

Analyzing Political Discourse on Instagram during the 2024 Municipal Elections in Northeastern Brazilian Capitals

Giovanna Valentina Esteves Xavier Teles de Andrade, Tarcísio Daniel Pontes Lucas

¹Universidade de Pernambuco (UPE), Campus Caruaru
Caruaru, PE – Brazil

`giovanna.andrade@upe.br, tarcisio.lucas@upe.br`

Abstract. *This study examines Instagram posts by mayoral candidates in the capitals of Brazil's Northeast region during the 2024 municipal elections. Using Natural Language Processing (NLP) and Subgroup Mining techniques, we analyzed the texts of candidates leading in the polls (first and second place) to identify linguistic patterns and behavioral trends throughout the campaign. The results reveal striking discursive differences between opposing candidates in each city, as well as across the political spectrum (right-wing and left-wing).*

Resumo. *Este estudo investiga as publicações realizadas no Instagram por candidatos às prefeituras das capitais da região Nordeste do Brasil durante as eleições municipais de 2024. Por meio da aplicação de técnicas de Processamento de Linguagem Natural e Mineração de Subgrupos, foram analisados os textos de candidatos que ocupavam as primeiras e segundas posições nas pesquisas eleitorais, com o objetivo de identificar padrões e comportamentos linguísticos ao longo da campanha. Os resultados revelam diferenças discursivas marcantes entre candidatos oponentes em cada cidade, bem como em relação ao espectro político (direita e esquerda).*

1. Introdução

O Brasil é um dos países que mais consome a rede social Instagram. Segundo dados do *Datareportal.com* de 2023, o Instagram tinha cerca de 114 milhões de usuários ativos no Brasil, o que representa aproximadamente 60% da população brasileira. Esse número posiciona o Brasil como o terceiro maior mercado consumidor de redes sociais no mundo [Pacete 2023].

Com isso, o Instagram no Brasil conquistou importância no cotidiano dos brasileiros, deixando de ser apenas uma rede de interação social para se transformar em um canal de comunicação direta entre marcas, influenciadores e consumidores. Isso revolucionou a mídia brasileira, ampliando a forma como as informações são disseminadas e consumidas.

Nesse cenário, não foi surpresa quando o segmento político passou a utilizar o Instagram como uma ferramenta estratégica. A flexibilidade da plataforma ao permitir a interação direta com o público, a produção de conteúdo visual e a segmentação precisa de audiências tornaram o Instagram um palco relevante para campanhas eleitorais e movimentações políticas. Assim, a adaptação da política à dinâmica das redes sociais reflete uma tendência global, onde o digital exerce cada vez mais influência nas decisões

eleitorais, mudando o modo como candidatos se relacionam com os eleitores e como as campanhas são estruturadas.

Diversos trabalhos têm analisado a comunicação de políticos brasileiros em redes sociais durante campanhas eleitorais, normalmente com maior destaque para eleições presidenciais [Soares and Recuero 2021, Rossini et al. 2023, Venâncio et al. 2024, de Lima-Santos et al. 2024, Barbabela et al. 2024]. Já no contexto municipal existe uma maior carência de trabalhos. Nesse contexto, [Bachini et al. 2022] analisou as campanhas oficiais dos candidatos às prefeituras de todas as capitais brasileiras em 2020 no Facebook. Já [Silva and Santos 2024] buscou identificar discurso de ódio na mesma eleição.

As eleições municipais diferem das presidenciais no Brasil, principalmente devido ao foco local. Enquanto as presidenciais envolvem campanhas nacionalizadas e ideológicas, as municipais priorizam problemas concretos, como saneamento e transporte). O perfil dos candidatos também muda. Nas eleições presidenciais, figuras nacionais e partidos tendem a ter maior peso, enquanto nas municipais, elites locais e alianças pragmáticas tendem a ser mais predominantes.

Dessa forma, faz-se necessário o surgimento de novos trabalhos que busquem analisar o discurso político em eleições municipais. Nesse contexto, este trabalho tem como objetivo analisar o discurso no *Instagram* dos principais candidatos a prefeitura das capitais do Nordeste em 2024 a utilizando Processamento de Linguagem Natural e Mineração de Subgrupos.

