



Programa de Pós-Graduação em

Computação Aplicada

Mestrado Acadêmico

Jonathan Pereira Camargo

Avaliação da integração de abordagens de
verificação de Fake News

São Leopoldo, 2022

JONATHAN PEREIRA CAMARGO

**AVALIAÇÃO DA INTEGRAÇÃO DE ABORDAGENS DE VERIFICAÇÃO DE FAKE
NEWS**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Computação Aplicada, pelo Programa de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos (UNISINOS).

Orientador(a): Prof. Dr. Sandro José Rigo

São Leopoldo

2022

C173a Camargo, Jonathan Pereira.
Avaliação da integração de abordagens de verificação
de fake news / Jonathan Pereira Camargo. – 2022.
67 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio
dos Sinos, Programa de Pós-Graduação em Computação
aplicada, 2022.

“Orientador: Prof. Dr. Sandro José Rigo.”

1. Classificação de texto. 2. Fact-checking. 3. Fake news.
5. Inteligência artificial. I. Título.

CDU 004

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Silvana Dornelles Studzinski – CRB 10/2524)

JONATHAN PEREIRA CAMARGO

**AVALIAÇÃO DA INTEGRAÇÃO DE ABORDAGENS DE VERIFICAÇÃO DE FAKE
NEWS**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Computação Aplicada, pelo Programa de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos (UNISINOS).

Aprovado em 7 de abril de 2022

BANCA EXAMINADORA

Prof. Dr. Sandro José Rigo – Orientador – UNISINOS

Profa. Dra. Marta Rosecler Bez – Avaliadora – FEEVALE

Prof. Dr. Rafael Kunst – Avaliador – UNISINOS

AGRADECIMENTOS À CAPES

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

AGRADECIMENTOS

Sou grato à uma série de pessoas, desde amigos, família, professores, colegas e até ex-colegas, que apesar de não serem citados aqui, contribuíram de alguma maneira ao desenvolvimento deste trabalho.

Em especial, agradeço pela grande disposição do Prof. Dr. Sandro Rigo em estar presente em todas as etapas deste trabalho, ao ter me acolhido e orientado com toda sua sensibilidade e inteligência.

Agradeço aos meus pais e à minha irmã, por me ensinarem o valor da educação e do conhecimento, principalmente em tempos como o que vivemos. Agradeço à minha namorada, que me proporciona reflexões com diferentes perspectivas. Agradeço ao meu queridíssimo amigo Matheus Nienow, que me indicou à inscrição do mestrado e me incentivou a buscar qualificação.

Que todos possamos partilhar da presença de pessoas que nos inspiram e motivam a sermos ainda melhores.

RESUMO

A agilidade com que se pode produzir, divulgar e consumir informações em virtude da internet não só democratizou o acesso à informação, como também gerou abundância desta, tendo como efeito colateral a facilitação da propagação de informações falsas ou tendenciosas, denominadas popularmente como *Fake News*. Devido ao alto volume e efemeridade dos dados, métodos automatizados com técnicas de inteligência artificial tornam-se essenciais no processo de verificação de *Fake News*. A partir da leitura do estado da arte, observa-se que as abordagens existentes possuem limitações de aplicabilidade em contextos específicos e que não há uma abordagem capaz de lidar com diversos contextos sem comprometer seus resultados. Com base nisto, este trabalho propõe a integração de dois métodos para avaliar a integração de métodos distintos de verificação de *Fake News* com o intuito de ampliar a abrangência de aplicação em contextos diferentes. O método de classificação de texto obteve acurácia de 95,33% com uso de *Random Forest*, enquanto o método de *fact-checking* com *question answering* foi capaz de responder adequadamente as perguntas elaboradas. Foi proposta uma metodologia de comparação para analisar qualitativamente os resultados dos experimentos, que permitiu identificar contribuições e trabalhos futuros. Os textos classificados como falso-negativos no experimento de classificação serviram de subsídio para a elaboração das perguntas testadas no experimento de *fact-checking* com *question answering*, validando a complementaridade entre os métodos.

