

IMPLEMENTAÇÃO DO K-MEANS PARA ANÁLISE EXPLORATÓRIA INICIAL DA ASSOCIAÇÃO ENTRE SOLIDÃO, POLÍTICA E MÍDIAS SOCIAIS NO BRASIL

Jenaina de Fatima dos Santos – jenainasantos60@gmail.com
Jesus Antonio García Sánchez – jesus@unifei.edu.br

Palavras-chave: Solidão, Comportamento político, Mídias sociais, Machine learning, K-means.

1. INTRODUÇÃO

A solidão configura-se como um fenômeno de impacto social agravado pelas remodelações políticas e tecnológicas atuais (Hertz, 2021). Embora países centrais avancem na associação entre solidão e questões políticas como extremismo (Peterson et al., 2022; Langenkamp, 2025), no Brasil persiste uma lacuna significativa de pesquisas, caracterizada pela limitação de instrumentos e análises restritas a grupos específicos (Barroso et al., 2015). Esta escassez é preocupante em um contexto periférico, podendo refletir severamente esse fenômeno, especialmente entre grupos minoritários (Hertz, 2021).

Diante deste cenário, este estudo é parte de uma dissertação de mestrado do Programa de Pós-graduação em Desenvolvimento, Tecnologias e Sociedade da Universidade Federal de Itajubá, orientado pela questão central de como evidências empíricas podem confirmar a associação entre solidão, comportamento político e mídias sociais, utilizando algoritmos de *Machine Learning* para analisar. O caráter exploratório do trabalho justifica-se pela necessidade de superar a individualização de problemas de solidão, seguindo o caminho aberto por estudos que analisam condições sociais estruturais, compreendendo a solidão como uma questão de justiça social (Barreto et al., 2024).

2. REFERENCIAL TEÓRICO

São poucos achados sobre o tema no cenário global. Best (2015) identifica que fatores como insônia e isolamento influenciam comportamentos políticos de risco, alertando para a idealização errônea de líderes como figuras solitárias. Um estudo feito no Canadá associou o bem-estar social a crenças neoliberais e a pandemia da COVID-19, apontando menores níveis de solidão (Card, Hepburn, 2022). O contexto da pandemia foi crucial para essa associação, com Zhang *et al.* (2022) documentando, via dados do *Twitter*, picos de discussão sobre solidão e políticas governamentais nos EUA. Lo Moro *et al.* (2023) associaram o aumento da ciberagressão na Itália ao uso de redes sociais, notando que temas políticos envolvendo grupos identitários geram mais conflitos online.

As manifestações de solidão também possuem nuances de gênero, como apontado por Mejova e Hommadova (2023), que observaram maior menção à solidão por mulheres no *Twitter*, enquanto os homens possivelmente sub-relatam o problema devido a estigmas sociais. No cenário brasileiro, o estudo de Josiowicz (2021) é citado como único a abordar o fenômeno, revelando que o *Twitter* funcionou simultaneamente como espaço de mobilização política durante a pandemia, servindo como uma forma de "leitura digital terapêutica" para os usuários.

3. MATERIAIS E MÉTODOS

Seguiu-se a adaptação do método de *Machine Learning (ML)* proposto por Géron (2021) para análise exploratória de dados subjetivos sobre solidão, comportamento político e mídias sociais, utilizando uma amostra de 176 participantes. Esta pesquisa foi submetida à aprovação do Comitê de Ética em Pesquisa antes da coleta de dados.

A implementação seguiu quatro etapas principais: 1) análise do panorama geral através de revisão sistemática na *Scopus*, que revelou apenas 6 estudos relevantes usando os termos "*loneliness*" AND "*politics*" AND "*social media*" presentes no título, resumo e palavras-chave; 2) obtenção de dados empíricos mediante questionário *on-line* aplicado no primeiro semestre de 2023, com 20 itens da Escala Brasileira de Solidão-UCLA (Barroso *et al.*, 2016) e 28 questões adaptadas sobre comportamento político e

