

FakeTrueBR: Um corpus brasileiro de notícias falsas

Juan Pablo Chavarro¹, Jonata Tyska Carvalho¹, Tarlis Tortelli Portela¹, e
Jonathan Cardoso Silva²

¹Departamento de Informática e Estatística – Universidade Federal de Santa Catarina
Caixa Postal 476 – 88.040-900 – Florianópolis – SC – Brazil

²Data Science Institute – London School of Economics and Political Science
London, U.K.

juan.pcs@posgrad.ufsc.br, jonata.tyska@ufsc.br, tarlis@tarlis.com.br,
j.cardoso-silva@lse.ac.uk

Abstract. *Currently, the large volume of fake news circulating on social media is a danger to society's perception of reality. Machine learning techniques have been useful in combating misinformation, but to generate good results they require balanced and high-quality training datasets. Since the main corpora publicly available for training fake news detection models are outdated or misaligned, this work proposes an innovative approach to recover true news from fake ones and improve their similarity and alignment. Thus, a dataset was developed that allows us to verify and classify the information we consume daily on the web through natural language processing. Additionally, the resulting corpus was evaluated using classical natural language processing techniques for text representation, such as BoW and BoW TF-IDF, along with various traditional classification methods. The results demonstrate that this dataset is effective for news classification, with an f1-score of 0.945 using a multi-layer perceptron. Therefore, this new corpus is a valuable resource in the fight against misinformation and for improving the quality of available online information.*

Resumo. *Atualmente, o grande volume de notícias falsas circulando nas redes sociais é um perigo para a percepção da realidade da sociedade. Técnicas de aprendizado de máquina têm sido úteis para o combate da desinformação, mas para gerar bons resultados necessitam de conjuntos de dados de treinamento balanceados e com boa qualidade. Uma vez que os principais corpus para o treinamento de modelos de detecção de notícias falsas disponíveis publicamente estão desatualizados ou desalinhados, neste trabalho foi proposto um enfoque inovador para recuperar as notícias verdadeiras a partir das falsas, e melhorar sua similaridade e alinhamento. Desta forma, foi desenvolvido um conjunto de dados que permite verificar e classificar a informação que consumimos diariamente na rede, por meio de processamento da linguagem natural. Além disso, o corpus resultante foi avaliado utilizando técnicas clássicas de processamento da linguagem natural para a representação de textos, como BoW e BoW TF-IDF, juntamente com vários métodos tradicionais de classificação. Os resultados obtidos demonstram que este conjunto de dados é efetivo para a classificação de notícias, com um f1-score de 0.945 usando um classificador de Multi-layer perceptron. Portanto, este novo corpus configura-se como um recurso valioso na luta contra a desinformação e para melhorar a qualidade da informação disponível online.*

1. Introdução

Avanços tecnológicos ocorridos nas últimas décadas têm alterado de forma significativa a forma como a humanidade se comunica e se organiza. Hoje em dia, as tecnologias de comunicação, especialmente a Internet, nos permitiram criar espaços como as redes sociais, onde podemos transmitir rapidamente qualquer notícia em massa, e acessá-la da mesma forma. Isso tem grandes vantagens, pois nos permite estar informados sobre qualquer evento ao redor do mundo, o que nos ajuda a transformar a cultura e a sociedade em que vivemos.

No entanto, uma das desvantagens trazida pelo advento e popularização das redes sociais é que qualquer usuário pode criar ou manipular a informação, distorcendo o conteúdo de acordo com seus interesses, e isso pode ter um efeito negativo sobre a opinião pública [Vosoughi et al. 2018]. Este fenômeno tornou-se evidente no Brasil durante os anos de 2021 e 2022, nos quais temas de grande importância como a pandemia de COVID-19 [Rocha et al. 2021] e as eleições presidenciais foram objetos de manipulação nas redes [Netlab 2022].