Palavras-chave: *fake news*; verificação de *fake news*; classificação de texto; *fact-checking*; inteligência artificial.

ABSTRACT

The agility with which information can be produced, disseminated and consumed by virtue of the internet not only democratized the access to information, but also generated an abundance of it, having as a side effect the facilitation of the propagation of false or biased information, popularly known as Fake News. Due to the high volume and ephemerality of the data, automated methods with artificial intelligence techniques become essential in the Fake News verification process. From the reading of the state of the art, it is observed that the existing approaches have limits of applicability in specific contexts and that there is no approach capable of dealing with different contexts without compromising their results. Based on this, this work proposes the integration of two methods to evaluate the integration of different methods of verification of Fake News in order to expand the scope of application in different contexts. The text classification method obtained an accuracy of 95.33% using Random Forest, while the fact-checking method with question answering was able to adequately answer the elaborated questions. A comparison methodology was proposed to qualitatively analyze the results of the experiments, which allowed the identification of contributions and future work. The texts classified as false negatives in the classification experiment served as a subsidy for the elaboration of the questions tested in the fact-checking experiment with question answering, validating the complementarity between the methods.

Key-words: fake news; fake news verification; text classification; fact-checking; artificial intelligence.

LISTA DE FIGURAS

Figura 1 - Modelo geral da aplicação	28
Figura 2 - Resultado após uma busca.....	29
Figura 3 - Visão geral do <i>Web Scrapping</i>	30
Figura 4 – Esquema de classificação de texto	34
Figura 5 – Texto do corpus SIRENE-News	36
Figura 6 – Texto do corpus Fake.br	37
Figura 7 - Notícia verdadeira no site G1.....	39
Figura 8 - Notícia verdadeira no site Estadão	40
Figura 9 – Matriz de confusão.....	42
Figura 10 – Resultados parciais do experimento de classificação de textos.....	43
Figura 11 - Visão geral do <i>Haystack</i>	46
Figura 12 – Resposta para a pergunta “Qual é a crítica de Zico ao Flamengo?”	48
Figura 13 – Metodologia de comparação	49
Figura 14 – Resultados do método de classificação de texto.....	50
Figura 15 – Exemplo de texto verdadeiro classificado como falso	51
Figura 16 - “Qual é a crítica de Zico ao Flamengo?”	52
Figura 17 - “O que o deputado Wladimir Costa divulgou no <i>whatsapp</i> ?”	53
Figura 18 - “Quem é Jair Bolsonaro?”	54
Figura 19 - “Qual foi a captação da poupança em 2017?”	55
Figura 20 - “Qual foi a inflação em 2017?”	55

SUMÁRIO

1 INTRODUÇÃO	10
1.1 Questão de pesquisa	12
1.2 Objetivos	12
1.2.1 Objetivo Geral	12
1.2.2 Objetivos Específicos	12
1.3 Metodologia	13
1.4 Estrutura do Texto	13
2 FUNDAMENTAÇÃO TEÓRICA	14
2.1 Fake News.....	14
2.2 Contexto.....	15
2.3 Métodos manuais.....	16
2.4 Métodos automatizados.....	16
2.5 Extração de Informação.....	18
2.6 Processamento de Linguagem Natural	19
3 TRABALHOS RELACIONADOS	21
3.1 Fake News Identification on Twitter with Hybrid CNN and RNN Models	21
3.2 What is Real or Fake? - Machine Learning Approaches for Ru-mor Verification using Stance Classification	22
3.3 Combining Neural, Statistical and External Features for Fake News Stance Identification	22
3.4 dEFEND: A System for Explainable Fake News Detection	23
3.5 Fake News Early Detection: A Theory-driven Model.....	23
3.6 SocialTruth Project Approach to Online Disinformation (Fake News) Detection and Mitigation.....	24
3.7 Análise crítica e comparação entre os trabalhos	24
4 PLATAFORMA WEB FAKE	27
4.1 Modelo.....	27
4.2 Módulos e serviços	28
4.3 Usabilidade e funcionamento.....	29
5 MATERIAIS E MÉTODOS	32
5.1 Visão geral	32
5.2 Classificação de textos.....	33

