

ANÁLISE DO ALCANCE DAS FAKE NEWS EM SAÚDE

João Victor Lima Teixeira¹
Larissa Deadame De Figueiredo Nicolete²

RESUMO

A pandemia causada pelo Covid-19 imergiu o mundo em um contexto de incertezas, preenchido por medo, necessidade de novas informações sobre o vírus, suas manifestações clínicas, transmissão, prevenção e tratamento. Tais dúvidas foram alocadas, cronologicamente, quando, globalmente, havia um processo de isolamento social em que a exposição aos meios digitais cresceu abruptamente. Assim, com o uso de mídias sociais para informação sobre o curso da doença, a disseminação de *Fake News* se fez um risco exponencial a saúde das pessoas. Dessa forma, o estudo objetivou coletar e analisar as principais *Fake News* na área da saúde, desmentidas por agências de *fact-checking*, estimando seu alcance nas redes. Para tal utilizou-se ferramentas de raspagem de dados no site da Agência Lupa o qual, por ser uma agência de checagem de fatos possui grande biblioteca de notícias conhecidamente falsas, tais amostras foram processadas de acordo com a metodologia *Knowledge Discovery in Databases* - KDD em base estatística de linguagem R adaptada com lexicometria no software Iramuteq. Ao fim do processo extraiu-se 348 notícias, que após tratamento foram compiladas, mostrando alguns termos de maior sensibilidade no sentido da identificação do perfil de *Fake News*, tais como "vacina", "hidroxicloroquina" e "tratamento". Desse modo, observou-se que temas com teor conspiracionista e negacionista tiveram prevalência elevada no presente estudo. Destaca-se que houve clara relação entre política e saúde, de forma que os presentes resultados são instigantes para novas pesquisas sobre o tema.

Palavras-chave: Fake News; Saúde; Fact-checking.

UNILAB, CEARÁ, Discente, 2016infor20@gmail.com¹

UNILAB, CEARÁ, Docente, larissanicolete@unilab.edu.br²

INTRODUÇÃO

Em meados de 1981, com o início da computação pessoal, que nos anos seguintes ganhou exponencial popularidade em vista dos computadores e das novas formas de contato digital (*Orkut e Facebook*), houve uma preocupação sobre as consequências da exposição constante a essas tecnologias. Isso acarretou em um crescimento de pesquisas que buscavam entender as mudanças decorrentes do consumo constante de grande volume de informações na vida dos seres humanos modernos.

Atualmente, com a ascensão das redes sociais, o consumo de informação diária cresceu em níveis alarmantes de forma que, segundo o *Global Digital Report*, em 2021, o brasileiro passou cerca de 10 horas e 08 minutos conectado na internet por dia.

Quanto a construção desse volume de informações, é interessante observar que a antiga teoria base do *Big Data* levava em consideração apenas três principais pontos: O volume, a variedade e a velocidade com que são criados e processados os dados. Cada um desses pontos com seu papel no planejamento adequado de uma rede de servidores, variando seu calibre de tráfego, capacidade de armazenamento e poder computacional. No entanto, ao decorrer do tempo, surgiu um fenômeno intitulado de *Fake News* que Ferreira (2018) define como “textos que são intencionalmente falsos, criados com objetivo de enganar o leitor”, elencando a necessidade de atentar-se ao valor e à veracidade dos dados.

Nesse cenário, identificar notícias falsas e de baixo valor vem sendo um dos grandes desafios devido ao forte volume de dados gerados por segundo no mundo, brecha que sustenta máquinas de propagação de *Fake News*; ferramentas de elevado poder de modular uma sociedade moderna em seus pensamentos, valores e até incitar atos intelectualmente incoerentes e perigosos.

A exemplo destes atos, observa-se que com a pandemia causada pelo Covid-19 criou-se um contexto de incertezas que fora preenchido por medo, necessidade de novas informações sobre o vírus, suas manifestações clínicas, transmissão, prevenção e tratamento. O que se traduziu em um ecossistema social perfeito para o uso de mídias sociais no intuito de proporcionar desinformação sobre o curso da doença e a disseminação de *Fake News*, a fim de afetar a autonomia e capacidade de decisão dos cidadãos como seu núcleo moral.

