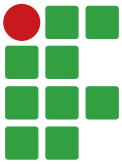

INSTITUTO FEDERAL
DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
Goiás

Bacharelado em Sistemas de Informação

Miguel Ângelo Pinheiro Fernandes

Aprendizado de máquina aplicado à detecção de Fake News em português do Brasil

Luziânia
2025

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

[ ] Tese	[ ] Artigo Científico
[ ] Dissertação	[ ] Capítulo de Livro
[ ] Monografia – Especialização	[ ] Livro
[x] TCC - Graduação	[ ] Trabalho Apresentado em Evento
[ ] Produto Técnico e Educacional - Tipo: _____	

Nome Completo do Autor: Miguel Ângelo Pinheiro Fernandes
Matrícula: 20201080080014
Título do Trabalho: Aprendizado de máquina aplicado à detecção de Fake News em português do Brasil

Autorização - Marque uma das opções

1. (x) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso aberto);
2. ( ) Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data ___/___/_____ (Embargo);
3. ( ) Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG (acesso restrito).

Ao indicar a opção **2 ou 3**, marque a justificativa:

( ) O documento está sujeito a registro de patente.
( ) O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.
( ) Outra justificativa: _____

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

Luziânia GO, 16/07/2025.
Local Data

gov.br Documento assinado digitalmente
MIGUEL ANGELO PINHEIRO FERNANDES
Data: 16/07/2025 13:28:01-0300
Verifique em https://validar.iti.gov.br

_____
Assinatura do Autor e/ou Detentor dos Direitos Autorais

Instituto Federal de Educação, Ciência e Tecnologia de Goiás
Bacharelado em Sistemas de Informação

Miguel Ângelo Pinheiro Fernandes

Aprendizado de máquina aplicado à detecção de Fake News em português do Brasil

Trabalho de Conclusão de Curso 2 apresentado à Coordenação do Curso de Bacharelado em Sistemas de Informação do Instituto Federal de Educação, Ciência e Tecnologia de Goiás - Câmpus Luziânia, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Daniel Rosa Canedo

Coorientador: Dra. Camila de Vasconcelos Tabares

Luziânia

2025

F363a Fernandes, Miguel Ângelo Pinheiro

Aprendizado de máquina aplicado à detecção de fake news em português do Brasil / Miguel Ângelo Pinheiro Fernandes. – Luziânia: IFG, 2025.
66 f. : il. color.

Orientador: Prof. Dr. Daniel Rosa Canedo.
Coorientador: Prof. Dra. Camila de Vasconcelos Tabares.
TCC (Graduação - Bacharelado em Sistemas de Informação).
Instituto Federal de Goiás - IFG, Câmpus Luziânia, 2025.

1. Computação 2. Notícias falsas 3. Aprendizado do computador I. Título. II. Canedo, Daniel Rosa, orient. III. Tabares, Camila de Vasconcelos, coorient.

CDD 004

Bibliotecário: Helio Lino Delfino – CRB-1/3031

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
REITORIA

Miguel Ângelo Pinheiro Fernandes

Aprendizado de Máquina Aplicado à Detecção de Fake News em Português do Brasil

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Bacharelado em Sistemas de Informação do Instituto Federal de Goiás - Campus Luziânia, como parte das exigências do curso de Bacharelado em Sistemas de Informação para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em 19 de fevereiro de 2025

Orientador: Prof. Dr. Daniel Rosa Canedo - IFG

Co-Orientadora : Profa. Dra. Camila de Vasconcelos Tabares

Professora: Profa. Me. Luila Moraes de Oliveira - IFG

Professora: Pofa. Me. Kalinka Martins da Silva - IFG

Luziânia - Goiás - Brasil
Julho - 2025

<Dedico este trabalho à minha mãe, pelo amor incondicional, pela paciência e por todo o apoio que sempre me deu. Sua força e dedicação foram essenciais para a realização deste TCC. Obrigado por sempre acreditar em mim. >

AGRADECIMENTOS

Gostaria de agradecer primeiramente a Deus, que me permitiu chegar ao fim dessa jornada. Afinal, é a conclusão de um ciclo muito importante na minha vida. Gostaria de deixar registrado, do fundo do meu coração, a gratidão que tenho pela minha mãe por ter ficado ao meu lado em todos os momentos da minha vida. Se hoje consigo finalizar esta etapa da minha vida, é por ela e por causa dela. Gostaria de agradecer aos meus colegas de curso, que tornaram essa jornada um pouco mais fácil. Foi um prazer dividir todas as responsabilidades e trabalhos do curso de BSI com Maria Laura, Hugo Kazita e Marco Filipe. Obrigado. E gostaria de finalizar mencionando meus orientadores, tanto o Professor Dr. Daniel Rosa Canedo quanto à Professora Dra. Camila de Vasconcelos Tabares , que são responsáveis pela finalização deste trabalho, cada um em um aspecto. Obrigado por acreditarem no projeto.

"O mundo não se divide em pessoas boas e más. Todos temos luz e trevas dentro de nós. O que importa é o lado o qual decidimos agir. Isso é o que realmente somos!."

(Sirius Black no livro "Harry Potter e a Ordem da Fênix".)

RESUMO

Fernandes, M. Â. P.. **Aprendizado de máquina aplicado à detecção de Fake News em português do Brasil. Luziânia, 2025. 66p. Trabalho de Conclusão de Curso 2. Instituto Federal de Educação, Ciência e Tecnologia de Goiás**

As Fakes News, têm se tornado cada vez mais presentes na sociedade, principalmente com o avanço da internet e das redes sociais. A facilidade de disseminação de informações na era digital tem sido um terreno fértil para a propagação de notícias falsas, que muitas vezes são compartilhadas sem a devida checagem da veracidade dos fatos. Nesse contexto, a inteligência artificial, mais especificamente o aprendizado de máquina, é apontada como uma possível solução para a detecção de Fake News. O aprendizado pode processar grandes quantidades de dados em um curto período, o que os torna uma ferramenta utilizada para a identificação de Fake News. O objetivo deste estudo é realizar uma revisão sistemática sobre Fake News e uma analise qualittiva e quantitativa dos algoritmos de aprendizado de máquina em identificar notícias falsas, e determinar a relevância da sua precisão na detecção de Fake News, contribuindo assim para mitigar os impactos negativos das Fake News na sociedade. Para isso, é realizado um levantamento bibliográfico sobre o tema 'Fake News e aprendizado de máquina', juntamente com a utilização da base de dados Fake.br, que contém 7.200 notícias em português. Como resultado preliminar deste estudo, foi observado que o algoritmo SVM alcançou a maior acurácia, com um valor de 0,95, destacando-se como o mais eficiente entre os modelos testados. Em comparação, a CNN apresentou uma acurácia de 0,89, enquanto o Naive Bayes obteve um desempenho intermediário, com acurácia de 0,92.

Palavras-chave: Fake News; Redes Sociais; Aprendizado de máquina;Fake.br; Inteligência artificial.

LISTA DE ILUSTRAÇÕES

Figura 1 – Distúrbio da informação elaborado por Wardl 19
Figura 2 – Elementos do distúrbio da informação elaborado por Wardle. 21
Figura 3 – Diagrama de aprendizado de máquina.. 33
Figura 4 – Aprendizado de máquina Supervisionado 35
Figura 5 – Aprendizado de máquina não supervisionado 36
Figura 6 – Exemplo Naive bayes Fake News 39
Figura 7 – Arquitetura de uma CNN 40
Figura 8 – Nuvem de palavras 46
Figura 9 – Matriz de confusão do algoritmo SVM. 56
Figura 10 – Curva ROC do algoritmo SVM 57
Figura 11 – Matriz de confusão do algoritmo CNN. 58
Figura 12 – Curva ROC do algoritmo CNN 58
Figura 13 – Matriz de confusão do algoritmo Naive Bayes. 59
Figura 14 – Curva ROC do algoritmo Naive Bayes 60

LISTA DE TABELAS

Tabela 1 – Estudos sobre Detecção de Fake News com Machine Learning - Parte 1 13
Tabela 2 – Estudos sobre Detecção de Fake News com Machine Learning - Parte 2 14
Tabela 3 – Conceitos sobre Fake News - Parte 1 16
Tabela 4 – Conceitos sobre Fake News - Parte 2 17
Tabela 5 – Conceitos sobre Fake News - Parte 3 18
Tabela 6 – Eventos Históricos Relacionados à Fake News - Parte 1 22
Tabela 7 – Eventos Históricos Relacionados à Fake News - Parte 2 23
Tabela 8 – Eventos Históricos Relacionados à Fake News - Parte 3 24
Tabela 9 – Efeito das Fake News na ação coletiva 27
Tabela 10 – Exemplo de stop words 30
Tabela 11 – Exemplo Lematização 31
Tabela 12 – Exemplo Tokenização e Vetorização 32
Tabela 13 – Representação de uma Matriz de Confusão. 43
Tabela 14 – Categorias do conjunto de dados Fake.BI 45
Tabela 15 – Desempenho dos algoritmos de aprendizado de máquina. 60
Tabela 16 – Resultados de trabalhos correlatos 61

SUMÁRIO

1 INTRODUÇÃO 11
1.1 Objetivos 12
1.1.1 Objetivo específico 1: 13
1.1.2 Objetivo específico 2: 13
1.1.3 Objetivo específico 3: 13
1.2 Trabalhos Correlatos 13
1.3 Contribuições esperadas do trabalho 15

2 FUNDAMENTAÇÃO TEÓRICA 16
2.1 Fake news 16
2.1.1 Definição 16
2.1.2 Ecossistema de desinformação 18
2.1.3 Evolução da disseminação de Fake News 22
2.1.4 Impactos na democracia 25
2.1.5 Consequências e impactos na sociedade 26
2.2 Processamento de linguagem natural 29
2.2.1 Definição 29
2.2.2 Remoção de StopWords 30
2.2.3 Lematização 31
2.2.4 Tokenização e Vetorização 31
2.3 Aprendizado de máquina 33
2.3.1 Definição 33
2.3.2 Aprendizado supervisionado 34
2.3.3 Aprendizado não supervisionado 35
2.4 Algoritmos de classificação 37
2.4.1 Máquinas de vetores de suporte (svm) 37
2.4.2 Naive bayes 38
2.4.3 Rede neural convolucional 39
2.5 Métricas de desempenho 41
2.5.1 Acurácia 41
2.5.2 Precisão 42
2.5.3 Recall 42
2.5.4 F1-score 42
2.5.5 Curva roc 43
2.5.6 Matriz de confusão 43

2.6 Banco de dados fake.br corpus . . . 44
2.6.1 Análise exploratória de dados . . . 45

3 METODOLOGIA . . . 47
3.0.1 Etapa 1 . . . 47
3.0.1.1 Fake News . . . 47
3.0.1.2 Aprendizado de Máquina e Fake News . . . 48
3.0.1.3 Estrutura dos Artigos Levantados . . . 48
3.1 Etapa 2 . . . 49
3.1.1 Pré-processamento dos Dados . . . 49
3.1.2 Implementação dos Algoritmos . . . 50
3.1.3 Avaliação dos Modelos . . . 50

4 PROCESSAMENTO E VALIDAÇÃO DOS ALGORITMOS . . . 52
4.1 Processo de treinamento e validação . . . 52
4.1.1 SVM (Support Vector Machine) . . . 53
4.1.2 CNN (Convolutional Neural Network) . . . 53
4.1.3 Naive Bayes . . . 54

5 RESULTADOS . . . 56
5.1 Comparação entre os Algoritmos . . . 56
5.2 Resultado dos Algoritmos . . . 60

6 CONCLUSÕES E TRABALHOS FUTUROS . . . 62

REFERÊNCIAS . . . 64

1 INTRODUÇÃO

As Fake News, ou notícias falsas, são informações enganosas que se apresentam no formato de conteúdo da mídia, mas carecem do mesmo rigor organizacional e intencionalidade. Embora associadas ao avanço da tecnologia e ao mundo digital, as Fake News não são um fenômeno novo; elas são, na verdade, uma forma antiga de disseminação de ideias falsas que podem atingir diferentes escalas, sejam regionais, nacionais ou até internacionais (ITUASSU et al., 2019). Segundo (LAZER et al., 2018), as redes sociais oferecem um terreno fértil para a propagação massiva dessas notícias falsas. No entanto, devido à sua vasta disseminação, ainda não há uma definição concreta e universalmente aceita do termo Fake News, como observa (COSTA; COSTA, 2022).

O impacto das Fake News vai além da desinformação; elas podem influenciar eleições, incitar violência, manipular opiniões públicas e desestabilizar instituições. Em um mundo cada vez mais conectado, onde as informações são consumidas e compartilhadas em questão de segundos, as Fake News se aproveitam da velocidade e da abrangência das plataformas digitais para alcançar um público vasto em pouco tempo. Este fenômeno gera um ciclo de desinformação, onde notícias falsas são amplamente compartilhadas antes que possam ser verificadas, exacerbando a polarização e a desconfiança na mídia tradicional e nas instituições públicas. A natureza viral das Fake News torna o combate a elas um desafio complexo, que exige abordagens inovadoras e eficazes.

Em resposta ao desafio crescente das Fake News, a aplicação de tecnologias de aprendizado de máquina surge como uma das abordagens mais promissoras para sua identificação e combate. Com o volume de informações na internet aumentando exponencialmente, torna-se cada vez mais difícil para os usuários discernir entre o que é verdadeiro e o que é falso. De fato, a quantidade de informações não necessariamente reflete sua qualidade ou veracidade, como ressalta (OLIVEIRA, 2020). Portanto, o uso de algoritmos avançados para filtrar e validar informações é não apenas relevante, mas crucial para garantir a integridade da informação consumida pelas pessoas.

O aprendizado de máquina se destaca pela sua capacidade de analisar grandes volumes de dados e identificar padrões que seriam impossíveis para humanos detectarem em tempo hábil. Em relação às Fake News, essa tecnologia pode ser empregada para desenvolver modelos que não apenas detectem a veracidade das informações, mas também identifiquem as fontes de desinformação e os mecanismos de disseminação. Esses modelos podem ser treinados para reconhecer características típicas de notícias falsas, como sensacionalismo, falta de fontes confiáveis e estrutura textual manipuladora. Além disso, com o uso de técnicas de processamento de linguagem natural (NLP), o aprendizado de máquina pode

avaliar o contexto em que a informação é apresentada, ajudando a distinguir entre informações genuínas e enganosas.

Este projeto, direcionado à análise quantitativa e qualitativa dos principais algoritmos utilizados na detecção de Fake News, busca aprofundar a compreensão da complexa interação entre notícias falsas, tecnologia e aprendizado de máquina. As Fake News, um fenômeno global, representam um desafio multifacetado: por um lado, a tecnologia acelera sua disseminação; por outro, fornece ferramentas indispensáveis para combatê-las de maneira eficiente. Nesse contexto, o aprendizado de máquina emerge como uma solução promissora, automatizando a identificação de padrões linguísticos e contextuais que caracterizam notícias falsas, enquanto responde de forma ágil e precisa a essa problemática.

A pesquisa realizada concentra-se na avaliação detalhada do desempenho de diferentes algoritmos de aprendizado de máquina aplicados à detecção de Fake News. Além de medir métricas como acurácia e eficácia, o estudo investiga as limitações desses modelos, especialmente em cenários onde a linguagem das notícias apresenta ambiguidade ou está fortemente influenciada por contextos culturais. A análise qualitativa desempenha um papel crucial ao examinar como os algoritmos lidam com nuances linguísticas, enquanto a análise quantitativa permite comparar objetivamente os resultados, oferecendo uma visão clara do desempenho de cada abordagem.

Para garantir um entendimento abrangente, o projeto explora algoritmos com diferentes características e complexidades, como Naive Bayes, Máquinas de Vetores de Suporte (SVM) e Redes Neurais Convolucionais (CNN). Cada modelo é analisado não apenas por sua capacidade técnica, mas também pela simplicidade de implementação e adaptabilidade a diversos contextos. Essa abordagem comparativa visa identificar não apenas as forças, mas também as fraquezas de cada método, permitindo uma avaliação equilibrada de sua aplicação prática.

Concluindo, ao realizar uma análise integrada — tanto quantitativa quanto qualitativa —, o projeto busca contribuir para um entendimento mais profundo das ferramentas disponíveis no combate às Fake News. Esse esforço não apenas evidencia o potencial do aprendizado de máquina como uma solução eficaz, mas também destaca suas limitações, fornecendo uma base sólida para futuras investigações e aprimoramentos. Assim, o estudo estabelece um ponto de partida para avanços na detecção automatizada de Fake News, com aplicações que vão além do contexto atual.

1.1 Objetivos

Avaliar, de forma qualitativa e quantitativa, a eficácia dos principais algoritmos de aprendizado de máquina na identificação de Fake News, destacando seus pontos fortes, limitações e relevância no contexto atual. Para alcançar este objetivo geral, foram definidos

os seguintes objetivos específicos:

1.1.1 Objetivo específico 1:

- Conduzir uma revisão sistemática abrangente sobre o fenômeno das Fake News e o uso de técnicas de aprendizado de máquina para sua detecção.

1.1.2 Objetivo específico 2:

- Empregar o conjunto de dados Fake.br como referência principal para a avaliação precisa de algoritmos de detecção de Fake News.

1.1.3 Objetivo específico 3:

- Conduzir testes comparativos para avaliar, de forma precisa, a eficácia dos algoritmos de aprendizado de máquina na distinção entre notícias verdadeiras e falsas, ressaltando seu desempenho e potencial impacto na identificação de Fake News.

1.2 Trabalhos Correlatos

Tabela 1 – Estudos sobre Detecção de Fake News com Machine Learning - Parte 1

Item	**Título**	**Autor(es)**
01	Fake News Detection Using Machine Learning Approaches: Este estudo analisa pesquisas sobre detecção de Fake News e utiliza modelos de aprendizado de máquina supervisionado para desenvolver um produto capaz de classificar notícias como verdadeiras ou falsas. Usando Python, scikit-learn e NLP, o processo envolve a extração de características e vetorização de dados textuais, com o objetivo de identificar as melhores características que resultem na maior precisão na detecção de fake news.	Z Khanam 1, BN Alwasel 1, H Sirafi 1 e M Rashid 2
02	Fake News Detection using Machine Learning Algorithms: Este artigo aborda a classificação binária de artigos de notícias online utilizando Inteligência Artificial, Processamento de Linguagem Natural e Aprendizado de Máquina. O objetivo é permitir que os usuários classifiquem as notícias como falsas ou verdadeiras e verifiquem a autenticidade dos sites que publicam essas notícias, mitigando os efeitos da disseminação de Fake News nas mídias sociais.	Uma Sharma, Sidarth Saran, Shankar M. Patil

Elaboração própria do autor.

