O USO DA INTELIGÊNCIA ARTIFICIAL NA IDENTIFICAÇÃO DE FAKE NEWS

THE USE OF ARTIFICIAL INTELLIGENCE IN FAKE NEWS DETECTION

Ryan Lucas Pires Campos [1]
Mário Henrique Akihiko da Costa Adaniya [2]

RESUMO

Este artigo possui como principal objetivo analisar métodos de Inteligência Artificial (IA) aplicados para a detecção de *fake news*, sendo uma complicação crescente com o desenvolvimento das redes sociais. Através de uma revisão sistemática da literatura, buscando os algoritmos mais eficientes para a detecção, foram analisados alguns algoritmos como BERT, EANN, Decision Tree, aplicados em datasets como PolitiFact, FakeNewsNet e Weibo, avaliando o desempenho a partir de métricas como F1-Score, recall, precisão e acurácia. Os resultados apresentam que algoritmos baseados em *deep learning* e redes neurais, como BERT e EANN, demonstram melhor eficiência na detecção de *fake news* em comparação com os outros métodos. Com base na análise e nos resultados obtidos, recomenda-se a combinação de diferentes abordagens de IA, pois demonstra uma maior eficiência para detecção de *fake news*.

Palavras-chave: fake news; detecção; inteligência artificial; redes sociais.

ABSTRACT

This article aims to analyze artificial intelligence (AI) methods applied to fake news detection, a growing challenge with the development of social media. Through a systematic literature review, we sought to identify the most efficient algorithms for detection. Various algorithms, including BERT, EANN, and Decision Tree, were evaluated using datasets such as PolitiFact, FakeNewsNet, and Weibo, with performance assessed based on metrics like F1-Score, recall, precision, and accuracy. The results indicate that deep learning and neural network-based algorithms, particularly BERT and EANN, demonstrate superior efficiency in detecting fake news compared to other methods. Based on the analysis and findings, it is recommended to combine different AI approaches, as this demonstrates greater efficiency in fake news detection.

Keywords: fake news; artificial intelligence; detection; social media.

[1] Discente do Centro Universitário Filadélfia de Londrina -UniFil
[2] Docente do Centro Universitário Filadélfia de Londrina -UniFil

1 INTRODUÇÃO

As *fake news*, ou notícias falsas, tornaram-se um grande desafio na atualidade, especialmente devido ao impacto das redes sociais. Essas notícias são criadas para enganar o público e podem influenciar decisões importantes em várias áreas, como saúde, política e economia (Allcott; Gentzkow, 2017). Elas aparecem de diferentes formas e são frequentemente divulgadas com o objetivo de manipular a opinião pública ou gerar confusão.

Os casos de disseminação de *fake news* não ocorrem somente nos dias atuais, no entanto, devido a evolução da tecnologia, houve também uma evolução significativa em relação às *fake news*. Antes de toda a evolução tecnológica, as informações falsas eram disseminadas através de meios tradicionais, como rádios e jornais, consequentemente limitando o alcance da disseminação, porém com o avanço tecnológico, diversos meios de comunicação foram criados, maximizando o alcance, velocidade e ocorrência de *fake news* (Aïmeur; Amri; Brassard, 2023). O desenvolvimento e avanço da tecnologia resultou em uma ascensão na disseminação de notícias falsas, permitindo a criação de *deepfake*, na qual é uma técnica que gera vídeos e imagens realistas de pessoas em situações falsas, permitindo também a síntese de voz, por consequência, gerando vídeos e áudios falsos que são convincentes.

No campo da saúde, por exemplo, informações falsas podem levar a comportamentos prejudiciais, como ocorreu durante a pandemia de COVID-19, na qual houve uma grande disseminação de notícias falsas sobre tratamentos não comprovados, consequentemente ocasionando práticas perigosas, como o uso de medicamentos ineficazes e a rejeição de vacinas seguras, comprometendo a saúde pública no geral (Sharma *et al*., 2020).

Na política, *fake news* têm o poder de influenciar eleições e decisões governamentais ao disseminar desinformação, assim como podem ser usadas para manipular a opinião pública, causando divisões sociais, manipular eleitores, e descredibilizar candidatos e partidos (Allcott; Gentzkow, 2017).

Além disso, na economia, a divulgação de notícias falsas sobre produtos, empresas, ou mercados financeiros podem alterar o valor das ações, provocar movimentos especulativos, e afetar as decisões tomadas sobre investimento (Clarke *et al*., 2020).

As redes sociais possuem um papel significativo na propagação de *fake news*, em razão de sua capacidade de potencializar o alcance das informações. Uma grande parte das plataformas existentes possuem algoritmos projetados para aumentar o engajamento dos usuários, podendo assim, priorizar informações sensacionalistas (Zimmer *et al*., 2019; Aïmeur; Amri; Brassard, 2023). A priorização de conteúdos sensacionalistas, favorece a disseminação de *fake news*, que tendem a gerar mais interações, consequentemente, ocasionando mais engajamento (Domenico *et al*., 2021).

Outros fatores que impactam a disseminação de *fake news* são as “eco chambers” e “bolhas de filtro”. As bolhas de filtro são formadas por algoritmos que personalizam os conteúdos apresentados aos usuários com base em suas interações passadas. Esses algoritmos acabam criando grupos de usuários que possuem as mesmas opiniões e preferências, conhecida como “eco chambers”, consequentemente, sendo um local propício para a disseminação de notícias falsas (Zimmer *et al*., 2019).

Para combater esse problema, muitas ferramentas de Inteligência Artificial (IA) foram criadas. Essas ferramentas usam técnicas como Machine Learning (ML) e Processamento de Linguagem Natural (PLN) para identificar *fake news*. Porém, com tantas opções disponíveis, é difícil saber qual delas funciona melhor (MEESAD, 2021).