OBJETIVOS

Geral: Analisar se as principais Fake News na área da saúde, desmentidas por agências de fact-checking, fizeram parte de cascatas de informação e qual foi o alcance nas principais redes sociais para realizar a análise diferencial do conjunto de dados.

Específicos:

- I. Através de mecanismos de busca estabelecer se é possível encontrar a Fake News ou se ela já foi retirada do ar;
- II. Averiguar quais usuários públicos compartilharam a notícia e seu alcance nas mídias através de visualizações, comentários e compartilhamentos;
- III. Denunciar uma Fake News que ainda esteja circulando em duas ou mais mídias sociais do nosso estudo e analisar o tempo de resposta para retirada do conteúdo;
- IV. Elaborar uma tabela com todas as Fake News na área da saúde que foram publicadas em agências de fact-checking participantes do Projeto Comprova, com assuntos relevantes tais como: tratamento precoce, vacina e uso de máscara;

METODOLOGIA

Nos últimos anos alguns estudos vêm buscando estratégias eficazes para detecção e confrontação de notícias

falsas, MONTEIRO, Rafael A. et al (2018) coletou 7.200 notícias, sendo que 50% eram conhecidamente *Fake News*, e de forma manual preencheram e alinharam um Banco de Dados (B.D) para usar na aprendizagem de máquina (Machine Learning) e desenvolver um programa capaz de buscar e detectar automaticamente estes dados inverídicos.

Fernandes e colaboradores (2020) se utilizaram das vantagens do *web crawler* para realizar coleta massiva de dados de agências *fake check*. Já GUPTA, Sandeep Kumar et al (2020) desenvolveram em linguagem *Python*, um vetorizador de contagem (*Count Vectorizer Web*).

Assim, o presente projeto vem trabalhando no sentido de cooperar no processo de coletar e analisar o alcance da propagação de rumores sobre saúde presentes nas agências de *fact-checking* de língua portuguesa para investigar a difusão diferencial de notícias verdadeiras, falsas e mistas (parcialmente verdadeiras, parcialmente falsas) usando um conjunto de dados abrangente de todas as cascatas de informações verificadas que se espalham nas redes sociais.

Inicialmente elaborou-se um fluxo para análise de BD, contabilizando manualmente todas as *Fake News* na área da saúde publicadas em agências de *fact-checking* que participam do Projeto Comprova. A fim de formar um conjunto de dados tabulados, houve a organização em subtópicos de pesquisa dentro da área da saúde (saúde, covid, políticas públicas, economia e outros).

Após isso, através de cursos online e programas de código aberto, iniciou-se um processo de capacitação de ciclo básico na linguagem de programação *Python*, envolvendo os principais temas como *Command Lines*, instrução em laços de repetição, utilização do *Scrapy*, *Natural Language Processing*, *Machine Learning*, noções básicas de *big data*, formas de baixar metadados e processamento neural, conhecimentos de base para progressão do projeto.

O foco foi, primeiramente, em apenas uma rede social, o *Facebook*, iniciando assim a construção das primeiras linhas de código do *web crawler*. Contudo, devido às limitações técnicas adaptou-se o endereçamento da coleta de dados para o site da Agência Lupa, visto que sua estrutura de metadados é mais intuitiva para a técnica de raspagem de dados.

Foi utilizado processo lógico organizacional conhecido como KDD, *Knowledge Discovery in Databases* (Descoberta de Conhecimento nas Bases de Dados), proposto por Fayyad et al. (1996). Um conjunto de ordens sequenciais que podem ser aplicadas e avaliadas a fim de trazer resultados concisos no que tange a geração de conhecimento a partir de um grande volume de informação, semelhante ao processamento de *Big Data*. As cinco etapas que compõem o KDD são: extração de dados, tratamento de dados, padronização de dados, mineração de dados e avaliação da informação.

Dessa forma o primeiro passo para essa metodologia foi coletar, extrair os dados através de *web scrapy* no site oficial da Agência Lupa, esse conjunto de metadados foi comprimido em formato *.csv* que posteriormente fora transformado em *.txt*, formato ideal para fase do tratamento de dados que, por sua vez, utilizou ferramentas de código aberto para retirada de qualquer caractere especial, além da intervenção manual para retirada de palavras que não são de interesse como “e”, “são”, “sim”, “não”, dentre outras. Também foram padronizadas palavras que se referiam a mesma coisa sob forma genérica para melhor processamento dos dados.