Um bom exemplo dos riscos trazidos pela má utilização destas tecnologias pode ser visto no relatório técnico elaborado pelo grupo NetLab da Universidade Federal do Rio de Janeiro (UFRJ) [Netlab 2022] em relação às eleições presidenciais do Brasil em 2022. Este relatório indica que durante o segundo turno houve um aumento significativo na quantidade diária de notícias falsas. As redes sociais, especialmente o Twitter, foram as principais ferramentas utilizadas para disseminar essas mensagens enganosas. Esse fenômeno representa um grande desafio para a sociedade brasileira, e também para o mundo de forma geral, já que a manipulação de informações pode gerar uma polarização política e afetar gravemente a democracia dos países. Embora os grandes conglomerados de mídia ainda dominem a cobertura de notícias, as informações disponíveis nas redes sociais estão em constante aumento [Newman et al. 2022], o que logicamente também contribui com o aumento da disseminação de notícias falsas.

Muitos portais de notícias e profissionais de mídia têm feito um esforço considerável para verificar a veracidade das notícias que se propagam online. No entanto, esse processo é tedioso e caro, e não é tão rápido quanto a velocidade com que as notícias falsas são criadas e se espalham. Assim, apenas a checagem manual não é suficiente para verificar todas as notícias, o que torna esse trabalho não a solução mais eficaz para lidar com o problema da desinformação. É a partir daqui, e com o objetivo de otimizar esse trabalho, que são propostas tecnologias de processamento de linguagem natural, que analisam a estrutura, o significado e a semântica das notícias para detectar, classificar e prever notícias falsas.

Para alcançar o mencionado acima, é essencial entender o fenômeno das notícias falsas e fornecer objetos de estudo, como bases de dados, para treinar as tecnologias de processamento de linguagem natural e melhorar sua precisão [Mishima and Yamana 2022]. Ou seja, o conjunto de dados é essencial para treinar e melhorar os modelos de detecção de notícias falsas, já que os modelos precisam aprender dos dados e padrões encontrados nos textos reais. Portanto, ter conjuntos de dados relevantes e confiáveis é fundamental para melhorar a capacidade das tecnologias de processamento de linguagem natural para detectar notícias falsas de forma eficaz em diferentes idiomas [Mishra et al. 2022].

No entanto, alguns dos conjuntos de dados disponíveis apresentam limitações, o que pode afetar negativamente a precisão e eficácia dos modelos de detecção de notícias falsas. Portanto, neste trabalho propõe-se a criação de um conjunto de dados automático utilizando BERT [Reimers and Gurevych 2019], a fim de obter uma representação do texto de alta qualidade, o que melhora o alinhamento entre cada notícia falsa e verdadeira. Isso último é o que realmente distingue esta construção dos conjuntos de dados existentes e a torna inovadora. Dessa forma, busca-se melhorar a qualidade e eficácia dos modelos de detecção de notícias falsas já existentes, garantindo que sejam treinados em um conjunto de dados representativo e equilibrado. Com essa abordagem, espera-se enriquecer os conjuntos de dados existentes e melhorar a precisão da detecção e classificação de notícias falsas escritas em português.

Na seção 2 apresentamos trabalhos relacionados à criação de corpus de notícias falsas, principalmente em língua portuguesa. Na seção 3 apresentamos a construção do corpus proposto. Na seção 4 abordamos a metodologia que seguirá o experimento e seus resultados que serão apresentados na seção 5. Por fim, na seção 6 apresentamos as conclusões de nosso trabalho.

2. Trabalhos relacionados

Com o objetivo de mitigar a propagação de notícias falsas na Internet, são coletados conjuntos de dados que possam ajudar a treinar algoritmos de aprendizado de máquina para identificar padrões e características que auxiliem a distinguir entre notícias falsas e verdadeiras.

Em [Jain et al. 2020], os autores utilizaram a combinação de dois conjuntos de notícias sobre as eleições dos EUA em 2016 para refinar os detectores de notícias falsas que existiam até aquele momento. Primeiro, o texto foi pré-processado para representá-lo por características estilométricas/linguísticas, através de Bag of Words (BoW TF) e Term Frequency-Inverse Document Frequency (BoW TF-IDF). Em seguida, esse conjunto de notícias foi submetido a modelos de aprendizado de máquina e sua eficácia foi avaliada. No entanto, esses resultados estão limitados apenas a notícias escritas em inglês. Portanto, pode-se dizer que o idioma em que as notícias são escritas é um fator determinante para sua classificação.