2. Fundamentação Teórica

2.1. Processamento de Linguagem Natural (PLN)

PLN é uma área que combina linguística e Inteligência Artificial, sendo fundamental para o desenvolvimento de sistemas capazes de realizar tarefas como tradução automática, análise de sentimentos e resposta a perguntas [Barbosa et al. 2017]. No contexto da extração de conhecimento, o PLN é amplamente utilizado para identificar padrões e extrair informações relevantes de grandes corpora de texto [Pinto 2015]. Tal extração de conhecimento pode ser feita com o auxílio de nuvens de palavras (do inglês *Word Clouds*).

A geração de nuvens de palavras com uso de PLN geralmente segue as seguintes etapas: pré-processamento textual (tokenização, remoção de stopwords, lematização e normalização), cálculo do TF-IDF (para ponderar a relevância das palavras no corpus, destacando termos frequentes em um documento, mas raros nos demais), seleção das palavras mais significativas (com base nos pesos TF-IDF) e, por fim, visualização por meio de uma nuvem de palavras, onde termos com maior peso TF-IDF aparecem em destaque (tamanho ou cor). Essa abordagem permite identificar rapidamente os conceitos-chave de um conjunto de textos, sendo útil para análise exploratória em pesquisas, redes sociais ou mineração de opinião [Mikolov et al. 2013, Devlin et al. 2019].

2.2. Mineração de Subgrupos

A Mineração de Subgrupos (do inglês *Subgroup Discovery*) é uma área da mineração de dados que visa identificar grupos de características relacionadas a presença exagerada

de uma classe [Herrera et al. 2011]. Dada uma base de dados onde cada linha representa um candidato e cada coluna os termos usados por estes nos discursos de campanha. Considerando que sabemos quem ganhou e quem perdeu a eleição, podemos utilizar mineração de subgrupos para identificar o conjunto de termos mais relacionados aos candidatos vencedores. A mineração de subgrupos já foi utilizada para descoberta de conhecimento em diversas áreas, como depressão [de Medeiros Azevedo and Lucas 2022], câncer [Carvalho Neto 2022] e bioinformática [Liu et al. 2015].

As métricas de avaliação desempenham um papel essencial na tarefa de minerar subgrupos, uma vez que define matematicamente a relevância de um subgrupo. Uma das métricas mais utilizadas para avaliar subgrupos ao longo de uma busca é a Acurácia Relativa ao Peso (WRAcc), definida pela fórmula:

$$WRAcc(s) = \frac{TP + FP}{|D|} \left(\frac{TP}{TP + FP} - \frac{|D^+|}{|D|} \right)$$

onde TP (True Positives) são os verdadeiros positivos e FP (False Positives) são os falsos positivos. $|D|$ é o número total de exemplos no conjunto de dados, e $|D^+|$ é o número total de exemplos positivos.

Um dos algoritmos mais usados para mineração de subgrupos em bases de alta dimensionalidade é o SSDP+ (*Simple Search for Discriminative Patterns Plus*), um algoritmo genético adaptativo com mecanismos autoajustáveis de mutação e cruzamento baseados no desempenho dos k melhores subgrupos encontrados ao longo da busca. O algoritmo incorpora ainda um operador de diversificação, garantindo resultados não redundantes [Lucas et al. 2018, Lucas et al. 2017]. O SSDP+ tem como principais parâmetros: métrica de avaliação (usada para avaliar a qualidade de cada subgrupo candidato) e k (número de subgrupos retornado) [Lucas et al. 2018]. Mais informações sobre a área de mineração de subgrupos podem ser encontrada em [Herrera et al. 2011].

3. Metodologia

A metodologia aplicada neste estudo pode ser dividida quatro etapas: (1) coleta de dados, (2) formação de bases de dados; (3) análise dos termos mais relevantes e (4) análise dos grupos de termos mais relevantes.