Para a fase de mineração de informações no corpus textual formado pelas 348 amostras, se fez necessário o software Iramueteq que realizar análises voltadas à geração de contagem de frequência de termos gerando um grafo de similitude que provê informações sobre o relacionamento semântico entre os termos, possibilitando assim verificar o fluxo da rede de conexões entre as temáticas abordadas no projeto.

RESULTADOS E DISCUSSÃO

No primeiro contato percebeu-se que existem barreiras estratégicas impostas por plataformas da empresa *META, Platforms, Inc* que dificultam as análises estatísticas em suas bases de dados, seja na criptografia de suas tags ou até mesmo no desligamento de *Bots* de busca massiva de dados.

Quanto a análise manual, foi possível observar pico de ocorrências notificadas no ano de 2021, fato que foi incentivado por movimentos negacionistas de ignorância estratégica, termo que McGoey (2012) define como “a habilidade de explorar o desconhecimento para ganhar mais poder”, um instrumento hoje internacionalizado que vem sendo usado por governos e instituições negacionistas que desacreditam na ciência e suas afirmações.

Ainda na fase de coleta de termos, decorrida nos primeiros meses do projeto, foi possível notar que as palavras “hidroxicloroquina” e “cloroquina fizeram-se bastante presentes nas notícias falsas, ato de incentivo à automedicação que culmina em elevado risco a vida de quem os segue. Isto comprova que mesmo sendo um banco construído em 2022, temas desmentidos diversas vezes pelas agências de *fact-Checking*, ainda podem ser garimpados por mecanismos de buscas direcionados, fator importante em futuros estudos retrospectivos. Notando essa tendência, realizou-se um mapeamento dos discursos de *Fake News* semelhante aos exemplos como CASTILHO, Isabela Ribeiro (2020), Vosoughi Roy e Aral (2018), Degenner e Forsé (1999): Wasserman e Faust (1994), com aplicação de softwares para análises textuais como *Iramuteq* e *Gephi 13*.

Seguindo a linha de fases do método KDD, alimentamos o software *Iramuteq* com o corpus textual, iniciando uma análise conhecida como nuvem de palavras, análise essa que mostra em escala de relevância gráfica a frequência dos termos encontrados no corpus textual. Percebeu-se alta frequência de palavras como “rede social”, “covid”, “vacina”, “hidroxicloroquina” e “tratamento”.

Para preservar a isenção do trabalho, alguns nomes obtidos foram suprimidos pela palavra “Governo” e também o ex-governadores, sendo identificados apenas como “Governador”, contudo devido às novas criptografias das principais redes sociais, que protegem os envios de dados pelo usuário, não foi possível determinar se agentes públicos apenas são referenciados nas *Fake News*, ou se eles também divulgaram notícias falsas como usuários das redes sociais.

Ao analisar a perigosa tendência no perfil geral das *Fake News* observa-se que se tornou comum associar saúde com política, especialmente nas redes sociais, conseqüentemente distorcendo informações terapêuticas e corroborando com a elevação da pseudociência e o declínio da importância dos tratamentos baseados em evidência, fato que nos impulsionou a prosseguir com uma análise de similitude.

O grafo de similitude baseia-se na frequência e relações semânticas entre os dados que compõem as sentenças do banco de dados que fora previamente preparado para essa análise, estabelecendo assim uma rede de palavras que segue os mesmos preceitos do método de nuvem, onde o tamanho refere-se à frequência total do termo e a largura das ligações entre os pontos é relacionada ao peso de ligação semântica entre os termos, assim como sua proximidade.

De forma intuitiva, o software organizou em comunidades os termos que estão ligados entre si, sendo possível observar que existem, no presente estudo, 3 grandes nichos: “rede social”, “covid” e “vacina”. A comunidade “rede social” apresenta como constituintes primários termos sobre política como “eleições”, “voto”, “governador” e “saúde”, mostrando, pela proximidade, intenso perfil associativo entre os termos.

