



COMPARATIVO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA E APRENDIZADO PROFUNDO NA DETECÇÃO DE FAKE NEWS

Victoria C. M. ATILIO¹; Vinícius A. SILVA²

RESUMO

Este trabalho propõe uma análise comparativa entre técnicas de aprendizado de máquina e aprendizado profundo para a detecção de fake news. Serão utilizados os algoritmos Naive Bayes, SVM e XGBoost como representantes do aprendizado de máquina, e os modelos BERT e BERTimbau como exemplos de aprendizado profundo. O corpus utilizado será o Fake.Br, composto por notícias em português brasileiro. A avaliação dos modelos será realizada utilizando métricas como acurácia, precisão, recall e F1-score. Espera-se que os modelos de aprendizado profundo apresentem desempenho superior, devido à sua capacidade de capturar nuances contextuais e linguísticas.

Palavras-chave:

Classificação; Processamento de Linguagem Natural; Redes Neurais;

1. INTRODUÇÃO

A propagação de informações falsas, popularmente conhecidas como fake news, tornou-se uma questão relevante no cenário contemporâneo, especialmente devido à velocidade e amplitude com que se espalham nas plataformas digitais. Embora a disseminação de notícias falsas tenha ocorrido ao longo da história, a era da internet potencializou esse fenômeno. Sites de notícias e redes sociais contribuem para a rápida difusão de informações, muitas vezes enganosas, criadas com o intuito de manipular a opinião pública ou promover agendas específicas (BBC, 2018).

O impacto das fake news pode ser observado em diversas esferas, como na saúde pública e na política. Por exemplo, o movimento antivacina influenciou diretamente a queda na cobertura vacinal no Brasil, o que resultou em uma maior vulnerabilidade a doenças antes controladas (Saad, 2023). Além disso, as eleições brasileiras de 2022 foram marcadas por uma avalanche de desinformações, como as acusações infundadas sobre a confiabilidade das urnas eletrônicas, que causaram instabilidade política e social (Prado, 2022). Esse cenário evidencia a necessidade de desenvolver mecanismos eficazes para detectar e mitigar a disseminação dessas notícias.

Nesse contexto, a detecção automática de fake news se torna uma ferramenta essencial para lidar com o volume massivo de dados gerados na internet. O Processamento de Linguagem Natural (PLN), combinado com técnicas de Aprendizado de Máquina (AM) e Aprendizado Profundo (AP), oferece soluções promissoras para identificar automaticamente informações falsas. O AM é uma área da Inteligência Artificial focada na criação de algoritmos que aprendem padrões a partir dos dados,

¹Discente do curso de Bacharelado em Ciência da Computação, IFSULDEMINAS – Campus Passos. E-mail: victoria.atilio@alunos.ifsuldeminas.edu.br.

²Docente, IFSULDEMINAS – Campus Passos. E-mail: vinicius.silva@ifsuldeminas.edu.br.

enquanto o AP, uma subárea do AM, se baseia em redes neurais profundas, capazes de capturar nuances complexas em textos (LeCun; Bengio; Hilton, 2015).

Este trabalho busca realizar um comparativo entre técnicas tradicionais de aprendizado de máquina, como Naive Bayes, SVM e XGBoost, e modelos avançados de aprendizado profundo, como BERT e BERTimbau, aplicados à tarefa de detecção de fake news. A análise será conduzida com base no corpus Fake.Br, um conjunto de dados em português brasileiro, e os modelos serão avaliados por métricas de desempenho padrão, como acurácia, precisão, recall e F1-score.

2. MATERIAL E MÉTODOS

Para este estudo, foi utilizado o corpus Fake.Br, criado por Monteiro et al. (2018), composto por 7.200 notícias escritas em português, sendo 3.600 verdadeiras e 3.600 falsas. As notícias verdadeiras foram extraídas de fontes confiáveis, como G1, Folha de S. Paulo e Estadão, enquanto as notícias falsas foram coletadas de websites reconhecidos por divulgar desinformação, como Diário do Brasil e A Folha do Brasil. A seleção cuidadosa das notícias visou garantir uma similaridade estilística entre os textos verdadeiros e falsos como pode ser visto na Tabela 1, o que aumenta a dificuldade da tarefa de classificação e, conseqüentemente, a importância dos modelos avaliados.

Tabela 1. Exemplo de notícias falsas e verdadeiras alinhadas.

Falsa	Verdadeira
<i>Michel Temer propõe fim do carnaval por 20 anos, “PEC dos gastos”. Michel Temer afirmou que não deve haver gastos com aparatos supérfluos sem pensar primeiramente na educação do Brasil. A medida pretende cancelar o carnaval de 2018</i>	<i>Michel Temer não quer o fim do Carnaval por 20 anos. Notícias falsas misturam proximidade dos festejos, crise econômica e medidas impopulares do governo do peemedebista</i>
<i>Acabou a mordomia! Ingresso mais barato pra mulher é ilegal. Baladas que davam meia entrada para mulher, ou até mesmo gratuidade, esto na ilegalidade agora. Acabou o preconceito com os homens nas casas de show de todo o Brasil</i>	<i>Ingresso feminino barato como marketing ‘não inferioriza mulher’, diz juíza do DF. Afirmação consta em decisão sobre preços diferentes para homens e mulheres em festa no Lago Paranoá. ‘Prática permite que mulher possa optar por participar de tais eventos sociais’, diz texto</i>

Fonte: Adaptado de Monteiro et al. (2018).

