

ESTRATÉGIAS COMPUTACIONAIS PARA A ANÁLISE DE CONTEÚDO EM REDES SOCIAIS NO COMBATE DA COVID-19

Felipe Carneiro Magrinelli¹; Amanda Zampieri de Souza²; Nelson Tenório³

¹Acadêmico do Curso de Engenharia de Software, Universidade Cesumar - UNICESUMAR, Maringá (PR). Bolsista do PIBIC/ICETI- UniCesumar. felipemagrinelli08@gmail.com

²Mestranda do Programa de Gestão do Conhecimento nas Organizações, Campus Maringá/PR, UNICESUMAR. Bolsista CAPES

³Orientador, PhD, Docente do Programa de Gestão do Conhecimento nas Organizações, UNICESUMAR. Pesquisador e Bolsista Produtividade em Pesquisa do Instituto Cesumar de Ciência, Tecnologia e Inovação (ICETI). nelson.tenoriojr@gmail.com

RESUMO

O objetivo geral desta pesquisa é identificar na literatura as estratégias computacionais de detecção de *fake news*. Como objetivos específicos deseja-se buscar na literatura a definição e as classificações de *fake news*, estudar métodos e ferramentas de análise de conteúdo baseados em mineração de dados e textos, e propor estratégias de detecção de *fake news*. A pesquisa é de abordagem mista e caráter descritiva, e para que fossem alcançados os objetivos propostos, primeiramente foi feita a identificação e estudo dos desafios atuais em relação ao combate das *fake news*. Posteriormente, foi realizada a coleta de dados a partir do repositório online Acm Library, utilizando a mineração de textos como método principal. Finalmente, a análise dos dados coletados foi realizada com a utilização do software *Orange*, por permitir uma análise exploratória e uma visualização interativa dos dados obtidos. Assim, espera-se que este artigo contribua como ponto inicial para que pesquisadores reflitam e aprimorem suas pesquisas com relação à detecção de *fake news*. Futuramente, pretende-se verificar os principais algoritmos utilizados nas estratégias encontradas.

PALAVRAS-CHAVE: Aprendizado de máquina; Ciência de Dados; Mineração de textos; Notícias falsas.

1 INTRODUÇÃO

Vivemos em uma era de 'pós-verdade'. Notícias que são enganosas, falsas ou estranhas, as chamadas *fake news*, competem por nossa atenção abertamente em uma rede tecnológica a que bilhões de pessoas possuem acesso irrestrito. O crescimento explosivo do uso de redes sociais para a disseminação de notícias falsas está erodindo as democracias, a justiça e a confiança das pessoas, aumentando assim a urgente demanda por detecção de *fake news* (ZHOU et al., 2019). De fato, as *fake news* se espalham como um vírus e se tornaram uma questão social importante e um desafio para democracias e também para as empresas de mídia social que trabalham no esforço de as conterem quando devidamente identificadas (MCCARTHY, 2018, LAKSHMANAN; SIMPSON; THIRUMURUGANATHAN, 2018). Nesse sentido, é difícil detectar esse conteúdo falso, uma vez que é possível que ocorra erros de postagens ou sátiras (GOLBECK et al., 2018, SHARMA et al., 2019).

Como um tópico interdisciplinar, o estudo de notícias falsas incentiva um esforço conjunto de especialistas em ciência da computação e da informação, ciência política, jornalismo, ciências sociais, psicologia e economia (SHU; BERNARD; LIU, 2019, ZHOU.; ZAFARANI, 2019). Atualmente, no contexto de uma pandemia da COVID-19, as *fake news* continuam existindo e se propagando. No entanto, neste momento elas possuem um efeito ainda mais devastador, pois têm o potencial de colocar milhares de pessoas em risco de morte. Assim, é imperativo que as informações propagadas sejam estudadas e que se faça um estudo qualitativo para propor estratégias de boa comunicação para a prevenção de circulação de *fake news* nas redes sociais para mitigar rumores falsos relacionados à COVID-19.

2 MATERIAIS E MÉTODOS

Para atingir os objetivos propostos, foi realizada uma pesquisa mista sobre estratégias tecnológicas para a identificação de *fake news* acerca da Covid-19 nas redes sociais.

A revisão bibliográfica foi feita no repositório online “Acm Library” (disponível em: <https://dl.acm.org/search/advanced>), utilizando as palavras chave: “*fake news*”; covid OR coronavirus; detection, e limitando a busca apenas para artigos do ano de 2020. Foram encontrados 45 resultados, e 32 foram utilizados (Apêndice 1).

Em seguida, os artigos foram exportados para a ferramenta de análise de dados Orange e o tratamento do texto utilizado consistiu em três etapas: etapa 1 – Transformação: todas as palavras foram colocadas em minúsculo, URLs e acentos foram removidos; etapa 2 – Tokenização: sinais de pontuação foram omitidos, e apenas palavras foram consideradas; etapa 3 – Filtragem: foram consideradas somente palavras que referenciam algoritmos, bases de dados de treinamento ou auxiliam nas estratégias de detecção de *fake news*.

Finalmente, para a apresentação e discussão dos resultados, foi gerada uma nuvem de palavras das estratégias levantadas nos artigos devidamente apresentados no Apêndice I (1) e nas estratégias identificadas com a leitura direcionada pela nuvem de palavras devidamente geradas por aquelas apresentadas no Apêndice I (2).