3.1. Coleta de dados

A coleta de dados foi feita utilizando web scraping. Nós extraímos os comentários das postagens feitas no *Instagram* do primeiro e segundo colocados nas eleições de 2024 das nove capitais nordestinas, totalizando 18 perfis públicos considerados.

Os dados extraídos passaram por uma etapa de pré processamento, passando por: conversão para letras minúsculas; remoção de pontuação, stopwords e termos irrelevantes para análise (URLs, menções (@), emojis, caracteres especiais, nomes dos candidatos, das cidades, frase “propaganda eleitoral”). Em seguida, fizemos a tokenização e vetorização dos textos utilizando TF-IDF.

3.2. Formação de bases de dados

A partir dos dados pré-processados, nós criamos três grupos de bases de dados, onde cada linha representa uma postagem e cada coluna o valor do TF-IDF associado aos termos considerados, foram elas:

1. Por políticos: uma base de dados para cada um dos 18 políticos considerados.
2. Eixo político: uma base para cada eixo político: centro, direita e esquerda.
3. Situação final: uma base com os termos utilizados por quem venceu e outra por quem perdeu as eleições.

Nós criamos ainda outros grupos de base de dados onde cada linha representa um dos 18 político e cada coluna um dos termos considerados por estes no conjunto das postagens. Nós preenchemos com o valor 1 quando o termo era considerado relevante para o político (TF-IDF maior que a mediana) e 0 caso contrário. Em seguida, geramos duas variações desta, uma tendo como classe se o político venceu ou perdeu a eleição e o seu eixo político (direita, esquerda ou centro).

3.3. Análise dos termos mais relevantes

A análise dos termos mais relevantes foi feita a partir do uso de imagens representando núvens de palavras, com o tamanho de cada termo proporcional ao valor do seu TF-IDF. Nós criamos uma núvem de palavras para cada político com o objetivo de comparar os discursos dos vencedores e perdedores em cada capital do Nordeste. Nós geramos núvens de palavras ainda para confrontar, de uma forma geral, os seguintes cenários: vitoriosos e derrotados; e por ideologia (direita, esquerda e centro).

3.4. Análise dos grupos de termos mais relevantes

A análise dos grupos de termos foi feita utilizando mineração de subgrupos buscando identificar os grupos de termos relevantes que melhor diferenciavam os vencedores dos derrotados e o viés de direita e esquerda. Nós utilizamos o algoritmo SSDP+ utilizando Q_g como métrica de avaliação e gerando os top-10 subgrupos, de forma que as características utilizadas nos primeiros cinco subgrupos não se repitam nos demais.

4. Resultados e Discussões

4.1. Candidatos e suas palavras

A análise do conteúdo discursivo dos 18 candidatos às prefeituras e dos seus principais traços de linguagem, assim como seus estilos e estratégias individuais nos mostra que, apesar das convergências, as campanhas revelam diferenças claras de posicionamento estratégico, refletidas na seleção de palavras, estilo de linguagem e ênfase temática.

As diferenças de retórica encontradas refletem não apenas os perfis pessoais dos candidatos, mas também suas estratégias, os públicos que desejam atingir e as narrativas que querem construir: de continuidade, renovação, pertencimento ou ruptura. Para todas as figuras desta seção, no lado esquerdo estão os vencedores e no lado direito os derrotados, como na Figura 1 e Figura 2, respectivamente.

4.1.1. Aracajú

Emília Corrêa (Figura 1) apresenta um discurso bastante personalizado. Termos como “calça”, “blusa”, “tênis” apontam para uma campanha centrada em sua figura e estilo. Já o uso de termos como “sistemão” e “basta” parece indicar uma tentativa de se posicionar como outsider, ou seja, alguém que não pertence aos grupos políticos considerados padrões.

frequente dos temas ”bolsonaro” e ”mudança”. O tema de saúde surgiu com força no seu discurso também, representado pelo uso dos temas ”saúde” e ”médico”.

