

Ada - ferramenta para detecção automática de notícias falsas

Caroline Barbosa de Oliveira Lira¹, Kamila Rios da Hora Rodrigues¹

¹ICMC, Universidade de São Paulo (USP)
Av. Trab. São Carlense, 400 - Centro, São Carlos – SP – Brasil

clira23@usp.br , kamila.rios@icmc.usp.br

Resumo. *As notícias falsas se tornaram um problema de segurança e saúde pública. As mídias sociais e a sociedade conectada, principalmente por meio de aplicativos de redes sociais, tornaram-se canais de divulgação, tornando a desinformação e as fake news muitas vezes indecifráveis devido a sua prática cada vez mais elaborada. O uso de Machine Learning (ML) supervisionado e Processamento de Linguagem Natural (NLP), tornaram-se meios indispensáveis para a verificação de conteúdo duvidoso. Nesse contexto, a ferramenta Ada para detecção automática foi desenvolvida no contexto dessa pesquisa e atualmente possui uma precisão de 82%. O corpus da ferramenta é em português, e foi construído tendo como prioridade usuários idosos, uma vez que estes são considerados o público que mais dissemina notícias falsas no Brasil. A ferramenta foi avaliada por um grupo de idosos em um estudo empírico, e os resultados indicam que cerca de 87,5% dos participantes usariam a ferramenta em seu dia a dia. Fizeram, no entanto, sugestões de melhorias no ADA, que estão sendo analisadas pela equipe responsável.*

1. Introdução

Fake News, desinformação, má informação, não são fenômenos novos. Informações manipuladas e conteúdo enganoso, aparecem na história da humanidade desde os primórdios. Desde propagandas de desinformação da Primeira Guerra Mundial (1914-1918) e Segunda Guerra Mundial (1939-1945), passando por fatos da Guerra do Vietnã (1955-1975), Guerra Fria (1947-1991), Guerra do Iraque (2003-2011), Rússia e Ucrânia (2014 e 2022), *Brexit*, acusações de interferência em processos eleitorais, escândalo da *Cambridge Analytica*, “*Pizzagate*” e COVID-19, todos os eventos foram marcados por desinformação e *fake news*, intensificando as crises contextuais [Posetti and Matthews 2018].

O conceito de *fake news* é complexo. Ela pode ser definida como notícias falsas ou fabricadas, sem fatos ou fontes confiáveis, e também como notícias com intenção de prejudicar. Também existe a má informação, definida como “informações potencialmente perigosas ou prejudiciais, informações inadequadas (...)”. A desinformação, por sua vez, “inclui todas as formas de informações falsas, imprecisas ou informações enganosas, projetadas, com intenção de prejudicar” [Belloir et al. 2022].

Algumas das características das notícias falsas, é que elas costumam estimular o emocional do leitor, o que faz com que afete a sua opinião [Sharma et al. 2020] e manipule o comportamento de um indivíduo sobre temas específicos, mas principalmente políticos. Outro ponto é que em diversas vezes os textos possuem erros ortográficos, misturando informações reais e fictícias [Eleitoral 2022].

Embora esse fenômeno não seja novo, o aumento de uma sociedade mais conectada e com plataformas de mídias sociais nas mãos, favorece a propagação desses conteúdos, saindo do controle das autoridades. As consequências afetam a vida das pessoas, produzem instabilidade em governos e refletem em sociedades menos democráticas [Carvalho 2020]. Há diversas ferramentas para identificação de *fake news* [de Queiroz et al. 2021]. No entanto, todas elas trazem alguma dificuldade para o público idoso. Assim, este trabalho se beneficiou das técnicas de Design Participativo para criar uma ferramenta em colaboração com idosos. A ferramenta aqui descrita faz uso do Processamento de Linguagem Natural (PLN) e Aprendizado de Máquina Supervisionado (AMS), para criar mecanismos automáticos e facilitar a detecção de notícias falsas, uma vez que essas técnicas são mais baratas e com maior precisão de acertos, do que a checagem humana, que pode conduzir, em alguns casos, ao viés político. Neste contexto, foi desenvolvida a ferramenta ADA¹. Trata-se de uma ferramenta para detecção automática de notícias falsas.

O desenvolvimento da ADA, para implementação dos recursos citados acima, iniciou no ano de 2022 e teve como objetivo inicial aprimorar a ferramenta *FakeCheck*, desenvolvida pelo grupo de pesquisa do professor Thiago Pardo da Universidade de São Paulo [Monteiro 2018]. Porém, no decorrer do estudo, foi observado que o *corpus* continha notícias antigas e que a interface gráfica da ferramenta não era de boa usabilidade e não apresentava recursos de acessibilidade. O intuito deste trabalho de pesquisa era realizar um estudo sobre *fakenews* e idosos. Neste sentido, a ferramenta apresentava diversas barreiras para esse público. Assim, um novo *corpus* foi treinado e uma nova interface gráfica de usuário foi projetada.