Semelhantemente, a comunidade “covid” expõe relação de proximidade entre os termos “pandemia”, “hidroxicloroquina” e “tratamento”, além de revelar um cruzamento especial de relações entre a referida comunidade e a comunidade “rede social”, ponto de importante destaque pois trata-se do termo “kit covid”, uma alternativa pseudo-terapêutica que fora amplamente divulgada no período de pandemia.

Tais notícias falsas têm elevada possibilidade de impacto, principalmente em populações mais sensíveis, tal como citado por DANTAS, Carolina (2021) onde, na época, embora 62% dos indígenas na Amazônia

estivessem vacinados contra Covid; Acre e Pará apresentaram aldeias com baixa adesão devido, em especial, a desinformação com relação a vacina, associando-a a teorias conspiracionistas nacionais e internacionais de cunho claramente político. Semelhantemente a essa associação, o perfil encontrado na comunidade “vacina” mostra claramente que a proporção do boato não esteve presente somente nas aldeias indígenas, mas também em nossa amostra, visto que os termos “5g”, “aborto”, “coronovac”, “sinovac” e “pfizer” apresentam relação muito elevada no presente estudo.

Nota-se, portanto, que as palavras ligadas às vacinas de Covid tiveram muito mais interações com as marcas dos fabricantes e palavras que não são relacionadas à saúde. Vale ressaltar que procuramos palavras-chave mais comumente associadas ao tema das vacinas, tais como: reação adversa, proteção, dose, idade, imunologia, contudo, nenhum destes temas foi prevalente.

CONCLUSÕES

Observou-se que temas com teor conspiracionista e negacionista tiveram prevalência elevada no presente estudo, em especial os relacionados à saúde. Destaca-se que houve clara relação entre política e saúde, de forma que os presentes resultados são instigantes para novas pesquisas sobre o tema, podendo inclusive focar nos temas com elevado destaque e impacto a verdade, a saúde e a vida, a fim de compreender e mapear o perfil das Fake News, podendo assim ter base para criar ferramentas e assim combater com eficácia sua disseminação no futuro.

AGRADECIMENTOS

Agradeço a Dra. Professora Larissa Nicolete, a FUNCAP e a UNILAB por dar a oportunidade de estar aqui hoje com este projeto tão importante no atual momento em que nos encontramos.

REFERÊNCIAS

- BAPTISTA, João. Ethos, pathos e logos. Análise comparativa do processo persuasivo das (fake) news. **Eikon**, v. 1, n. 7, 2020.
- BAGDADI, Golnaz; TOWHIDKHAH, Farzad; ROSTAMI, Reza. Um modelo matemático da interação entre controladores de atenção de baixo para cima e de cima para baixo em resposta a um alvo e um distrator em seres humanos. **Pesquisa de Sistemas Cognitivos**, v. 58, p. 234-252, 2019.
- MONTEIRO, Rafael A. et al. Contribuições para o estudo das fake news em português: Novo corpus e resultados de detecção automática. In: **Conferência Internacional sobre Processamento Computacional da Língua Portuguesa. Springer, Cham**, 2018. p. 324-334
- CASTILHO, Isabela Ribeiro. O uso de fake news nas eleições de 2018: uma análise das notícias falsas checadas pelas agências Lupa e Aos Fatos sobre o jornalismo. 2020.
- DE SOUZA, Marcelo Pereira et al. A linguistic-based method that combines polarity, emotion and grammatical characteristics to detect fake news in portuguese. In: **Proceedings of the Brazilian Symposium on Multimedia and the Web**. 2020. p. 217-224.
- MCGOEY, Linsey. Desconhecidas estratégicas: para uma sociologia da ignorância. *Economia e sociedade*, v. 41, n. 1, pág. 1-16, 2012.
- VISHWAKARMA, Dinesh Kumar; VARSHNEY, Deepika; YADAV, Ashima. Detecção e análise de veracidade de notícias falsas via sucateamento e autenticação da pesquisa na web. **Pesquisa de Sistemas Cognitivos**, v.

58, p. 217-229, 2019.

KAUSAR, Md Abu; DHAKA, VS; SINGH, Sanjeev Kumar. Rastreador da Web: uma revisão. **International Journal of Computer Applications**, v. 63, n. 2, 2013.