Este trabalho tem como objetivo analisar métodos existentes de IA aplicados para detecção de fake news e realizar a avaliação do desempenho em diferentes datasets. Com base nos resultados obtidos, os algoritmos BERT e EANN apresentaram melhor desempenho na maioria dos datasets em que foram aplicados. A análise contribui para ao oferecer uma comparativa entre os métodos analisados, ajudando na escolha das abordagens mais eficazes para identificar fake news.

O artigo está estruturado da seguinte forma: a seção de Fundamentação Teórica apresenta os conceitos e estudos relevantes para detecção de fake news; a

Metodologia descreve o processo de seleção e avaliação dos algoritmos utilizados; no Desenvolvimento, são apresentados os resultados dos algoritmos aplicados nos datasets; a seção de Discussão dos Resultados analisa o desempenho comparativo dos métodos; por fim, a Conclusão resume os principais resultados obtidos e sugere questões para trabalhos futuros.

2 MÉTODOS E MODELOS ANALISADOS

Nesta seção, serão apresentados os principais conceitos e métodos utilizados na detecção de fake news por meio da IA. A partir dessa fundamentação será possível compreender como diferentes abordagens contribuem para melhorar a eficiência na identificação de fake news.

2.1 APRESENTAÇÃO DOS ALGORITMOS

Durante a análise dos estudos selecionados, foram observados diversos métodos de IA utilizados para a detecção de *fake news*, sendo assim, é necessário ter entendimento do funcionamento de cada um. Nessa seção serão apresentados brevemente cada método selecionado.

- **K-Nearest Neighbors (KNN):** O *K-Nearest Neighbors* (KNN) é um algoritmo de classificação e regressão, que realiza a previsão com base na proximidade entre os dados. Sendo assim, o conceito básico é que as instâncias semelhantes estão próximas uma das outras, ou seja, similaridade medida pela distância entre os pontos (Chumachenko *et al*., 2022).
- **Decision Tree:** A *Decision Tree* é um algoritmo supervisionado, na qual cria uma árvore de decisão, começando pela raiz e se ramifica a partir da criação de nós. A comparação entre os dados originais dos atributos com os conjuntos de dados, define o nó consecutivo (Chumachenko *et al*., 2022).
- **Random Forest:** O *Random Forest* é uma técnica que realiza a média dos resultados de várias Árvores de Decisão (*Decision Trees*), ou seja,

diferentemente da *Decision Tree* que possui o resultado somente de uma árvore, na *Random Forest* é calculada a média de várias árvores, sendo considerado uma técnica mais robusta e eficaz (Qadees; Hannan, 2023).

- **Gated Recurrent Unit (GRU):** O *Gated Recurrent Uni*t (GRU) é um tipo de *Recurrent Neural Network* (RNN) projetada para lembrar de informações importantes do passado, utilizando assim, duas portas que auxiliam na decisão do que manter e do que descartar (Seabe; Moutsinga; Pindza, 2023).
- **Rede Neural Convolucional (Convolutional Neural Network):** A *Convolutional Neural Network* (CNN) são algoritmos utilizados em *deep learning*, como por exemplo, reconhecimento de imagens, reconhecimento de voz, entre outras áreas. Esse algoritmo consegue identificar padrões, sem a necessidade de intervenção humana, utilizando conexões locais e compartilhamento de pesos, consequentemente reduzindo a quantidade de parâmetros, acelerando assim o treinamento e melhorando a generalização do modelo (Alzubaidi *et al*., 2021).

- **Stochastic Gradient Descent (SGD):** A *Stochastic Gradient Descent* (SGD) é um algoritmo eficiente de otimização bastante utilizado, principalmente em tarefas de *machine learning*, na qual envolvem grande conjuntos de dados dispersos, como processamento de linguagem natural e classificação de texto. A capacidade e simplicidade de escalar para conjuntos de dados de grande porte são características importantes do SGD, na qual é uma abordagem relativamente comum entre especialistas de *machine learning* (Qadees; Hannan, 2023).

- **Passive-Aggressive:** O algoritmo *Passive-Aggressive* é uma técnica de classificação binária, sendo eficaz em situações quando há novos dados entrando continuamente. Esse algoritmo atribui pesos às previsões, sendo assim, quando uma previsão está correta, o algoritmo ajusta de forma "passiva" e quando está incorreta, o algoritmo ajusta de forma "agressiva", alterando rapidamente os pesos para corrigir os erros (Gurmessa, 2020).

- **Logistic Regression:** A *Logistic Regression* é uma técnica preditiva binária, ou seja, que podem ter apenas duas opções de saída, como "sim" ou "não". Basicamente apresenta como algo que queremos prever se relaciona com outras variáveis que influenciam essa previsão, sabendo disso, a regressão logística é utilizado quando o resultado é categórico, sendo bastante utilizado em *machine learning*, em situações que é necessário classificar dados em duas categorias (Qadees; Hannan, 2023).

- **Naive-Bayes:** *Naive-Bayes* é um método de *machine learning* usado para classificação de dados, utilizado principalmente em classificação de texto. Ele se baseia no teorema de Bayes, uma fórmula famosa de estatística, na qual é usada para calcular probabilidade de algo acontecer, considerando diferentes características dos dados (Qadees; Hannan, 2023).

- **Event Adversarial Neural Networks (EANN):** O modelo EANN é utilizado para detectar notícias falsas que possuem características multimodais, ou seja, informações textuais e visuais. O modelo possui três componentes principais: um extrator de características multimodais, um detector de notícias falsas e um discriminador de eventos. Com base nessas informações, o EANN é capaz de distinguir melhor eventos, ao mesmo tempo que identifica postagens falsas (Wang *et al*., 2018).

- **Bidirectional Encoder Representations from Transformers (BERT):** BERT é um modelo de *deep learning* que foi desenvolvido para pré-treinar representações bidirecionais de texto não rotulado. O diferencial desse modelo é a capacidade de capturar informações contextuais tanto do lado esquerdo quanto do lado direito de uma palavra em uma frase, sendo principalmente utilizado em processamento de linguagem natural (Guo *et al*., 2023).