Tabela 2 – Estudos sobre Detecção de Fake News com Machine Learning - Parte 2

Item	**Título**	**Autor(es)**
03	Fake Detect: A Deep Learning Ensemble Model for FakeNews Detection: Este estudo aborda a necessidade de um sistema automatizado para classificar Fake News, dada a crescente disseminação de informações enganosas nas mídias sociais. Utilizando o conjunto de dados "LIAR"e técnicas de aprendizado de máquina, o trabalho visa eliminar boatos e ajudar a identificar a confiabilidade das fontes de notícias, destacando a importância de tal abordagem no contexto da crescente digitalização global.	Nida Aslam, Irfan Ullah Khan, Farah Salem Alotaibi, Lama Abdulaziz Aldaej, Asma Khaled
04	Fake News Detection Using Machine Learning Ensemble Methods: Este estudo explora a evolução e os desafios da disseminação de Fake News nas mídias sociais e outras plataformas digitais. Aborda técnicas computacionais, incluindo aprendizado de máquina, para detecção e classificação de Fake News. Métodos como processamento de linguagem natural (NLP) e análise de redes sociais são empregados para identificar e diferenciar notícias falsas de verdadeiras. O estudo também destaca a importância de abordagens híbridas, combinando características textuais e contextuais para melhorar a precisão na detecção de Fake News.	Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, Muhammad Ovais Ahmad
05	Fake news detection in Urdu language using machine learning: Este estudo aborda a detecção de fake news em notícias em urdu, destacando a falta de robustez em abordagens existentes para conjuntos de dados multi-domínio e o problema de traduções automáticas sem verificação manual. O trabalho propõe um classificador utilizando técnicas como TF-IDF, BoW e n-grams, com empilhamento de características, combinando vetores de texto e verbos extraídos. Modelos de machine learning como SVM, k-NN, RF e ET foram utilizados, com o empilhamento aplicado para melhorar a precisão da detecção.	Muhammad Shoaib Farooq 1, Ansar Naseem 1, Furqan Rustam 2, Imran Ashraf 3

Elaboração própria do autor.

Nas tabelas 1 e 2 apresenta, na primeira coluna, os títulos de cada trabalho; na segunda coluna, um resumo sobre o conteúdo de cada trabalho; e na terceira coluna, os autores relacionados a cada estudo. De modo geral, o quadro desempenha um papel crucial na construção deste Trabalho de Conclusão de Curso (TCC). Ele reúne e sintetiza os principais estudos relacionados ao tema da detecção de Fake News utilizando aprendizado de máquina, destacando as metodologias e abordagens utilizadas por diferentes autores

em pesquisas anteriores. A inclusão deste quadro é fundamental, pois oferece uma base sólida sobre a qual o presente trabalho foi desenvolvido. A análise dos trabalhos correlatos permitiu identificar lacunas nas abordagens existentes, bem como extrair insights valiosos que direcionaram a escolha das técnicas e ferramentas aplicadas nesta pesquisa. Assim, este quadro não apenas contextualiza a relevância do tema, mas também justifica a originalidade e a contribuição deste TCC ao campo de estudo, servindo como ponto de partida para a construção teórica e prática do texto.

1.3 Contribuições esperadas do trabalho

Este trabalho visa contribuir academicamente com as temáticas de combate a propagação de Fake News, abordando temas de grande relevância nos últimos anos e que provavelmente continuarão a ser debatidos por muito tempo. As Fakes News, ou notícias falsas, se tornaram uma ameaça global, impactando diversas esferas da sociedade, incluindo a política, a saúde e a economia. A análise crítica desses fenômenos e a busca por soluções são essenciais para entender e enfrentar os desafios que a desinformação por meio das Fake News representam.

Mitigar o compartilhamento massivo de desinformação na era digital é uma questão de grande relevância para a sociedade. Este estudo tem como foco realizar uma análise qualitativa e quantitativa da eficácia de algoritmos de aprendizado de máquina na identificação de Fake News, fornecendo uma base sólida para compreender seu desempenho e limitações. Ao contribuir com dados e insights precisos, espera-se promover um ambiente informativo mais confiável e fomentar a conscientização crítica frente à disseminação de informações falsas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Fake news

Este capítulo está organizado da seguinte forma: inicialmente, abordaremos a definição do termo Fake News, destacando as dificuldades e complexidades envolvidas na conceitualização desse conceito. Em seguida, explorarando o ecossistema da informação, detalhando como ocorre a criação e o compartilhamento dessas notícias, analisando os mecanismos e os atores envolvidos nesse processo. Conectando a evolução das Fake News ao longo dos anos, desmistificando a ideia equivocada de que elas são um problema exclusivamente contemporâneo e mostrando como esse fenômeno tem raízes históricas e se transformou ao longo do tempo. Posteriormente, discutiremos os impactos das Fake News nas instituições democráticas, examinando as consequências e os desafios que elas impõem à governança e à integridade dos processos democráticos. Por fim, analisaremos os efeitos das Fake News na sociedade, incluindo os impactos sociais, culturais e psicológicos, abordando como a disseminação de desinformação afeta a confiança pública e a coesão social. Esta estrutura permite uma compreensão aprofundada e organizada do fenômeno das Fake News, seus antecedentes históricos, sua dinâmica atual e suas implicações para a sociedade e a democracia.

2.1.1 Definição

O artigo de (COSTA; COSTA, 2022) destaca a complexidade em definir de forma clara o termo Fake News, o qual é frequentemente utilizado em diferentes contextos no dia a dia das pessoas. É comum observar que, em certas situações, esse termo é empregado de forma menos grave, enquanto que em outras, é associado a práticas criminosas. Independentemente da conceituação, a disseminação de Fake News deve ser tratada com cautela, uma vez que em nenhuma análise foi considerada como algo positivo ou como uma forma legítima de expressão de opinião.

Tabela 3 – Conceitos sobre Fake News - Parte 1

Item	**Conceito**	**Autor(es)**
01	Fake News representa informações de várias vertentes que são apresentadas como reais, mas são claramente falsas, fabricadas, ou exageradas ao ponto em que não mais correspondem à realidade; além disso, a informação opera no interesse expresso de enganar ou confundir um alvo ou audiência imaginada.	Reilly, 2018, citado por Menezes, 2018, p. 49

Tabela 4 – Conceitos sobre Fake News - Parte 2

Item	**Conceito**	**Autor(es)**
02	O termo Fake News é agora comumente aplicado para histórias enganosas, espalhadas de forma maliciosa por fontes que se fingem legítimas.	Torres e Gerhard, 2018, citado por Menezes, 2018, p. 49
03	Fake News se apresentam como sites que deliberadamente publicam farsas, propagandas e desinformação que se pretende como notícias verídicas, usualmente utilizando redes sociais para dirigir tráfego online e ampliar seu efeito.	Tan e Ang, 2017, citado por Menezes, 2018, p. 49
04	Fake news são coisas inventadas, manipuladas para parecerem notícias jornalísticas críveis, facilmente espalhadas online para audiências amplas propensas a acreditar nas ficções.	Klein e Wueller, 2017, citado por Menezes, 2018, p. 49
05	Fake news são notícias falsas nas quais existe uma ação deliberada para enganar os consumidores. Não coincide com o conceito de false news, que parte de incompetência ou irresponsabilidade.	Menezes, 2018, p. 40
06	Divulgação de informação inverídica com potencial para exercer influência difusa em qualquer grupo social ou pessoa, incluindo o compartilhamento em redes sociais.	Alves e Maciel, 2020
07	Criar, divulgar ou compartilhar, no ano eleitoral, fatos sabidamente inverídicos sobre pré-candidatos, candidatos ou partidos, capazes de influenciar o eleitorado.	Alves e Maciel, 2020
08	Divulgar informação ou notícia falsa que possa modificar a verdade com relação à saúde, segurança pública, economia ou processo eleitoral.	Alves e Maciel, 2020
09	Notícias falsas capazes de provocar atos de hostilidade e violência contra o governo.	Alves e Maciel, 2020
10	Fake news são notícias intencionalmente falsas e passíveis de verificação, criadas para difundir informações falsas com objetivo econômico ou ideológico.	Pacheco et al., 2022
11	São histórias falsas que, ao manterem a aparência de notícias jornalísticas, são disseminadas para influenciar posições políticas ou como uma piada.	Justen et al., 2022
12	Fake News são conteúdos inverídicos, distorcidos ou fora de contexto, espalhados como notícias reais para promover propositalmente a desinformação do público.	De Melo Lobo, 2018

Tabela 5 – Conceitos sobre Fake News - Parte 3

Item	**Conceito**	**Autor(es)**
13	Notícias que divulgam conteúdo informacional falso para atrair público e tornarem-se virais, enganando pela plausibilidade e transformando parte do público em propagadores.	Dos Anjos et al.
14	Informações falsas na forma de artigos, imagens ou vídeos apresentadas como reais, manipulando a opinião pública.	Ionos, 2020
15	Fake news são informações formadas por dizeres distorcidos e/ou incompletos, que promovem a desinformação, intencionalmente deliberada.	Wardle; Derakhshan, 2019
16	Fake news agem contra a democracia em toda parte do planeta, sendo uma nova modalidade de mentira.	Bucci, 2019
17	Fake news são "informações de combate", disseminadas com objetivo de convencimento e fortalecimento de posições em disputas narrativas polarizadas.	Ribeiro; Ortellado, 2018
18	Fake news visam desconstruir o próximo e causar impacto, manipulando informações e a opinião pública.	Fonseca; Ravache, 2021

Fonte: Adaptado do estudo de Da Costa (2022).

As Tabelas 3, 4 e 5 , onde a primeira coluna apresenta os itens, a segunda coluna lista os conceitos, e a terceira coluna fornece os respectivos autores. O quadro como um todo é organizado para refletir as perspectivas de diversos autores sobre a definição e compreensão de Fake News, oferecendo uma visão abrangente sobre a complexidade da conceitualização de Fake News. Ele evidencia que não é possível afirmar que uma única definição está correta, pois cada uma aborda apenas uma perspectiva limitada do fenômeno. Isso ressalta a necessidade de um estudo que consiga dimensionar e definir de maneira abrangente o que é uma Fake News. Se no ambiente acadêmico existe uma pluralidade de definições fragmentadas, como se pode esperar que uma pessoa leiga consiga identificar uma Fake News de forma eficaz? Essa diversidade de interpretações contribui para a criação de um ecossistema de desinformação, onde a identificação e a compreensão de Fake News se tornam ainda mais desafiadoras.

2.1.2 Ecossistema de desinformação

De acordo com(WARDLE; DERAKHSHAN, 2017), a conceituação de "notícias falsas"é insuficiente para descrever a complexidade do fenômeno da produção, disseminação e consumo de informações enganosas. Em vez disso, o autor propõe uma nova abordagem que o compreende como um ecossistema de desinformação. (WARDLE; DERAKHSHAN, 2017) argumenta que, embora o impacto histórico de rumores e conteúdo fabricado tenha sido amplamente documentado, a tecnologia social contemporânea tem criado uma

realidade: a poluição da informação em escala global. Este fenômeno é caracterizado por uma complexa rede de motivações que impulsionam a criação, a disseminação e o consumo de informações "poluídas".

O termo ecossistema de desinformação tem ganhado força na literatura acadêmica e no discurso público, pois ajuda a descrever a complexidade do fenômeno e a necessidade de abordá-lo de maneira abrangente. A abordagem reconhece que a produção e a disseminação de informações enganosas são influenciadas por diversos fatores, incluindo a tecnologia, as motivações dos criadores e divulgadores, as dinâmicas sociais e políticas e as percepções e expectativas dos consumidores.

Nesse sentido, (WARDLE; DERAKHSHAN, 2017) propõe uma nova estrutura conceitual para analisar o distúrbio da informação, identificando três tipos diferentes: informação incorreta, desinformação e má informação. Com base nas dimensões de dano e falsidade, é possível descrever as diferenças entre esses três tipos de informações.

Figura 1 – Distúrbio da informação elaborado por Wardl

Na Figura 1 apresentada por (WARDLE; DERAKHSHAN, 2017) , o conceito de "desordem informacional"é organizado em três categorias principais: informação incorreta, desinformação, e má-informação, que se sobrepõem em uma interseção central chamada "des-informação".

1. Informação incorreta refere-se a quando informações falsas são compartilhadas sem a intenção de prejudicar, muitas vezes ocorrendo em casos de boatos ou enganos inocentes. Isso é representado no lado esquerdo do diagrama. 2. Desinformação, situada na interseção central do diagrama, é o ponto onde informações falsas são conscientemente

compartilhadas com a intenção de causar danos. A interseção entre os conjuntos "falso"e "nocivo"exemplifica a complexidade desse tipo de conteúdo, que pode incluir falso contexto, conteúdo impostor, ou conteúdo fabricado.

3. Má-informação, localizada no lado direito do diagrama, ocorre quando informações verdadeiras são compartilhadas com o intuito de causar danos. Isso pode incluir o vazamento de informações, assédio, ou discurso de ódio. Essa representação visual ajuda a esclarecer como diferentes formas de conteúdo podem ser categorizadas com base em sua veracidade e nas intenções por trás de sua disseminação. A interseção entre essas categorias reflete a complexidade das motivações e estratégias utilizadas na criação e propagação de informações enganosas, o que é crucial para a elaboração de políticas públicas e estratégias mais eficazes no combate à desinformação.

Nesse sentido, (WARDLE; DERAKHSHAN, 2017) fez uma constatação importante de que as Fake News não são apenas notícias falsas jogadas na internet sem motivo aparente. Pelo contrário, sua fabricação ocorre de acordo com a necessidade do agente que criou a desinformação, visando alcançar seus objetivos, sejam eles políticos, econômicos ou de outra natureza. Isso destaca a complexidade do fenômeno e a necessidade de ferramentas capazes de identificar e combater a disseminação dessas notícias falsas. Para entender a desordem da informação, é necessário analisar separadamente seus três elementos – o agente, a mensagem e o intérprete.

A Figura 2 apresenta uma categorização detalhada dos componentes envolvidos na disseminação de Fake News segundo Wardle, que são organizadas em três áreas principais: Agente, Mensagem e Intérprete. Para ilustrar essa estrutura teórica, o caso da "Bruxa do Guarujá"é um exemplo emblemático de como a desinformação pode levar a consequências trágicas.

Em maio de 2014, Fabiane Maria de Jesus foi brutalmente linchada e morta por moradores do Guarujá, São Paulo, após boatos se espalharam nas redes sociais, acusando-a falsamente de sequestrar crianças para rituais de magia negra. A origem dessa tragédia foi uma postagem na página do Facebook "Guarujá Alerta", que alertava a população local sobre a presença de uma mulher supostamente envolvida nesses crimes. A informação falsa se espalhou rapidamente, provocando pânico na comunidade e culminando na morte injusta de Fabiane. Esse caso trágico expõe a perigosa combinação de desinformação e ação coletiva impulsionada pelo medo e pela histeria (MARCH et al., 2022).

No que tange ao Agente, o principal difusor da desinformação foi a página do Facebook "Guarujá Alerta", caracterizada como um tipo de ator não oficial. Essa plataforma, sem qualquer autoridade formal, desempenhou um papel central na propagação do boato que culminou em um linchamento. A disseminação ocorreu em uma organização dispersa, composta por uma rede de indivíduos conectados pelas redes sociais, sem uma estrutura hierárquica clara. A motivação para espalhar o boato foi predominantemente social e

Figura 2 – Elementos do distúrbio da informação elaborado por Wardle.

psicológica, impulsionada pelo medo coletivo e por reações emocionais, sem indicativos de ganhos financeiros ou políticos. O nível de automação foi humano, com indivíduos compartilhando postagens e comentando nas redes sociais, ampliando o alcance do boato. A audiência desejada eram membros específicos da comunidade local, alertados sobre um perigo fabricado. A intenção de prejudicar era evidente, pois o objetivo era incitar pânico e mobilizar a população contra um suposto perigo, sendo também clara a intenção de enganar, uma vez que a informação divulgada era totalmente falsa.

Em relação à Mensagem, esta teve uma duração de curto prazo, mas com impacto devastador, resultando no linchamento de Fabiane Maria de Jesus em poucos dias. A

precisão da informação era enganosa e manipuladora, pois o boato, embora totalmente falso, foi apresentado de maneira a parecer autêntico. A motivação por trás da mensagem, apesar de aparentemente tentar "proteger"a comunidade, era ilegal, dado que resultou em atos de violência. O tipo de impostor utilizado na mensagem envolveu a marca, ao empregar imagens e um retrato falado para identificar falsamente a vítima como uma impostora. O alvo da mensagem era individual, centrado na figura de Fabiane Maria de Jesus, injustamente identificada e atacada pela comunidade.

No que concerne ao Intérprete, os moradores do Guarujá interpretaram a mensagem de forma hegemônica, aceitando-a sem questionamento, e ao mesmo tempo oposicionista em relação à vítima, vista como uma ameaça. A ação tomada foi a de compartilhar a mensagem em apoio à falsa narrativa, replicando a desinformação nas redes sociais e culminando no trágico evento de linchamento. Essa análise do caso da "Bruxa do Guarujá"evidencia a complexidade das Fake News e demonstra como elas podem ser analisadas sob diferentes perspectivas para melhor compreender e combater a desinformação. O caso também destaca a necessidade de que as estratégias de combate a essas notícias falsas evoluam de maneira constante para lidar com os desafios que surgem na era da informação digital. Surge então a questão: Fake News é um problema contemporâneo ou uma manifestação modernizada de uma prática antiga?

2.1.3 Evolução da disseminação de Fake News

Segundo (DARNTON, 2023) a história das Fake News remonta à antiguidade, com exemplos de boatos e rumores que circulavam em sociedades pré-modernas, verificando que notícias falsas não são novidade e têm raízes antigas. Elas não estão ligadas à tecnologia moderna, mas sim às crenças e valores individuais, independentemente da época. Entender este contexto histórico é fundamental para compreender a influência das Fake News na sociedade. A verdade é que o compartilhamento de informações falsas sempre existiu e é importante estarmos atentos a elas, independentemente da época e da tecnologia disponível.

Tabela 6 – Eventos Históricos Relacionados à Fake News - Parte 1

Título	**Fake News**
Cerca de 44 aC – campanha de difamação de Marco Antônio	A campanha de propaganda de Otaviano contra Antônio implantou slogans dignos do Twitter gravados em moedas para manchar a reputação de Antônio.
1835 – A Grande Farsa da Lua	O New York Sun publicou seis artigos sobre a descoberta de vida (inexistente) na lua, alegando recontar as descobertas do astrônomo Sir John Herschel.

Fonte: Adaptado do estudo "A short guide to the history of 'Fake News' and disinformation"de Darnton (2018).

