



## **Notícias para monitoramento de Políticas Públicas: análise descritiva sobre o discurso midiático nos 20 anos do Programa Bolsa Família**

**Autores/as:** Angélica Barbosa Neres Santana (INCT Combate à Fome, Universidade de São Paulo, USP; angelicabnsantana@gmail.com); Jorge Louçã (INCT Combate à Fome, Instituto Universitário Lisboa, ISCTE; jorge.l@iscte-iul.pt); Eric Tokuda (INCT Combate à Fome, Universidade de São Paulo, USP; tokudaek@usp.br); Alexandre Cláudio Botazzo Delbem (INCT Combate à Fome, Universidade de São Paulo, USP; acbd@icmc.usp.br); Flavia Mori Sarti (INCT Combate à Fome, Universidade de São Paulo, USP; flamori@usp.br).

### **Resumo:**

#### *Introdução*

A pandemia e as crises econômicas recentes representam substanciais desafios impostos ao alcance do compromisso de erradicação da fome, insegurança alimentar e má nutrição, estabelecido na Agenda 2030 para Desenvolvimento Sustentável (FAO, 2022). O custo de uma alimentação saudável tem aumentado ao longo dos últimos anos, resultando em incremento da ocorrência de insegurança alimentar grave em vários países, inclusive no Brasil (FAO, 2023; REDE PENSSAN, 2022; IBGE, 2024). Diversas ações têm sido empreendidas no país para enfrentamento do quadro de insegurança alimentar, incluindo-se transferência de renda no contexto do Programa Bolsa Família (PBF), implementado desde meados da década de 2000. Um dos principais objetivos do PBF reside na promoção da segurança alimentar e nutricional entre famílias vulneráveis, expostas a situações de pobreza e pobreza extrema (BELIK, 2012; BRASIL, 2010; CARVALHO et al., 2020).

Entretanto, mudanças institucionais ocorridas desde final da década de 2010 comprometeram o resultado dos esforços de construção de sistema de proteção social baseado em políticas nacionais de promoção da alimentação saudável e da segurança alimentar e nutricional no Brasil (SARTI et al., 2017; COSTA et al., 2023). Conseqüentemente, o país regressou ao Mapa da Fome nos últimos anos (IBGE, 2023), destacando-se o incremento da insegurança alimentar grave concomitantemente à persistência da epidemia de obesidade como desafio global (WHO, 2021). O PBF foi sujeito a mudanças no período da pandemia, incluindo-



se alteração do nome para Auxílio Brasil em 2021 e retorno à designação original em 2023 (COSTA et al., 2023). O movimento pela reconstrução do sistema nacional de políticas públicas direcionadas ao combate à fome tem sido elemento fundamental no restabelecimento dos compromissos social, político e econômico para concretização de mudanças estruturais nos padrões de alimentação da população, visando a promoção de estilos de vida saudáveis (SARTI et al., 2017).

Para alcançar esses objetivos, é fundamental utilizar a ciência social computacional, que tem o potencial de transformar a maneira como as políticas públicas são desenvolvidas e implementadas. Esta abordagem integra métodos computacionais avançados, permitindo a interação entre diferentes áreas do conhecimento e a realização de análises mais profundas e abrangentes. Com isso, torna-se possível detectar comportamentos indesejados ou tendências na sociedade, como a disseminação de fake news, informar respostas políticas adequadas e monitorar padrões socioeconômicos, como os observados durante a pandemia de COVID-19, a fim de identificar o momento oportuno para intervenções necessárias (BERTONI et al., 2022; MONTI et al., 2022; OLIVEIRA et al., 2022)

### *Objetivo*

Diante dos desafios à implementação de ações para promoção da segurança alimentar e nutricional no Brasil, torna-se fundamental adotar estratégias eficazes para comunicação de resultados de monitoramento e avaliação de políticas públicas à população. Este estudo tem como objetivo identificar e descrever elementos do discurso midiático publicado na imprensa brasileira sobre o Programa Bolsa Família, a partir das notícias expressas nas mídias físicas e websites no período de 2003 a 2023. O propósito é avaliar a viabilidade do uso das notícias como ferramenta para o monitoramento das políticas públicas de segurança alimentar e nutricional.