Com a apresentação dos algoritmos analisados que foram utilizados na detecção de fake news, é possível compreender as técnicas fundamentais que

embasam este estudo. Na próxima seção, serão apresentados e detalhados os critérios e os processos adotados para a seleção e avaliação dos métodos apresentados em diferentes datasets, garantindo assim, uma análise comparativa.

3 METODOLOGIA

Este artigo foi realizado utilizando uma abordagem qualitativa, baseando-se em uma revisão sistemática da literatura, tendo como principal objetivo realizar uma análise das ferramentas de IA mais utilizadas para a detecção de *fake news*. O estudo seguiu uma metodologia que consiste em levantamento bibliográfico, análise de dados, seleção e agrupamento dos dados coletados.

Levantamento Bibliográfico - Esta etapa consistiu em uma revisão bibliográfica da literatura existente sobre o uso da IA na identificação de *fake news*. Entre os estudos revisados, haviam artigos científicos, revistas, revisões sistemáticas, contendo informações sobre métodos e algoritmos que são mais utilizados para esses casos de *fake news*. Foi utilizado ferramentas como *Google Scholar* e *Semantic Scholar* para identificar estudos relevantes para a composição deste artigo, selecionando pesquisas publicadas entre 2015 e 2024, devido ao desenvolvimento na área, aumentando assim, a relevância do tema durante esse período.

Durante essa fase, foram levantadas as seguintes questões de pesquisa definidas no Quadro 1, com base nessas questões, foi possível definir filtros de pesquisa utilizando as palavras-chave “Fake News” “Artificial intelligence” “Detection”, resultando inicialmente em 35.900 estudos envolvendo esses assuntos. Esses filtros permitiram buscar estudos que possuem como principal objetivo a detecção de fake news utilizando IA.

Quadro 1 – Questões de Pesquisa

Questão	Descrição
Q1	Quais os métodos de IA mais utilizados para detecção de *Fake news*?
Q2	Quais *datasets* são utilizados nessas aplicações?
Q3	Quais dos métodos observados possuem a melhor eficiência?

Coleta e Seleção de Dados - Essa etapa envolveu uma coleta de informações sobre a aplicação de algoritmos de IA para a detecção de informações falsas, assim como, a seleção de *datasets* que contêm exemplos de *fake news* e notícias verdadeiras.

A seleção dos dados foi feita com base nos critérios de inclusão estabelecidos pelas questões de pesquisa definidas no Quadro 1 e nos seguintes critérios de exclusão definidos no Quadro 2, com a aplicação desses critérios, houve uma grande redução nos estudos em relação ao número inicial de estudos, na qual 46 foram selecionados para análise detalhada até o momento. Esses estudos foram selecionados, pois fornecem respostas às questões de pesquisa que foram definidas, abordando métodos, algoritmos e datasets utilizados para o objetivo de detectar fake news.

Quadro 2 – Critérios de Exclusão

Critério	Descrição
E1	Estudos que não focam especificamente em algoritmos para detecção de *fake news*.
E2	Estudos não disponíveis em Inglês ou Português
E3	Artigos que não mencionem *datasets* específicos para a detecção de *fake news*

Análise e Agrupamento dos Dados - Os dados foram organizados e classificados de acordo com os campos de IA mencionados, como redes neurais, ML e PLN, e seus respectivos algoritmos, como por exemplo árvores de decisão. A análise comparativa envolveu a avaliação da eficiência, precisão e aplicabilidade dos métodos em diferentes contextos. Técnicas de análise de conteúdo e categorização temática foram utilizadas para examinar as informações obtidas.

4 DESENVOLVIMENTO

A metodologia apresentada anteriormente foi executada em várias etapas, sendo assim, foram analisadas detalhadamente os estudos durante a revisão bibliográfica, seguindo os requisitos apresentados. Cada estudo foi classificado com base nas técnicas utilizadas de IA para detectar *fake news*, considerando várias técnicas e organizando os algoritmos de acordo com suas características e desempenho.

Na etapa de coleta de dados, foram utilizados *datasets* mencionados na literatura, como PolitiFact (Galli *et al*., 2022; Guo *et al*., 2023), FakeNewsNet (GUO *et al*., 2023), Weibo (Jing *et al*., 2023; Wang *et al*., 2018), Twitter (Jing *et al*., 2023; Wang *et al*., 2018), FakeCovid (Iwendi *et al*., 2022) e LIAR (Galli *et al*., 2022). Esses *datasets* contém grande quantidade de notícias verdadeiras e falsas, para que o algoritmos consiga identificar quais notícias são verdadeiras e falsas.

O foco da comparação está na eficiência de cada algoritmo utilizado, sendo utilizado métricas como acurácia, precisão, recall e f1-score.

4.1 MÉTRICAS

Durante a comparação entre os algoritmos, foram utilizadas as seguintes métricas: acurácia, precisão, recall e f1-score. As métricas apresentadas oferecem uma perspectiva diferente sobre a eficiência do algoritmo, sendo assim, é necessário entender a importância de cada um.

Acurácia

É a proporção de previsões corretas feita pelo algoritmo em relação ao total de previsões realizadas. Sendo calculada da seguinte forma:

$$Acurácia = \frac{Número\ de\ Previsões\ Corretas}{Número\ Total\ de\ Previsões} \quad (1)$$

A acurácia indica a frequência em que o modelo classifica corretamente os dados, porém em um conjunto de dados desbalanceado, pode haver enganos (MARIANO, 2021).