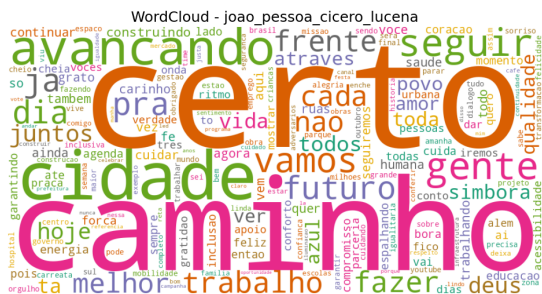


Figura 5. Cícero Lucena.

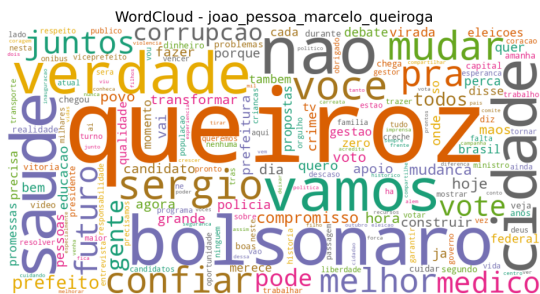


Figura 6. Marcelo Queiroga.

4.1.4. Maceió

No discurso de JHC (Figura 7) os termos ”não” e ”pra” foram usando com muita frequência. No entanto, tais termos isolados são pouco informativos. Já os temas ”ja”, ”vamos”, ”aqui” e ”trabalho” parecem sugerir um posicionamento de quem já atua de forma intensa pela cidade.

Já Rafael Brito (Figura 8) adota um discurso com vocabulário mais coloquial, usando temas como ”levada”, ”vms” e ”pra”, mesclando com termos que tentam passar confiança, como ”serio”, ”verdade”, ”vamos” e ”junto”.



Figura 7. JHC.

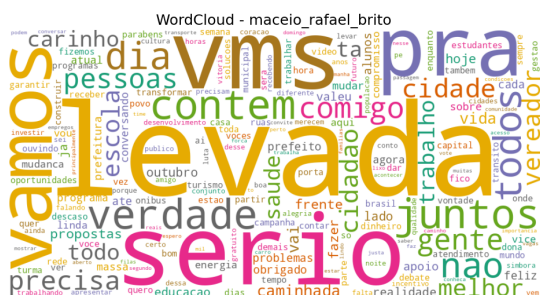
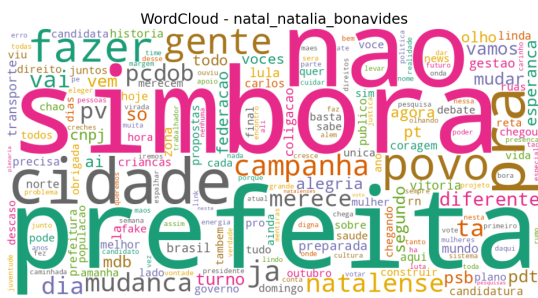


Figura 8. Rafael Brito.

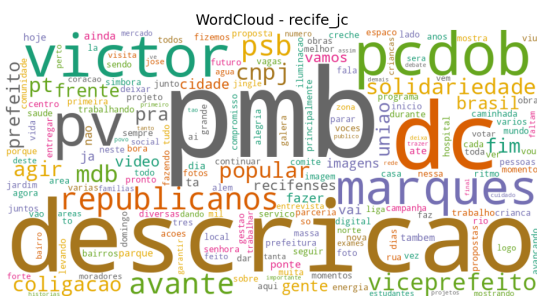
4.1.5. Natal

Paulinho Freire (Figura 9) apresenta uma narrativa que parece querer posicionar o candidato como alguém que já faz pela cidade, através dos termos ”cidade”, ”avancando”, ”frente”, ”dia” e invoca os eleitores a continuarem apostando nele, através do grande destaque dado ao termo ”vamos”. Já o discurso de Natália Bonavides (Figura 10) é marcado por informalidade e identificação com o povo. Termos como ”simbora”, ”gente”, ”povo” e ”prefeita” constroem uma imagem de proximidade e empatia.