Tabela 7 – Eventos Históricos Relacionados à Fake News - Parte 2

Título	**Fake News**
1899-1902 – A Guerra dos Bôeres	A propaganda perpetua o estereótipo do "Boer" durante este conflito na África do Sul. Foi popularizado pelo exército britânico para influenciar a opinião pública britânica a apoiar uma guerra impopular.
1914-1918 – Primeira Guerra Mundial	A propaganda desempenhou um papel crucial no esforço de recrutamento, apelando ao nacionalismo e ao patriotismo: "Seu país precisa de VOCÊ"; "Papai, o que VOCÊ fez na Grande Guerra?
1917 – A fábrica alemã de cadáveres	Durante a Primeira Guerra Mundial, a propaganda britânica demonizou os alemães e espalhou a desinformação de que as forças alemãs usavam cadáveres de soldados para produzir farinha de ossos e comida de porco. Essa desinformação teve implicações na credibilidade dos relatos do Holocausto durante a Segunda Guerra Mundial.
1933 - Ministério do Reich de Esclarecimento Público e Propaganda estabelecido	Joseph Goebbels estabeleceu o Ministério de Esclarecimento Público e Propaganda do Reich para disseminar mensagens nazistas de ódio contra os judeus, utilizando todos os meios disponíveis, incluindo teatro e imprensa. A propaganda nazista foi fundamental para motivar os perpetradores do Holocausto e garantir a aquiescência de milhões de espectadores à perseguição e ao assassinato em massa.
1938 - Drama de rádio Guerra dos Mundos	O drama de rádio "Guerra dos Mundos"nos EUA enganou muitos ouvintes, que acreditaram que a Terra estava sendo atacada. Isso prenunciou as respostas do século 21 às sátiras de notícias. No entanto, ninguém envolvido com o programa esperava enganar qualquer ouvinte.
1939-1945 – Segunda Guerra Mundial	Edward Herzstein, em seu livro *The War that Hitler Won* (1978), descreveu a campanha de propaganda nazista como "a mais infame campanha de propaganda da história".
1947-1991 – A Guerra Fria	Durante este período, a radiodifusão internacional foi aproveitada para influenciar as populações a tomar partido.
1983 – Entrevista no Dia da Mentira	O historiador Joseph Boskin, da Universidade de Boston, inventou uma história sobre um bobo da corte que se tornou rei, esperando que o repórter percebesse a brincadeira, mas a história foi publicada como verdadeira.
1996 – Começa o Daily Show	A sátira de notícias e o programa de TV autodenominado de "notícias falsas" começaram nos EUA, dando lugar ao surgimento de notícias satíricas como um gênero que se tornou "algum tipo de corretivo e substituto para o jornalismo convencional".
2003-2011 – Guerra do Iraque	Antes da invasão do Iraque em 2003, o *New York Times* publicou uma série de artigos sobre armas de destruição em massa, que ficaram conhecidas como "Armas de distração em massa".
2005 – Começa o Relatório Colbert	O programa satírico de entrevistas noturnas na televisão começou nos EUA, sendo chamado de "desenvolvimento de longo prazo" na sátira política.

Fonte: Adaptado do estudo "A short guide to the history of 'Fake News' and disinformation"de Darnton (2018).

Tabela 8 – Eventos Históricos Relacionados à Fake News - Parte 3

Título	**Fake News**
2010 – Jornal estatal egípcio publica fotos manipuladas	O jornal *Al-Ahram* 'photoshopou' uma foto para colocar o presidente egípcio Hosni Mubarak na frente de líderes mundiais, o que foi revelado por um blogueiro egípcio.
2011 – Guerra Civil Síria	Durante a guerra na Síria, houve desinformação disseminada, mas o *New York Times* verificou que o governo sírio usou uma bomba de cloro, apesar das negações do governo de Assad.
2015 – Âncora de TV egípcia transmite imagens de videogame	Um âncora de TV transmitiu imagens de um videogame russo, elogiando a intervenção russa na Síria como se fosse real.
2016 – Votos	Durante a campanha presidencial de 2016 nos EUA, houve desinformação disseminada por ambos os lados, incluindo alegações falsas feitas por Donald Trump e proliferação de sites de notícias falsas.
2016-2017 – Fazendas de trolls e notícias falsas com fins lucrativos	Durante as eleições presidenciais dos EUA em 2016, fazendas de trolls e sites de notícias falsas geraram lucros com a disseminação de desinformação usando publicidade automatizada como o Google AdSense.
2016 – Hacker colombiano revela manipulação de eleições	Andres Sepulveda revelou como hackeou eleições na América Latina, manipulando redes sociais e instalando spyware, sendo condenado por suas ações.
2016 – Ministro da Defesa do Paquistão reage a notícia falsa	O ministro paquistanês twittou ameaças de retaliação nuclear após ler uma história falsa sobre Israel, que foi desmentida.
2017 – Envolvimento da Rússia nas eleições dos EUA	Agências de inteligência dos EUA relataram que a Rússia usou "trolls"para influenciar as eleições de 2016, com participação de Google, Facebook e Twitter.

Fonte: Adaptado do estudo "A short guide to the history of 'Fake News' and disinformation"de Darnton (2018).

As tabelas 6, 7 e 8 ilustram a evolução das Fake News ao longo dos anos, desde 44 a.C. até 2017. Esse panorama histórico revela que a disseminação de notícias falsas não é uma prática recente, mas uma estratégia antiga que se modernizou ao longo do tempo. Mesmo com as mudanças nos meios de comunicação e nas tecnologias utilizadas para espalhar desinformação, é possível identificar consistentemente os componentes principais: Agente, Mensagem e Intérprete. Além disso, a mensagem pode ter diversas intenções, desde a propagação de mentiras até objetivos mais graves, como o enfraquecimento das democracias. Esse último ponto deixa claro que as Fake News têm o potencial de minar a confiança pública nas instituições democráticas, polarizar a sociedade e desestabilizar governos. Assim, a compreensão e o combate à desinformação são essenciais para preservar

a integridade e a saúde das sociedades democráticas.

2.1.4 Impactos na democracia

Assim como as Fake News evoluíram nos últimos anos, houve um processo evolutivo na forma de fazer política. O uso de algoritmos, automação e curadoria humana tornou-se fundamental para gerenciar e distribuir informações enganosas por meio das redes sociais (ITUASSU et al., 2019). Esses avanços tecnológicos não apenas transformaram a maneira como as notícias são compartilhadas, mas também revolucionaram a forma como campanhas políticas são conduzidas, influenciando eleitores de maneiras sem precedentes.

Na eleição de Obama em 2008 nos EUA, por exemplo, pode-se ver um grande avanço no uso de tecnologia para arrecadação de fundos, chegando a mais de US 400 milhões em doações. Já nas eleições presidenciais de 2016, a campanha de Hillary Clinton gastou US 258 milhões em anúncios de TV, enquanto a plataforma de Trump, utilizando principalmente o Facebook, investiu menos da metade disso, US 100 milhões. Esse dado é significativo, pois Trump obteve uma vitória surpreendente, evidenciando a eficácia das novas estratégias digitais (ITUASSU et al., 2019).

Em 2016, a campanha eleitoral de Donald Trump se destacou por utilizar Fake News como principal estratégia de campanha, com ênfase na rede social Facebook. Esse uso sistemático de notícias falsas tornou a campanha de Trump um marco na história política recente, amplificado pelo uso intensivo de automação de algoritmos para a propagação de Fake News e a captação de eleitores. Infelizmente, essa prática também foi utilizada no Brasil durante a eleição que elegeu o presidente Jair Bolsonaro.

A conexão entre as redes sociais e os resultados das eleições nos Estados Unidos é evidente ao se analisar o contexto da campanha. Um exemplo marcante é o Facebook, que permitiu o vazamento de dados de cerca de 50 milhões de usuários para a empresa britânica Cambridge Analytica. Esta empresa utilizou essas informações para favorecer a candidatura de Donald Trump, traçando estratégias para captar o eleitorado de forma mais eficaz, o que destaca a influência das redes sociais no processo eleitoral (MORONI, 2018).

Outro ponto relevante sobre a eleição presidencial de 2016 nos EUA foi o uso dos chamados Dark Posts, um mecanismo de compartilhamento de informações falsas não vinculado ao site oficial da campanha. Essas informações eram compartilhadas de forma não oficial com grupos específicos de eleitores, visando garantir a disseminação de Fake News sem associá-las diretamente à imagem de Donald Trump (ITUASSU et al., 2019).

Silva (2020) observa que em 2016 o termo Fake News teve um aumento significativo de 365% em sua menção, coincidente com as eleições presidenciais dos Estados Unidos naquele ano, marcando uma escalada na utilização de notícias falsas nos meios digitais. Em

pesquisa realizada por (GRINBERG et al., 2019), foi observado que 27% das pessoas que tiveram contato com Fake News durante a campanha eleitoral verificaram essas notícias novamente nas semanas que antecederam as eleições, enquanto apenas 2,6% buscaram essas informações em sites tradicionais e confiáveis.

Esse cenário revela uma fragilidade muito forte na democracia, pois quando o maior símbolo da democracia, que é uma eleição, é movido e financiado por notícias falsas, os impactos na sociedade são profundos. A massa de eleitores exposta a informações falsas tende a acreditar em mentiras, o que pode distorcer a percepção da realidade e influenciar decisões cruciais, enfraquecendo a integridade do processo democrático e comprometendo a confiança pública nas instituições.

2.1.5 Consequências e impactos na sociedade

As redes sociais funcionam como comunidades onde os usuários têm a oportunidade de compartilhar suas opiniões. No entanto, esse ambiente pode favorecer a formação de "câmaras de eco", que amplificam e disseminam informações incorretas e desinformação. Esse fenômeno pode resultar na propagação viral de desinformação, aumentando significativamente o alcance e o impacto de notícias falsas na internet (BOVET; MAKSE, 2019).

Segundo (MORONI, 2018), as Fake News possuem padrões informacionais que podem afetar hábitos sociais já estabelecidos, o que pode impactar diretamente alguns fatores, como a ação coletiva. A ação coletiva refere-se à habilidade de agir em conjunto com outros indivíduos para atingir um objetivo comum, sem perder de vista as próprias convicções e interesses pessoais. No entanto, é importante ressaltar que as emoções compartilhadas decorrentes do contágio emocional causado pelas Fake News podem influenciar a ação coletiva (MORONI, 2018).

Inicialmente, as Fake News podem causar um contágio emocional que se propaga rapidamente, afetando a percepção dos usuários e levando à disseminação de padrões coletivos de ação baseados em emoções compartilhadas. Isso, por sua vez, pode afetar a prospectividade, flexibilidade e coordenação emergente, como pode ser observado na tabela a seguir:

A tabela 9 mostra que quando a prospetividade, a flexibilidade e a coordenação emergente de uma pessoa são comprometidas, o compartilhamento de Fake News tende a aumentar. Isso ocorre porque os usuários que espalham notícias falsas frequentemente não conseguem reconhecer a falsidade das informações, já que o compartilhamento está ligado às suas opiniões e valores pessoais. O raciocínio é o seguinte: "Se eu acredito que algo está alinhado com minhas crenças e valores, como isso poderia ser falso?"Esse fenômeno resulta em uma inversão de valores, onde o compartilhamento de fake news pode

Tabela 9 – Efeito das Fake News na ação coletiva

Conceito	**Definição**	**Efeito com a utilização de Fake News**
Prospectividade	Propriedade antecipatória da ação	A prospectividade pode ajudar a combater as Fakes News, permitindo que as pessoas compreendam de forma mais clara e objetiva as possibilidades e consequências de suas ações no compartilhamento de informações em um determinado contexto.
Flexibilidade	Propriedade que permite a habilidade para encerrar uma ação em contextos que se mostram adversos	A habilidade de encerrar uma ação em contextos adversos pode ser aplicada no combate às Fake News, evitando a disseminação de informações falsas.
Coordenação emergente	Propriedade que possibilita a ação sincronizada entre indivíduos, os quais podem agir de modo simulado ou involuntário	No contexto das Fake News, por exemplo, quando uma notícia falsa é compartilhada por pessoas, outras acabam compartilhando a mesma notícia, mesmo sem terem confirmado previamente, criando assim uma coordenação emergente na disseminação da Fake News.

Elaboração própria adaptado do estudo de Morini (2019).

ser confundido com uma forma legítima de expressão pessoal. Além disso, essa situação provoca um desgaste nas relações entre a sociedade e o Estado, pois a propagação de desinformação mina a confiança pública nas instituições e prejudica a legitimidade dos processos democráticos. Em resposta a esses desafios, a criação e a implementação de leis que combatam a desinformação e regulamentem a responsabilidade das plataformas digitais tornam-se cruciais para preservar a integridade da informação e a coesão social.

Desde então, muitos projetos de lei foram criados em todo o mundo para combater a disseminação de Fake News. Por exemplo, na França, a lei que proíbe a disseminação de notícias falsas durante as eleições é conhecida como "Loi relative à la lutte contre la diffusion de fausses nouvelles (NOUVELLES, 2018). Já no Brasil, a Lei Geral de Proteção de Dados Pessoais (LGPD) (LGPD, 2018), estabelece regras para o tratamento de dados pessoais na internet e prevê sanções para as plataformas digitais que não apagarem conteúdos falsos ou ofensivos em até 24 horas após a denúncia. Na União Europeia, foi criada a Diretiva do Parlamento Europeu e do Conselho 2019/790 sobre direitos autorais na era digital, que visa proteger os cidadãos contra a disseminação de notícias falsas.

Com o objetivo de combater a desinformação digital e aumentar a transparência na internet, foi incluído para tramitação o Projeto de Lei n° 2630, de 2020 (DEPUTADOS,

2020), que visa regulamentar e estabelecer normas sobre a transparência de redes sociais e serviços de mensagens privadas, além de definir a responsabilidade dos provedores.

No entanto, ao se realizar uma consulta pública sobre a PL 2630, que questionava a opinião das pessoas sobre a criminalização das Fake News, foi obtido um resultado contundente: 20% dos votos a favor contra 70% contrários ao conteúdo do projeto de lei (ATIVIDADE LEGISLATIVA, 2021).

Na consulta pública realizada no Twitter, pôde-se observar que muitas pessoas que se opunham ao projeto de lei argumentam que sua aprovação seria uma forma séria de censura por parte do Estado em relação à população. Existe uma parcela da sociedade que acredita que as Fake News são uma forma de experimentar a liberdade de expressão. No entanto, a fim de entender os valores políticos e seu impacto nas decisões políticas atuais, mais pesquisas estão sendo conduzidas (FERREIRA e TABARES, 2020).

No entanto, o combate às Fake News representa um desafio multifacetado, que exige um equilíbrio delicado entre proteger a liberdade de expressão e assegurar a veracidade das informações compartilhadas nas redes sociais. O crescente papel dos dados na sociedade moderna intensifica esse desafio, a disseminação de Fake News não é apenas uma questão de desinformação, mas também está frequentemente associada a interesses financeiros e políticos.

De acordo com (LAZER et al., 2018), as Fake News são criadas com o intuito de influenciar opiniões políticas e econômicas, e esse fenômeno é impulsionado pelo potencial lucrativo que essas informações falsas podem gerar. A natureza altamente viral das Fake News, aliada ao valor significativo dos dados para as plataformas digitais, torna a regulamentação e a monitorização uma tarefa complexa, que deve considerar tanto a proteção da liberdade individual quanto a integridade das informações disseminadas.

Diante dos riscos significativos que as Fake News representam para a sociedade, surgem iniciativas cada vez mais destacadas para identificar e mitigar o compartilhamento de informações falsas. Nos últimos anos, o processamento de linguagem natural (PLN) tem emergido como uma ferramenta fundamental na luta contra o ecossistema da desinformação(**??**). A principal vantagem do PLN é sua capacidade de converter a linguagem humana, em todas as suas nuances e formas, em uma linguagem computacional que pode ser processada e analisada por sistemas automáticos.

Isso se torna especialmente relevante, pois as Fake News, embora se apresentem como "notícias", são, na verdade, dados manipulados e distorcidos. Uma vez que essas notícias são essencialmente informações, a questão crucial é se é possível identificar padrões dentro desses dados para discernir a veracidade das informações. Isso permite uma detecção mais eficiente e precisa das Fake News, ajudando a conter a propagação de desinformação.

2.2 Processamento de linguagem natural

A aplicação do Processamento de Linguagem Natural (PLN) na língua portuguesa apresenta desafios específicos devido às suas características e nuances distintas em comparação com outros idiomas. Diferente de línguas como o inglês, onde a estrutura das palavras e frases tende a ser mais direta e regular, o português é uma língua com várias em flexões verbais, concordâncias de gênero e número, além de apresentar uma complexa estrutura de frases com uma ordem que pode ser mais flexível. Essa riqueza morfológica e sintática requer que as ferramentas de PLN sejam adaptadas para lidar com a variabilidade e as sutilezas do português (FISCHER et al., 2022).

A formação de palavras por meio de processos como a derivação e a composição, assim como a existência de verbos irregulares e construções sintáticas elaboradas, impõe um nível adicional de dificuldade na tarefa de processamento automático da língua. Ademais, expressões idiomáticas e regionalismos podem modificar o significado das frases, tornando o entendimento contextual e semântico ainda mais desafiador para algoritmos de PLN. Assim, o desenvolvimento de modelos eficazes para o português demanda uma abordagem cuidadosa que considere essas peculiaridades, garantindo que as ferramentas sejam capazes de compreender e gerar texto com precisão e naturalidade (ALMEIDA, 2023).

2.2.1 Definição

O Processamento de Linguagem Natural (PLN) é considerado uma ferramenta fundamental na luta contra o ecossistema da desinformação (**??**). A principal utilidade do PLN é sua capacidade de transformar a linguagem humana, em todas as suas formas, em linguagem computacional e processável por um computador.

(SILVA, 2020) faz um paralelo, que assim como aprendemos o estudo da língua portuguesa, um algoritmo de PLN também pode ser treinado por meio de exemplos. Inicialmente, começamos com o aprendizado das palavras, suas definições e como usá-las corretamente em diferentes contextos. Em seguida, aprendemos a gramática da língua, como as palavras se juntam para formar frases e como usar as regras gramaticais para comunicar nossas ideias de forma clara e eficaz.

(GARCIA, 2023) destaca que a linguagem humana é ampla e rica em abreviaturas, regionalismos e múltiplos significados para a mesma palavra, dependendo do contexto em que é empregada. Para entender o contexto de uma frase, é necessário conhecer as palavras e todas as suas conexões para evitar ambiguidades. Este é um dos maiores desafios do PLN, fazer com que os algoritmos identifiquem nuances como ambiguidades e ironias, especialmente prevalentes na língua portuguesa.