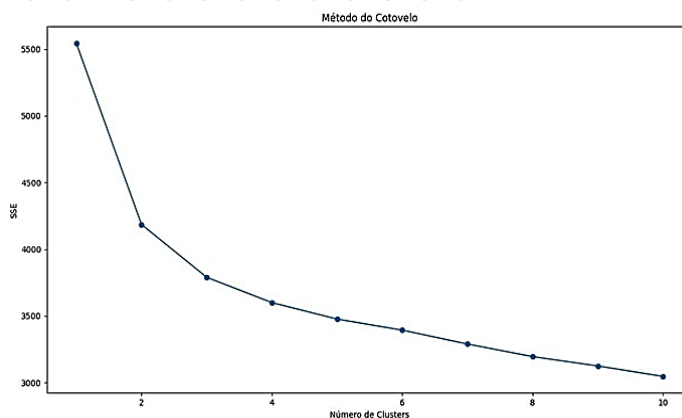
uso de médias durante as eleições de 2022; 3) pré-processamento em Python 3.12.4 utilizando bibliotecas como Pandas e Scikit-learn, tratando 150 instâncias com 63 variáveis ; 4) seleção e treinamento do algoritmo *K-means* não supervisionado para clusterização dos dados.

A implementação do *K-means* definiu $K=4$ *clusters* através do método do cotovelo, considerada ideal para minimizar a soma dos quadrados dentro do agrupamento. Para visualização dos resultados de saída, utilizou-se *PCA* para projeção dos dados em espaço 2D. O *Random Forest* foi utilizado em seguida para identificação das variáveis mais relevantes na associação. A visualização incluiu gráficos de distribuição de severidade da solidão e faixa etária por *cluster*, dado seu impacto significativo nos agrupamentos. Todo o processo para implementação do modelo foi desenvolvido no *Visual Studio Code*.

4. RESULTADOS E DISCUSSÃO

Ao atribuir o $K=4$ como o número de *clusters* de modo que partir deste ponto, a dispersão foi mantida baixa. A Figura 1 mostra o resultado de aplicação do método do cotovelo, onde pode ser observado a dispersão *SSE* no eixo vertical e o número de *clusters* no eixo horizontal, onde o modelo não conseguiu encontrar um número de *clusters* adequado, pois a curva não atinge um patamar horizontal. Isso indica ainda a necessidade de mais instâncias ou a redução do número de variáveis ou atributos.

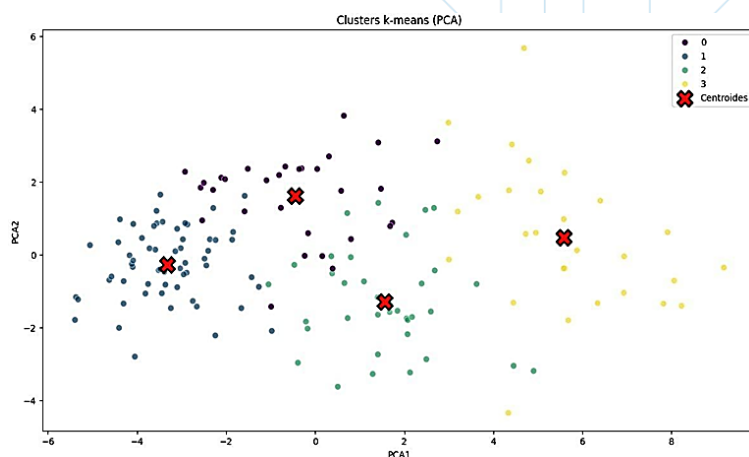
Figura 1- gráfico do método do Cotovelo



Fonte: elaborado pelos autores

Na Figura 2 mostra os resultados do método *PCA*, considerando todo o conjunto de variáveis presentes no *dataset* inicial, onde observou-se que os *clusters* não foram detectados ou agrupados devidamente. Evidencia-se a alta dimensionalidade dos dados pelo número de características apresentadas.