Precisão

É a proporção de previsões positivas corretas em relação ao total de previsões positivas realizadas pelo algoritmo. Sendo calculada da seguinte forma:

$$Precisão = \frac{Número\ de\ Verdadeiras\ Positivas}{Número\ de\ Verdadeiras\ Positivas + Número\ de\ Falsas\ Positivas} \quad (2)$$

A precisão indica a capacidade do algoritmo classificar corretamente, e não indicar que uma informação negativa seja positiva (Mariano, 2021).

Recall

É a proporção de previsões positivas corretas em relação ao total de dados positivas reais dentro de um conjunto de dados. Sendo calculada da seguinte forma:

$$Recall = \frac{Número\ de\ Verdadeiras\ Positivas}{Número\ de\ Verdadeiras\ Positivas + Número\ de\ Falsas\ Negativas} \quad (3)$$

Mede a capacidade do algoritmo classificar corretamente todos os dados positivos dentro do conjunto (Mariano, 2021).

F1-Score

O F1-Score combina a precisão e o recall em uma única métrica. Sendo calculada da seguinte forma:

$$F1 - Score = 2 * \frac{Precisão * Recall}{Precisão + Recall} \quad (4)$$

Sendo assim, bastante utilizado para situações onde deseja-se analisar a capacidade do modelo identificar tanto dados positivos quanto a capacidade de não classificar errado dados negativos como positivo (Mariano, 2021).

4.2 APRESENTANDO OS DATASETS

Durante a análise dos estudos selecionados foram observados diversos *datasets* implementados, sendo assim, é necessário entender as características e utilidades principais de cada *dataset*. Cada um desses *datasets* contém um conjunto de informações falsas e verdadeiras, permitindo que haja uma comparação entre os métodos aplicados de IA. Os *datasets* analisados foram:

- **PolitiFact:** Esse conjunto de dados é extraído do site PolitiFact, na qual é uma plataforma de checagem de fatos. Esse dataset é bastante utilizado em estudos sobre detecção de *fake news*. O *dataset* contém notícias verificadas por especialistas, sendo informações falsas ou verdadeiras. As informações desse conjunto de dados, possui 6 tipos de rótulos: Verdadeiro, majoritariamente verdadeiro, meia verdade, majoritariamente falso, falso e totalmente falso. Possui também vários outros detalhes como fonte das informações, data em que foi publicada, entre outros. As colunas deste *dataset* incluem:

Quadro 3 – Informações do Dataset PolitiFact

Coluna	**Tipo**	**Descrição**
verdict	Categórico (String)	Veredicto da verificação de fatos: True, Mostly True, Half True, Mostly False, False, Pants on Fire.
statement_originator	Categórico (String)	Autor da declaração verificada.
statement	Texto (String)	Declaração verificada.
statement_date	Data (Date)	Data da declaração (YYYY-MM-DD).
statement_source	Categórico (String)	Fonte da declaração: discurso, TV, notícias, blog, etc..
factchecker	Categórico (String)	Verificador da alegação.
factcheck_date	Data (Date)	Data da publicação da verificação (YYYY-MM-DD).
factcheck_analysis_link	URL (String)	Link para o artigo verificado.

Fonte: Adaptado pelos autores.

- **FakeNewsNet:** Esse *dataset* é uma união entre duas bases, PolitiFact e GossipCop, oferecendo artigos de notícias classificados como verdadeiros ou falsos. O GossipCop possui como foco a verificação de notícias envolvendo figuras públicas e celebridades. Esse conjunto de dados, possui tanto informações textuais quanto visuais, sendo um dataset multimodal (Shu *et al*., 2018). As colunas deste *dataset* incluem:

Quadro 4 – Informações do Dataset FakeNewsNet

Coluna	**Tipo**	**Descrição**
id	Categórico (String)	Identificador único para cada notícia.
url	URL (String)	URL do artigo na web que publicou a notícia.
title	Categórico (String)	Título do artigo de notícias.
tweet_ids	Categórico (String)	IDs dos tweets que compartilham a notícia, listados em separado por tabulação.

Fonte: Adaptado pelos autores.

- **LIAR:** Esse *dataset* foi extraído a partir de declarações políticas verificadas pelo PolitiFact, tendo como principal utilidade classificar a veracidade de informações públicas. O *dataset* LIAR possui como foco declarações curtas, ou seja, é estruturado para análises rápidas. Os rótulos desse conjunto de dados são semelhantes aos rótulos do *dataset* PolitiFact (Wang, 2017). As colunas deste *dataset* incluem:

Quadro 5 – Informações do Dataset LIAR

Coluna	Tipo	Descrição
id	Categórico (String)	ID da declaração ([ID].json).
label	Categórico (String)	A etiqueta (verdadeiro, falso, etc.).
statement	Texto	A declaração em si.
subjects	Categórico (String)	O(s) assunto(s) abordado(s).
speaker	Categórico (String)	O orador que fez a declaração.
speaker's job title	Categórico (String)	Cargo do orador.
state info	Categórico (String)	Informações sobre o estado.
party affiliation	Categórico (String)	A afiliação política do orador.
total credit history count	Numérico	Contagem total do histórico de crédito, incluindo a declaração atual.
barely true counts	Numérico	Contagem de declarações "pouco verdadeiras".
false counts	Numérico	Contagem de declarações "falsas".
half true counts	Numérico	Contagem de declarações "meio verdadeiras".
mostly true counts	Numérico	Contagem de declarações "majoritariamente verdadeiras".
pants on fire counts	Numérico	Contagem de declarações "totalmente falsas".
context	Texto	O contexto (local/venue) da declaração.

Fonte: Adaptado pelos autores.