De acordo com o exposto, é possível perceber a importância de construir conjuntos de dados robustos que contribuam para a detecção e classificação de notícias falsas no Brasil. Vários pesquisadores já se adiantaram nessa tarefa, como podemos ver em [Silva et al. 2020]. Nesse corpus, os autores criaram um conjunto de dados de notícias em português chamado *FakeBr*, composto por 7.200 notícias publicadas no Brasil entre 2016 e 2018, metade delas falsas e a outra metade verdadeiras. A qualidade desse corpus se deve ao fato de que as notícias falsas foram coletadas manualmente de portais de verificação de notícias, enquanto as notícias verdadeiras foram coletadas de forma semiautomática de agências de notícias confiáveis usando um rastreador da web impulsionado por notícias falsas, obtendo assim 40.000 notícias verdadeiras. A partir disso, foram analisados diferentes modelos de aprendizado de máquina, como *logistic regression* (LR), *support vector machines* (SVM), *decision trees* (DT), *random forest* (RF), *bootstrap aggregating* (bagging), e *adaptive boosting* (AdaBoost), além de transformar o texto com técnicas de processamento de linguagem natural (NLP), como FastText, Word2Vec

e BoW TF-IDF. Finalmente, uma verificação manual das notícias associadas foi realizada para garantir sua qualidade, tornando-o um dos melhores corpus até o momento. No entanto, ressalta-se que o tamanho do texto influencia a efetividade do classificador e recomenda-se trabalhar com textos cruzados para maior efetividade.

A partir do *FakeBr* e juntamente com outros trabalhos, são apresentados os conjuntos de dados *FakeRecogna*¹ [Garcia et al. 2022] e *FakePedia Corpus*² [Charles et al. 2022].

O dataset *FakeRecogna* consiste em 11.902 notícias em português brasileiro, coletadas automaticamente entre os anos de 2019 e 2021 a partir de notícias verdadeiras de sites verificados e notícias falsas de agências de verificação. As notícias foram submetidas a diferentes experimentos utilizando 6 classificadores, e foi encontrado que os melhores foram o *Multi-Layer Perceptron* (MLP) e *Support Vector Machines* (SVM), utilizando as representações de texto Bag-of-Words (BoW) e FastText. Apesar de ser um dos corpora mais recentes, devido à sua construção, *FakeRecogna* tem um equilíbrio entre notícias falsas e verdadeiras em termos de proporção, mas elas não apresentam um alinhamento entre si.

O *FakePedia Corpus* é um conjunto de dados composto por 12.398 notícias do Brasil entre os anos de 2013 e 2021, das quais metade são falsas e metade são verdadeiras. Para construir este corpus, primeiro foram coletadas as notícias falsas de portais de verificação (*Boatos.org*, *Lupa* e *Aos Fatos*), em seguida foram realizadas buscas avançadas no Google para encontrar as 5 notícias reais que mais se assemelhavam a cada notícia falsa. Utilizando Word2Vec e avaliando a similaridade com base na distância de cosseno, para prever a notícia real mais semelhante das 5 selecionadas, foram obtidas notícias reais que correspondiam a cada notícia falsa. Para avaliar o corpus, foi realizado um experimento de classificação usando o modelo LSTM, que obteve resultados satisfatórios. No entanto, no corpus *FakePedia*, é possível encontrar notícias com uma similaridade bastante baixa, como pode ser visto na Tabela 2.

Também existem conjuntos de dados de mensagens de notícias falsas divulgadas em redes sociais, como por exemplo, *FakeWhatsApp.BR*³ [Cunha 2021] e *FakeNews-SetGen*⁴ [da Silva et al. 2020]. Estes capturaram mensagens em português brasileiro de notícias falsas do WhatsApp e Twitter, respectivamente. Embora tenham bons resultados ao detectar as mensagens falsas por meio de técnicas de classificação, esses dados estão protegidos por políticas de segurança, o que faz que o conjunto de dados não seja suficientemente grande para realmente generalizar o funcionamento dos classificadores em outros conjuntos.

O conjunto de dados *Fake News Multilabel*⁵ [de Moraes et al. 2019], criado com o mesmo objetivo dos corpus já mencionados, foi utilizado para classificar notícias em diferentes categorias. O autor realizou uma extração de atributos textuais que permitiu identificar a mistura das classes de notícias falsas/legítimas e satíricas/objetivas. Dessa forma, foi possível distinguir se uma história de notícias tinha a intenção de enganar ou

¹<https://github.com/Gabriel-Lino-Garcia/FakeRecogna>

²<https://github.com/andersoncordeiro/Fakepedia-Corpus>

³<https://github.com/cabrau/FakeWhatsApp.Br>

⁴<https://github.com/kamplus/FakeNewsSetGen>

⁵<https://github.com/hugoabonizio/fake-news-multilabel>

se simplesmente se tratava de uma sátira.