4.1.6. Recife

João Campos (Figura 11) foca seu discurso na composição partidária e institucional. A repetição de siglas como “pmb”, “dc”, “pcdob” e “psb” evidencia uma campanha centrada em alianças políticas e suporte eleitoral. Já o candidato GMN (Figura 12) prioriza uma linguagem institucional e patriótica, com termos como “ordem”, “progresso”, “pl” e “coligação”. Também há a presença de palavras como “precisa” e “não”, que podem sugerir uma retórica crítica. O candidato também fez uso significativo do termo “bolsomaro”.



4.1.7. Salvador

Bruno Reis (Figura 13) adota uma abordagem voltada à coligação política. A presença maciça de siglas partidárias mostra foco na estrutura política de apoio. Kleber Rosa (Figura 14) também destacou sua identidade política com forte presença de termos como “psol”, “rede”, “vice” e “federacao”.

4.1.8. São Luís

O discurso de Eduardo Braide (Figura 15) se destaca por uma abordagem técnico-administrativa. Termos como “led”, “asfalto”, “iluminação” e “trabalho” apontam para uma comunicação voltada à entrega de serviços e infraestrutura urbana. Parece haver também uma tentativa de reforçar sua imagem como gestor eficiente, priorizando resultados concretos e melhorias visíveis para a cidade.

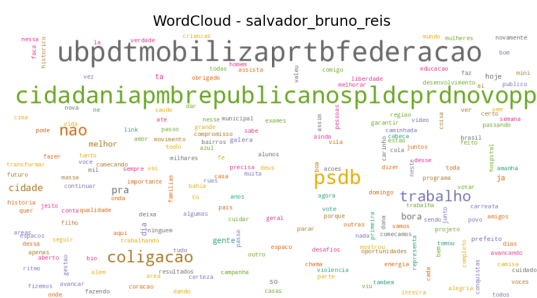


Figura 13. Bruno Reis.

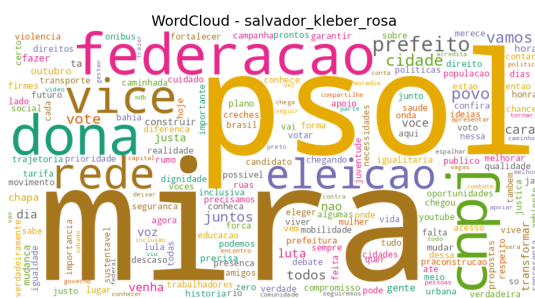


Figura 14. Kleber Rosa.

Já Duarte Jr. (Figura 16) adota um tom propositivo e mobilizador. A presença frequente de termos como “agora”, “vamos”, “juntos” e “resolver” revela uma linguagem voltada à ação imediata e à construção coletiva.



Figura 15. Eduardo Braide.

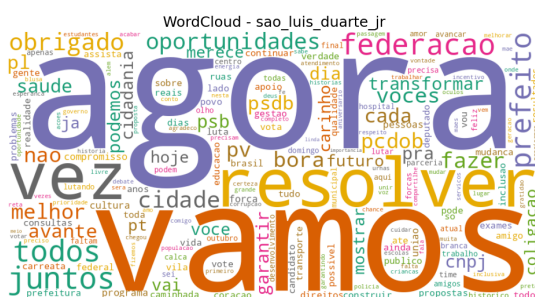


Figura 16. Duarte Jr.

4.1.9. Teresina

Silvio Mendes (Figura 17) enfatiza sua conexão com os teresinenses e utiliza uma linguagem de compromisso e responsabilidade. O uso de palavras como “vote”, “voltar”, “cuidar”, “amor” e “compromisso” reforça sua proposta de retorno à gestão e continuidade do cuidado com a cidade.

Fábio Novo (Figura 18) utiliza uma linguagem positiva e de construção coletiva. Expressões como “vamos”, “juntos”, “fazer” e “futuro” aparecem em seu discurso. Já o forte uso do termo “capital” pode indicar sua visão estratégica da cidade dentro do estado.



Figura 17. Silvio Mendes.



Figura 18. Fábio Novo.