- **Weibo:** O conjunto de dados Weibo foi extraído da rede social chinesa chamada Weibo, na qual é utilizado para estudar a disseminação de *fake news*. Esse *dataset* contém informações em chinês, com rótulos que definem se aquela notícia é verdadeira ou falsa (Jing *et al*., 2023). As colunas deste *dataset* incluem:

Quadro 6 – Informações do Dataset Weibo

Coluna	Tipo	Descrição
id_weibo	Categórico (String)	ID único de cada postagem.
data_publicacao	Data (String)	Data e hora da postagem.
conta_autor	Categórico (String)	Conta do autor.
conteudo_texto	Texto (String)	Texto da postagem em chinês.
imagem_weibo	URL (String)	URL da imagem, se houver.
video_weibo	URL (String)	URL do vídeo, se houver.
inclinacao_emocional	Numérico (Int)	Inclinação emocional (0 = Neutro, 1 = Positivo, -1 = Negativo).

Fonte: Adaptado pelos autores.

- **Twitter:** Esse dataset contém 7.021 informações falsas e 5.974 verdadeiras, cada uma contendo textos e imagens. Esse conjunto de dados foi obtido e divulgado pela iniciativa MediaEval Benchmarking, tendo como objetivo avaliar o desempenho de estratégias multimodais (Jing *et al*., 2023). As colunas deste dataset incluem:

Quadro 7 – Informações do Dataset Twitter

Coluna	Tipo	Descrição
tweetId	Categórico (String)	ID único de cada tweet.
tweetText	Texto (String)	Texto do tweet.
userId	Categórico (String)	ID do usuário que postou o tweet.
imageId(s)	Categórico (String)	ID(s) da imagem associada ao tweet, se houver.
username	Categórico (String)	Nome de usuário do autor do tweet.
timestamp	Data (String)	Data e hora da publicação do tweet.
label	Categórico (String)	Classificação do tweet como "fake"ou "real".

Fonte: Adaptado pelos autores.

- **FakeCovid:** As informações deste *dataset* foram coletadas em várias plataformas como Facebook, Twitter, The New York Times, Harvard Health Publishing e a Organização Mundial de Saúde (ONU), entre outros. Contendo várias notícias falsas e reais sobre a COVID-19, tendo 1.164 informações, sendo elas 586 verdadeiras e 578 falsas (Iwendi *et al*., 2022). As colunas deste dataset incluem:

Quadro 8 – Informações do Dataset FakeCovid

Coluna	**Tipo**	**Descrição**
ID	Categórico (String)	Identificador único para cada notícia verificada.
ref_category_title	Categórico (String)	O título da categoria à qual o artigo pertence.
ref_url	URL (String)	O URL do artigo.
verifiedby	Categórico (String)	O nome do site de checagem de fatos que verificou o artigo.
country	Categórico (String)	O país de origem do artigo.
class	Categórico (String)	Se o artigo é verdadeiro, falso ou misto.
published_date	Data (Date)	A data em que o artigo foi publicado.
country1	Categórico (String)	O primeiro país relacionado ao artigo.
country2	Categórico (String)	O segundo país relacionado ao artigo.
country3	Categórico (String)	O terceiro país relacionado ao artigo.
country4	Categórico (String)	O quarto país relacionado ao artigo.
article_source	Categórico (String)	A fonte original do artigo.
ref_source	Categórico (String)	A fonte da referência.
source_title	Categórico (String)	O título da fonte do artigo.
content_text	Texto (String)	O texto completo do artigo.
lang	Categórico (String)	O idioma do artigo.

Fonte: Adaptado pelos autores.

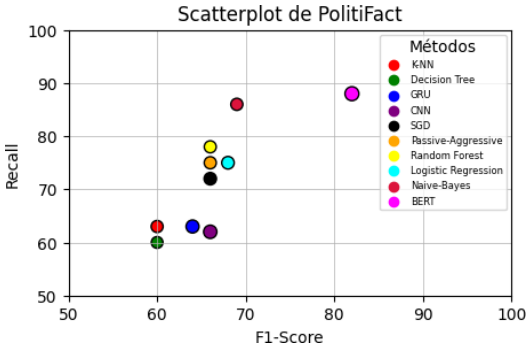
4.4 APRESENTAÇÃO DOS RESULTADOS

Os resultados obtidos a partir da análise dos estudos são apresentados na Tabela 9, mostrando cada técnica utilizada e os respectivos datasets utilizados por cada um, assim como os resultados de cada métrica utilizada para avaliar o desempenho do algoritmo.

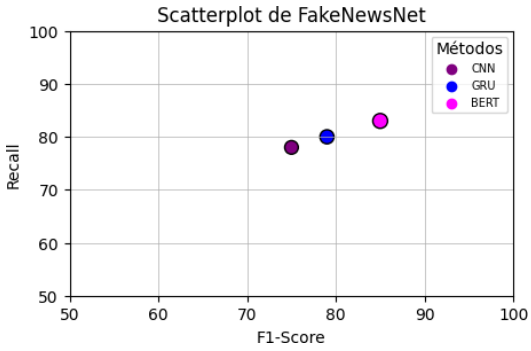
Dataset	**Método**	**Acurácia**	**Precisão**	**Recall**	**F1-Score**	**Fonte**
PolitiFact	K-NN	0,53	0,57	0,63	0,60	(GALLI *et al*., 2022)
	Decision Tree	0,54	0,58	0,60	0,60	(GALLI *et al*., 2022)
	GRU	0,68	0,67	0,63	0,64	(GUO *et al*., 2023)
	CNN	0,66	0,70	0,62	0,66	(GUO *et al*., 2023)
	SGD	0,58	0,61	0,72	0,66	(GALLI *et al*., 2022)
	Passive-Aggressive	0,55	0,58	0,75	0,66	(GALLI *et al*., 2022)
	Random Forest	0,55	0,57	0,78	0,66	(GALLI *et al*., 2022)
	Logistic Regression	0,59	0,62	0,75	0,68	(GALLI *et al*., 2022)
	Naive-Bayes	0,58	0,58	0,86	0,69	(GALLI *et al*., 2022)
	BERT	0,78	0,77	0,88	**0,82**	(GUO *et al*., 2023)
FakeNewsNet	CNN	0,74	0,73	0,78	0,75	(GUO *et al*., 2023)
	GRU	0,79	0,78	0,80	0,79	(GUO *et al*., 2023)
	BERT	0,84	0,87	0,83	**0,85**	(GUO *et al*., 2023)
Weibo	SVM	0,64	0,74	0,57	0,65	(JING *et al*., 2023)
	GRU	0,70	0,67	0,79	0,73	(JING *et al*., 2023)
	CNN	0,74	0,74	0,76	0,74	(JING *et al*., 2023)