Devido à complexidade e ao custo da criação de um corpus público de notícias relevantes, como *FakeBr*, que utiliza verificações manuais e está desatualizado. Além disso, os corpus mais recentes apresentam uma fraca ou nula alinhamento entre notícias falsas e verdadeiras. Apresentamos a construção automática de um corpus que utiliza o modelo Sentence-BERT [Reimers and Gurevych 2019] para a representação do texto, o que nos permite obter uma maior similaridade entre notícias falsas e verdadeiras e uma melhor alinhamento.

3. Criação do corpus FakeTrue

A criação de um conjunto de dados relevante para treinar classificadores, como o *FakeBr* [Silva et al. 2020], é uma tarefa complexa, especialmente se o conjunto precisar ser mantido alinhado.

De acordo com Monteiro et al. [Monteiro et al. 2018], a alinhamento entre notícias verdadeiras e falsas é importante tanto para estudos linguísticos quanto para aprendizado de máquina, pois instâncias positivas e negativas são necessárias para validar padrões linguísticos e de aprendizado de máquina, de acordo com a abordagem adotada. Seguindo as orientações apresentadas em [Rubin et al. 2015] para criar um corpus relevante de notícias falsas, apresentamos o corpus de notícias falsas em português brasileiro *FakeTrueBR*, juntamente com suas respectivas notícias verdadeiras que as desmentem, construído de forma semelhante à utilizada em [Silva et al. 2020].

3.1. Construção do FakeTrueBR

Devido ao fato de que a grande maioria das notícias com conteúdo enganoso são propagadas através das redes sociais e de diferentes portais, é desejável desmentir mensagens enganosas dessas plataformas, já que as notícias falsas podem ser prejudiciais para a sociedade e para a tomada de decisões informadas. No entanto, a coleta e verificação dessas notícias é difícil e cara; por isso, foi utilizado um verificador de notícias falsas conhecido como *Boatos.org*⁶, famoso por desmentir informações enganosas que são consumidas diariamente pela Internet. O *Boatos.org*, ao contrário de outras agências, apresenta em cada uma de suas publicações o conteúdo explícito da mensagem falsa difundida e a justificativa de por que é falsa.

Por isso, Para a criação do nosso corpus, criamos um *crawler* em Python e BeautifulSoup para extrair mensagens falsas do Boatos.org. No total, foram obtidas 4.600 mensagens falsas, das quais 2.083 mensagens atendiam ao requisito de ter mais de 300 caracteres. É importante mencionar que, se toda a publicação do portal fosse considerada, isso reduziria o grau de falsidade das notícias, já que ela contém a razão de sua falsidade.

O próximo passo foi encontrar notícias reais que contrapusessem as 2.083 mensagens das notícias falsas. Para isso, foram extraídas palavras-chave dos títulos de cada notícia falsa, que resumem seu conteúdo principal. Em seguida, foi utilizado um *crawler* semelhante ao anterior, mas adaptado para procurar nas fontes confiáveis *GI*⁷ e *Folha*⁸

⁶<https://www.boatos.org/>

⁷<https://g1.globo.com/>

⁸<https://www.folha.uol.com.br>

notícias verdadeiras que contivessem as palavras-chave extraídas. Esse processo resultou na recuperação de 3.032 notícias verdadeiras, sendo no máximo 10 por notícia falsa.

Depois, foram utilizados embeddings gerados pelo Sentence-BERT multilíngue⁹ [Reimers and Gurevych 2019] para representar o texto das notícias e calcular a medida de similaridade léxica do cosseno. Dessa forma, foi selecionada uma única notícia verdadeira correspondente a cada notícia falsa previamente coletada.

Essa representação do texto permitiu obter resultados mais precisos, pois ela tem um bom desempenho em tarefas de similaridade semântica textual [Andrade Junior et al. 2021], e com isso garantiu-se que cada par de notícias, falsa e verdadeira, estão relacionados tematicamente. Em comparação com outros conjuntos de dados, como se apresenta nas tabelas 1 e 2, essa relação é notável.