### *Métodos*

O estudo é parte da pesquisa no contexto do Instituto Nacional de Ciência e Tecnologia (INCT): “*Combate à Fome: Estratégias e Políticas Públicas para a Realização do Direito Humano à Alimentação Adequada - Abordagem Transdisciplinar de Sistemas Alimentares com*



*Apoio de Inteligência Artificial*”<sup>1</sup>, que conta com desenvolvimento do trabalho interdisciplinar entres os eixos de políticas públicas e inteligência artificial.

A identificação e extração de publicações em notícias foram realizadas a partir de levantamento baseado em monitoramento de mídias, em parceria com o Instituto Fábrica de Ideias, que realiza clipagem de notícias a mais de 20 anos. O escopo da extração de dados foi baseado na busca de manchetes nacionais contendo palavras-chave “Bolsa Família” ou “Auxílio Brasil” no período de 2003 a 2023.

O processamento dos dados, especialmente a remoção de duplicatas, desempenha um papel essencial nesse processo. A duplicação de dados pode levar a resultados enganosos, superestimação de tendências e conclusões errôneas. Portanto, é essencial implementar etapas rigorosas de engenharia de dados, como a remoção de observações com formatos inesperados, eliminação de caracteres desnecessários, e remoção de textos e títulos idênticos ou semelhantes, considerando critérios como comprimento e temporalidade. Essas medidas garantem que o conjunto de dados utilizado para análises seja preciso, consistente e livre de redundâncias.

Dados tabulares caracterizam-se por apresentar um formato padrão para cada um dos campos. Quando uma observação apresenta um dado diferente do padrão, frequentemente adota-se um marcador para este valor anômalo, como *N/A*. Estes marcadores normalmente surgem em tabelas devido a resultados numéricos não-representáveis, como divisão por zero ou a raiz quadrada de um número negativo, ou através da marcação intencional de dados ausentes. A presença deste tipo de valores pode prejudicar análises subsequentes, pois muitos algoritmos de processamento de dados não podem lidar diretamente com eles. Portanto, a primeira etapa consiste em identificar e remover todas as linhas ou colunas que contenham estes valores anômalos. Isso é feito para garantir que o conjunto de dados estejam completo e consistente antes de prosseguir para etapas subsequentes de processamento e análise.

Valores ausentes podem ser representados de outras maneiras, como espaços em branco, valores nulos específicos para cada conjunto de dados, ou marcações distintas como

---

<sup>1</sup> <https://www.fsp.usp.br/inct-combate-a-fome/>



“NAN”, “*Não Informado*”. Para garantir que os dados sejam uniformes e confiáveis, as observações com campos contendo este tipo de valores foram descartados.

Em conjuntos de dados textuais, especialmente aqueles que envolvem textos coletados da web, os dados podem conter caracteres inesperados, como por exemplo caracteres de controle e de linguagem de marcação que são irrelevantes para análise textual e podem interferir nas análises posteriores. Estes caracteres foram sistematicamente removidos dos dados textuais.

Em seguida, observações com textos da matéria idênticos foram identificados e removidos. Apenas informações únicas foram mantidas no banco de dados. Além do texto, o título também foi utilizado como critério para determinação de repetições. Observações contendo títulos iguais e com um comprimento mínimo especificado também foram removidos. Este comprimento foi baseado no número de 25 caracteres. A estabelecimento de um tamanho mínimo dos títulos é uma medida conservadora que visa evitar a remoção de notícias diferentes *com títulos curtos iguais*, como “Notícias da semana”.

Dos dados restantes, a remoção de textos semelhantes foi realizada usando uma abordagem baseada em frequência de palavras (*Term Frequency-Inverse Document Frequency*) (SPARCK JONES, 1972) para medir a importância de cada palavra em relação aos documentos. Além disso, foram considerados critérios como a similaridade de comprimento e a diferença temporal de pelo menos três meses entre textos semelhantes. Essa etapa foi fundamental para eliminar repetições que não são idênticas, mas contêm informações semelhantes o suficiente para serem consideradas redundantes.