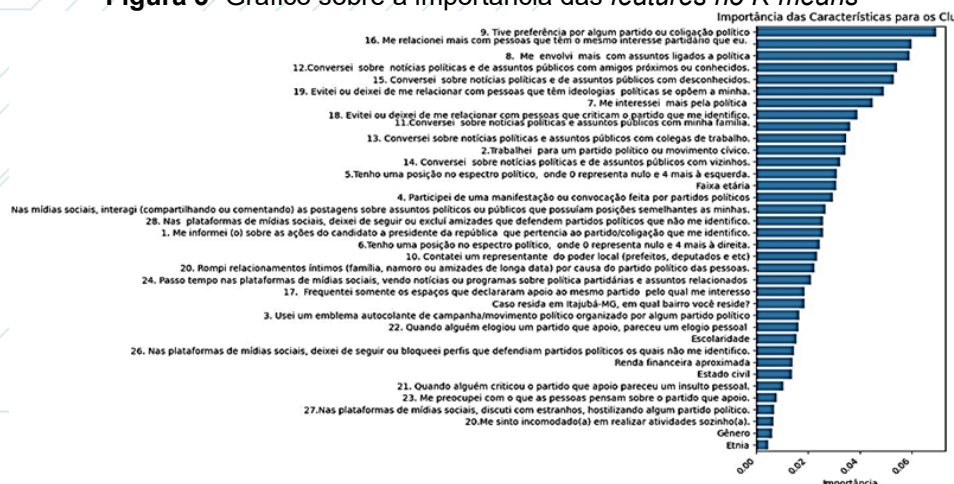
Figura 2- Gráfico *PCA* do *K-means*



Fonte: elaborado pelos autores.

Na Figura 3 mostram-se os resultados obtidos através do *random forest*. As variáveis relacionadas ao comportamento político, como a preferência por partido político, tiveram mais influência na formação dos *clusters*, quando comparadas às variáveis sociodemográficas (gênero e etnia). Isso pode indicar viés da seleção amostra com falta de critérios bem definidos e estas características não precisam ser consideradas para implementação de outros algoritmos.

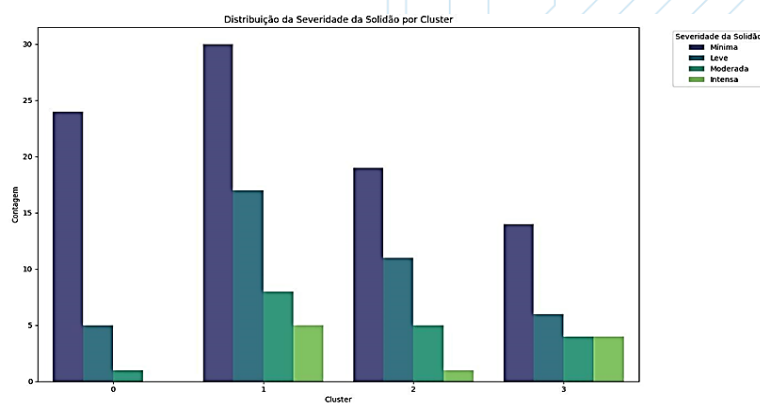
Figura 3- Gráfico sobre a importância das *features* no *K-means*



Fonte: elaborado pelos autores.

Na Figura 4, analisa-se a distribuição da intensidade da solidão por *cluster* por meio de um gráfico, a solidão mínima está representada em azul escuro, a solidão leve em azul claro, a solidão moderada em verde escuro e a solidão intensa em verde claro. A análise da severidade da solidão entre os *clusters* não revelou um padrão claro de distribuição. Embora cada *cluster* apresente uma proporção diferente de casos de solidão mínima e intensa, não foi possível identificar uma tendência consistente de como a intensidade da solidão influencia o agrupamento.

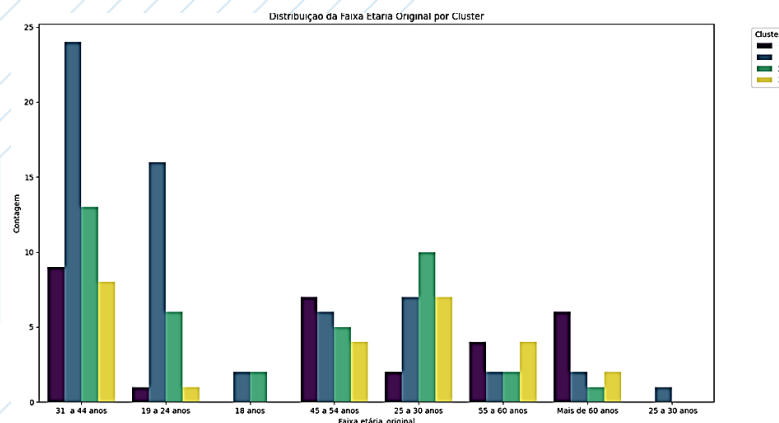
Figura 4- Gráfico de distribuição da severidade da solidão por cluster



Fonte: Elaborada pelos autores.