Assim como na construção de uma frase, a desconstrução dessa frase para a linguagem computacional ocorre em etapas pré-definidas, conforme a necessidade de

cada algoritmo. Essas etapas incluem o tratamento das strings, removendo informações desnecessárias, como pontuação e stopwords, tokenização, que divide as notícias em unidades menores, como palavras ou subpalavras, essenciais para análises subsequentes. Posteriormente, codificações baseadas em palavras e, em seguida, vetorização, que mapeiam cada palavra em um espaço vetorial de alta dimensão. A seguir, serão abordados cada uma dessas fases de PLN que foram utilizados.

2.2.2 Remoção de StopWords

A remoção de stopwords é uma técnica crucial na PLN e na mineração de texto, projetada para eliminar palavras comuns e frequentes que não agregam valor significativo à análise textual. Estas palavras, que incluem artigos, preposições e outras expressões gramaticais, frequentemente não carregam informações essenciais sobre o conteúdo principal dos textos. De acordo com (SHARMA; SARAN; PATIL, 2020), essas palavras são consideradas "ruído"no contexto de análise textual, uma vez que sua presença não contribui para a semântica do documento, mas pode impactar a eficácia dos modelos de PLN.

As vantagens da remoção de stopwords são amplamente reconhecidas na literatura acadêmica e têm um impacto significativo na qualidade da análise textual. A técnica reduz o volume de dados a ser processado, o que melhora a eficiência dos algoritmos e diminui o tempo necessário para a execução das análises. Além disso, ao focar nas palavras relevantes, a remoção de stopwords aumenta a precisão dos resultados obtidos e facilita a extração de informações valiosas de grandes volumes de texto (RICARDO et al., 1999). Observe na quadro a seguir:

Tabela 10 – Exemplo de stop words

Texto Original	**Texto Após Remoção de Stop Words**
"Os cientistas estavam estudando novos métodos para descobrir uma solução mais eficaz."	"cientistas estavam estudando novos métodos descobrir solução mais eficaz"
"Eles descobriram que a descoberta poderia melhorar os resultados dos experimentos."	"descobriram descoberta poderia melhorar resultados experimentos"

Como pode ser observado na 10 , a remoção de stopwords é uma técnica que pode ser implementada por meio de listas pré-definidas ou através de métodos de aprendizado automático que identificam e filtram essas palavras. No quadro, a coluna "Texto Original"exibe frases completas, onde todas as palavras, incluindo as de baixo valor informativo, estão presentes. Já a coluna "Texto Após Remoção de Stopwords"demonstra como o texto é simplificado após a eliminação dessas palavras, deixando apenas os termos mais relevantes. Essa técnica permite que os algoritmos de aprendizado de máquina concentrem

seus esforços nas palavras de maior valor informativo, aumentando a precisão dos resultados obtidos durante a análise de grandes volumes de texto.

2.2.3 Lematização

A lematização visa reduzir palavras a suas formas básicas ou "lemmas", ou seja, a forma dicionarizada e normalizada das palavras. Diferentemente da stemming, que realiza uma redução mais agressiva, a lematização busca preservar o significado da palavra dentro do seu contexto gramatical. Como descrito por Bird e Klein (2009), a lematização utiliza dicionários lexicais e regras morfológicas para converter variações de uma palavra em sua forma canônica, considerando o papel gramatical da palavra na frase.

As vantagens da lematização são amplamente reconhecidas na literatura de PLN e têm um impacto significativo na análise textual e na construção de modelos de aprendizado de máquina. Ao normalizar palavras para suas formas base, a lematização reduz a dimensionalidade dos dados e melhora a precisão dos modelos analíticos. Além disso, facilita a recuperação de informações e a análise semântica ao alinhar palavras com significados semelhantes, independentemente das variações morfológicas (KIBBLE, 2013).

Tabela 11 – Exemplo Lematização

Texto Após Remoção de Stop Words	**Texto Após Lematização**
"cientistas estavam estudando novos métodos descobrir solução eficaz"	"cientista estar estudar novo método descobrir solução eficaz"
"descobriram descoberta poderia melhorar resultados experimentos"	"descobrir descobrir poder melhorar resultado experimento"

A tabela 11 mostra que o processo de lematização envolve a análise sintática e semântica das palavras para identificar sua forma base. Por exemplo, as palavras "correndo", "correu"e "corre"são reduzidas ao lema "correr". Isso é realizado por meio de algoritmos que analisam a estrutura gramatical e o contexto, assegurando que a redução não apenas remova sufixos, mas também mantenha o significado subjacente.

2.2.4 Tokenização e Vetorização

A tokenização envolve a segmentação de um texto em unidades menores chamadas "tokens". Esses tokens podem ser palavras, frases, caracteres ou outras unidades significativas, dependendo da natureza do texto e dos objetivos analíticos. O objetivo principal da tokenização é transformar o texto em uma forma estruturada que facilite a análise e a interpretação por algoritmos e modelos de PLN (SHARMA; SARAN; PATIL, 2020).

A precisão na tokenização é vital, pois pode influenciar diretamente a qualidade da análise subsequente e o desempenho dos modelos (BIRD; KLEIN; LOPER, 2009).A

tokenização é amplamente utilizada em várias tarefas de PLN, incluindo análise de sentimentos, tradução automática e recuperação de informações. Por exemplo, na recuperação de informações, uma tokenização eficaz pode melhorar a indexação e a busca por termos relevantes dentro de grandes volumes de texto (SALTON, 1983).

Já na vetorização ocorre a conversão do texto em uma representação numérica, permitindo que algoritmos manipulem e processem informações textuais de forma estruturada. O objetivo principal da vetorização é transformar unidades textuais, como palavras ou documentos, em vetores numéricos que possam capturar suas características semânticas e contextuais. Este processo é fundamental porque a maioria dos modelos de aprendizado de máquina exige dados numéricos para realizar análises e previsões (MANNING, 2008).

Durante a vetorização, o texto é convertido em uma forma que permite a análise quantitativa e estatística, facilitando a extração de padrões e insights significativos. Existem várias abordagens para vetorização, que variam em complexidade e na forma como representam a informação textual. Algumas técnicas simples baseiam-se em contagens de palavras ou frequências, enquanto métodos mais avançados tentam capturar informações contextuais e semânticas das palavras, observe no Quadro 6:

Tabela 12 – Exemplo Tokenização e Vetorização

Texto Após Lematização	**Tokens**	**Vetores de Exemplo**
"cientista estudar novo método descobrir solução eficaz"	["cientista", "estudar", "novo", "método", "descobrir", "solução", "eficaz"]	[0.21, 0.43, 0.67, ...], [0.54, 0.22, 0.67, ...], ...
"descobrir descobrir poder melhorar resultado experimento"	["descobrir", "descobrir", "poder", "melhorar", "resultado", "experimento"]	[0.34, 0.22, 0.56, ...], [0.45, 0.78, 0.33, ...], ...

Na Tabela 12 é possível perceber que após todas essas etapas, o conjunto de dados está pronto para a fase de treinamento. Contudo, é importante destacar que o pré-processamento dos dados pode variar de um algoritmo para outro. Embora não seja possível discutir sobre a conceitualização das fases de pré-processamento, as técnicas específicas devem ser adaptadas para melhor se adequar ao escopo do estudo e ao tipo de modelo utilizado. Por exemplo, ao utilizar um algoritmo como Naive Bayes, o foco pode estar na remoção de stopwords e na transformação de palavras para suas raízes (stemming). Já para uma rede neural convolucional (CNN), pode ser mais relevante a tokenização e a padronização do comprimento das sequências. Essas adaptações asseguram que os dados estejam otimizados para a análise e para a construção de modelos preditivos eficazes, garantindo a máxima eficiência e precisão no processo de detecção de Fake News.

2.3 Aprendizado de máquina

Este capítulo está organizado da seguinte maneira: inicialmente, apresenta-se uma definição abrangente do que é aprendizado de máquina, seguida por uma discussão sobre seus desdobramentos e aplicações dentro do escopo desta pesquisa. Em seguida, são detalhados os conceitos e fundamentos dos algoritmos de aprendizado de máquina utilizados neste estudo: SVM (Support Vector Machine), CNN (Convolutional Neural Network) e Naive Bayes. Cada algoritmo será explicado em termos de suas características, funcionamento e relevância para a classificação de Fake News.

2.3.1 Definição

Segundo (GARCIA, 2023) o aprendizado de máquina é uma área da inteligência artificial que se concentra em desenvolver sistemas capazes de aprender e melhorar com experencias anteriores, sem a necessidade de programação explícita, ajudando na tomada de decisão com base em dados. Esses modelos computacionais, gerados através do aprendizado de máquina, são capazes de classificar padrões desconhecidos após serem treinados com um conjunto de dados. Esses dados podem ser supervisionados, ou não supervisionados (ALTIERI; APRO, 2022).

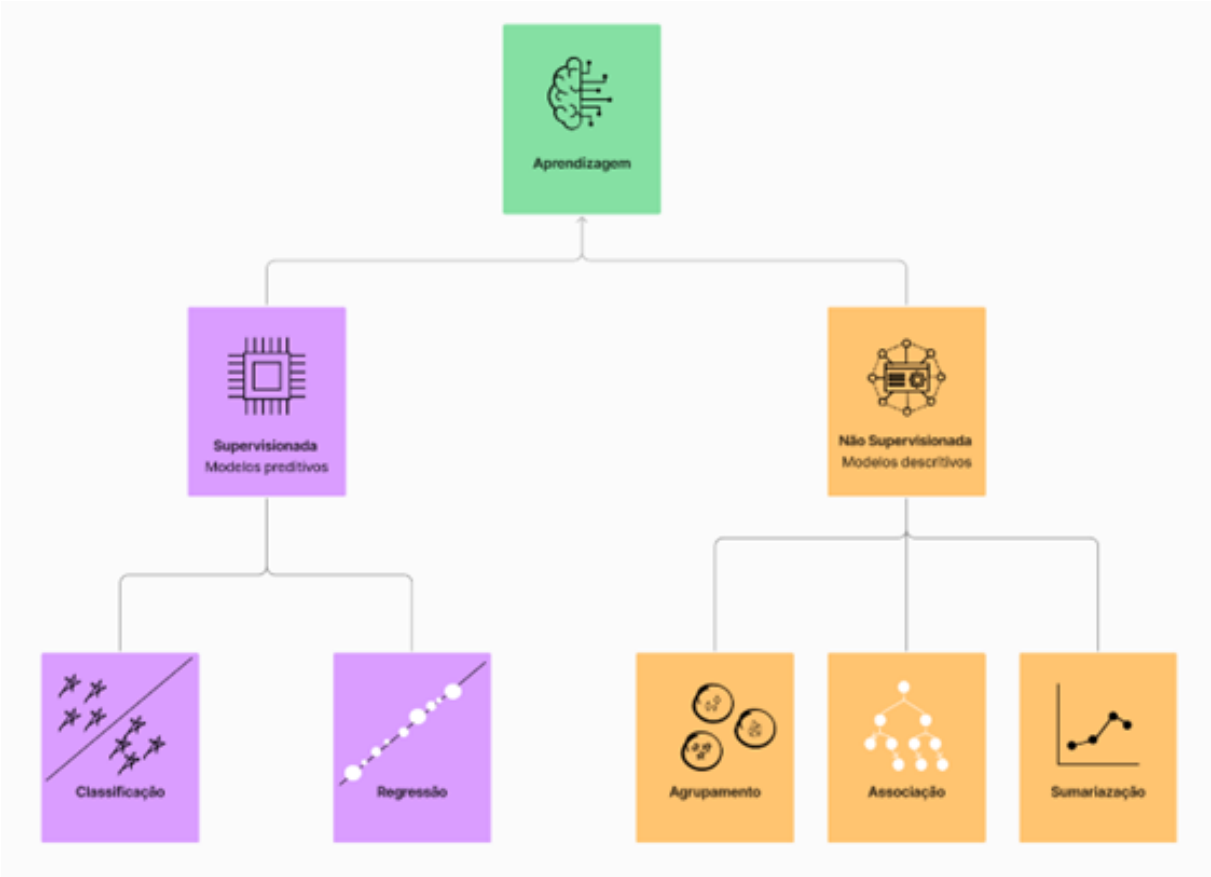

Figura 3 – Diagrama de aprendizado de máquina..

A Figura 3. ilustra a categorização dos tipos de aprendizado em inteligência artificial, destacando duas abordagens principais: aprendizado supervisionado e não supervisionado. O aprendizado supervisionado, descrito como "Modelos preditivos", inclui técnicas de classificação e regressão, no qual o modelo é treinado com dados rotulados para prever resultados. Já o aprendizado não supervisionado, referido como "Modelos descritivos", envolve métodos como agrupamento, associação e sumarização, utilizados para identificar padrões e relações em dados não rotulados. Essas abordagens formam a base para o desenvolvimento de sistemas inteligentes capazes de interpretar e analisar grandes volumes de dados.

2.3.2 Aprendizado supervisionado

Segundo (GARCIA, 2023), no aprendizado supervisionado, o modelo trabalha com dados já rotulados, tentando descobrir padrões nos dados de entrada para ser capaz de classificar novas instâncias. O aprendizado supervisionado possui dois modelos principais: classificação e regressão. Silva (2022) observa que a classificação é usada quando os resultados são categorias discretas (como "verdadeira"ou "falsa"), enquanto a regressão é usada quando os resultados são valores contínuos (como o tempo de leitura em minutos).

Classificação: A classificação consiste em categorizar itens em grupos distintos. Por exemplo, ao desenvolver um modelo capaz de identificar se uma notícia é verdadeira ou falsa, os dados de saída são discretos, pois o resultado será necessariamente uma das duas categorias: "verdadeira"ou "falsa". Este é um problema de classificação, visto que o objetivo é atribuir cada notícia a uma dessas categorias específicas e mutuamente exclusivas. Neste trabalho, adota-se esse modelo para a verificação de notícias, buscando classificar informações de forma eficaz e precisa, contribuindo assim para a identificação e mitigação da disseminação de Fake News.

Regressão: A regressão é uma técnica utilizada para prever valores contínuos. Por exemplo, considere a tarefa de prever o tempo que um usuário gastará lendo uma notícia online. Nesse caso, os dados de saída são contínuos, uma vez que o tempo de leitura pode variar dentro de um intervalo, como 5,2 minutos ou 10,5 minutos. Este é um problema de regressão, pois o objetivo é prever um valor numérico contínuo com base em variáveis preditivas. A aplicação dessa técnica permite a modelagem de fenômenos onde as variáveis de interesse variam de maneira contínua, sendo essencial para a análise de dados que visam estimar tendências ou comportamentos, como no caso do tempo de engajamento de usuários com conteúdo online.

A Figura 4. mostra o processo de aprendizado supervisionado em algoritmos de aprendizado de máquina. No Passo 1, o algoritmo é alimentado com dados de entrada e saída que foram previamente categorizados ou "rotulados". Por exemplo, notícias falsas são fornecidas ao algoritmo para que ele possa aprender a identificar padrões e características

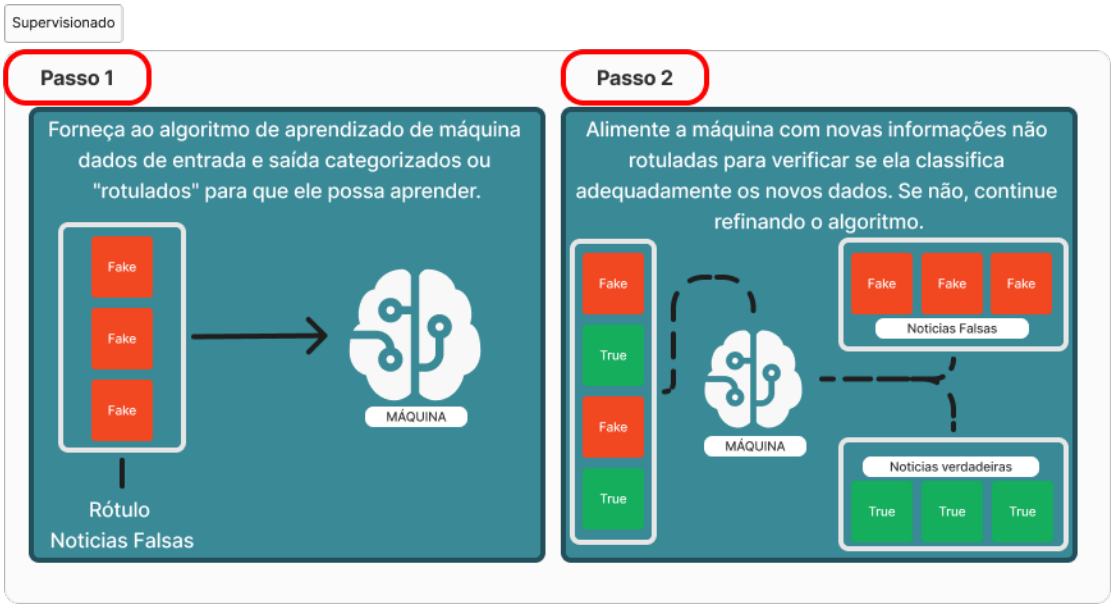

Figura 4 – Aprendizado de máquina Supervisionado

desses dados. No Passo 2, a máquina é alimentada com novas informações não rotuladas para verificar se ela consegue classificar adequadamente os novos dados. Se o algoritmo conseguir identificar corretamente as notícias falsas e verdadeiras, ele está funcionando conforme esperado. Caso contrário, é necessário continuar refinando o algoritmo para melhorar sua precisão.

2.3.3 Aprendizado não supervisionado

De acordo com Barbosa (2022), no aprendizado de máquina não supervisionado, não existem atributos meta, mas sim um conjunto de dados X. Observa-se que, nesse modelo, não há entrada de dados rotulados, mas sim o reconhecimento de padrões sem o envolvimento de um atributo de destino. Isso significa que o sistema tenta identificar estruturas ocultas nos dados sem orientação prévia. As principais técnicas utilizadas são agrupamento, associação e sumarização (GARCIA, 2023):

Agrupamento: O agrupamento (clustering) é uma técnica de mineração de dados que consiste em dividir um conjunto de dados em grupos ou clusters, de modo que os elementos dentro de cada grupo sejam mais semelhantes entre si do que em relação aos elementos de outros grupos. Essa técnica é amplamente utilizada em diversas áreas, como segmentação de mercado, onde ajuda a identificar diferentes perfis de clientes com base em características semelhantes, e em análise de imagens, para agrupar pixels em regiões de interesse. Um exemplo clássico de aplicação do agrupamento é na análise de clientes de um banco, onde pode-se identificar grupos de clientes com comportamentos financeiros similares, auxiliando na personalização de produtos e serviços. O agrupamento

é especialmente útil quando se deseja explorar a estrutura natural dos dados sem rótulos predefinidos.