Estes critérios foram adotados de modo a permitir que os dados repetidos pudessem ser identificados e removidos de modo automático. Convém avaliar a eficácia do processamento adotado, e para isso foi realizada uma verificação manual de uma amostra dos dados. Todas as notícias repetidas foram identificadas manualmente na amostra. Notícias repetidas foram determinadas baseadas na comparação *entres os textos das notícias*. Considerando a mesma amostra de dados, o procedimento de identificação de repetições automático previamente proposto foi aplicado, de modo a permitir a comparação entre o resultado da abordagem automática e o resultado esperado, a partir da inspeção manual. O nível de correspondência



entre os dois resultados é um indicativo da efetividade dos critérios adotados com relação à similaridade textual e à espaço de tempo entre notícias repetidas.

### *Resultados/Discussão*

A cobertura do levantamento de informações incluiu veículos de mídia impressa e digital nacionais, totalizando 497 diferentes veículos em quatro categorias:

- Jornais nacionais (impressos com fac-símile);
- Revistas nacionais (impressos com fac-símile);
- Principais meios de comunicação online;
- Sites de notícias regionais (26 estados e Distrito Federal).

A base de dados original resultou em 447.574 notícias publicadas entre dezembro de 2002 e dezembro de 2023, contendo características quanto ao tipo de veículo de mídia, local de publicação (Unidade da Federação ou Brasil, dependendo da característica de distribuição do veículo de mídia), data de publicação, título e conteúdo das notícias. Após o processamento de dados, o número de notícias legíveis resultou em 273.040, representando 61% do total de discursos midiáticos publicado sobre PBF ao longo dos 20 anos.

Os procedimentos descritos acima foram adotados para limpeza da base de dados. A efetividade deste processo de limpeza foi avaliada a partir de análise manual. Inicialmente, uma amostra de 200 observações em um período de seis meses, entre janeiro e junho de 2009 foi aleatoriamente escolhida. Foi realizada a comparação entre todas as combinações de pares de notícias desta amostra. A verificação manual levou a 38 notícias repetidas, das quais 16 idênticas. A identificação automática, por sua vez, levou a 30 notícias repetidas, sendo todas estas repetições foram também identificadas como repetições pela verificação manual. Desse modo, verificamos a remoção automática a partir dos critérios adotados corresponde a uma correta remoção de 30 das 38 notícias repetidas, correspondendo a 79% de dados repetidos. Na verificação manual também se observou que todas as repetições ocorreram em no máximo 5 dias e o período máximo de 3 meses adotado cobre a totalidade das repetições neste caso.

Entre os tipos de mídias e meios de comunicação onde os dados foram localizados e extraídos, 83% são originados da web, 14% de jornais, 2% de blogs e 1% de revistas. Após o processamento dos dados, observamos que a maior quantidade de repetições foi encontrada nos



meios web, representando 42% dos dados duplicados. A repetição de notícias na web aumenta a percepção de veracidade, mesmo para informações falsas ou implausíveis, ressaltando a necessidade de identificar e excluir duplicidades para o monitoramento eficaz de políticas públicas (FAZIO et al., 2019).

A frequência de notícias disponíveis apresentou um aumento progressivo ao longo dos anos, passando de 451 em 2003 para 24.221 em 2023, com o maior número de publicações, 47.106, ocorrendo em 2022, o que representa 17% de todos os dados. O mesmo comportamento foi observado com as repetições de notícias. Após o processamento para remoção de duplicatas, verificou-se uma diminuição média de 35% dos dados em cada ano, atingindo um pico de aproximadamente 70% em 2012. Este aumento está em linha com as mudanças na produção de notícias digitais, amplificação na disseminação de informações através das mídias digitais (LAZER, et al., 2018). Além disso, é importante destacar que, entre 2018 e 2023, o Programa Bolsa Família sofreu alterações de nome, resultando em grandes discussões na mídia (SORDI, 2023).