Tabela 1. Exemplo de notícia falsa e sua contraparte verdadeira, pertencentes ao corpus construído.

Verdadeiro	Falso
Circula pelas redes sociais uma mensagem que diz que a China busca aprovação de seu Tribunal Superior para matar mais de 2 mil pacientes com coronavírus e evitar a disseminação do vírus é, a embaixada da China no Brasil diz que a mensagem é totalmente falsa, o Ministério da Saúde Brasileiro também afirma que a informação é falsa, já que não há nenhum registro de audiência na suprema corte popular da China sobre esse tema, a mensagem falsa circula também em outros idiomas e já foi alvo de agências de checagem mundo afora a origem da mensagem é um site que tem a fama de divulgar informações falsas.	“China busca aprovação do tribunal para matar mais de 2 pacientes com coronavírus para evitar maior disseminação do vírus”. Fevereiro 6, 2022. o mais alto nível de tribunal em China, o Supremo Tribunal Popular, deve aprovar na sexta-feira a matança em massa de pacientes com coronavírus na China, como forma de controlar a propagação do vírus mortal. O Estado diz ao tribunal que a China está prestes a perder seus profissionais de saúde para o coronavírus, pois pelo menos 2 profissionais de saúde contraem o vírus diariamente. O Estado argumenta que os pacientes com coronavírus admitidos em hospitais só têm suas mortes adiadas e infectam muitos outros enquanto recebem atendimento no hospital.

O conjunto de dados final obtido consiste em 3582 notícias em português, publicadas entre 2017 e 2023 (1791 verdadeiras e 1791 falsas), armazenadas em um arquivo csv¹⁰ e distribuídas em 5 categorias, que são: Brasil, política, entretenimento, saúde e mundo. Além de coletar notícias falsas e verdadeiras, foram extraídos certos metadados como a data de publicação, o link e o autor de cada notícia. Isso foi feito com o objetivo de enriquecer as informações do conjunto de dados.

4. Metodologia

Nesta seção, é descrito o design dos experimentos e as ferramentas utilizadas para detectar notícias falsas por meio de algoritmos de aprendizado de máquina baseados apenas no texto, a fim de avaliar a aplicabilidade do corpus criado.

⁹<https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2>

¹⁰<https://github.com/jpchav98/FakeTrue.Br>

Tabela 2. Exemplos de erros de similaridade entre notícias falsas e suas contrapartes verdadeiras apresentadas em outros corpus.

Verdadeiro	Falso
Significado da marca de roupas GAP é gay e orgulhoso.	Afro Presença: Quais ações precisamos para não aumentar o gap para negros?
Soldado passou três anos na guerra e encontrou esposa grávida no retorno.	A construção do Homem Rousseauiano - Educar para Regenerar

Primeiramente, foi realizado um pré-processamento do texto que incluiu normalização e remoção de palavras vazias e termos que poderiam introduzir vies. Em seguida, foi apresentada uma nuvem de palavras que resume a frequência dos termos tanto no conjunto de notícias falsas quanto no conjunto de notícias verdadeiras, conforme mostrado na Figura 1.



Figura 1. Wordcloud notícias verdadeiras e falsas.

4.1. Representação de texto

Foram realizados experimentos utilizando duas técnicas distintas de representação de texto, Bag-of-Words com frequência (BoW TF) e Bag-of-Words com frequência e ponderação inversa de documentos (BoW TF-IDF), utilizando unigramas, bigramas e trigramas.

4.2. Classificadores

Os experimentos foram conduzidos utilizando vários métodos de classificação, incluindo *logistic regression* (LR), *Random Forest* (RF), *support vector machines* (SVM), *multilayer perceptron* (MLP), *Naive Bayes* (NB), *decision tree* (DT), *k-nearest neighbor* (KNN), *stochastic gradient descent* (SGD) e *Gradient Boosting* (GB). Todos esses métodos foram implementados utilizando as ferramentas fornecidas pela biblioteca scikit-learn.

4.3. Medidas de Avaliação

A avaliação dos classificadores foi realizada utilizando métricas de precisão, acurácia, recuperação e F1-score (1), que é calculado como a média harmônica entre precisão e recuperação, e resume o desempenho de um modelo de classificação [Geron 2019].

$$F1 = 2 \times \frac{\text{precisão} \times \text{recuperação}}{\text{precisão} + \text{recuperação}} \quad (1)$$

Os resultados médios obtidos após a realização de uma validação cruzada de 5-folds são apresentados na próxima seção.