Associação: A associação é uma técnica utilizada para descobrir relações ou padrões frequentes entre conjuntos de itens em grandes bases de dados. Essa técnica é particularmente útil em áreas como o comércio eletrônico e análise de transações financeiras. Por exemplo, na análise de cestas de compras, a associação pode identificar padrões como a frequência com que determinados produtos são comprados juntos, como pão e manteiga. Esse tipo de informação pode ser utilizado para otimizar a disposição de produtos em uma loja ou para sugerir itens complementares durante compras online. A técnica de associação é valiosa para revelar conexões ocultas entre dados, que podem ser exploradas para melhorar a estratégia de vendas e marketing.

Sumarização: A sumarização é uma técnica que visa reduzir grandes volumes de informação a um formato mais compacto, mantendo as partes mais relevantes do conteúdo original. Ela é particularmente aplicada em processamento de linguagem natural, onde longos textos são condensados em resumos que capturam a essência da informação. Existem dois principais tipos de sumarização: a extrativa, que seleciona as frases mais importantes do texto original, e a abstrata, que gera novas frases que sintetizam o conteúdo central. Um exemplo de aplicação de sumarização é na geração automática de resumos de notícias, permitindo que os leitores acessem rapidamente os pontos-chave das informações. A sumarização é especialmente útil em contextos onde a rapidez e a eficiência na transmissão da informação são cruciais.

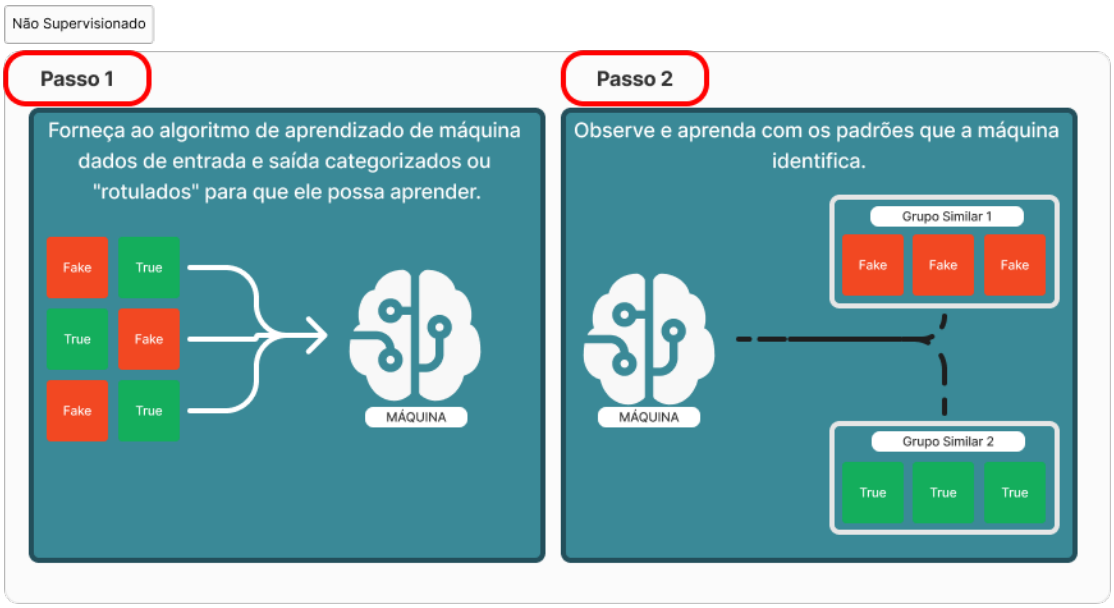

Figura 5 – Aprendizado de máquina não supervisionado

Na Figura 5 temos o processo de aprendizado não supervisionado em algoritmos

de aprendizado de máquina. No Passo 1, o algoritmo é alimentado com dados de entrada que não foram previamente categorizados ou "rotulados". Esses dados são fornecidos ao algoritmo para que ele possa identificar padrões e similaridades por conta própria. No Passo 2, o algoritmo analisa os dados fornecidos e agrupa as informações em categorias baseadas nos padrões que ele identifica. Por exemplo, ele pode agrupar dados semelhantes em "Grupo Similar 1", que contém notícias falsas, e "Grupo Similar 2", que contém notícias verdadeiras. Esse processo permite que o algoritmo descubra estruturas ocultas nos dados sem a necessidade de supervisão humana.

2.4 Algoritmos de classificação

Após a revisão bibliográfica realizada para a construção deste projeto, identificou-se que, dentre os artigos selecionados, três algoritmos se destacaram: SVM (Support Vector Machine), CNN (Convolutional Neural Network) e Naive Bayes. Esses algoritmos foram escolhidos para implementação devido à sua relevância e eficácia comprovada na literatura. Todos os algoritmos pertencem à categoria de aprendizado supervisionado, sendo que os dados de entrada são provenientes do corpus Fake.br, o qual está dividido em notícias verdadeiras e falsas.

2.4.1 Máquinas de vetores de suporte (svm)

O SVM (Support Vector Machine) é uma técnica poderosa para realizar classificações lineares em bases de dados de pequeno a médio porte, sendo especialmente relevante no combate às Fake News. No contexto da classificação de notícias falsas, o SVM utiliza um conjunto de dados de treinamento, no qual cada notícia é representada como um vetor de características derivadas do texto, como a frequência de palavras ou termos específicos.

O objetivo do SVM é encontrar o hiperplano que maximiza a margem entre as classes de notícias verdadeiras e falsas, garantindo assim uma melhor generalização do modelo. Para atingir esse objetivo, o SVM resolve um problema de otimização, no qual busca minimizar a função de perda — que penaliza classificações incorretas — enquanto maximiza a margem entre as classes. Esse processo resulta em um classificador robusto, capaz de identificar padrões sutis e complexos nos dados textuais, contribuindo significativamente para a detecção eficaz de Fake News (VALKENBORG et al., 2023).

(SILVA, 2020) destaca que, na sua versão linear, o SVM oferece duas abordagens principais: margem rígida e margem flexível. A margem rígida exige uma separação clara entre as classes, sem permitir erros de classificação. Esse tipo de abordagem é ideal em situações onde os dados são perfeitamente separáveis. No entanto, em muitos cenários do mundo real, como na detecção de Fake News, os dados podem ser ruidosos ou sobrepostos. Para lidar com esses casos, a margem flexível permite alguns erros de classificação,

ajustando-se melhor a dados que não são perfeitamente separáveis e aumentando a robustez do modelo no combate à desinformação.

Uma vantagem significativa do SVM é a sua capacidade de operar com alta eficácia mesmo quando o número de características (dimensionalidade) é muito maior do que o número de exemplos de treinamento. Isso é particularmente útil na análise de textos, onde cada palavra pode ser considerada uma característica. No combate às Fake News, essa capacidade é crítica, pois permite ao SVM lidar com grandes conjuntos de dados textuais com inúmeras variáveis, mantendo a precisão na classificação. Além disso, o SVM pode ser combinado com técnicas de kernel para lidar com classificações não lineares, ampliando ainda mais seu alcance e aplicabilidade em cenários mais complexos.

Por fim, embora o SVM seja tradicionalmente utilizado para classificações binárias, ele pode ser estendido para problemas de múltiplas classes por meio de abordagens como "um contra todos"(one-vs-all) ou "um contra um"(one-vs-one). Isso permite que o SVM seja utilizado em uma variedade de contextos, além da detecção de Fake News, como na categorização de tópicos, análise de sentimentos e outras aplicações de processamento de linguagem natural (PLN). A flexibilidade e a robustez do SVM fazem dele uma ferramenta valiosa no arsenal contra a desinformação, contribuindo para a criação de soluções eficazes e confiáveis na era digital.

2.4.2 Naive bayes

De acordo Palmeir (2023), o Naive Bayes é um algoritmo de classificação baseado no Teorema de Bayes, que tem como foco determinar a probabilidade de uma classe a partir da probabilidade inversa. Sua classificação é realizada por meio de cálculos estatísticos, levando em conta a probabilidade de eventos passados com base nos dados de entrada. Ele é particularmente eficiente em termos de tempo de execução, sendo geralmente mais rápido do que outros algoritmos de classificação. Além disso, é considerado de fácil aplicação devido à sua simplicidade e eficiência em termos de processamento (GARCIA, 2023). A fórmula do Teorema de Bayes é:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \tag{2.1}$$

Khanam (2021) explica melhor a fórmula:

P(AB) é a probabilidade de A dado B.

P(BA) é a probabilidade de B dado A.

P(A) é a probabilidade a priori de A.

P(B) é a probabilidade a priori de B.

No cenário de classificação de Fake News, é calculada a probabilidade de um artigo ser verdadeiro ou falso com base em suas características textuais. Durante a fase de treinamento, o algoritmo analisa um conjunto de dados rotulados, onde os artigos já foram classificados como verdadeiros ou falsos, para calcular as probabilidades de ocorrência de certas palavras ou frases em cada categoria. Quando um novo artigo precisa ser classificado, o Naive Bayes avalia a probabilidade de cada palavra do artigo pertencer à classe "verdadeiro"ou "falso"e combina essas probabilidades para determinar a categoria final do artigo. Esse processo permite que o Naive Bayes faça previsões rápidas e eficientes sobre a veracidade de novos artigos, baseando-se nas características observadas durante o treinamento. A fórmula do Teorema de Bayes no cenário das Fake News:

$$P(Fake \mid "vacina", "câncer", "segurança") = \frac{P("vacina", "câncer", "segurança").P(Fake)}{P("vacina", "câncer", "segurança")}$$

Figura 6 – Exemplo Naive bayes Fake News

Onde:

P(Fake"vacina", "câncer", "segurança): A probabilidade de que a notícia seja "Fake"dado que contém as palavras "vacina", "câncer "e "segurança".

P("vacina", "câncer ", "segurança “Fake): A probabilidade de que uma notícia contenha as palavras "vacina", "câncer "e "segurança"dado que é uma notícia "Fake".

P(Fake): A probabilidade a priori de que uma notícia seja "Fake". Isso pode ser calculado com base na proporção de notícias "Fake"no conjunto de dados.

P("vacina", "câncer ", "segurança"): A probabilidade a priori de que uma notícia contenha as palavras "vacina", "câncer "e "segurança". Isso pode ser calculado com base na frequência dessas palavras no conjunto de dados, independentemente da classificação da notícia.

2.4.3 Rede neural convolucional

As Redes Neurais Convolucionais (CNNs) foram desenvolvidas para imitar o sistema visual humano, sendo especialmente eficazes em tarefas que requerem o reconhecimento e a interpretação de objetos, como a classificação, detecção e segmentação de imagens (KEITA 2024). Existem diferentes tipos de CNNs adaptadas para variadas formas de dados: Conv1D é projetada para dados unidimensionais, como sequências de texto ou séries temporais; Conv2D é utilizada para dados bidimensionais, como imagens; e Conv3D é aplicada a dados tridimensionais, como vídeos ou dados volumétricos. Neste trabalho,

utilizou-se o tipo Conv1D, devido ao foco na classificação de texto. Observe a Figura 7 a Arquitetura das CNN:

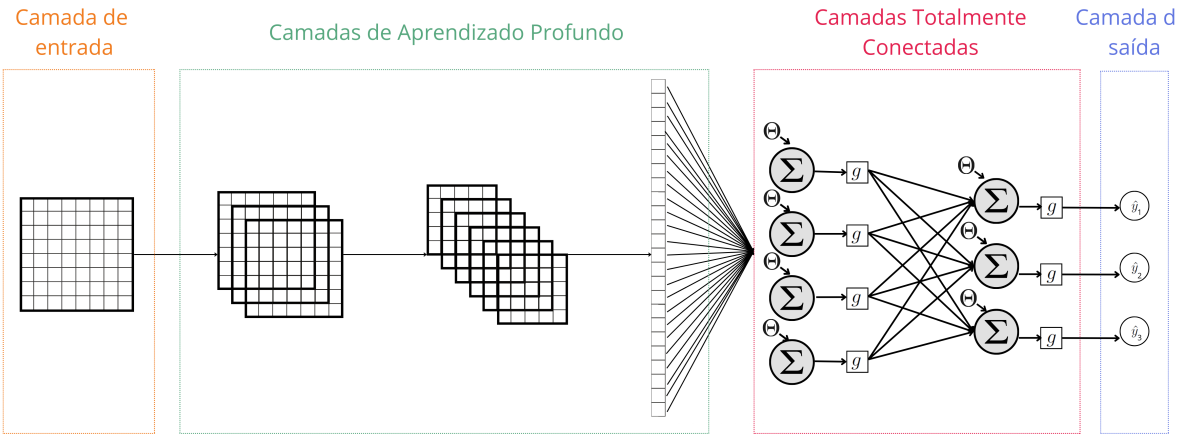

Figura 7 – Arquitetura de uma CNN

A Figura 7. ilustra a arquitetura de uma Rede Neural Convolucional (CNN), composta por quatro partes principais, cada uma desempenhando um papel essencial no processamento e classificação dos dados de entrada. A primeira parte é composta pelas camadas convolucionais. Essas camadas aplicam filtros (ou kernels) sobre os dados de entrada, o que permite a extração de características locais, como bordas, texturas e padrões. Cada filtro é responsável por capturar uma característica específica, e o resultado da convolução é um mapa de características que destacam essas propriedades nos dados de entrada. À medida que os dados percorrem as várias camadas convolucionais, a rede é capaz de identificar padrões cada vez mais complexos, fundamentais para o processo de classificação (SINGH; SINGH; MISHRA,).

Em seguida, temos a Unidade Linear Retificada (ReLU), que é uma função de ativação amplamente utilizada em redes neurais modernas. A função ReLU introduz não-linearidades ao modelo, o que é crucial para que a rede possa aprender e representar padrões complexos. Sem essas não-linearidades, a rede se comportaria como um modelo linear, limitando sua capacidade de resolver problemas não-lineares (CONG; ZHOU, 2023).

As camadas de pooling constituem a terceira parte da arquitetura. O pooling é uma técnica de redução de dimensionalidade que resume as informações contidas nos mapas de características, reduzindo a sua resolução espacial. Isso não apenas diminui a quantidade de parâmetros e computação na rede, mas também aumenta a robustez das variações de escala e posição nos dados de entrada. Em outras palavras, o pooling ajuda a rede a se concentrar nas características mais relevantes, ignorando variações menores que poderiam interferir na classificação (SINGH; SINGH; MISHRA,).

Finalmente, as camadas totalmente conectadas representam a última parte da arquitetura da CNN. Nessas camadas, cada neurônio está conectado a todos os neurônios

da camada anterior, permitindo a integração de todas as características extraídas pelas camadas convolucionais e de pooling. Essas camadas são responsáveis por realizar a tarefa final, seja ela classificação ou regressão, traduzindo as informações processadas em uma saída que pode ser interpretada, como a classe de um objeto ou um valor numérico (CONG; ZHOU, 2023). Conforme descrito por (**??**), essa estrutura é fundamental para o sucesso das redes neurais convolucionais em tarefas de visão computacional e outras áreas onde a extração e interpretação de características locais dos dados são cruciais.

2.5 Métricas de desempenho

Neste capítulo, serão abordadas as principais métricas utilizadas para avaliar os resultados obtidos, que incluem: Acurácia, Precisão, Recall, F1-score, Curva ROC e Matriz de Confusão. Cada uma dessas métricas desempenha um papel crucial na análise do desempenho do modelo, fornecendo uma visão detalhada sobre sua capacidade de classificação, a qualidade das previsões e a eficácia na distinção entre classes. A compreensão dessas métricas é essencial para interpretar corretamente os resultados e melhorar a performance dos modelos de aprendizado de máquina.

Lembrando que para a interpretação das métricas de desempenho, os seguintes termos são fundamentais: VP (Verdadeiros Positivos), que representa o número de casos positivos corretamente identificados pelo modelo; VN (Verdadeiros Negativos), que indica os casos negativos corretamente identificados; FP (Falsos Positivos), que corresponde aos casos negativos incorretamente classificados como positivos; e FN (Falsos Negativos), que são os casos positivos que foram classificados incorretamente como negativos

2.5.1 Acurácia

A acurácia é uma métrica fundamental para avaliar o desempenho de modelos de classificação. Ela representa a proporção de previsões corretas em relação ao total de previsões realizadas (SILVA 2022). E é definida como:

$$\text{Acurácia} = \frac{VP + VN}{VP + FP + FN + VN} \tag{2.2}$$

Em outras palavras, a acurácia mede a porcentagem de amostras corretamente classificadas pelo modelo. Em tarefas de classificação de Fake News, uma alta acurácia indica que o modelo é eficiente em identificar corretamente as notícias verdadeiras e falsas, mas pode não refletir a capacidade do modelo em lidar com desequilíbrios de classe, onde uma classe pode ser mais prevalente que a outra.

2.5.2 Precisão

A precisão é uma métrica que avalia a proporção de previsões positivas que são realmente corretas. Em contextos de classificação, a precisão é particularmente importante quando o custo de falsos positivos é alto. No caso da classificação de Fake News, uma alta precisão indica que, entre as notícias que o modelo classificou como falsas, a maioria realmente é falsa.

$$\text{Precisão} = \frac{VP}{VP + FP} \tag{2.3}$$

A precisão é calculada como a razão entre o número de verdadeiros positivos e a soma de verdadeiros positivos com falsos positivos, sendo uma métrica crucial para garantir que o modelo não esteja classificado incorretamente em muitas notícias verdadeiras como falsas.

2.5.3 Recall

O recall, também conhecido como sensibilidade ou taxa de verdadeiros positivos, mede a proporção de exemplos positivos que foram corretamente identificados pelo modelo. Em tarefas como a classificação de Fake News, um alto recall indica que o modelo é eficaz em identificar a maioria das notícias falsas, minimizando a quantidade de falsos negativos (AHMAD et al., 2020).

$$\text{Recall} = \frac{VP}{VP + FN} \tag{2.4}$$

O recall é calculado como a razão entre o número de verdadeiros positivos e a soma de verdadeiros positivos com falsos negativos, e é essencial para avaliar a capacidade do modelo de detectar todas as ocorrências de uma classe específica.

2.5.4 F1-score

O F1-score é uma métrica que combina precisão e recall em uma única medida, fornecendo um equilíbrio entre as duas. É especialmente útil quando há um desequilíbrio entre as classes (AHMAD et al., 2020).

$$F1\text{-}Score = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{2.5}$$

O F1-score é a média harmônica da precisão e do recall, oferecendo uma visão mais completa do desempenho do modelo ao considerar tanto a capacidade de identificar corretamente as notícias falsas quanto a capacidade de evitar falsos positivos. Um F1-score alto indica que o modelo tem um bom equilíbrio entre precisão e recall.