A maioria dos dados são representados por notícias veiculadas em mídias nacionais, correspondendo a 45% do total. Ao analisar as publicações por estado e Distrito Federal (DF), observa-se um maior número delas em São Paulo (5,7%), Pernambuco (4,9%), Piauí (4,1%), e de forma igualitária no Paraná e Rio de Janeiro (3,8% cada). Entretanto, ao categorizar por regiões, o Nordeste se sobressai com 36% das notícias identificadas por região, indicando uma cobertura midiática local destacada. Essa tendência é congruente com aspecto da fome, pois há impacto predominante do Programa Bolsa Família na região nordestina, particularmente em aspectos relacionados à melhora na alimentação (SPERANDIO, et al., 2017).

### *Considerações Finais*

Os resultados apresentados destacam a eficácia do uso de notícias como uma ferramenta para o monitoramento e avaliação de políticas públicas, especialmente quando seguidos de processos, como exemplificado, de extração e processamento de dados. Este estudo é o ponto inicial na exploração desses resultados, utilizando técnicas avançadas de processamento de linguagem natural, como o LDA (*Latent Dirichlet Allocation*) (BLEI et al., 2003), para análises descritivas mais profundas. O objetivo futuro é identificar estruturas temáticas através dos



tópicos emergentes nas notícias, além de realizar análises de sentimentos para identificar opinião pública positiva, neutra ou negativa em relação ao Programa Bolsa Família. Essas abordagens têm o potencial de servir como indicadores robustos para um monitoramento contínuo das políticas públicas voltadas a segurança alimentar e nutricional, contribuindo para uma gestão mais eficaz e adaptável às necessidades emergentes da população beneficiária.

### **Referência bibliográfica:**

- BELIK W. A política brasileira de segurança alimentar e nutricional: concepção e resultados. *Segur Aliment Nutr* 2012; 19(2):94-110.
- BERTONI E, FONTANA M, GABRIELLI L, SIGNORELLI S, VESPE M. *Handbook of Computational Social Science for Policy*. 1st ed. Ispra, Italy:Springer; 2022. <https://doi.org/10.1007/978-3-031-16624-2>
- BIRD S, KLEIN E, LOPER E. *Natural language processing with Python - analyzing text with the natural language toolkit*. Disponível em: <https://www.nltk.org/book/> [Acesso 17 jun. 2024].
- BLEI DM, NG AY, JORDAN MI. Latent Dirichlet Allocation. *Journal of machine Learning research*. 2003;3(Jan):993-1022.
- BRASIL. *Fome Zero: uma história brasileira*. Ministério do Desenvolvimento Social e Combate à Fome. vol. I, Brasília, 2010.h
- CARVALHO TFB, SÁ TS, RUAS JPP, VIEIRA MA, SAMPAIO CA. Estado nutricional e segurança alimentar de famílias beneficiadas pelo Programa Bolsa Família: revisão integrativa. *Rev Fun Care Online* 2020; 12:593-602.
- COSTA DM, MAGALHÃES R, CARDOSO MLM. Do Bolsa Família ao Auxílio Brasil: desafios e alcances a partir de uma pesquisa avaliativa baseada na teoria do programa. *Cad Saúde Pú* 2023; 39(7):e00207922.
- DIAZ AKR, LIMA AP, SILVA AM, COSTA FHS, PAGNOSSIM JLM, PERES SM. Relatório técnico: uma análise comparativa das ferramentas de pré-processamento de dados textuais: NLTK, PreText e R. PPgSI-001/2018. Disponível em: [http://ppgsi.each.usp.br/arquivos/RelTec/PPgSI-001\\_2018.pdf](http://ppgsi.each.usp.br/arquivos/RelTec/PPgSI-001_2018.pdf) [Acesso em 20 jun. 2024].



