol. 159, n. 2, pp. 569–591.

Lundberg, S. M.; Lee, S.-I. (2017), "A Unified Approach to Interpreting Model Predictions", *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774.

Martins, D. C. et al. (2024), "Smart K-Means: novo método para representação de fenômenos sociais multidimensionais", in *Anais do LVI Simpósio Brasileiro de Pesquisa Operacional*, Fortaleza. DOI: 10.59254/sbpo-2024-193783.

Martins, D. C.; Laudares, S.; Ekel, P. I.; D'Angelo, M. F. S. V.; Libório, M. P. (2026), "A New Method for Representing Multidimensional Phenomena and Its Application to Social Exclusion in Cities", *Measurement: Interdisciplinary Research and Perspectives*, publicação online antecipada. DOI: 10.1080/15366367.2026.2653237.

Martinuci, O. S.; Libório, M. P. (Org.) (2022), *Desigualdades intraurbanas: metodologias para produção e análise de indicadores compostos*, 1ª ed., Curitiba: Editora CRV, 218 p.

Mazziotta, M.; Pareto, A. (2016), "On a Generalized Non-compensatory Composite Index for Measuring Socioeconomic Phenomena", *Social Indicators Research*, vol. 127, n. 3, pp. 983–1003.

Molnar, C. (2022), *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2ª ed.

Shannon, C. E. (1948), "A Mathematical Theory of Communication", *Bell System Technical Journal*, vol. 27, n. 3, pp. 379–423.

Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. (2007), "Bias in random forest variable importance measures: Illustrations, sources and a solution", *BMC Bioinformatics*, vol. 8, art. 25.