MEESAD, Phayung. Detecção de notícias falsas tailandesas com base na recuperação de informações, processamento de linguagem natural e aprendizado de máquina. *SN Ciência da Computação*, v. 2, n. 6, pág. 1-17, 2021.

VOGEL, Inna; JIAN, Pedro. Detecção de notícias falsas com o novo conjunto de dados alemão "GermanFakeNC". In: **Conferência Internacional sobre Teoria e Prática de Bibliotecas Digitais**. Springer, Cham, 2019. p. 288-295

BAGDADI, Golnaz; TOWHIDKHAH, Farzad; ROSTAMI, Reza. Um modelo matemático da interação entre controladores de atenção de baixo para cima e de cima para baixo em resposta a um alvo e um distrator em seres humanos. **Pesquisa de Sistemas Cognitivos**, v. 58, p. 234-252, 2019.

FERNANDES, Andreza et al. Detector de fake news para analisar notícias sobre o COVID-19 divulgadas em redes sociais. 2020. Disponível em: <https://insightlab.ufc.br/pesquisadores-desenvolvem-detector-de-fake-news-para-noticias-sobre-o-covid-19/>. Acesso em: 14 mar. 2022.

MOURA, Hysla Magalhães de et al. Perception of Death among University Students: Contributions of the Terror Management Theory. **Psicologia: Ciência e Profissão**, v. 40, 2020.

ZHANG, Chaowei et al. Detectar notícias falsas para reduzir os riscos de desinformação usando abordagens analíticas. **Revista Europeia de Pesquisa Operacional**, v. 279, n. 3, pág. 1036-1052, 2019.

DE BARCELOS, Thainá do Nascimento et al. Análise de fake news veiculadas durante a pandemia de COVID-19 no Brasil. **Revista Panamericana de Salud Pública**, v. 45, p. e65, 2021.

BONDIELLI, Alessandro; MARCELONI, Francisco. Uma pesquisa sobre notícias falsas e técnicas de detecção de boatos. **Ciências da Informação**, v. 497, p. 38-55, 2019.

Q. Bai, Z. Xu, Y. Liu, X. Miao e Q. Xiu, "The Web Crawler Based on Python to Obtain Public Emotional data Analysis - Take The Weibo Hot Rearch data as An Example," 2021 **7th International Conference on Sistemas e Informática (ICSAI)**, 2021, pp. 1-6, doi: 10.1109/ICSAI53574.2021.966

DANTAS, Carolina. 62% dos indígenas na Amazônia estão vacinados contra Covid; AC e PA têm aldeias com baixa adesão. 2021.

CASTRO, Suzana da Silva et al. As narrativas por trás das fake news: uma ameaça que tem interferido na decisão de vacinar. 2020.

GUPTA, Sandeep Kumar et al. Detecção de problemas de fake news e sua avaliação por meio de inteligência artificial. In: **Conferência Internacional de Negócios e Tecnologia**. Springer, Cham, 2020. p. 94-101.

DAVENPORT, Thomas H. Ecologia da informação: porque só a tecnologia não basta para o sucesso na era da informação. 1998. "Digital no Brasil: todas as estatísticas que você precisa em 2021".

DataReportal - Global Digital Insights, <https://datareportal.com/reports/digital-2021-brazil>. Acesso em 12 de março de 2022. BATESINI, Flávio Junior. O poder de manipulação das fake news na sociedade. 2020.

"Estudo revela grau de conhecimento científico dos brasileiros". Grupo de Institutos, Fundações e Empresas - GIFE, 2014. Disponível em: <https://gife.org.br/estudo-revela-grau-de-conhecimento-cientifico-dos-brasileiros/>. Acesso em 01 de Março de 2022.

MINTZBERG, Henry; AHLSTRAND, Bruce; LAMPEL, Joseph. **Safári da estratégia**. Bookman Editora, 2009.

GUPTA, Anshul et al. Uma análise exploratória sobre o desdobramento de notícias falsas durante a pandemia de COVID-19. In: **Sistemas Inteligentes: Inovações em Computação**. Springer, Cingapura, 2022. p. 259-272.



VIII SEMANA UNIVERSITÁRIA

A Universidade pós-isolamento social: desafios, expectativas e perspectivas