O corpus foi organizado em seis categorias principais: política, TV e celebridades, sociedade e notícias do cotidiano, ciência e tecnologia, economia e religião. A categoria política, por sua relevância no cenário nacional, é a mais frequente no conjunto de dados. Além disso, o corpus oferece uma versão pré-processada dos textos, em que foram removidas stopwords, acentuação e diacríticos, facilitando a vetorização e análise subsequente.

Os métodos empregados para a detecção de fake news neste trabalho incluem tanto técnicas de Aprendizado de Máquina quanto de Aprendizado Profundo. Para os algoritmos de aprendizado de

máquina, como Naive Bayes, SVM e XGBoost, foi utilizada a vetorização de textos por meio do método TF-IDF (Term Frequency-Inverse Document Frequency). A tokenização foi realizada com a ferramenta CountVectorizer, transformando os textos em uma matriz de frequência normalizada. Esses algoritmos foram selecionados por sua eficiência em tarefas de classificação de textos.

Já para os modelos de aprendizado profundo, foi utilizado o BERT (Bidirectional Encoder Representations from Transformers) e o BERTimbau, uma versão do BERT treinada especificamente para a língua portuguesa. Esses modelos foram aplicados diretamente ao texto, utilizando suas próprias técnicas de tokenização e representação de palavras. O BERT é capaz de capturar relações semânticas mais complexas nos textos, o que o torna uma ferramenta poderosa para a detecção de nuances entre notícias verdadeiras e falsas.

O ambiente de desenvolvimento utilizado foi o Google Colab, devido à sua capacidade de executar código Python com suporte a GPU, o que acelera o treinamento dos modelos de aprendizado profundo. Para os modelos de aprendizado de máquina, foi empregada a biblioteca scikit-learn, enquanto para os modelos de aprendizado profundo foi utilizada a biblioteca Transformers, do Hugging Face, junto com o PyTorch.

A avaliação dos modelos foi feita utilizando 80% dos dados para treinamento e 20% para teste, sendo que, para os modelos baseados em BERT, foi realizada uma validação adicional com 10% dos dados de treinamento. As métricas utilizadas para avaliar o desempenho dos modelos foram acurácia, precisão, recall e F1-score, sendo essas amplamente aceitas para tarefas de classificação.

3. RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta os resultados dos modelos de Aprendizado de Máquina (AM) e Aprendizado Profundo (AP) utilizados para a detecção de fake news.

Tabela 2. Resultados obtidos.

Classificador	F1-Score	Precisão	Recall	Acurácia
Naive Bayes	0.797	0.957	0.683	82,71%
SVM	0.966	0.966	0.966	96,66%
XGBoost	0.956	0.943	0.969	95,56%
BERT	0.898	0.972	0.834	90,56%
BERTimbau	0.947	0.951	0.944	94,79%

Fonte: Autora (2024).

Os resultados indicam que, embora fosse esperado que os modelos de AP tivessem um desempenho superior, os modelos de AM se saíram melhor, com exceção do Naive Bayes. O SVM apresentou a melhor performance geral, seguido pelo XGBoost. O BERTimbau, um modelo de

aprendizado profundo treinado especificamente para português, também teve um bom desempenho, mas ficou atrás dos modelos de AM.

Os resultados deste estudo estão em linha com os achados de Brasil (2021), que testou os algoritmos BERT, Naive Bayes e XGBoost em dois corpus: Fake.br (em português) e Fake News (em inglês). Para o corpus em português, foi utilizado o BERTimbau, e para o corpus em inglês o BERT-large. No corpus em português, o XGBoost se destacou como o modelo mais eficaz, com uma acurácia de 96%, mostrando também uma superioridade dos algoritmos de AM para dados em português. Em contrapartida, no corpus em inglês, o BERT-large se destacou, com 98% de acurácia.

Resultados disponíveis em: <<https://github.com/adena-welt/deteccao-de-fake-news>>

4. CONCLUSÃO

Os modelos de Aprendizado de Máquina, especialmente o SVM e o XGBoost, superaram os modelos de Aprendizado Profundo na detecção de fake news em português, com exceção do Naive Bayes. Isso sugere que, apesar das expectativas, os métodos tradicionais ainda são altamente eficazes para essa tarefa. O estudo destaca a importância de considerar diferentes abordagens para obter os melhores resultados em tarefas de classificação de textos, e reforça a eficácia dos modelos de aprendizado de máquina em contextos específicos.

REFERÊNCIAS

BBC News. **As mensagens falsas usadas no século 16 para tentar sabotar o reinado do espanhol Felipe 2º**. Disponível em: Acesso em 27 de março 2024.

Approaches. **Journal of Information & Knowledge Management**, [S.I], 2024.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep Learning. **Nature**. v. 521, p.436-44, maio 2015.

PRADO, Samantha. **Como as fake news têm impactado as eleições de 2022**. Disponível em: <<https://diplomatie.org.br/como-as-fake-news-tem-impactado-as-eleicoes-de-2022/>>. Acesso em 1 de abril de 2024.

MONTEIRO, Rafael Augusto et al. Contributions to the study of fake news in portuguese: New corpus and automatic detection results. *Computational Processing of the Portuguese Language*, Springer International, p. 324–334, 2018.

SAAD, Maria Amélia. **Projeto indica como reverter queda na cobertura vacinal**. Disponível em: <<https://portal.fiocruz.br/noticia/projeto-indica-como-reverter-queda-na-cobertura-vacinal>>. Acesso em 1 de abril de 2024.

BRASIL, Lucas Cordeiro. **Comparação Entre Modelos com Diferentes Abordagens para Classificação de Fake News**. 2021. 13f. Trabalho de Conclusão de Curso (Graduação) – Universidade Federal de Campina Grande, Campina Grande, 2021.