2.5.5 Curva roc

A Curva ROC (Receiver Operating Characteristic) é uma ferramenta gráfica que avalia a capacidade de um modelo de classificação em distinguir entre classes. As fórmulas para estas taxas são:

Taxa de Verdadeiros Positivos (TPR), que é equivalente ao recall para a classe positiva:

$$TPR = \frac{VP}{VP + FN} \tag{2.6}$$

Taxa de Falsos Positivos (FPR):

$$FPR = \frac{FP}{FP + VN} \tag{2.7}$$

A área sob a curva ROC (AUC-ROC) é uma métrica importante, ela plota a taxa de verdadeiros positivos contra a taxa de falsos positivos em diferentes limiares de decisão (OLIVEIRA, 2020).

2.5.6 Matriz de confusão

A matriz de confusão é uma ferramenta visual que mostra o desempenho do modelo de classificação em relação às classes verdadeiras e previstas. Ela apresenta a contagem de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos, permitindo uma análise detalhada dos erros de classificação (SOUSA et al., 2022).

Tabela 13 – Representação de uma Matriz de Confusão.

Sample	**Predicted True**	**Predicted False**
True	True Positive (TP)	False Negative (FN)
False	False Positive (FP)	True Negative (TN)

Na classificação de Fake News, a matriz de confusão ajuda a identificar onde o modelo está cometendo erros, como classificar erroneamente notícias verdadeiras como falsas, e vice-versa. Essa métrica é essencial para entender os pontos fortes e fracos do modelo e para ajustar estratégias de melhoria.

Embora este trabalho aborda diversas métricas de desempenho, como Acurácia, Precisão, Recall, F1-score, Curva ROC e Matriz de Confusão, a implementação prática dessas análises foi simplificada através do uso das bibliotecas sklearn.metrics e seaborn. Essas bibliotecas proporcionam funções pré-construídas que automatizam o cálculo dessas métricas, eliminando a necessidade de computar manualmente cada valor. Por exemplo, sklearn.metrics oferece uma ampla gama de ferramentas para avaliar o desempenho

dos modelos, enquanto seaborn facilita a visualização dos resultados através de gráficos intuitivos. Assim, a utilização dessas bibliotecas não só economiza tempo, mas também garante a precisão e a consistência dos resultados apresentados. Esta abordagem permite focar mais na interpretação dos resultados e nas melhorias do modelo, sem se preocupar com a complexidade dos cálculos subjacentes.

2.6 Banco de dados fake.br corpus

Nesse contexto, no qual informações são tratadas como dados, a escolha de um banco de dados é essencial para investigar a possibilidade de se extrair padrões de notícias verdadeiras e falsas. Um banco de dados robusto e bem estruturado permite o armazenamento e a gestão eficiente desses dados, facilitando o processo de análise e identificação de padrões específicos associados a Fake News. Isso possibilita uma abordagem sistemática para detectar e mitigar a disseminação de desinformação, tornando-se um componente fundamental para o sucesso desse processo investigativo.

O banco de dados que foi utilizado para construção desse estudo é o Fake.Br Corpus, desenvolvido por Monteiro et al. (2018). Este banco de dados contém notícias verdadeiras e falsas, todas em língua portuguesa. A coleta das notícias falsas foi realizada manualmente na web para garantir que elas eram realmente falsas, seguida de uma coleta semiautomática das notícias verdadeiras correspondentes às falsas. É necessário haver um equilíbrio entre instâncias positivas e negativas para validar os padrões (SANTOS; MONTEIRO; PARDO, 2018).

No estudo realizado por (SANTOS; MONTEIRO; PARDO, 2018) foram coletadas um total de 7.200 notícias, com uma divisão equitativa entre notícias verdadeiras e falsas, durante um intervalo de tempo determinado entre janeiro de 2016 a janeiro de 2018. Os dados coletados na internet estão, portanto, limitados a esse período específico.

Foi realizada uma análise e coleta manual de todas as notícias falsas disponíveis em um determinado período, provenientes de quatro sites: Diário do Brasil, A Folha do Brasil, The Journal Brasil e Top Five TV. Em seguida, foi filtrado as notícias que apresentavam meias verdades, deixando apenas aquelas que eram completamente falsas. Já as notícias verdadeiras foram coletadas de forma semiautomática, integrando o corpus de dados (GRILLO et al., 2022).

Para coletar notícias verdadeiras e relacionadas às falsas, utilizou-se um crawler para pesquisar palavras-chave nas principais agências de notícias do Brasil, resultando em 40.000 notícias. Em seguida, aplicou-se uma medida de similaridade lexical para selecionar notícias verdadeiras mais parecidas com as falsas, seguida de uma verificação manual. É importante ressaltar que algumas notícias verdadeiras negavam explicitamente a falsa correspondência, enquanto outras apenas tratavam do mesmo tema (SANTOS;

MONTEIRO; PARDO, 2018). O banco está dividido da seguinte maneira:

Tabela 14 – Categorias do conjunto de dados Fake.BI

Categoria	**Quantidade**	**Porcentagem**
Política	4180	58,0%
TVs e Celebridades	1544	21,4%
Sociedade e Notícias diárias	1276	17,7%
Ciência e Tecnologia	112	1,5%
Economia	44	0,7%
Religião	44	0,7%
Total	**7200**	**100%**

Na Tabela 14 é possível observar que as Fake News se espalham pelos mais diversos meios, mas é na política que está concentrado 58% dos dados. Isso evidencia a tendência das Fake News de se concentrarem em assuntos relacionados à política, utilizando pautas eleitorais como catalisadores para a disseminação de informações falsas. A propagação dessas notícias pode influenciar a opinião pública e potencialmente impactar os resultados eleitorais, ressaltando a importância de estratégias eficazes para combater a desinformação, especialmente em tempos de eleições. Entender como as notícias são organizadas dentro do Fake.Br Corpus é o primeiro passo para realizar uma análise precisa.

2.6.1 Análise exploratória de dados

Após analisar a estrutura do corpus "Fake br", foi observado que ele possui uma pasta dedicada a notícias verdadeiras, com cada notícia armazenada individualmente em formato Txt. Da mesma forma, as notícias falsas também estão organizadas em suas próprias pastas. Diante disso, foi decidido criar um banco de dados para facilitar consultas futuras dentro desse vasto conjunto de notícias, evitando a necessidade de acessar cada arquivo txt individualmente. Outro ponto importante é que existem metadados associados a cada notícia, os quais serão incorporados ao banco para reforçar a veracidade das notícias. No entanto, esses metadados não serão utilizados no processamento de dados seguinte, uma vez que este já foi realizado apenas com as notícias verdadeiras e falsas, após a análise inicial do banco, para confirmar o'que foi apresentado na Tabela 1 foi gerado uma nuvem de palavras:

A nuvem de palavras na Figura 8. gerada a partir do banco de dados Fake BR, após a análise inicial, demonstra que as informações apresentadas pela tabela de categorias se confirmam, uma vez que a maioria das palavras estão relacionadas à política. Isso confirma o equilíbrio do banco de dados, pois tanto nas notícias falsas quanto nas verdadeiras o tema política se sobressai em relação aos outros temas.

Figura 8 – Nuvem de palavras

3 METODOLOGIA

O presente Trabalho de Conclusão de Curso (TCC) está dividido em duas fases. A primeira fase, focada na teoria, abrange a revisão bibliográfica e a fundamentação teórica de todo o trabalho a ser desenvolvido. Nessa etapa, foram discutidos de forma conceitual toda estrutura do trabalho. Os algoritmos selecionados para o estudo, incluem Naive Bayes, Convolutional Neural Network (CNN) e Support Vector Machine (SVM).

Na segunda fase, com base no embasamento teórico previamente estabelecido, será realizada a análise qualitativa e quantitativa dos algoritmos selecionados. Essa etapa incluirá a aplicação dos modelos, seu treinamento e a avaliação detalhada das métricas de desempenho. Os resultados obtidos serão analisados para identificar a eficácia de cada algoritmo na identificação de Fake News.

A abordagem metodológica descrita na primeira fase visa estabelecer uma base sólida para as implementações práticas subsequentes. Cada aspecto teórico foi abordado com o objetivo de fornecer um entendimento claro e detalhado das etapas que serão seguidas na fase prática. Com a conclusão desta segunda fase, espera-se integrar de maneira eficaz os conceitos discutidos e aplicá-los nas analises posteriores.

3.0.1 Etapa 1

Revisão Sistemática e Coleta de Dados resultando no levantamento de artigos e identificação das bases de consulta científica. Neste tópico, são descritos os procedimentos utilizados para a identificação de artigos relacionados aos temas "Fake News"e "Aprendizado de Máquina"em bases de dados acadêmicas e no Google Acadêmico.

3.0.1.1 Fake News

Para a identificação de artigos sobre "Fake News", foi utilizado o Periódico Capes como principal fonte de consulta. Os critérios de busca aplicados foram:

- Seleção de artigos que abordassem a temática "Fake News";
- Consideração do período de publicação entre 2017 e 2024;
- Restrição a periódicos revisados por pares;
- Foco em artigos de acesso aberto.

Foi realizada uma priorização dos artigos com base na relevância para a área de pesquisa. Esse processo resultou na identificação de 70 artigos que atendiam aos critérios estabelecidos e eram pertinentes ao tema "Fake News".

3.0.1.2 Aprendizado de Máquina e Fake News

Além disso, foi realizada uma busca adicional no Google Acadêmico e no Periódico Capes para investigar a intersecção entre "Aprendizado de Máquina"e "Fake News", com foco em artigos no formato PDF. Os critérios adotados para essa busca foram:

- Identificação de artigos que abordassem simultaneamente "Aprendizado de Máquina"e "Fake News";
- Consideração do período de publicação entre 2017 e 2024;
- Foco em artigos disponíveis no formato PDF;
- Prioridade dada a artigos relevantes para a área específica da pesquisa.

Essa busca resultou na identificação de 396 artigos que atendiam aos critérios definidos e eram relevantes para o estudo da relação entre "Aprendizado de Máquina"e "Fake News".

3.0.1.3 Estrutura dos Artigos Levantados

Os artigos selecionados foram organizados em dois blocos principais para uma melhor compreensão e fundamentação do estudo. O primeiro bloco reúne artigos que abordam exclusivamente o tema "Fake News", oferecendo uma base sólida para entender o fenômeno em questão. O segundo bloco inclui artigos que exploram a aplicação de técnicas de "Aprendizado de Máquina"na detecção de Fake News, incluindo aqueles que combinam "Aprendizado de Máquina"e "Fake News"como foco principal.

A estrutura em dois blocos foi escolhida para garantir que a discussão sobre Fake News fosse adequadamente fundamentada antes da introdução e análise dos algoritmos de aprendizado de máquina. Durante a triagem inicial dos artigos, observou-se que muitos estavam fora do escopo de nossa pesquisa, seja por focarem predominantemente na COVID-19, seja por utilizarem técnicas e algoritmos incompatíveis com a base de dados Fake BR, que já é rotulada.

Após a aplicação dos critérios de filtragem, foram selecionados 50 artigos que se destacaram pela relevância e qualidade, proporcionando uma base teórica robusta para a construção da linha narrativa deste trabalho. Esses artigos foram essenciais para mapear o desenvolvimento do fenômeno das Fake News, desde sua origem e disseminação até os impactos sociais e políticos causados pela desinformação.

Além disso, os artigos selecionados contribuíram significativamente para a análise crítica das soluções tecnológicas, especialmente aquelas baseadas em técnicas de aprendizado de máquina, demonstrando como essas abordagens podem ser eficazes na detecção e mitigação de Fake News. Dessa forma, a seleção criteriosa desses artigos permitiu uma abordagem abrangente e fundamentada, que sustenta a proposta central deste estudo: o uso do aprendizado de máquina como uma ferramenta estratégica no combate à desinformação.

Os dados utilizados são provenientes do *Fake.Br Corpus*, uma base de dados pública e amplamente reconhecida que contém notícias rotuladas como verdadeiras ou falsas. Este corpus é essencial para garantir a qualidade e confiabilidade dos resultados, sendo uma referência importante para estudos em língua portuguesa. A escolha do *Fake.Br Corpus* foi motivada pela sua abrangência e pelo fato de ser amplamente utilizado em pesquisas acadêmicas, o que facilita a comparação dos resultados obtidos com estudos anteriores.

3.1 Etapa 2

Após a definição de toda a estrutura do trabalho, os tópicos teóricos serão colocados em prática, começando pela fase de pré-processamento dos dados. Em seguida, realizando a implementação dos algoritmos selecionados, seguida pela avaliação dos modelos utilizando métricas de desempenho específicas.

3.1.1 Pré-processamento dos Dados

A etapa seguinte envolve o pré-processamento e a limpeza dos dados. Utilizando técnicas de Processamento de Linguagem Natural (NLP), como:

- **Remoção de caracteres especiais**: Eliminação de caracteres que não contribuem para a análise do texto.
- **Normalização de texto**: Padronização de palavras para uma forma base.
- **Tokenização**: Separação do texto em unidades significativas (tokens).
- **Remoção de stopwords**: Eliminação de palavras comuns que não agregam valor semântico (ex.: "e", "de", "a").

Essas técnicas são essenciais para transformar os dados brutos em um formato adequado para a análise. A remoção de caracteres especiais, por exemplo, evita que elementos irrelevantes influenciem os resultados dos algoritmos. A normalização de texto e a tokenização são etapas fundamentais para garantir que as palavras sejam tratadas de maneira consistente pelos modelos de aprendizado de máquina.

Além disso, a remoção de stopwords ajuda a focar a análise nas palavras que realmente carregam significado semântico. Essa etapa do pré-processamento é crucial para melhorar a precisão e a eficiência dos modelos, uma vez que reduz a dimensionalidade dos dados e elimina ruídos que poderiam prejudicar o desempenho dos algoritmos.

3.1.2 Implementação dos Algoritmos

Com os dados pré-processados, a linguagem de programação Python será utilizada para explorar técnicas de aprendizado de máquina e treinar os modelos. Serão empregados pacotes e bibliotecas populares, como:

- **Scikit-learn**: Para implementar o algoritmo Naive Bayes e SVM.
- **TensorFlow e Keras**: Para implementar a CNN.

A escolha dessas bibliotecas se deve à sua robustez e ampla utilização na comunidade acadêmica e industrial. O Scikit-learn é uma biblioteca consolidada para aprendizado de máquina, oferecendo uma interface simples e eficiente para a implementação de algoritmos como Naive Bayes e SVM. Já o TensorFlow e o Keras são ferramentas poderosas para a construção e treinamento de redes neurais, permitindo a implementação de arquiteturas complexas como a CNN.

O objetivo é construir modelos capazes de distinguir com precisão notícias verdadeiras de falsas, com base nas características textuais. Para isso, serão realizados diversos experimentos, ajustando os hiperparâmetros dos algoritmos e avaliando seu desempenho em diferentes cenários. A implementação será documentada detalhadamente, permitindo a replicação dos experimentos e a validação dos resultados obtidos.

3.1.3 Avaliação dos Modelos

Após o treinamento, os modelos serão testados para avaliar sua precisão e desempenho. Serão utilizadas as seguintes métricas:

- **Acurácia**: Proporção de previsões corretas.
- **Precisão**: Proporção de verdadeiros positivos sobre o total de positivos preditos.
- **Recall**: Proporção de verdadeiros positivos sobre o total de positivos reais.
- **F1-Score**: Média harmônica entre precisão e recall.

A avaliação dos modelos é uma etapa crucial para garantir a validade e a robustez dos resultados. A acurácia fornece uma visão geral da performance dos modelos, enquanto

a precisão e o recall ajudam a entender a capacidade dos algoritmos em identificar corretamente as classes positivas e negativas. O F1-Score, por sua vez, oferece uma medida equilibrada do desempenho, combinando precisão e recall em uma única métrica.

Os resultados serão visualizados através da matriz de confusão e da curva ROC, permitindo uma análise detalhada do desempenho dos modelos. A matriz de confusão é uma ferramenta poderosa para identificar os acertos e erros dos modelos, enquanto a curva ROC ajuda a avaliar a capacidade de discriminação dos algoritmos em diferentes limiares de decisão.

Os resultados indicam que:

- **SVM**: Demonstrou a maior acurácia (0.95), com valores idênticos de precisão, recall e F1-score (0.95) para ambas as classes, indicando um desempenho equilibrado e robusto.
- **CNN**: Obteve uma acurácia de 0.88, com precisão de 0.95 para a classe 0 e 0.82 para a classe 1, mostrando uma ligeira diminuição na capacidade de classificar corretamente as verdadeiras (1).
- **Naive Bayes**: Apresentou uma acurácia de 0.92, com alta precisão para ambas as classes (0.88 para a classe 0 e 0.98 para a classe 1), mas com uma variação no recall, especialmente na classe 1 (0.86), resultando em um F1-score mais equilibrado.

Esses resultados destacam o SVM como o algoritmo mais eficaz e equilibrado para esta tarefa específica de classificação binária, seguido pelo Naive Bayes e, por fim, a CNN. A análise dos resultados preliminares fornece insights valiosos para o aprimoramento dos modelos e a definição das próximas etapas do projeto.

4 PROCESSAMENTO E VALIDAÇÃO DOS ALGORITMOS

A partir deste capítulo, inicia-se a parte prática deste Trabalho de Conclusão de Curso (TCC), na qual são descritas detalhadamente as tecnologias e métodos aplicados no pré-processamento e no treinamento dos algoritmos. Serão apresentados gráficos de curva ROC e matrizes de confusão para analisar o desempenho dos modelos, permitindo a observação das características específicas de cada algoritmo no contexto da classificação de fake news. A seção encerra com uma comparação dos três algoritmos, evidenciando as métricas de desempenho relevantes, como acurácia, precisão, e F1-score.

4.1 Processo de treinamento e validação

O pré-processamento de dados começa com a importação das bibliotecas e pacotes necessários para manipulação e análise textual, como Pandas, NumPy e NLTK. A biblioteca Pandas é fundamental para manipular dados em formato tabular, permitindo a leitura e organização em DataFrames e facilitando o gerenciamento de grandes volumes de dados textuais. Já o NLTK (Natural Language Toolkit) oferece um conjunto robusto de ferramentas para processamento de linguagem natural, incluindo a remoção de stopwords (palavras irrelevantes) e a aplicação de stemming, técnica que reduz as palavras à sua raiz. Essas etapas são fundamentais para eliminar variações desnecessárias e manter o foco nas palavras mais relevantes para a análise.