A precisão alcançada na detecção da ferramenta ADA atual foi de 82%, com um *corpus* com 2549 notícias, em Português, rotuladas manualmente. Adotou-se um classificador *Naive Bayes*, além de Regressão Logística, *Decision Tree* e *Random Forest* [Ahmed et al. 2021]. Assim, o usuário obtém informações se a notícia é *True* (Verdadeira) ou *Fake* (Falsa).

Para o desenvolvimento da ferramenta ADA, foi considerado o Design Participativo (DP) [Bødker et al. 2022, Benyon 2011] conceito da área de Interação Humano-Computador (IHC), que busca tornar o usuário o ator mais importante do processo de desenvolvimento, tendo suas dificuldades e características bem analisadas na etapa de design da solução. Neste trabalho, oficinas participativas com usuários alvo foram conduzidas tanto na etapa de design quanto de avaliação da ADA.

Este artigo descreve como ferramenta ADA pode ser um recurso auxiliar para as pessoas na identificação automática de notícias falsas. Uma prova de conceito da ferramenta foi realizada com usuários idosos. Os resultados dessa prova apontam para a aceitação desse público quanto ao uso da ferramenta, mas revela importantes mudanças que devem ser feitas na sua interface para facilitar a interação dos usuários alvo.

2. Metodologia

Este trabalho foi conduzido em cinco etapas para a criação da ferramenta ADA, a saber: 1) Definição do conjunto de dados, 2) Pré-Processamento dos dados, 3) Aprendizado de

¹A ferramenta foi nomeada de Ada, em homenagem a Ada King, condessa de Lovelace, a primeira programadora da história, com um papel importante na história da Ciência da Computação.

Máquina Supervisionado, 4) Construção de uma interface gráfica de usuário e 5) Prova de Conceito. Este trabalho dá enfoque nas etapas 1, 4 e 5.

2.1. Definição do conjunto de dados

Os dados utilizados para o *corpus* da ferramenta foram adquiridos de um conjunto de 2549 notícias, entre elas, verdadeiras e falsas, extraídas de mídias sociais nacionais e internacionais. Considerou-se temas relevantes de política, política internacional, saúde, entretenimento e justiça. As notícias estavam divididas em: 46,7% são falsas e 53,3% são verdadeiras. Também foram adicionados artigos acadêmicos da área médica, para ajudar no treinamento do conteúdo relacionado à saúde.

Destaca-se que também foi realizada a contagem de palavra-chave para identificar quais assuntos estão mais presentes no *corpus*, para que houvesse controle do conteúdo das notícias por palavras-chave, ou seja, a palavra de maior relevância no texto. Além disso, para a construção do *corpus*, houve a seleção de 147 estudos acadêmicos, realizados em países como Índia, Canadá, Estados Unidos, Japão, Hungria e Brasil, para análise de viabilidade da detecção automática de notícias falsas.

2.2. Construindo uma interface gráfica de usuário para a ADA

Após criação do *corpus* da ADA, a etapa seguinte foi o desenvolvimento de uma interface gráfica de usuário que permitisse que pessoas inserissem trechos de notícias e a ferramenta retornasse a informação com o diagnóstico de ser uma notícia falsa ou verdadeira.

Técnicas de Design Participativo (DP) foram empregadas para o desenvolvimento da ADA. Para Schuler e Namioka “o design participativo representa uma nova abordagem para o projeto de sistemas computacionais na qual as pessoas destinadas a usar o sistema desempenham um papel crítico na sua concepção” [Schuler and Namioka 1993]. Na primeira etapa de construção da interface, houve um levantamento de estudos da literatura nas áreas de Psicologia e Neurociência, sobre as dificuldades cognitivas e comportamentais de idosos (público alvo da ferramenta), incluindo com a interação com as *fake news*. Após esse estudo foi dada início à etapa de criação da interface gráfica. No protótipo inicialmente criado estavam disponíveis as páginas “Home” e “Verificar”. Ao digitar ou colar um texto, e clicar em ‘verificar’, os usuários são direcionados para a resposta “true” e “fake”, da página verificar.

A Figura 1 ilustra partes da ferramenta com a tela em que o usuário pode consultar uma notícia e, em seguida, receber o resultado, neste caso *Fake*, abaixo das instruções sobre o uso. O usuário pode realizar consultas novas, depois do recebimento do resultado. É possível observar o ícone no canto superior esquerdo de acessibilidade. No entanto, esse recurso ainda está em fase de implementação. O usuário pode, ainda, realizar quantos consultas desejar e não há limite de caracteres, todavia, textos maiores, possuem maior probabilidade de acerto.

2.3. Prova de Conceito

Após o desenvolvimento de uma versão da interface gráfica da ADA, um grupo de idosos, frequentadores de um curso de letramento digital para a terceira idade da Universidade de São Paulo (USP), foi convidado a interagir com a ferramenta e fornecer suas principais impressões. Destaca-se que durante o curso de letramento esses idosos têm aulas sobre

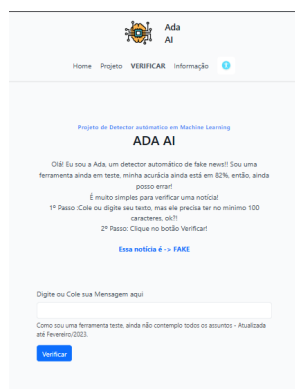


Figura 1. Design Gráfico Atual da Ferramenta Ada.

identificação de notícias falsas e como se prevenir. Destaca-se ainda, que essas atividades foram previstas e aprovadas em Comitê de Ética em Pesquisa em projeto referente ao curso de letramento, com número CAAE: 57875016.3.0000.5390.