5. Experimentos e resultados.

Foram realizados vários testes para analisar o desempenho de vários algoritmos de aprendizado de máquina ao utilizar diferentes representações de texto, como a técnica BoW e BoW TF-IDF, no corpus criado. A tabela 3 apresenta os resultados mais destacados para cada representação vetorial, considerando diferentes tipos de n-gramas. Em particular, a implementação de TF-IDF com trigramas e um classificador de *Multi Layer Perceptron* (MLP) obteve um F1-score médio de 0.945.

Tabela 3. Resultados experimentais dos melhores classificadores por representação de texto.

	n-grams	Classificador	Accuracy	Precision	Recall	F1-score
BoW TF	unigram	NB	0.890	0.819	1.000	0.901
	bigram	MLP	0.886	0.815	1.000	0.898
	trigram	MLP	0.882	0.809	1.000	0.894
BoW TF-IDF	unigram	NB	0.890	0.819	1.000	0.901
	bigram	MLP	0.931	0.881	0.996	0.935
	trigram	MLP	0.942	0.901	0.993	0.945

Após uma análise detalhada dos resultados das 54 classificações, pode-se concluir que os 10 melhores resultados do experimento utilizaram principalmente a técnica de representação BoW TF-IDF, conforme apresentado na tabela 4. Apesar de ser uma técnica padrão e amplamente utilizada, conseguiu obter resultados significativos ao considerar diferentes n-gramas e ao utilizar os classificadores LR, SVM, NB, SGD e MLP. As pontuações médias F1 obtidas por esses classificadores variaram entre 0.9 e 0.945, o que indica alta eficácia na classificação de textos.

Tabela 4. 10 melhores resultados experimentais.

	n-grams	Classificador	Accuracy	Precision	Recall	F1-score
BoW TF-IDF	trigram	MLP	0.942	0.901	0.993	0.945
	bigram	MLP	0.931	0.881	0.996	0.935
	trigram	SGD	0.931	0.883	0.993	0.935
	trigram	LR	0.931	0.883	0.993	0.935
	trigram	SVM	0.929	0.883	0.990	0.933
	bigram	LR	0.920	0.864	0.996	0.926
	bigram	SGD	0.918	0.862	0.996	0.924
	bigram	SVM	0.918	0.862	0.996	0.924
BoW TF	unigram	NB	0.890	0.819	1.000	0.901
BoW TF-IDF	unigram	NB	0.890	0.819	1.000	0.901

6. Conclusão

Neste estudo, desenvolvemos um corpus de notícias falsas obtidas de um portal de verificação que fornece o texto completo da notícia falsa sem incluir a verificação correspondente. Ao utilizar uma técnica de representação de texto baseada em embeddings Sentence-BERT, simplifica-se a tarefa de relacionar a notícia falsa com uma notícia verdadeira, o que aumenta a confiabilidade da similaridade entre elas. Como resultado,

alcançamos um corpus atualizado e de fácil atualização, que abrange diversas categorias de notícias falsas e é apresentado de forma limpa.

Foram realizados experimentos utilizando duas técnicas simples e populares no processamento de linguagem natural, BoW Tf e BoW Tf-IDF, com diferentes n-gramas e modelos de classificação de machine learning, obtendo resultados interessantes com uma pontuação F1 de 0.945 na melhor classificação.

Este é um corpus de notícias que apresenta uma melhor alinhamento entre as notícias falsas e suas correspondentes notícias verdadeiras. Além disso, é um corpus atualizado que abrange o período de 2017 a 2023 e apresenta as notícias falsas de maneira limpa, ou seja, sem incluir o motivo pelo qual são falsas. Essas são as principais vantagens do nosso corpus em comparação com os atuais, tornando-o um conjunto útil para tarefas de processamento de linguagem natural e pesquisa na área de detecção e classificação de notícias falsas.

No entanto, existem algumas limitações, como a necessidade de verificar manualmente que cada notícia falsa tenha uma notícia verdadeira que negue seus fatos, e o possível viés no comprimento das notícias verdadeiras em comparação com as notícias falsas.