Na sequência, o texto passa por uma normalização, onde caracteres especiais são removidos e as palavras são convertidas para minúsculas, garantindo consistência e reduzindo ruídos. Com o texto uniformizado, é possível prosseguir para a vetorização, que transforma o conteúdo textual em uma representação numérica compreensível para o algoritmo de aprendizado de máquina.

Para essa vetorização, utiliza-se o TfidfVectorizer (Term Frequency-Inverse Document Frequency), que atribui pesos aos termos de acordo com a frequência relativa em que aparecem nos documentos. Essa técnica reduz a influência de palavras muito frequentes e comuns, como artigos e preposições, destacando os termos mais significativos. Esse processo de vetorização é essencial para a classificação, pois permite que o algoritmo diferencie com maior precisão textos de fake news de textos verdadeiros, otimizando sua capacidade preditiva e tornando o modelo mais eficaz.

4.1.1 SVM (Support Vector Machine)

O treinamento do algoritmo SVM (Support Vector Machine) para a detecção de fake news seguiu uma metodologia rigorosa para garantir a precisão e a robustez do modelo. Primeiramente, foi realizada a leitura e preparação dos dados, onde cada notícia foi carregada em um DataFrame. Cada texto no dataset foi associado a um rótulo binário indicando se a notícia era verdadeira (1) ou falsa (0), formando a base para a classificação supervisionada. Para assegurar a qualidade e integridade dos dados, utilizou-se a função isnull().sum() para detectar valores ausentes, seguido pela aplicação de info(), que permite verificar o tipo de dados em cada coluna e o número de entradas não nulas. Esse procedimento garante que os dados estejam completos e prontos para serem processados, eliminando problemas que poderiam comprometer o desempenho do modelo.

Com a integridade dos dados verificada, a base foi dividida em conjuntos de treino e teste utilizando a função train_test_split da biblioteca Scikit-Learn, com uma divisão de 80% para treino e 20% para teste. Essa divisão é essencial para medir a capacidade de generalização do modelo, evitando que ele seja avaliado apenas em dados com os quais já teve contato. Essa prática também evita problemas de overfitting, pois permite que o modelo aprenda com uma porção significativa dos dados e seja testado em novas amostras, simulando o uso do modelo em dados reais futuros. A separação em treino e teste também permite ajustar o modelo em função de seus erros, garantindo um balanço entre acurácia e capacidade de generalização.

Na configuração do SVM, foi escolhido o kernel linear, uma função de transformação que determina como os dados são separados no espaço multidimensional. O kernel linear é apropriado para problemas em que as classes são linearmente separáveis, o que é comum em problemas de classificação de texto, onde padrões de linguagem específicos podem distinguir claramente uma classe da outra. Além disso, o kernel linear apresenta menor complexidade computacional em comparação a outros kernels, como o radial basis function (RBF), o que o torna uma escolha eficiente para grandes conjuntos de dados textuais. O modelo também foi configurado com um valor elevado para o hiperparâmetro C, definido como 1000. Esse parâmetro controla o rigor do modelo em relação aos erros de classificação, onde valores altos para C indicam que o modelo terá uma menor tolerância a erros, aumentando a precisão nas previsões ao custo de uma menor flexibilidade. Com C=1000, o SVM maximiza a margem entre as classes e ajusta-se aos dados com maior precisão, minimizando os erros de classificação.

4.1.2 CNN (Convolutional Neural Network)

Primeiramente, foi realizada a preparação dos dados textuais com a tokenização e o preenchimento (padding) das sequências. Para tal, o código utilizou o Tokenizer, configurado para considerar um máximo de 10.000 palavras. Após a tokenização, cada

texto foi convertido em uma sequência de inteiros correspondentes aos índices das palavras no vocabulário, limitados a um tamanho máximo de 100 palavras por sequência. Esse limite reduz a variabilidade entre tamanhos de texto, garantindo que todas as entradas tenham o mesmo comprimento, essencial para o processamento em redes neurais. A vetorização das palavras foi feita com um vetor de embedding de dimensão 50, permitindo que cada palavra fosse representada por um vetor em um espaço de menor dimensão, capturando relações semânticas entre palavras.

A CNN foi então construída com a biblioteca Keras, em uma arquitetura sequencial que incluiu camadas específicas para análise textual. A camada inicial foi uma camada de Embedding, responsável por converter as palavras em representações densas. Seguindo essa camada, foi usada uma camada Conv1D com 128 filtros e um kernel de tamanho 5, ativada pela função ReLU, para capturar padrões locais nos dados textuais, como grupos de palavras que caracterizam fake news. Após essa camada, foi aplicada uma camada GlobalMaxPooling1D, que reduz a dimensionalidade dos dados ao capturar os valores máximos dos mapas de características gerados, essencial para manter informações relevantes de forma compacta. Por fim, a rede incluiu uma camada densa intermediária com 10 neurônios e função de ativação ReLU, seguida de uma camada de saída com um único neurônio e ativação sigmoide, utilizada para realizar a classificação binária entre notícias verdadeiras e falsas.

Para o treinamento da CNN, o modelo foi compilado com o otimizador adam e a função de perda binary_crossentropy, uma escolha comum em problemas de classificação binária. A métrica de acurácia foi definida para avaliar o desempenho do modelo durante o treinamento. O treinamento foi realizado ao longo de 5 épocas, com um tamanho de lote de 32, e uma divisão interna de 20% dos dados de treino para validação. A cada época, o modelo ajustou seus parâmetros com base no feedback das previsões e nas discrepâncias com os rótulos reais, resultando em uma melhoria gradual da acurácia.

4.1.3 Naive Bayes

O processo de treinamento do modelo Naive Bayes com o classificador BernoulliNB foi desenvolvido com o objetivo de identificar padrões que distinguissem fake news de notícias verdadeiras em um dataset previamente processado e vetorizado. O Bernoulli Naive Bayes é especialmente eficaz para problemas de classificação binária, como o presente caso, no qual cada notícia é rotulada como verdadeira ou falsa. Esse classificador utiliza uma abordagem probabilística baseada na presença ou ausência de características (palavras) específicas, sendo particularmente vantajoso para dados textuais em que as características podem ser representadas em uma estrutura binária. Abaixo, descrevem-se detalhadamente as etapas de preparação dos dados, configuração do modelo e avaliação dos resultados obtidos.

Inicialmente, o pré-processamento dos dados incluiu a vetorização textual com o CountVectorizer, uma técnica essencial para transformar o texto em uma matriz de características numéricas. Nesse caso, o CountVectorizer converteu cada notícia em um vetor de frequência binária, onde cada coluna da matriz representa uma palavra única do vocabulário total e é preenchida com valores 1 ou 0, indicando a presença ou ausência de cada termo no texto. Esse formato é ideal para o classificador BernoulliNB, que calcula a probabilidade de uma notícia pertencer a uma determinada classe (falsa ou verdadeira) com base na presença de palavras relevantes. A vetorização binária é uma abordagem eficiente, especialmente para o Bernoulli Naive Bayes, que utiliza a ocorrência de palavras como preditores independentes, o que se alinha com a suposição de independência condicional do modelo.

Após a vetorização, os dados foram divididos em conjuntos de treino e teste, com 80% das amostras dedicadas ao treinamento e os 20% restantes reservados para a avaliação do modelo. Essa divisão permite que o modelo aprenda com uma porção significativa dos dados enquanto preserva uma parte para verificar sua capacidade de generalização. O conjunto de treino serve para o ajuste dos parâmetros internos do BernoulliNB, permitindo que o modelo aprenda as probabilidades condicionais de cada palavra para as classes “verdadeira” e “falsa”. Esse processo de aprendizado baseia-se na frequência com que palavras específicas aparecem em cada classe, permitindo ao modelo associar características linguísticas típicas a notícias falsas ou verdadeiras.

Durante o treinamento, o BernoulliNB ajusta seu cálculo de probabilidades com base nos dados de entrada, construindo uma tabela de frequência para cada palavra no contexto de cada classe. Assim, o modelo aprende a calcular a probabilidade de uma notícia pertencer a uma classe específica com base na presença de palavras-chave associadas àquela classe. Essa abordagem probabilística segue a fórmula de Bayes, na qual cada termo contribui de forma independente para o cálculo da probabilidade final. O uso de uma matriz binária permite ao BernoulliNB operar com eficiência, focando em palavras cuja ocorrência ou ausência seja mais significativa para a diferenciação entre classes. Essa estratégia reduz o impacto de palavras menos informativas, favorecendo o aprendizado de padrões que realmente distinguem notícias verdadeiras de Fake News.

5 RESULTADOS

5.1 Comparação entre os Algoritmos

Após o treinamento dos algoritmos, a geração de gráficos e a obtenção das métricas de desempenho para cada modelo tornam-se etapas indispensáveis. Esses elementos são essenciais para conduzir análises comparativas, tanto quantitativas quanto qualitativas, permitindo avaliar a eficácia de cada abordagem na identificação de fake news. Por meio dessas análises, é possível identificar os pontos fortes e limitações de cada algoritmo, bem como compreender o impacto de diferentes configurações, como o ajuste de hiperparâmetros e a escolha de técnicas de pré-processamento. Essa etapa também possibilita uma visualização clara dos resultados, facilitando a comunicação das conclusões de maneira acessível e fundamentada.

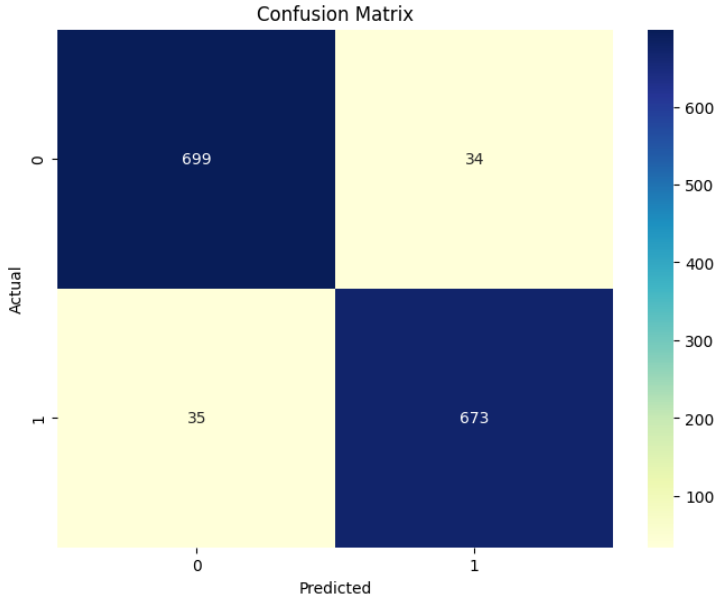

Figura 9 – Matriz de confusão do algoritmo SVM.

A matriz de confusão do algoritmo SVM Figura 9 apresentou um desempenho consistente na classificação das instâncias de fake news. Foram corretamente classificadas 699 instâncias da classe 0 (falsas) e 673 instâncias da classe 1 (verdadeiras). Contudo, o modelo gerou 34 falsos positivos (casos da classe 0 classificados incorretamente como 1) e 35 falsos negativos (casos da classe 1 classificados incorretamente como 0). Esses resultados evidenciam um desempenho equilibrado e preciso do SVM, com uma taxa de erro reduzida para ambas as classes. O baixo índice de falsos positivos e falsos negativos destaca a capacidade robusta do SVM em distinguir entre instâncias verdadeiras e falsas,

caracterizando-o como uma ferramenta eficaz e confiável para a tarefa de classificação binária no contexto da detecção de fake news.

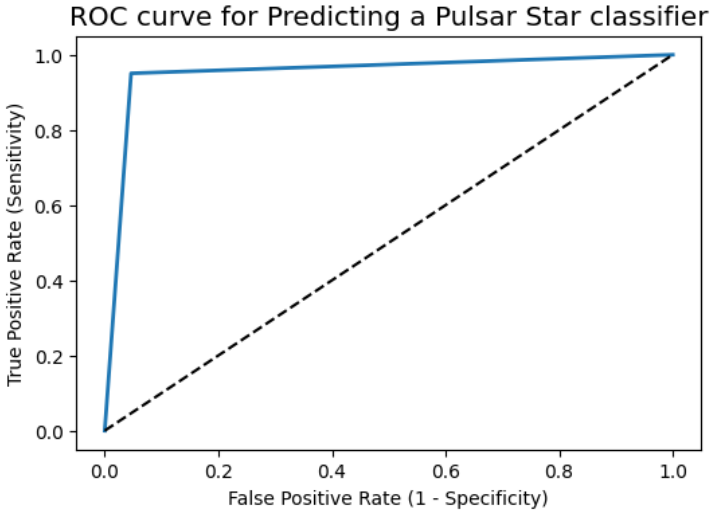

Figura 10 – Curva ROC do algoritmo SVM

A curva ROC (Receiver Operating Characteristic) apresentada na Figura 10 reflete o desempenho do modelo SVM na tarefa de classificação binária. A proximidade da curva com o eixo vertical, seguida de uma extensão quase horizontal ao longo do topo, indica uma alta taxa de sensibilidade (True Positive Rate), o que significa que o modelo é eficiente em identificar corretamente instâncias positivas (notícias verdadeiras). Simultaneamente, a baixa taxa de falsos positivos (False Positive Rate) sugere que o modelo apresenta uma elevada especificidade, ou seja, consegue distinguir com precisão entre instâncias positivas e negativas. A inclinação acentuada da curva no início da ROC, seguida por uma estabilização próxima ao valor máximo da taxa de sensibilidade, é indicativa de um bom equilíbrio entre sensibilidade e especificidade. Em termos práticos, essa configuração da curva implica que o SVM possui um desempenho robusto, minimizando tanto os falsos positivos quanto os falsos negativos. Uma curva ROC que se aproxima do canto superior esquerdo, como a observada, indica um modelo com uma AUC (Area Under Curve) próxima de 1, o que é ideal e denota um alto poder discriminativo do modelo SVM na tarefa de detecção de fake news.

A matriz de confusão da CNN na Figura 11 demonstra que o modelo classificou corretamente 584 instâncias da classe 0 (notícias falsas) e 679 instâncias da classe 1 (notícias verdadeiras). Contudo, a CNN apresentou 149 falsos positivos, nos quais instâncias da classe 0 foram incorretamente classificadas como pertencentes à classe 1. Além disso, houve 29 falsos negativos, indicando casos em que instâncias da classe 1 foram classificadas erroneamente como 0. Esses valores refletem que, embora a CNN tenha uma alta taxa de acertos gerais, existe uma discrepância no tratamento das classes, com um número considerável de falsos positivos. Esse comportamento sugere uma tendência da rede em

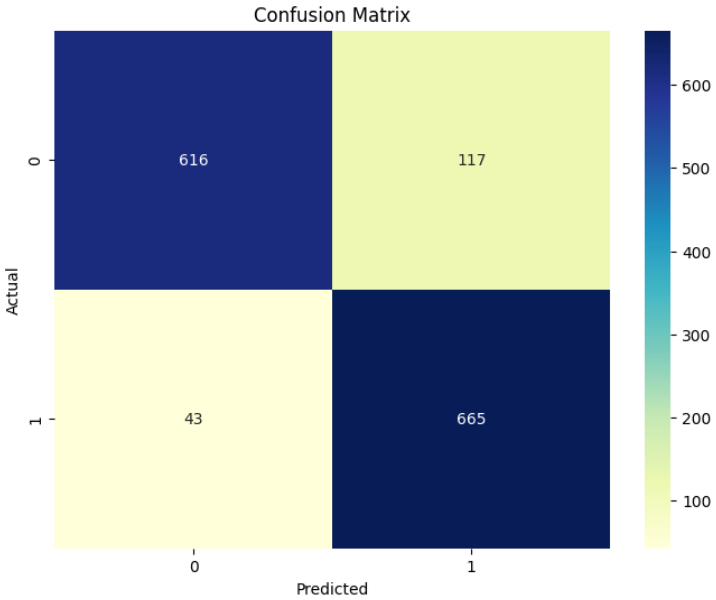

Figura 11 – Matriz de confusão do algoritmo CNN.

classificar instâncias da classe 0 como 1, o que pode reduzir a especificidade do modelo e, potencialmente, afetar sua confiabilidade em cenários onde é fundamental minimizar o número de falsos positivos. A presença de muitos falsos positivos implica que o modelo poderia erroneamente identificar notícias falsas como verdadeiras, o que é problemático em aplicações de detecção de fake news. No entanto, a taxa relativamente baixa de falsos negativos indica que a CNN possui uma boa sensibilidade, ou seja, é eficiente em identificar instâncias verdadeiras.

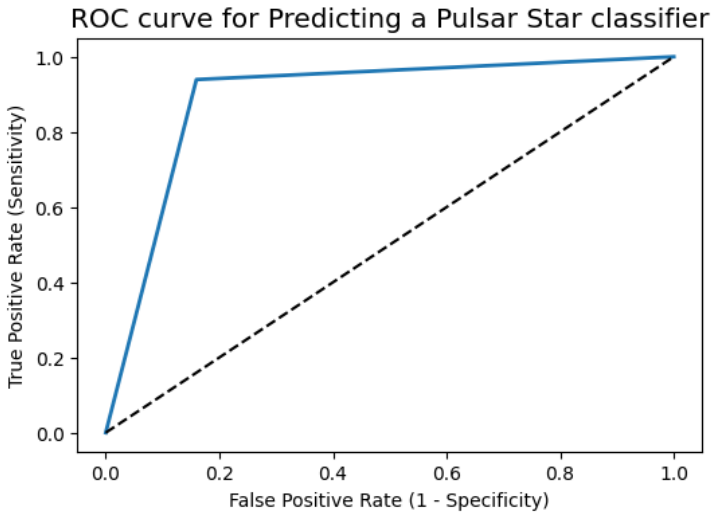

Figura 12 – Curva ROC do algoritmo CNN

A curva ROC apresentada na Figura 12 indica o desempenho do classificador CNN na identificação de pulsars. Observamos que a curva rapidamente se aproxima do eixo Y, sugerindo que o modelo alcança uma alta taxa de verdadeiros positivos (sensibilidade)

com poucos falsos positivos. Essa característica é desejável, pois demonstra a capacidade do modelo em identificar corretamente pulsars, mantendo a precisão. Após o ponto inicial de inclinação, a curva segue uma trajetória quase linear até o ponto (1,1), o que reforça a alta precisão do classificador, mesmo em níveis mais altos de sensibilidade. A linha pontilhada diagonal indica a performance de um classificador aleatório; o fato de a curva ROC estar consistentemente acima dessa linha evidencia que o modelo CNN possui um desempenho significativamente melhor que o acaso, mostrando-se eficaz para o problema de classificação de pulsars.