	EANN	0,83	0,85	0,81	**0,83**	(WANG *et al*., 2018)
Twitter	SVM	0,53	0,49	0,50	0,50	(JING *et al*., 2023)
	CNN	0,55	0,51	0,60	0,55	(JING *et al*., 2023)
	GRU	0,63	0,58	0,81	0,68	(JING *et al*., 2023)
	EANN	0,71	0,82	0,64	**0,72**	(WANG *et al*., 2018)
FakeCovid	K-NN	0,62	0,73	0,40	0,51	(IWENDI *et al*., 2022)
	Decision Tree	0,68	0,71	0,63	0,66	(IWENDI *et al*., 2022)
	Random Forest	0,83	0,83	0,85	**0,84**	(IWENDI *et al*., 2022)
LIAR	CNN	0,54	0,49	0,27	0,35	(GALLI *et al*., 2022)
	Decision Tree	0,56	0,59	0,62	0,60	(GALLI *et al*., 2022)
	K-NN	0,56	0,59	0,65	0,62	(GALLI *et al*., 2022)
	BERT	0,62	0,58	0,60	0,63	(GALLI *et al*., 2022)
	Random Forest	0,59	0,59	0,79	0,67	(GALLI *et al*., 2022)
	SGD	0,62	0,63	0,74	0,68	(GALLI *et al*., 2022)
	Logistic Regression	0,63	0,63	0,76	0,68	(GALLI *et al*., 2022)
	Naive-Bayes	0,60	0,59	0,87	**0,70**	(GALLI *et al*., 2022)

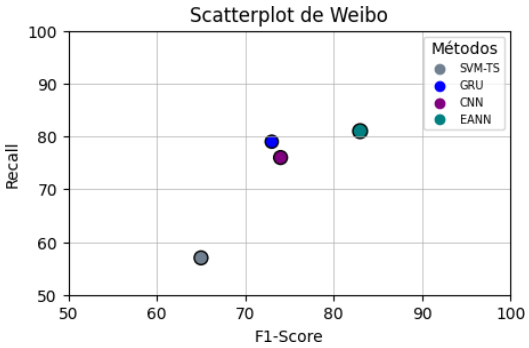
Figura 1 - Scatterplot dos Datasets



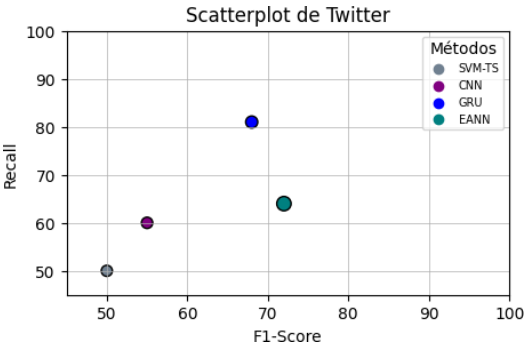
(a) Scatterplot PolitiFact



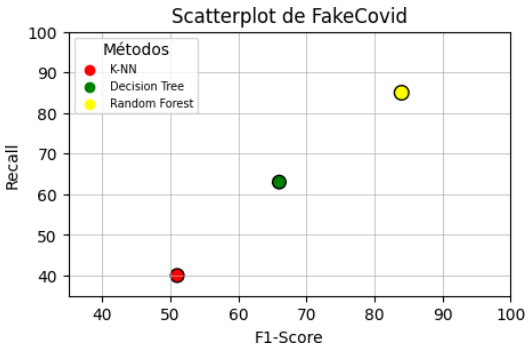
(b) Scatterplot FakeNewsNet



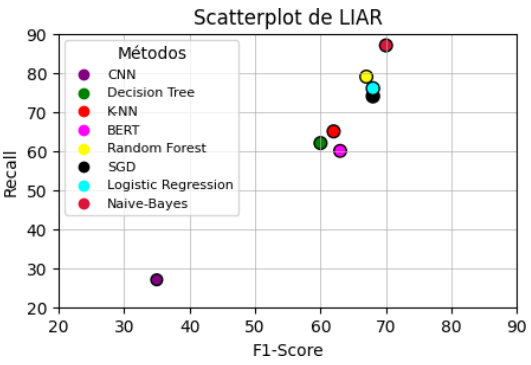
(a) Scatterplot Weibo



(b) Scatterplot Twitter



(a) Scatterplot FakeCovid



(b) Scatterplot LIAR

5 DISCUSSÃO DOS RESULTADOS

De acordo com os resultados obtidos, apresentados na Tabela 9 e nos gráficos *scatterplot* (Figura 1), na qual fornecem informações detalhadas de cada algoritmo aplicado a um determinado *dataset* de detecção de *fake news*, foram utilizadas duas métricas principais: o F1-Score e o Recall. O F1-Score foi escolhido para equilibrar a precisão e o recall, sendo assim, oferecendo uma visão geral da performance dos algoritmos. O Recall foi usado para medir a capacidade dos algoritmos identificarem corretamente as fake news. Nos gráficos, o eixo X representa o F1-Score, o eixo Y representa o Recall, e o tamanho das bolhas estão de acordo com o valor da precisão obtida.