4.2. Candidatos vitoriosos e derrotados

A análise nas postagens revela que tanto os vitoriosos como os derrotados das eleições compartilham muitos termos centrais voltados principalmente à coletividade, ao cotidiano e a mobilização, são eles: "cidade", "vamos", "dia", "juntos" e "não". Isso nos mostra que existe um vocabulário comum, independente do seu resultado, entre os candidatos durante as campanhas. Entretanto, há diferenças relevantes entre os vocabulários específicos de ambos os subgrupos (vitoriosos e derrotados).

O vencedor tende a dar ênfase em capacidade de execução, governabilidade e ações concretas, por meio de palavras como: "trabalho", "coligação", "pode" e "já". Já o derrotado, utiliza-se muito termos como: "gente", "povo", "fazer" e "melhor", que levam a um apelo e uma proximidade emocional e dando foco a promessas de melhoria. Tais termos podem indicar, na realidade, que os vencedores tendem a ser candidatos que já estão no poder e tem diversas obras recentes para explorar, enquanto os derrotados são políticos que tentam apresentar caminhos alternativos através de promessas.



Figura 19. Vencedores.

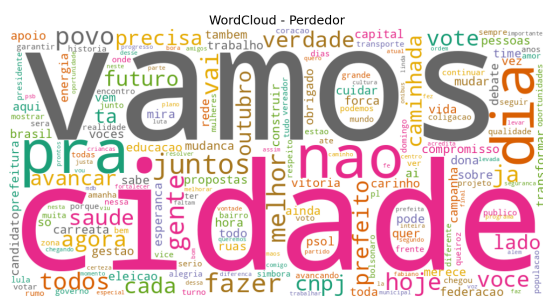


Figura 20. Perdedores

A Tabela 1 apresenta os grupos de termos que juntos melhor diferenciaram os discursos dos vencedores e derrotados, v indica o número de candidatos vitoriosos e d o de derrotados. Note que todos os grupos de termos estão associados exclusivamente a candidatos vencedores ou derrotados. Ao analisar os grupos de termos usados em conjunto pelos candidatos (Tabela 1), fica mais clara as diferenças nas abordagens. O discurso dos vencedores estavam mais associados a coisas que estão sendo feitas, um governo ou projeto em andamento, representada em termos como "durante", "trabalhando", "dando", "continuidade" e "ações". Já os derrotados usaram termos associados a mudança, como "medo", "mudar", "lutar" e "esperança". Ambos trouxeram pautas específicas, como educação, transporte e zeladoria, representadas pelos termos "ensino", "ônibus" e "praças". Os termos apresentados talvez indiquem mais uma vez um enviesamento no sentido que candidatos que estão no poder tendem a ser reeleitos ou formarem um sucessor.

4.3. Candidatos e suas orientações políticas

A análise por orientação política revela distintos padrões discursivos entre os candidatos à prefeitura nas capitais nordestinas. Com base nas núvens de palavras podemos notar que os candidatos de esquerda (Figura 22) apresentaram um vocabulário fortemente marcado por termos coletivos e mobilizadores, como “juntos” e “vamos”, sugerindo um discurso de união popular. Já os candidatos de direita (Figura 21) enfatizaram aspectos relacionados à ação prática, governabilidade e crítica, utilizando termos como “trabalho”, “coligação”

Tabela 2. Top-10 grupos de características associadas aos candidatos de direita e esquerda

Direita				Esquerda			
ID	Descrição	d	e/c	ID	Descrição	e	d/c
31	deus, abandono	6	0	41	arte, professores	7	0
32	deus, pronto	6	0	42	arte, juventude	7	0
33	deus, pouco, vale	6	0	43	arte, moradia	7	0
34	deus, jogo	6	0	44	arte, condicoes	7	0
35	coisa, fe, mae	6	0	45	juventude, plenaria	7	0
36	acao, outros, presente	6	0	46	conversa, time	7	0
37	conto, fica, maior	6	0	47	meio, promover	6	0
38	liberdade, preciso	5	0	48	meio, serio	6	0
39	estao, presente	5	0	49	construir, espalhar	6	0
40	desafios, parar	5	0	50	popular, virar	6	0

comunicação, com a direita recorrendo a valores como "deus" e "fe" e a esquerda a termos como "artes", "professor" e "juventude".

