

# Rumo à Detecção Automática de Notícias Falsas em Plataformas Digitais: Propriedades, Limitações e Aplicações

## Resumo Estendido - CTDSI/CTCCSI 2021

### Towards Automatic Fake News Detection in Digital Platforms: Properties, Limitations, and Applications\*

#### Extended Abstract - CTDSI/CTCCSI 2021

Julio C. S. Reis, Fabrício Benevenuto

<sup>1</sup>Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)  
Belo Horizonte – MG – Brasil

{julio.reis, fabricio}@dcc.ufmg.br

**Resumo.** Neste trabalho, investigamos o potencial das soluções automáticas para identificar notícias falsas disseminadas em plataformas digitais. Particularmente, exploramos novos conjuntos de dados e atributos para detecção de notícias falsas e avaliamos o desempenho de previsões de abordagens de aprendizado de máquina. Também quantificamos a informatividade dos atributos para detecção de notícias falsas e apresentamos uma explicação dos fatores que contribuem para as decisões dos modelos. Por fim, propomos e implementamos uma nova estratégia para ajudar os checadores de fatos a identificarem notícias com maior probabilidade de serem falsas, incorporando nossa abordagem em um sistema real.

**Abstract.** In this work, we investigate the potential of automatic solutions to identify fake news disseminated on digital platforms. Particularly, we explore new datasets and features for fake news detection to assess the prediction performance of machine learning approaches. We also quantify the informativeness of features for fake news detection, and present an explanation of factors contributing to model decisions. Finally, we propose and implement a new strategy to help fact-checkers identify news stories that are more likely to be fake, incorporating our approach into a real system.

## 1. Introdução

As plataformas digitais, incluindo sistemas de mídia social e aplicativos de mensagens, são usadas ativamente por mais de um terço da população mundial<sup>1</sup>. Essas plataformas mudaram drasticamente a maneira como as notícias são produzidas, disseminadas e consumidas em nossa sociedade.

---

\*Trabalho relacionado à tese de doutorado defendida em 03 de Novembro de 2020 no Programa de Pós-Graduação em Ciência da Computação (PPGCC) da Universidade Federal de Minas Gerais (UFMG). Ordem dos autores: Doutorando e Orientador.

<sup>1</sup><http://www.businessinsider.com/the-social-media-demographics-report-2017-8>

Parte das razões para essa mudança é inerente à natureza dessas plataformas uma vez que: (i) muitas vezes é mais oportuno e barato produzir e consumir notícias nestes ambientes em comparação com meios de comunicação tradicionais, como jornais ou televisão; e (ii) é mais fácil compartilhar, comentar e discutir notícias com amigos ou outros leitores em plataformas digitais, o que melhora a comunicação e a interação entre os usuários [Shu et al. 2017]. Conseqüentemente, as plataformas digitais estão moldando a maneira como os usuários consomem informações. Atualmente, cerca de 62% dos usuários americanos e 66% dos brasileiros já consomem notícias a partir das plataformas digitais<sup>23</sup>.

Apesar dos inúmeros benefícios que esses sistemas trazem para a nossa sociedade, um problema fundamental hoje é que as plataformas digitais se tornaram espaços amplamente abusados por campanhas de desinformação que afetam a credibilidade de todo o ecossistema de notícias. O surgimento de notícias falsas nesses ambientes evoluiu rapidamente para um fenômeno mundial, onde a falta de estratégias escaláveis de verificação de fatos é preocupante. Por exemplo, estima-se que pelo menos 800 pessoas tenham morrido em todo o mundo em decorrência de notícias falsas relacionadas ao coronavírus entre janeiro e março de 2020<sup>4</sup>. Assim, soluções automáticas para detecção de notícias falsas podem ser usadas por jornalistas e equipes de checagem de fatos como uma ferramenta auxiliar na identificação de notícias com alta probabilidade de serem falsas. Neste contexto, este trabalho visa investigar abordagens práticas para a detecção automática de notícias falsas disseminadas em plataformas digitais.