Em trabalhos futuros, planejamos expandir o corpus com mais notícias de diversas fontes, aumentar o limite de similaridade para melhorar a confiabilidade da alinhamento das notícias e desenvolver classificações utilizando modelos pré-treinados para o processamento de linguagem natural. Além disso, neste contexto de notícias falsas, trabalharemos em conjuntos de dados relacionados às redes sociais para tarefas de detecção e propagação de notícias falsas.

Referências

- Andrade Junior, J. E., Cardoso-Silva, J., and Bezerra, L. C. (2021). Comparing contextual embeddings for semantic textual similarity in portuguese. In *Intelligent Systems: 10th Brazilian Conference, BRACIS 2021, Virtual Event, November 29–December 3, 2021, Proceedings, Part II 10*, pages 389–404. Springer.
- Charles, A. C., Ruback, L., and Oliveira, J. (2022). Fakepedia corpus: A flexible fake news corpus in portuguese. In *Computational Processing of the Portuguese Language: 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21–23, 2022, Proceedings*, pages 37–45. Springer.
- Cunha, L. C. C. d. (2021). Fakewhatsapp. br: detecção de desinformação e desinformadores em grupos públicos do whatsapp em pt-br.
- da Silva, F. R. M., Freire, P. M. S., de Souza, M. P., de AB Plenamente, G., and Goldschmidt, R. R. (2020). Fakenewssetgen: A process to build datasets that support comparison among fake news detection methods. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, pages 241–248.
- de Moraes, J., Abonizio, H., Tavares, G., da Fonseca, A., and Barbon, S. (2019). Deciding among fake, satirical, objective and legitimate news: A multi-label classification system. In *Anais do XV Simpósio Brasileiro de Sistemas de Informação*, pages 167–174, Porto Alegre, RS, Brasil. SBC.

- Garcia, G. L., Afonso, L. C., and Papa, J. P. (2022). Fakerecogna: A new brazilian corpus for fake news detection. In *Computational Processing of the Portuguese Language: 15th International Conference, PROPOR 2022, Fortaleza, Brazil, March 21–23, 2022, Proceedings*, pages 57–67. Springer.
- Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor-Flow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, Inc.
- Jain, M. K., Gopalani, D., Meena, Y. K., and Kumar, R. (2020). Machine learning based fake news detection using linguistic features and word vector features. In *2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, pages 1–6. IEEE.
- Mishima, K. and Yamana, H. (2022). A survey on explainable fake news detection. *IEICE TRANSACTIONS on Information and Systems*, 105(7):1249–1257.
- Mishra, S., Shukla, P., and Agarwal, R. (2022). Analyzing machine learning enabled fake news detection techniques for diversified datasets. *Wireless Communications and Mobile Computing*, 2022.
- Monteiro, R. A., Santos, R. L., Pardo, T. A., De Almeida, T. A., Ruiz, E. E., and Vale, O. A. (2018). Contributions to the study of fake news in portuguese: New corpus and automatic detection results. In *Computational Processing of the Portuguese Language: 13th International Conference, PROPOR 2018, Canela, Brazil, September 24–26, 2018, Proceedings 13*, pages 324–334. Springer.
- Netlab (2022). Acompanhamento multiplataforma da desinformação durante as eleições 2022. <http://www.netlab.eco.ufrj.br/blog/acompanhamento-multiplataforma-da-desinformacao-durante-as-eleicoes-2022?categoryId=164721>. Relatório técnico.
- Newman, N., Fletcher, R., Kalogeropoulos, A., Nielsen, R. K., Alves, T., Kus, M., Steemers, J., Fletcher, C., and Vacchiano, G. (2022). Digital news report 2022. Technical report, Reuters Institute for the Study of Journalism, University of Oxford.
- Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.
- Rocha, Y. M., de Moura, G. A., Desidério, G. A., de Oliveira, C. H., Lourenço, F. D., and de Figueiredo Nicolete, L. D. (2021). The impact of fake news on social media and its influence on health during the covid-19 pandemic: A systematic review. *Journal of Public Health*, pages 1–10.
- Rubin, V. L., Chen, Y., and Conroy, N. K. (2015). Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4.
- Silva, R. M., Santos, R. L., Almeida, T. A., and Pardo, T. A. (2020). Towards automatically filtering fake news in portuguese. *Expert Systems with Applications*, 146:113199.
- Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.