Da mesma forma, na Figura 5, é visualizada a distribuição da severidade da solidão através da clusterização por faixa etária, onde foi apontado que a faixa etária de 31 a 44 anos existem mais casos de solidão leve e solidão intensa.

Figura 5- Gráfico agrupamento da intensidade da solidão por faixa etária



Fonte: Elaborada pelos autores.

4. CONSIDERAÇÕES FINAIS

Até o momento foi possível identificar o comportamento dos dados da estrutura da clusterização através do *K-means*. Este algoritmo apresentou limitações para a análise de padrões complexos de associação entre as variáveis com os dados disponibilizados. Estes dados possuem alta densidade e variabilidade de características de uma relação não linear entre as variáveis. Portanto, melhorias no pré-processamento seletivo dos dados precisam ser feitas, descartando variáveis pouco influentes na associação principal, viabilizando seleção de modelos ML mais assertivos para o tipo de dados disponibilizado.

AGRADECIMENTOS

Agradecemos a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior por subsidiar esta pesquisa.

REFERÊNCIAS

BARROSO, Sabrina *et al.* Evidências de validade da Escala Brasileira de Solidão UCLA. **Jornal Brasileiro de Psiquiatria**, v. 65, n. 1. Disponível em: <https://www.scielo.br/j/jbpsiq/a/MjfQxLfddpKdQpKmZMB64DN/abstract/?lang=pt>. Acesso em: 3 de dez. de 2024.

Best JB. The politician: a public health problem? **Med J Aust**, v. 202, n. 8, 4 maio 2015. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/25929513/>. Acesso em: 3 de dez. de 2024.

CARD, Kiffer.; HEPBURN, Kirk. Is neoliberalism killing us? A cross sectional study of the impact of neoliberal beliefs on health and social wellbeing in the midst of the COVID-19 pandemic. **International Journal of Social Determinants of Health and Health Services**, v. 53, n. 3, 2022. Disponível em: <https://journals-sagepub-com.ez38.periodicos.capes.gov.br/doi/10.1177/00207314221134040>. Acesso em: 2 de jul. de 2024.

GÉRON, Aurélien. **Mãos à obra: aprendizado de máquina com Scikit-Learn, Keras & TensorFlow**. 2. ed. Rio de Janeiro: Alta Books, 2021.

HERTZ, N. **O século solitário: como restaurar a conexão humana em um mundo fragmentado**. 1a ed. Rio de Janeiro: Record, 2021.

JOSIOWICZ, Alejandra. humanidades digitais e leitura no Twitter: um placebo sanador em tempos de COVID-19. **Estudos Históricos**, v. 34, n. 73, 2021. Disponível: <https://www.scielo.br/j/eh/a/dhbtntyJQdTFmV8Fvpkwc98t>. Acesso em: 3 de maio de 2025.

LANGENKAMP, Alexander. Linking social deprivation and loneliness to right-extreme radicalization and extremist antifeminism. **Current Opinion in Behavioral Sciences**, v. 63, 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2352154625000440?via%3Dihub#bib31>. Acesso em: 20 de maio de 2025.

LO MORO, Giuseppina et al. Exploring cyberaggression and mental health consequences among adults: An Italian nationwide cross-sectional study. **International journal of environmental research and public health**, v. 20, n. 4. Disponível: <https://www.mdpi.com/1660-4601/20/4/3224>. Acesso em: 5 de maio de 2025.

MEJOVA, Yelena; HOMMADOVA, Anya. Gender in the disclosure of loneliness on Twitter during COVID-19 lockdowns. **Frontiers in Digital Health**, v. 5, 2023. Disponível em: <https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2023.1297983/full>. Acesso em: 4 de maio de 2024.

PETERSON, Delaney et al. Loneliness is positively associated with populist radical right support. **Social Science & Medicine**, v. 366, 2025. Disponível em: <https://www.sciencedirect.com/science/article/pii/S027795362500005X?via%3Dihub#bib47>. Acesso em: 20 de jul. de 2025.

ZHANG, Senqi et al. The COVID-19 pandemic and mental health concerns on Twitter in the United States. **Health data science**, v. 2022, 2022. Disponível em: <https://spj-science.org.ez38.periodicos.capes.gov.br/doi/10.34133/2022/9758408>. Acesso em: 25 de abril de 2025.