Vale ressaltar que as análises apresentadas podem ser aprofundadas ou mesmo contraposta a medida que este trabalho alcançar especialistas em políticas dos respectivos municípios envolvidos. Trabalhos futuros poderão ampliar este escopo de análise para eleições de outros níveis (como estaduais e federais), bem como para cidades menores do país, permitindo comparações geográficas e contextuais.

Referências

- Bachini, N., Rosa, K. C., Costa, A. L. V., and Silva, R. N. d. F. (2022). Comunicação política no ambiente digital: uma análise das campanhas eleitorais municipais de 2020 no facebook. *Opinião Pública*, 28(3):750–786.
- Barbabela, E., Feres Júnior, J., and Machado, M. (2024). Elections without boundaries: The use of facebook in the 2022 brazilian presidential elections. In *Democracy Under Attack: Social Media and Disinformation in Brazilian Elections*, pages 53–74. Springer.
- Barbosa, J., Vieira, J. P. A., Santos, R., Junior, G. V. M., MUNIZ, M. d. S., and Moura, R. S. (2017). Introdução ao processamento de linguagem natural usando python. *III Escola Regional de Informatica do Piauí*, 1:336–360.
- Carvalho Neto, G. d. V. (2022). Mineração de subgrupos aplicada a dados de metilação de pacientes com câncer de mama. Master's thesis, Universidade Federal de Pernambuco.
- de Lima-Santos, M.-F., Gonçalves, I., Quiles, M. G., Mesquita, L., Ceron, W., and Lorena, M. C. C. (2024). Visual political communication on instagram: a comparative study of brazilian presidential elections. *EPJ Data Science*, 13(1):72.
- de Medeiros Azevedo, F. C. and Lucas, T. D. P. (2022). Depressão entre jovens brasileiros: uma investigação baseada em mineração de subgrupos. *Research, Society and Development*, 11(1):e10511124547–e10511124547.

- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.
- Herrera, F. et al. (2011). An overview on subgroup discovery: foundations and applications. *Knowledge and information systems*, 29:495–525.
- Liu, X., Wu, J., Gu, F., Wang, J., and He, Z. (2015). Discriminative pattern mining and its applications in bioinformatics. *Briefings in bioinformatics*, 16(5):884–900.
- Lucas, T., Silva, T. C., Vimieiro, R., and Ludermit, T. B. (2017). A new evolutionary algorithm for mining top-k discriminative patterns in high dimensional data. *Applied Soft Computing*, 59:487–499.
- Lucas, T., Vimieiro, R., and Ludermit, T. (2018). Ssdp+: A diverse and more informative subgroup discovery approach for high dimensional data. In *2018 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8. IEEE.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Pacete, L. G. (2023). Brasil é o terceiro maior consumidor de redes sociais em todo o mundo. <https://forbes.com.br/forbes-tech/2023/03/brasil-e-o-terceiro-pais-que-mais-consome-redes-sociais-em-todo-o-mundo/Site>. Acessado em 06 de junho de 2025.
- Pinto, S. C. S. (2015). Processamento de linguagem natural e extração de conhecimento. Master’s thesis, Universidade de Coimbra (Portugal).
- Rossini, P., Mont’Alverne, C., and Kalogeropoulos, A. (2023). Explaining beliefs in electoral misinformation in the 2022 brazilian election: The role of ideology, political trust, social media, and messaging apps. *Harvard Kennedy School Misinformation Review*, 4(3).
- Silva, B. P. and Santos, F. (2024). Identificação de discurso de Ódio nas eleições municipais de 2020. *Boletim de Conjuntura (BOCA)*, 19(55):227–247.
- Soares, F. B. and Recuero, R. (2021). Hashtag wars: Political disinformation and discursive struggles on twitter conversations during the 2018 brazilian presidential campaign. *Social Media+ Society*, 7(2):20563051211009073.
- Venâncio, O. R., Ferreira, C. H., Almeida, J. M., and da Silva, A. P. C. (2024). Unraveling user coordination on telegram: A comprehensive analysis of political mobilization during the 2022 brazilian presidential election. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 1545–1556.