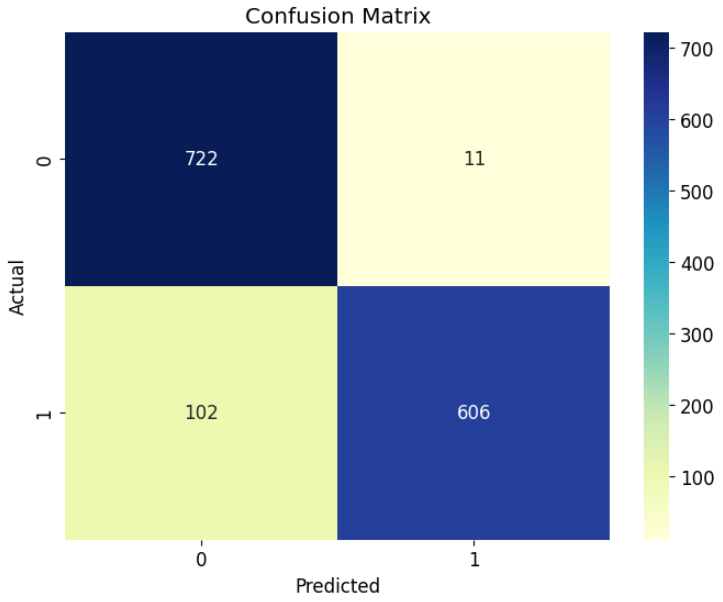

Figura 13 – Matriz de confusão do algoritmo Naive Bayes.

Na matriz de confusão gerada pelo algoritmo Naive Bayes apresenta na Figura 13 quatro métricas principais relacionadas à classificação de notícias verdadeiras (1) e falsas (0). No quadrante superior esquerdo, temos 722 notícias falsas corretamente classificadas como falsas (verdadeiros negativos). No quadrante superior direito, há 11 notícias falsas incorretamente classificadas como verdadeiras (falsos positivos). No quadrante inferior esquerdo, 102 notícias verdadeiras foram erroneamente classificadas como falsas (falsos negativos). No quadrante inferior direito, 606 notícias verdadeiras foram corretamente classificadas como verdadeiras (verdadeiros positivos). Esses resultados indicam que o algoritmo possui alta precisão na identificação de notícias verdadeiras e uma precisão significativa na detecção de notícias falsas, apesar de uma quantidade considerável de falsos negativos.

Observa-se que a curva ROC da Figura 14 está significativamente acima da linha de referência, indicando que o modelo possui uma boa capacidade de discriminação entre as classes. A área sob a curva (AUC) próxima de 1 sugere que o modelo tem um excelente desempenho na distinção entre as classes, com alta sensibilidade e especificidade, sendo

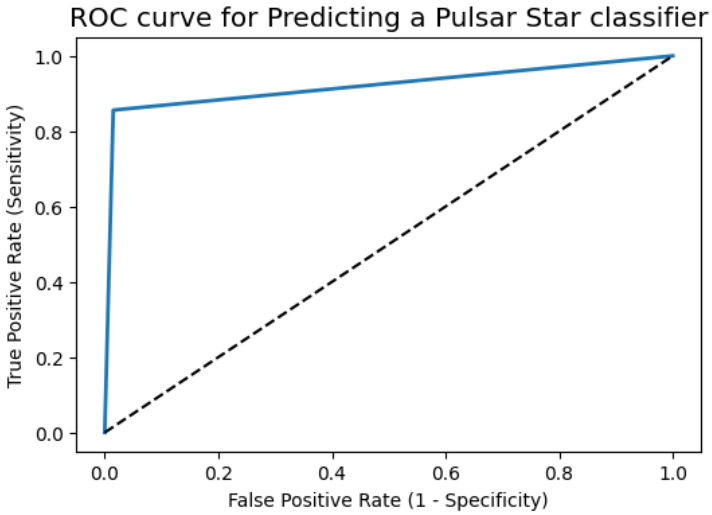

Figura 14 – Curva ROC do algoritmo Naive Bayes

eficiente em minimizar tanto falsos positivos quanto falsos negativos. Essa avaliação é fundamental para validar a eficácia do Naive Bayes em aplicações astronômicas.

5.2 Resultado dos Algoritmos

Algoritmo	**Acurácia**	**Classe**	**Precisão**	**Recall**	**F1-score**
SVM	0.95	0	0.95	0.95	0.95
		1	0.95	0.95	0.95
CNN	0.89	0	0.93	0.84	0.89
		1	0.85	0.94	0.89
Naive Bayes	0.92	0	0.88	0.98	0.93
		1	0.98	0.86	0.91

Tabela 15 – Desempenho dos algoritmos de aprendizado de máquina.

A Tabela Figura 15 apresenta os resultados de três algoritmos de aprendizado de máquina - SVM (Support Vector Machine), CNN (Convolutional Neural Network) e Naive Bayes - para a classificação de valores verdadeiros (1) e falsos (0). O algoritmo SVM demonstrou a maior acurácia, com 0.95, e valores idênticos de precisão, recall e F1-score (0.95) para ambas as classes, indicando um desempenho equilibrado e robusto na classificação de verdadeiras e falsas. A CNN obteve uma acurácia de 0.88, com precisão de 0.95 para a classe 0 e 0.82 para a classe 1, refletindo uma ligeira diminuição na capacidade de classificar corretamente as verdadeiras (1). O Naive Bayes apresentou uma acurácia de 0.92, com alta precisão para ambas as classes, sendo 0.88 para a classe 0 e 0.98 para a classe 1, mas uma ligeira variação no recall, especialmente na classe 1 (0.86), resultando em um F1-score mais equilibrado para ambas as classes.

Esses resultados ressaltam o desempenho superior do SVM, que se mostrou o algoritmo mais eficaz e equilibrado para a tarefa específica de classificação binária de fake news. A análise revelou que o SVM obteve os melhores índices de precisão e consistência, destacando-se pela capacidade de generalizar bem nos diferentes conjuntos de dados testados. Assim, o SVM foi escolhido como a principal abordagem para a aplicação web, garantindo uma detecção confiável e eficiente para o usuário final.

Tabela 16 – Resultados de trabalhos correlatos

Artigo	**Algoritmo**	**Acurácia**	**Precisão**	**Recall**	**F1-score**
Garcia (2023)	SVM	92,70%	89,20%	97,20%	93,00%
	CNN	95,72%	96,04%	95,97%	96,00%
	Naive Bayes	70,60%	63,00%	99,00%	77,03%
Silva (2022)	SVM	0,79997	0,76196	-	0,81345
Silva (2021)	SVM	0,9569	0,9606	-	0,9569
GRILLO (2022)	CNN	0,9630	0,9576	0,9667	0,9621
SHARMA (2020)	Naive Bayes	0,60	0,59	0,92	0,72
(ADIBA et al., 2020)	Naive Bayes	0,874	0,86	0,89	0,87
(AJAO, 2018)	CNN	80,38	43,94	39,53	39,70

A análise dos resultados permite uma comparação consistente com estudos correlatos, evidenciando as diferentes abordagens metodológicas e os pré-processamentos aplicados. Embora essas variações sejam justificadas pelas especificidades de cada pesquisa, elas não comprometem a análise comparativa, mas ampliam a compreensão das tendências gerais e do desempenho relativo dos algoritmos.

Na Tabela 16 , observa-se o uso de algoritmos como SVM, CNN e Naive Bayes, amplamente discutidos na literatura, com desempenhos variáveis conforme o contexto e os dados analisados. Esses resultados demonstram que cada técnica apresenta méritos adaptáveis a diferentes cenários. Nesta pesquisa, os valores obtidos alinham-se aos encontrados em trabalhos anteriores, evidenciando a eficiência dos métodos utilizados.

Mesmo com configurações heterogêneas, como variações no pré-processamento e nos critérios de avaliação, as métricas de acurácia e F1-score apresentaram tendências consistentes. Isso reforça a robustez dos algoritmos analisados e a confiabilidade dos resultados obtidos, destacando a validade do estudo em comparação com a literatura.

Assim, a análise dos resultados confirma o desempenho competitivo do método proposto neste trabalho. Essa consistência evidencia sua relevância como uma contribuição significativa ao campo de estudo, posicionando-o em sintonia com os avanços relatados na literatura e demonstrando seu potencial para oferecer soluções robustas e eficazes.

6 CONCLUSÕES E TRABALHOS FUTUROS

Os resultados deste estudo evidenciam a complexidade do problema das Fake News e os desafios envolvidos na sua identificação. A propagação de desinformação é um fenômeno multifacetado e dinâmico, influenciado por diversos fatores como o contexto social, político e tecnológico. A tarefa de classificar notícias como verdadeiras ou falsas não é simples, uma vez que a linha entre a verdade e a mentira pode ser tênue, e as técnicas de manipulação de informação estão em constante evolução. Nesse cenário, os algoritmos de aprendizado de máquina demonstraram um desempenho promissor, com o SVM alcançando a maior acurácia (0,95), seguido pelo Naive Bayes e a CNN. Esses resultados, embora positivos, foram obtidos em um ambiente controlado com um banco de dados limitado, o que aponta para a necessidade de validação em contextos mais dinâmicos e em tempo real.

Embora o estudo tenha mostrado que os algoritmos, especialmente o SVM, apresentam boa capacidade de distinguir entre notícias verdadeiras e falsas, é importante destacar que esses resultados são oriundos de uma base de dados específica, com recorte temporal limitado. Isso significa que o ambiente de teste não reflete integralmente as variáveis complexas e em constante mudança do cenário digital, onde novas informações e padrões de manipulação de dados surgem continuamente. Portanto, embora os resultados indicativos sejam promissores, é preciso realizar mais testes utilizando bases de dados mais robustas, diversificadas e atualizadas para verificar a real eficácia dos modelos em contextos mais amplos e dinâmicos.

A identificação de Fake News, como foi possível observar neste estudo, não depende apenas de técnicas matemáticas e algoritmos poderosos, mas também da constante atualização dos dados e do entendimento de como a desinformação evolui. Isso envolve, entre outros fatores, a análise de novas formas de manipulação de notícias, o estilo de escrita que pode enganar os leitores e as fontes de onde a informação se origina. A melhoria da acurácia dos algoritmos, portanto, não está somente no refinamento das métricas, mas também na adaptação dos modelos às mudanças rápidas do ambiente digital.

Com isso, o estudo revela a importância de continuar investindo no aprimoramento das abordagens de aprendizado de máquina para detecção de Fake News, mantendo um ciclo contínuo de testes, ajustes e validações. No futuro, a construção de um software que utilize esses dados processados, permitindo a verificação de notícias em tempo real, seria um passo significativo para tornar as ferramentas de combate à desinformação mais acessíveis e eficazes. Esse desenvolvimento, porém, exige não apenas algoritmos precisos, mas também uma integração constante de novos dados e uma capacidade de adaptação

rápida às mudanças nas tendências da desinformação. Em última análise, o estudo da identificação de Fake News se mostra como um campo em constante evolução, desafiador e essencial para a construção de um ambiente informativo mais saudável e confiável.

REFERÊNCIAS

ADIBA, F. I.; ISLAM, T.; KAISER, M. S.; MAHMUD, M.; RAHMAN, M. A. Effect of corpora on classification of fake news using naive bayes classifier. *International Journal of Automation, Artificial Intelligence and Machine Learning*, v. 1, n. 1, p. 80–92, 2020. Citado na página 61.

AHMAD, I.; YOUSAF, M.; YOUSAF, S.; AHMAD, M. O. Fake news detection using machine learning ensemble methods. *Complexity*, Wiley Online Library, v. 2020, n. 1, p. 8885861, 2020. Citado na página 42.

AJAO, S. Fake news identification on twitter with hybrid cnn and rnn models. In: *Proceedings of the 9th international conference on social media and society*. [S.l.: s.n.], 2018. p. 226–230. Citado na página 61.

ALMEIDA, R. F. S. de. Building portuguese language resources for natural language processing tasks. 2023. Citado na página 29.

ALTIERI, M. P.; APRO, P. H. O uso de machine learning na classificação de textos com ênfase em fake news. Universidade Presbiteriana Mackenzie, 2022. Citado na página 33.

BIRD, S.; KLEIN, E.; LOPER, E. *Natural language processing with Python: analyzing text with the natural language toolkit*. [S.l.]: "O'Reilly Media, Inc.", 2009. Citado na página 31.

BOVET, A.; MAKSE, H. A. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, Nature Publishing Group UK London, v. 10, n. 1, p. 7, 2019. Citado na página 26.

CONG, S.; ZHOU, Y. A review of convolutional neural network architectures and their optimizations. *Artificial Intelligence Review*, Springer, v. 56, n. 3, p. 1905–1969, 2023. Citado 2 vezes nas páginas 40 e 41.

COSTA, M. A. F. da; COSTA, M. d. F. B. da. Fake news–contribuições para os processos de ensino. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, v. 8, n. 11, p. 1–14, 2022. Citado 2 vezes nas páginas 11 e 16.

DARNTON, R. *A Short Guide to the History of Fake News and Disinformation*. 2023. Disponível em: <https://www.icfj.org/sites/default/files/201807/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf>. Acesso em: 18 fev. 2023. Citado na página 22.

DEPUTADOS, C. dos. *PL 2630/2020*. 2020. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735>. Acesso em: 20 set. 2021. Citado na página 28.

FISCHER, M.; HAQUE, R.; STYNES, P.; PATHAK, P. Identifying fake news in brazilian portuguese. In: SPRINGER. *International Conference on Applications of Natural Language to Information Systems*. [S.l.], 2022. p. 111–118. Citado na página 29.

GARCIA, G. L. Detecção de fake news utilizando aprendizado de máquina. Universidade Estadual Paulista (Unesp), 2023. Citado 5 vezes nas páginas 29, 33, 34, 35 e 38.

GRILLO, G. J. Y.; ROSATI, G. T. B.; HORA, J. M. da; SOARES, P. H. de S.; RODRIGUES, K. R. da H. Retificai: Um verificador de notícias falsas com base em ia e aprendizado de máquina. In: SBC. *Simpósio Brasileiro de Fatores Humanos em Sistemas Computacionais (IHC)*. [S.l.], 2022. p. 146–150. Citado na página 44.

GRINBERG, N.; JOSEPH, K.; FRIEDLAND, L.; SWIRE-THOMPSON, B.; LAZER, D. Fake news on twitter during the 2016 us presidential election. *Science*, American Association for the Advancement of Science, v. 363, n. 6425, p. 374–378, 2019. Citado na página 26.

ITUASSU, A.; LIFSCHITZ, S.; CAPONE, L.; MANNHEIMER, V. Campanhas online e democracia: As mídias digitais nas eleições de 2016 nos estados unidos e 2018 no brasil. *O Brasil vai às urnas. Londrina: Syntagma*, p. 15–48, 2019. Citado 2 vezes nas páginas 11 e 25.

KIBBLE, R. Introduction to natural language processing. *London: University of London*, 2013. Citado na página 31.

LAZER, D. M.; BAUM, M. A.; BENKLER, Y.; BERINSKY, A. J.; GREENHILL, K. M.; MENCZER, F.; METZGER, M. J.; NYHAN, B.; PENNYCOOK, G.; ROTHSCHILD, D. et al. The science of fake news. *Science*, American Association for the Advancement of Science, v. 359, n. 6380, p. 1094–1096, 2018. Citado 2 vezes nas páginas 11 e 28.

LGPD. *Lei nº 13.709, de 14 de agosto de 2018*. 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Acesso em: 18 jan. 2023. Citado na página 27.

MANNING, C. D. *Introduction to information retrieval*. [S.l.]: Cambridge university press, 2008. Citado na página 32.

MARCH, N. D.; FONTOLAN, M.; GITAHY, L.; PERON, A. E. R. Redes, pânico e a "bruxa do guarujá". *Observatório da Imprensa*, v. 1171, 2022. Disponível em: <https://www.observatoriodaimprensa.com.br/desinformacao/redes-panico-e-a-bruxa-do-guaruja/>. Acesso em: 10 mar. 2023. Citado na página 20.

MORONI, J. Possíveis impactos de fake news na percepção-ação coletiva. *Complexitas–revista de Filosofia temática*, v. 3, n. 1, p. 130–160, 2018. Citado 2 vezes nas páginas 25 e 26.

NOUVELLES, L. relative à la lutte contre la diffusion de fausses. *n° 2018-1202 du 22 décembre 2018*. 2018. JORF n°0297 du 23 décembre 2018, texte n° 1. Disponível em: <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559>. Acesso em: 18 jan. 2023. Citado na página 27.

OLIVEIRA, N. R. D. Aplicações eficientes do processamento de linguagem natural: Identificação de notícias falsas, sumarização automática de textos e construção de ontologias. 2020. Citado 2 vezes nas páginas 11 e 43.

RICARDO, B.-Y. et al. *Modern information retrieval*. [S.l.]: Pearson Education India, 1999. Citado na página 30.

SALTON, G. Introduction to modern information retrieval. *McGrawHill Book Co*, 1983. Citado na página 32.

SANTOS, R. L.; MONTEIRO, R. A.; PARDO, T. A. The fake. br corpus-a corpus of fake news for brazilian portuguese. In: *Latin American and Iberian Languages Open Corpora Forum (OpenCor)*. [S.l.: s.n.], 2018. p. 1–2. Citado 2 vezes nas páginas 44 e 45.

SHARMA, U.; SARAN, S.; PATIL, S. M. Fake news detection using machine learning algorithms. *International Journal of creative research thoughts (IJCRT)*, v. 8, n. 6, p. 509–518, 2020. Citado 2 vezes nas páginas 30 e 31.

SILVA, C. V. M. Análise exploratória e experimental sobre detecção inteligente de fake news. Pós-Graduação em Ciência da Computação, 2020. Citado 2 vezes nas páginas 29 e 37.

SINGH, K.; SINGH, D.; MISHRA, N. Convolutional neural networks and its architecture. *International Journal of Health Sciences*, ScienceScholar, n. I, p. 9183–9190. Citado na página 40.

SOUSA, F.; BARBOSA, A.; OLIVEIRA, C.; BRAGA, R. Detecção de fake news em língua portuguesa combinando redes neurais convolucionais e algoritmos de aprendizagem de máquina. In: SBC. *Anais do XL Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. [S.l.], 2022. p. 336–348. Citado na página 43.

VALKENBORG, D.; ROUSSEAU, A.-J.; GEUBBELMANS, M.; BURZYKOWSKI, T. Support vector machines. *American Journal of Orthodontics and Dentofacial Orthopedics*, Elsevier, v. 164, n. 5, p. 754–757, 2023. Citado na página 37.

WARDLE, C.; DERAKHSHAN, H. *Information disorder: Toward an interdisciplinary framework for research and policymaking*. [S.l.]: Council of Europe Strasbourg, 2017. v. 27. Citado 3 vezes nas páginas 18, 19 e 20.