3 RESULTADOS E DISCUSSÕES

Como resultado desta pesquisa, observa-se um conjunto de estratégias baseadas em diferentes técnicas computacionais conforme apresenta a nuvem de palavras da Figura 1.



Figura 1: Nuvem de palavras
Fonte: Dados da pesquisa

Cinco principais estratégias encontradas:

Fact-checking: a estratégia refere-se à tarefa de assegurar a veracidade de informações e afirmações factuais presentes na mídia. Atualmente, recorre-se à tecnologia para fazer estas verificações.

Dataset: conjuntos de dados tabulados e utilizados em muitas situações para o treinamento de algoritmos de Machine Learning e análise de dados.

Web Scraping: processo de extrair e combinar conteúdos de interesse da web de forma sistemática através das tags HTML do site. É comparativamente mais rápida do que

a utilização de uma *Api*, mas é mais vulnerável à alterações feitas nos websites (DONGO et al., 2020).

Api: Protocolos de *Api* (*Application Programming Interfaces*) definem sequências legais de chamadas de métodos em objetos de softwares terceiros (BIERHOF et al., 2009). No caso desta pesquisa, a *Api* da rede social *Twitter* foi a mais utilizada no auxílio de detecção de fake News.

Machine learning: é um conjunto de métodos de pesquisa que tenta automatizar as as outras fases da pesquisa. É uma estratégia que utiliza técnicas matemáticas para detectar padrões ou gerar modelos a partir de uma base de informações com grande quantidade de dados (EDGAR; MANZ, 2017).

Quadro 1: Relação das palavras com a quantidade de vezes que apareceram

Palavra	Quantidade
<i>fact</i>	379
<i>web</i>	329
<i>dataset</i>	281
<i>analysis</i>	272
<i>learning</i>	262
<i>checking</i>	186
<i>api</i>	175
<i>classification</i>	165
<i>rumor</i>	158
<i>machine</i>	149
<i>network</i>	142
<i>scraping</i>	125
<i>sentiment</i>	105
<i>fang</i>	90
<i>neural</i>	83
<i>deep</i>	61
<i>nlp</i>	49
<i>ai</i>	35
<i>binary</i>	33
<i>pkg</i>	4

Fonte: Dados da pesquisa

4 CONSIDERAÇÕES FINAIS

Dado o objetivo desse artigo a saber, identificar na literatura as estratégias computacionais de detecção de *fake news* por meio de uma pesquisa utilizando métodos mistos, foram identificadas cinco principais estratégias, sendo elas: *Fact-Checking*, *Dataset*, *Web Scraping*, *Api* e *Machine learning*. Diante disso, espera-se que este artigo contribua como ponto inicial para que pesquisadores reflitam e aprimorem suas pesquisas com relação *fake news*. Como trabalho futuro, pretende-se avançar esta pesquisa no sentido de identificar os principais algoritmos utilizados pelas cinco estratégias aqui identificadas com o intuito de contribuir no combate à *fake news*.

REFERÊNCIAS

BIERHOFF, K.; ALDRICH, J.; GARLAN, D. **API Protocol Compliance in Object-**

Oriented Software. 2009.

DONGO, I.; CADINALE, Y.; AGUILERA, A.; et al. Web Scraping versus Twitter API: A Comparison for a Credibility Analysis. Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services. **Anais...**, iiWAS '20. p.263–273, 2020. New York, NY, USA: Association for Computing Machinery. Disponível em: <<https://doi.org/10.1145/3428757.3429104>>. .

EDGAR, T. W.; MANZ, D. O. **Research Methods for Cyber Security.** 2017.

GOLBECK, J. ... VISNANSKY, G. *Fake news* vs satire: A dataset and analysis. WebSci 2018 - Proceedings of the 10th ACM Conference on Web Science, p. 17–21, 2018.

LAKSHMANAN, L. V. S.; SIMPSON, M.; THIRUMURUGANATHAN, S. Combating *fake news*: A data management and mining perspective. Proceedings of the VLDB Endowment, v. 12, n. 12, p. 1990–1993, 2018.

MCCARTHY, L. “Let’s *Fake news*”. Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction. **Anais...**New York, NY, USA: ACM, 18 mar. 2018 Disponível em: <https://dl.acm.org/doi/10.1145/3173225.3173318>

SHARMA, K.; QIAN, F.; JIANG, H.; RUCHANSKY, N.; ZHANG, M.; LIU, Y. Combating *fake news*: A survey on identification and mitigation techniques. ACM Transactions on Intelligent Systems and Technology, v. 10, n. 3, 2019.

SHU, K.; BERNARD, H. R.; LIU, H. Studying *Fake news* via Network Analysis: Detection and Mitigation. p. 43–65, 2019.

ZHOU, X.; JAIN, A.; PHOHA, V. V.; ZAFARANI, R. *Fake news* Early Detection: A Theory-driven Model. p. 3207–3208, 2019.

ZHOU, X.; ZAFARANI, R. *Fake news* Detection: An Interdisciplinary Research. Companion Proceedings of The 2019 World Wide Web Conference on - WWW '19. **Anais...**New York, New York, USA: ACM Press, 2019 Disponível em: <http://arxiv.org/abs/1708.07104>