5.3 Extração de informação e <i>Fact-checking</i>	34
5.4 Conjunto de dados e recursos	35
5.5 Metodologia e métricas de avaliação	37
6 EXPERIMENTOS E RESULTADOS	38
6.1 Experimento de classificação de texto	38
6.1.1 Conjunto de dados	38
6.1.2 Pré-processamento	40
6.1.3 Classificação	41
6.1.4 Resultados	41
6.2 Experimento de <i>fact-checking</i> com <i>Question Answering</i>	43
6.3.1 Conjunto de dados	44
6.3.2 <i>Haystack</i>	45
6.3.3 Descrição do método	46
6.3.4 Resultados	47
6.3 Plataforma de integração	48
6.4 Metodologia de comparação	49
7 CONCLUSÃO	57
7.1 Contribuições	58
7.2 Limitações do trabalho	58
7.3 Trabalhos futuros.....	59
REFERÊNCIAS	60

1 INTRODUÇÃO

A informação atua como o epicentro de diversas situações econômicas, políticas e sociais, sendo assim, exerce influência desde as mais simples e rotineiras dinâmicas da sociedade, até às mais complexas e elementares. A agilidade com que se pode produzir, divulgar e consumir informações em virtude da internet, não só democratizou o acesso à informação, como também gerou abundância desta, tendo como efeito colateral a facilitação da propagação de informações falsas ou tendenciosas (DELMAZO; VALENTE, 2018; DE OLIVEIRA et al., 2021; SHU et al., 2017).

O surgimento de sites e plataformas especializadas em checagem de fatos evidencia a importância atribuída ao tema das *Fake News*. Atualmente, mesmo portais tradicionais incorporaram seções com notícias sobre checagem de fatos. Isto indica não só como o tema se tornou relevante, mas também como há procura e consumo deste tipo de informação em uma camada convencional (SHARMA et al. 2019).

Uma possível definição sobre *Fake News* é de que estas são notícias falsas ou com conteúdo enganoso, capazes de gerar confusões como a crença em informações falsas ou descrença em informações verdadeiras (SHU et. al 2017).

Consumir notícias nas plataformas de mídias sociais digitais é mais barato e mais oportuno em comparação com a mídia tradicional, como jornais ou televisão. Além disso, a agilidade encontrada na internet facilita ações como compartilhamento, comentários e discussões com amigos e outras pessoas da rede, propiciando um ambiente fértil para a proliferação de *Fake News*. A alta taxa de consumo de informações em plataformas digitais combinada com a baixa confiança nas instituições de notícias potencializa a disseminação de *Fake News*. (DEMARTINI et al., 2020; SHU et al., 2017; TANDOC; LIM; LING, 2018; WARDLE et al., 2017).

Diante de tantos dados dispersos nas plataformas digitais, métodos automatizados com a utilização de Inteligência Artificial desempenham um papel fundamental na verificação de *Fake News*. Por exemplo, Cordeiro et al. (2019) faz o uso de *deep learning e machine learning*, comparando resultados entre uma rede neural *Long Short Term Memory* (LSTM), um algoritmo de Regressão Logística e um algoritmo de *Support Vector Machine* (SVM). Em outro exemplo, Cui et al. (2019) utilizam Redes Neurais Recorrentes (*Recurrent Neural Networks - RNN*) como base para o sistema dEFEND, dedicado a identificar a autenticidade das notícias e também

indicar a explicação do processo de detecção. Estudos exploratórios de inteligência computacional evidenciam que a hibridização de recursos linguísticos e probabilísticos, incluindo também redes neurais, podem resultar em um melhor desempenho do algoritmo de classificação de *Fake News*, como pode ser observado em (AJAO; BHOWMIK; ZARGARI, 2018; BHATT et al., 2018; CHORAS et al., 2019; ZHOU et al., 2020).