## 2. Principais Resultados e Contribuições

Inicialmente, nós pesquisamos um grande número de trabalhos recentes e relacionados como uma tentativa de implementar atributos propostos na literatura para a detecção de notícias falsas. Isso nos possibilitou propor novos atributos, explorar conjuntos de dados rotulados disponíveis de diferentes contextos (i.e. política e saúde) e construir um novo conjunto de dados<sup>5</sup> para avaliar o desempenho de previsão das atuais abordagens de aprendizado de máquina supervisionadas na realização desta tarefa. Nossos resultados revelam que esses modelos computacionais propostos possuem um grau útil de poder discriminativo para detectar notícias falsas disseminadas em plataformas digitais.

Em seguida, nós propusemos um arcabouço computacional para quantificar a informatividade de atributos para detecção de notícias falsas. Como parte do arcabouço proposto, apresentamos uma explicação dos fatores que contribuem para as decisões dos modelos, promovendo assim o raciocínio cívico, complementando nossa capacidade de avaliar o conteúdo digital e chegar a conclusões justificadas. Também analisamos atributos e modelos que permanecem úteis para detectar notícias falsas em diferentes cenários: eleições nos Estados Unidos e no Brasil. Por fim, propomos e implementamos em um sistema real<sup>6</sup>, um novo mecanismo que, conforme resultados experimentais, reduziu significativamente o número de notícias que jornalistas e checadores de fatos precisam ler antes de encontrar uma história falsa.

<sup>2</sup><http://www.pewresearch.org/fact-tank/2016/07/07/modern-news-consumer/>

<sup>3</sup><https://reutersinstitute.politics.ox.ac.uk/risj-review/statistic-week-how-brazilian-voters-get-their-news>

<sup>4</sup><https://www.bbc.com/news/world-53755067>

<sup>5</sup>Disponível em: <https://doi.org/10.5281/zenodo.3779157>

<sup>6</sup><http://www.whatsapp-monitor.dcc.ufmg.br/>

### 3. Considerações Finais

Os principais resultados deste trabalho geraram as seguintes publicações: [Reis et al. 2020b, Reis et al. 2020a, Reis et al. 2019b, Reis et al. 2019a, Reis et al. 2017, Reis et al. 2016]. Vale ressaltar que este trabalho foi premiado pela Edição de 2018 do *Google Research Awards for Latin America*.

Por fim, parte dos resultados obtidos neste trabalho inspirou a construção de sistemas que foram incorporados ao projeto Eleições Sem Fake<sup>7</sup>, que foi amplamente utilizado durante as eleições brasileiras de 2018. Além disso, o Ministério Público de Minas Gerais (MPMG) está financiando um projeto, de dois anos (2021-2022), cujo objetivo é implantar o arcabouço para detecção automática de notícias falsas proposto neste trabalho de forma integrada com ferramentas de monitoramento de informações disseminadas em plataformas digitais.

### Agradecimentos

Este trabalho foi financiado pelo MPMG, por meio do projeto Capacidades Analíticas, CNPq, CAPES e Google.

### References

- Reis, J., Miranda, M., Bastos, L., Prates, R., and Benevenuto, F. (2016). Uma análise do impacto do anonimato em comentários de notícias online. In *Anais do Simpósio Brasileiro de Sistemas Colaborativos (SBSC)*, pages 46–60.
- Reis, J. C., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019a). Explainable machine learning for fake news detection. In *Proc. of the Int’l ACM Conference on Web Science (WebSci)*, pages 17–26.
- Reis, J. C., Correia, A., Murai, F., Veloso, A., and Benevenuto, F. (2019b). Supervised learning for fake news detection. *IEEE Intelligent Systems*, 34(2):76–81.
- Reis, J. C., Kwak, H., An, J., Messias, J., and Benevenuto, F. (2017). Demographics of news sharing in the us twittersphere. In *Proc. of the ACM Conference on Hypertext and Social Media (HYPERTEXT)*, pages 195–204.
- Reis, J. C., Melo, P., Garimella, K., Almeida, J. M., Eckles, D., and Benevenuto, F. (2020a). A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In *Proc. of the Int’l AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 903–908.
- Reis, J. C., Melo, P., Garimella, K., and Benevenuto, F. (2020b). Can whatsapp benefit from debunked fact-checked stories to reduce misinformation? *Harvard Kennedy School Misinformation Review*.
- Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1):22–36.

---

<sup>7</sup><http://www.eleicoes-sem-fake.dcc.ufmg.br/>