- **PolitiFact:** Os valores do algoritmo BERT se destacam em relação aos outros métodos, obtendo um F1-Score de 0,82, Recall de 0,88 e uma precisão de 0,77. Analisando o *Scatterplot* do PolitiFact, há uma distância considerável entre a bolha do BERT e os outros métodos. Alguns métodos como *Naive-Bayes* e *Logistic Regression* vieram logo após o BERT, tendo F1-Scores de 0,69 e 0,68 respectivamente. Os resultados demonstram que algoritmos baseados em *deep learning* são mais eficazes para esse conjunto de dados.
- **FakeNewsNet:** Igualmente com o *dataset* do PolitiFact, o algoritmo BERT se destacou, tendo um F1-Score de 0,85, Recall de 0,83 e uma precisão de 0,77. A bolha do BERT se destacou novamente em relação aos outros nos gráficos *scatterplot*. Os resultados reforçam que os algoritmos baseados em *deep learning* são mais eficazes na detecção de *fake news*.
- **Weibo:** Neste conjunto de dados, o algoritmo EANN demonstrou ser o mais eficiente, obtendo um F1-Score de 0,83, Recall de 0,81 e precisão de 0,85. A bolha deste *dataset* no respectivo gráfico *scatterplot* está isolado das demais, demonstrando uma maior eficiência em relação aos outros métodos.
- **Twitter:** Novamente o algoritmo EANN se demonstrou superior, obtendo um F1- Score de 0,72, Recall de 0,64 e uma precisão de 0,82. Nesse caso o algoritmo GRU obteve o maior Recall, tendo o valor de 0,81. Sendo assim, leva-

se em consideração o maior F1-Score, pois é composta pela combinação do Recall e precisão. Analisando o gráfico scatterplot é possível identificar que o algoritmo EANN está mais distante em relação ao eixo X (F1-Score), porém no eixo Y (Recall) o algoritmo GRU está isolado do restante dos métodos analisados.

- **FakeCovid:** Neste conjunto de dados, o algoritmo *Random Forest* se destacou, obtendo um F1-Score de 0,84, Recall de 0,85 e uma precisão de 0,83. Nesse *dataset* o *Random Forest* se mostrou muito eficiente em relação aos outros algoritmos analisados, sendo assim, é possível identificar no gráfico *scatterplot* do respectivo conjunto de dados que a bolha do algoritmo está isolado do restante, tanto em F1-Score quanto em relação ao Recall, assim como no tamanho que se refere a precisão.
- **LIAR:** Neste *dataset* o algoritmo *Naive-Bayes* é o mais eficiente com F1-Score de 0,70, Recall de 0,87 e uma precisão de 0,59. Porém nesse caso, o *Logistic Regression* e o SGD também estão com valores muito altos, obtendo F1-Score de 0,68 e precisão de 0,63, sendo assim, os dois algoritmos possuem uma precisão maior que o *Naive-Bayes*. Analisando o gráfico *scatterplot* do conjunto de dados, é possível identificar os três algoritmos facilmente, tendo o *Naive-Bayes* mais isolado e um conjunto de três algoritmos aglomerados, sendo eles o *Logistic Regression*, o SGD e o *Random Forest*.

Com base nos resultados de todos os datasets, juntamente com os gráficos *scatterplots* (Figura 1) e os resultados da Tabela 9, é possível identificar que os **BERT** e **EANN** são os mais eficientes no geral, obtendo F1-Score, Recall e precisão consideravelmente superiores aos outros métodos aplicados. Os algoritmos **K-NN** e **Decision Tree** são os que tiveram um desempenho inferior na maioria dos *datasets* que foram aplicados, especialmente no PolitiFact e LIAR. Portanto, com os resultados analisados, é possível constatar que modelos baseados em redes neurais e *deep learning* possuem uma relevância crescente em relação a detecção de *fake news*.

6 CONCLUSÃO

O objetivo principal deste artigo foi analisar os métodos de IA mais utilizados para a detecção de *fake news*, analisando assim os algoritmos com melhor desempenho e os principais *datasets* aplicados. Realizando uma revisão sistemática da literatura, foi possível identificar que os algoritmos BERT e EANN se destacaram na maioria dos conjuntos de dados que foram empregados, embora não tenham sido aplicados em todos os *datasets* considerados. Utilizando as métricas F1-Score, recall e precisão para avaliar os modelos, esses algoritmos obtiveram o melhor desempenho em relação aos outros métodos que foram analisados.

O algoritmo **BERT** que possui a capacidade de aprendizado contextual bidirecional, demonstrou um grande desempenho quando é aplicado em grandes *datasets*, possibilitando a detecção de padrões sutis em *fake news*, assim como foi demonstrado por Guo *et al*. (2023). Já o **EANN**, um modelo adversarial, foi possível captar a diferença entre informação verdadeira e falsas em diferentes contextos e plataformas de mídia social, possuindo como destaque a capacidade de trabalhar com informações multimodais, na qual inclui não somente texto, mas imagens e vídeos, portanto sendo um diferencial dos outros métodos, assim como foi demonstrado por Wang *et al*. (2018).

Em relação aos datasets, **PolitiFact** e **LIAR** são amplamente utilizados, pois seus dados são verificados por especialistas, facilitando a detecção automática de *fake news*. Esses conjuntos de dados possuem grande volume de dados, permitindo que os métodos que são aplicados, sejam treinados e testados com dados robustos e variados.

6.1 TRABALHOS FUTUROS

Com base na análise dos resultados obtidos, para o desenvolvimento de trabalhos futuros, seria interessante o desenvolvimento de novos datasets que incluem a língua portuguesa, com o propósito de detectar fake news, assim como o desenvolvimento para outras linguagens, para facilitar o entendimento da IA sobre as

diferentes formas que existem de fake news. Além disso, seria interessante realizar novas combinações de algoritmos de IA, melhorando assim, a precisão da IA.

REFERÊNCIAS

AÏMEUR, E.; AMRI, S.; BRASSARD, G. Fake news, disinformation and misinformation in social media: a review. **Social Network Analysis and Mining**, v. 13, 2023. Disponível em: https://api.semanticscholar.org/CorpusID:256701768. Acesso em: 5 mar. 2025.