- FAZIO, Lisa K.; BRASHIER, Nadia M.; ELISEEV, Nadya; MARSH, Elizabeth J. Repetition increases perceived truth even for known falsehoods. *Collabra: Psychology*, v. 5, n. 1, p. 15-22, 2019. Disponível em: <https://online.ucpress.edu/collabra/article/doi/10.1525/collabra.12345/12456/Repetition-Increases-Perceived-Truth-Even-for>. [Acesso em: 13 jul. 2024.]
- FONSECA A, BANDYOPADHYAY S, LOUÇÃ J, MANJALY J. Caste in the news: a computational analysis of Indian newspapers. *Social Media + Society* 2019. 5(4).
- FOOD AND AGRICULTURE ORGANIZATION (FAO). The state of food security and nutrition in the world 2022. Rome: FAO, 2022.
- FOOD AND AGRICULTURE ORGANIZATION (FAO). The state of food security and nutrition in the world 2023. Rome: FAO, 2023.
- HADI M, TASHI Q, QURESHI R, SHAH A, MUNEEER A, IRFAN M, et al. Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *TechRxiv*, 2023.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Inquérito Nacional sobre Insegurança Alimentar no Brasil 2022. Rio de Janeiro: IBGE, 2023.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Pesquisa Nacional por Amostra de Domicílios: síntese de indicadores, 2024. Rio de Janeiro: IBGE, 2024.
- JURAFSKY D, MARTIN JH. *Speech and language processing: an introduction to natural language processing, speech recognition, and computational linguistics*. 3rd. ed., 2023. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/> [Acesso em 17 jun. 2024].
- LAZER, D. et al. The science of fake news. *Science*, v. 359, n. 6380, p. 1094-1096, 2018.
- MONTI C, AIELLO LM, DE FRANCISCI MORALES G et al. The language of opinion change on social media under the lens of communicative action. *Sci Rep* 2022, 12:17920.
- OLIVEIRA AD, CASEMIRO JP, BRANDÃO AL, PINTO AM. Monitoramento e avaliação da segurança alimentar e nutricional: um olhar sobre as publicações oficiais. *Cien Saúde Colet* 2022, 27(2):631-640.
- QI K, OWUSU EK, SIU MF, CHAN PCA. A systematic review of construction labor productivity studies: clustering and analysis through hierarchical latent dirichlet allocation. *Ain Shams Engineering Journal* 2024. In press. doi: 10.1016/j.asej.2024.102896



REDE BRASILEIRA DE PESQUISA EM SOBERANIA E SEGURANÇA ALIMENTAR E NUTRICIONAL (REDE PENSSAN). II VIGISAN - Inquérito Nacional sobre Insegurança Alimentar no Contexto da Pandemia da COVID-19 no Brasil. São Paulo: Fundação Friedrich Ebert, Rede PENSSAN, 2022.

SARTI FM, HADDAD MR, SANTANA ABC. Histórico das políticas públicas de alimentação e nutrição em saúde no Brasil. In: Nutrição e saúde pública: produção e consumo de alimentos. Barueri: Manole; 2017.

SORDI, D. Empobrecimento, fome e pandemia: o Auxílio Emergencial, o fim do Programa Bolsa Família e o Auxílio Brasil, 2019-2022. História, Ciências, Saúde-Manguinhos, v. 30, n. suppl 1, 2023. DOI: 10.1590/S0104-59702023000100032.

SPARCK JONES K. A statistical interpretation of term specificity and its application in retrieval. Journal of documentation, 1972. 28(1), p.11-21.

SPERANDIO, N., PRIORE, SE, FRANCESCHINI, SCC, RODRIGUES, CT. Impacto do Programa Bolsa Família no consumo de alimentos: estudo comparativo das regiões Sudeste e Nordeste. Cien. Saúde Colet., v. 22, n. 6, 2017.

WORLD HEALTH ORGANIZATION (WHO). Overweight and obesity. Geneva: WHO, 2021. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> [Acesso em 17 jun. 2024].

**Palavras-chave:** Segurança alimentar e nutricional; Bolsa Família; políticas públicas.

**Fonte de financiamento/apoio:** O presente estudo foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processo 406774/2022-6) e pela Reitoria da Universidade de São Paulo. Agradecimentos ao Centro de Inteligência Artificial (C4A-IUSP), ao apoio da Fundação de Amparo à Pesquisa de São Paulo (#2019/07665-4, #2020/16578-5), à IBM Corporation e ao CEPID-CeMEAI ICMC-USP (#2013/07375-0).

**Conflito de interesses:** Não há conflito de interesse a declarar.