Oito idosos se voluntariaram para o estudo: 5 mulheres e 3 homens, com idade entre 73 e 64 anos. Os participantes, um a um, testaram com os pesquisadores o uso da ADA e, posteriormente, poderiam consultar na ferramenta sobre qualquer assunto de seu interesse. Ao realizar a consulta, eram questionados se estavam satisfeitos com a resposta, momento em que forneciam *feedback* sobre a ferramenta. Para finalizar, os participantes foram convidados a responder a um formulário com o teste de usabilidade. No questionário de usabilidade, os participantes foram questionados se usariam o sistema com frequência: 50% responderam ‘concordam totalmente’ e 37,5% ‘concordam’.

Uma preocupação para elaboração do sistema, voltado principalmente para idosos, é a limitação quanto ao conhecimento e dificuldade no uso de tecnologia, o que poderia comprometer o uso da ferramenta ADA, porém, ao serem questionados se o sistema era fácil de usar: 62,5% responderam concordar totalmente e 25% afirmaram concordar. Outro ponto foi que os participantes, ao serem questionados se imaginam que outras pessoas aprenderiam a usar rapidamente a ADA, 62,5% concordam totalmente e 37,5% não concordam. No desenvolvimento do protótipo, foi considerado que o sistema deveria ser claro nos comandos e instruções, bem como no local onde estava posicionada a resposta da verificação: 62,5% concordaram totalmente que as funções implementadas estava bem integradas e 37,5% concordaram.

3. Considerações Finais

Embora os resultados sejam positivos, e alguns apontamentos estejam dentro do esperado, nem todas as páginas da ferramenta estavam em funcionamento, o que impactou na experiência dos usuários idosos mais curiosos e que queriam interagir com todas as funcionalidades da ferramenta. Além disso, com um *corpus* mais robusto, contendo notícias mais atuais, os resultados seriam mais precisos do que a atual acurácia de 82%. Apesar disso, em um *corpus* de 2549 notícias, foi possível observar com clareza as dificuldades dos idosos na verificação de notícias falsas, e quais eram os temas de maior dúvida, principalmente, a dificuldade em entender as próprias funções dos dispositivos, ficando evidente que, uma ferramenta para detecção de notícia teria pouca adesão se fosse necessário digitar o endereço de *url*.

Referências

- Ahmed, A. A. A., Aljabouh, A., Donepudi, P. K., and Choi, M. S. (2021). Detecting fake news using machine learning: A systematic literature review. *arXiv preprint arXiv:2102.04458*.
- Belloir, N., Ouerdane, W., Pastor, O., Frugier, É., and de Barmon, L.-A. (2022). A conceptual characterization of fake news: a positioning paper. In *Research Challenges in Information Science: 16th International Conference, RCIS 2022, Barcelona, Spain, May 17–20, 2022, Proceedings*, pages 662–669. Springer.
- Benyon, D. (2011). Interação humano-computador. *Tradução de Heloisa Coimbra de Souza. 2a. ed. Sao Paulo: Person Prentice Hall*, page 464.
- Bødker, S., Dindler, C., Iversen, O. S., and Smith, R. C. (2022). What is participatory design? In *Participatory Design*, pages 5–13. Springer.
- Carvalho, L. B. (2020). A democracia frustrada: fake news, política e liberdade de expressão nas redes sociais. *InternetLab*.
- de Queiroz, M. A. A., Souza, C. S., and de Paula, C. F. (2021). Uma análise sobre a eficácia de programas e aplicativos na detecção de fake news. *Revista Multitexto*, 9(01 (jan-jul)):66–76.
- Eleitoral, T. S. (2022). Pílulas contra a desinformação: erros de português podem sinalizar que a notícia é falsa. *International Journal of Engineering Research Technology (IJERT)*.
- Monteiro, R.A., S. R. P. T. d. A. T. R. E. V. (2018). O.a. (2018). contribuições para o estudo das fake news em português: Novos corpus e resultados da detecção automática. in: , et al. processamento computacional da língua portuguesa. propor 2018. notas de aula em ciência da computação. *Springer*.
- Posetti, J. and Matthews, A. (2018). A short guide to the history of ‘fake news’ and disinformation. *International Center for Journalists*, 7(2018):2018–07.
- Schuler, D. and Namioka, A. (1993). *Participatory design: Principles and practices*. CRC Press.
- Sharma, U., Saran, S., and Patil, S. M. (2020). Fake news detection using machine learning algorithms. *International Journal of Creative Research Thoughts (IJCRT)*, 8(6):509–518.