ALLCOTT, H.; GENTZKOW, M. **Social media and fake news in the 2016 election**. CSN: Politics (Topic), 2017. Disponível em: https://api.semanticscholar.org/CorpusID: 32730475. Acesso em: 5 mar. 2025.

ALZUBAIDI, L. *et al*. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. **Journal of Big Data**, v. 8, 2021. Disponível em: https://api.semanticscholar.org/CorpusID:232434552. Acesso em: 5 mar. 2025.

CHUMACHENKO, D. *et al*. Investigation of statistical machine learning models for covid-19 epidemic process simulation: Random forest, k-nearest neighbors, gradient boosting. **Comput**., v. 10, p. 86, 2022. Disponível em: https://api.semanticscholar. org/CorpusID:249215597. Acesso em: 5 mar. 2025.

CLARKE, J. *et al*. Fake news, investor attention, and market reaction. **Information Systems Research, INFORMS**, v. 32, n. 1, p. 35–52, 2020.

DOMENICO, G. D. *et al*. Fake news, social media and marketing: A systematic review. **Journal of Business Research**, v. 124, p. 329–341, 2021.

GALLI, A. *et al*. A comprehensive benchmark for fake news detection. **Journal of Intelligent Information Systems**, v. 59, p. 237 – 261, 2022. Disponível em: https://api.semanticscholar.org/CorpusID:247596445. Acesso em: 5 mar. 2025.

GUO, Q. *et al*. Tiefake: Title-Text Similarity and Emotion-Aware Fake News Detection. *In*: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2023, China. **Proceedings** […]. China: IEEE, 2023. p. 1–7. Disponível em: https://api.semanticscholar.org/CorpusID: 258212604. Acesso em: 5 mar. 2025.

GURMESSA, D. K. Afaan oromo fake news detection using natural language processing and passive-aggressive. **United International Journal for Research & Technology**, v. 2, n. 2, 2020. Disponível em: https: //api.semanticscholar.org/CorpusID:233238227. Acesso em: 5 mar. 2025.

HOSEA, I. *et al*. A machine learning approach to fake news detection using support vector machine (svm) and unsupervised learning model. **Advances in Multidisciplinary and scientific Research Journal Publication**, 2023. Disponível em: https://api.semanticscholar.org/CorpusID:259545801. Acesso em: 5 mar. 2025.

IWENDI, C. *et al*. Detecting covid-19-related fake news using feature extraction. **Frontiers in Public Health**, v. 9, 2022. Disponível em: https://api.semanticscholar.org/CorpusID:245652860. Acesso em: 5 mar. 2025.

JING, J. *et al*. Multimodal fake news detection via progressive fusion networks. **Information Processing Management**, v. 60, n. 1, p. 103120, 2023. ISSN 0306-4573. Disponível em: https://www.sciencedirect.com/science/article/pii/S0306457322002217. Acesso em: 5 mar. 2025.

MARIANO, D. Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade e f-score. **BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional**, 2021. Disponível em: https://api.semanticscholar.org/CorpusID:240545621. Acesso em: 5 mar. 2025.

MEESAD, P. Thai fake news detection based on information retrieval, natural language processing and machine learning. **Sn Computer Science**, v. 2, 2021. Disponível em: https://api.semanticscholar.org/CorpusID:237278806. Acesso em: 5 mar. 2025.

MOLINA, A. C.; BERENGUEL, O. L. Deepfake: A evolução das fake news. **Research, Society and Development**, 2022. Disponível em: https://api.semanticscholar.org/CorpusID:248823207. Acesso em: 5 mar. 2025.

QADEES, M.; HANNAN, A. Cross comparison of covid-19 fake news detection machine learning models. *In*: INTERNATIONAL CONFERENCE ON OPEN SOURCE SYSTEMS AND TECHNOLOGIES (ICOSST), 17., 2023, Malaysia. **Proceedings** […]. Malaysia: IEEE, 2023. p. 1–7. Disponível em: https://api.semanticscholar.org/CorpusID:266691292. Acesso em: 5 mar. 2025.

SEABE, P. L.; MOUTSINGA, C. R. B.; PINDZA, E. Forecasting cryptocurrency prices using lstm, gru, and bi-directional lstm: A deep learning approach. **Fractal and Fractional**, v.7, n.2, 2023. Disponível em: https://api.semanticscholar.org/CorpusID:257060096. Acesso em: 5 mar. 2025.

SHARMA, K. *et al*. Covid-19 on social media: Analyzing misinformation in twitter conversations. **ArXiv,** arXiv:2003.12309, mar. 2020. Disponível em: https://api.semanticscholar.org/CorpusID:218966796. Acesso em: 5 mar. 2025.

SHU, K. *et al*. Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. **ArXiv**, abs/1809.01286,

set. 2018. Disponível em: https://api.semanticscholar.org/CorpusID:52164698. Acesso em: 5 mar. 2025.

WANG, W. Y. Liar, liar pants on fire: A new benchmark dataset for fake news detection. *In*: ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 55., 2017, Vancouver, Canadá. **Proceedings** […]. Vancouver, Canadá: ACL, 2017. Disponível em: https://api.semanticscholar.org/CorpusID:10326133. Acesso em: 5 mar. 2025.

WANG, Y. *et al*. Eann: Event adversarial neural networks for multi-modal fake news detection. *In*: SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY & DATA MINING, 24., 2018, London. **Proceedings** […]. London: ACM, 2018. p. 849-857. Disponível em: https://api.semanticscholar.org/CorpusID:46990556. Acesso em: 5 mar. 2025.

ZIMMER, F. *et al*. Fake news in social media: Bad algorithms or biased users? **Journal of Information Science Theory and Practice,** v. 7, p. 40–53, 2019. Disponível em: https://api.semanticscholar.org/CorpusID:202779877. Acesso em: 5 mar. 2025.