Devido às diferenças nas possibilidades de abordagem deste problema de identificação de notícias falsas, os trabalhos nesta área também são agrupados em categorias de métodos baseados em conteúdo e de métodos baseados em propagação (SHU et al., 2017; ZHOU et al., 2020). A primeira categoria, os métodos baseados em conteúdo, consiste em realizar a classificação de *Fake News* com base nas informações encontradas no conteúdo da notícia. Isto possibilita identificar uma *Fake News* antes mesmo de sua propagação e possivelmente conter seus impactos. Porém, métodos baseados em conteúdo precisam de algoritmos treinados no tema do texto a ser classificado para que obtenham uma melhor precisão (BHATT et al., 2018; CUI et al., 2019; SHU et al., 2017; ZHOU et al., 2020).

A segunda categoria, os métodos baseados em propagação, tem como objetivo utilizar as informações de propagação da notícia em redes de comunicação, tais como redes sociais. Neste são usados aspectos diversos, tais como, por exemplo, quem são os atores que propagam a notícia e como os usuários reagem à sua recepção, podendo classificar as reações dos usuários a partir das informações geradas por interações como comentários e compartilhamentos (AJAO; BHOWMIK; ZARGARI, 2018; CONROY; RUBIN; CHEN, 2015; CORDEIRO et al., 2019; SHU et al., 2017). Isto possibilita identificar uma *Fake News* sem conhecimento prévio do tema. Os métodos baseados em propagação necessitam de uma amostragem com um número significativo de dados, o que significa que a notícia deve já ter sido propagada para ser classificada com precisão. Destaca-se que cada uma das categorias possui vantagens e desvantagens, sendo que a integração destas permite obtenção de melhores resultados e maior cobertura de situações.

Os códigos das abordagens de verificação de *Fake News* desenvolvidas durante a construção deste trabalho estão públicos e disponíveis na plataforma *GitHub*¹. Influenciado pela formação em comunicação, o autor deste trabalho visa a

¹ Abordagens de Verificação de Fake News. 2022. Disponível em: <https://github.com/Jthnn/Abordagens-Verificacao-de-Fake-News>. Acesso em: 20 mai. 2022.

publicização das informações com o intuito de contribuir para com a comunidade em geral.

1.1 Questão de pesquisa

Através da leitura do estado da arte, observa-se que as abordagens de verificação de *Fake News* possuem limitações de aplicabilidade em contextos específicos e que não há uma abordagem capaz de lidar com diversos contextos sem comprometer seus resultados. Seu desempenho varia de acordo com o contexto ao qual é aplicado, o que pode limitar as possibilidades de verificação à determinados assuntos ou modelos de texto.

Diante deste cenário, a questão de pesquisa endereçada neste trabalho é definida como: Como a integração de abordagens distintas pode ampliar a abrangência de resultados de verificação de *Fake News* em contextos distintos?

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo geral deste trabalho é avaliar a integração de métodos distintos de verificação de *Fake News* com o intuito de ampliar a abrangência de aplicação em contextos diferentes.

1.2.2 Objetivos Específicos

Para alcançar o objetivo geral proposto, foram definidos os seguintes objetivos específicos:

- a) Realizar a leitura do estado da arte e destacar métodos aplicados em diferentes contextos;
- b) Desenvolver uma abordagem de verificação de *Fake News* com um método de classificação de informação;
- c) Desenvolver uma abordagem de verificação de *Fake News* com um método de *fact-checking* e *question answering*;
- d) Propor uma metodologia de comparação para analisar os resultados.

1.3 Metodologia

Para a realização deste trabalho foi empregado o método de pesquisa experimental conforme Wazlawick (2009, p.42). Foi realizada a pesquisa do estado da arte de trabalhos com abordagens de verificação de *Fake News* com métodos de inteligência artificial, considerando também a abrangência de diferentes aplicações como o emprego de métodos baseados em propagação e métodos baseados em conteúdo.

Foi realizada uma pesquisa bibliográfica para o desenvolvimento da fundamentação teórica que busca definir e contextualizar o termo *Fake News*, assim como, estudar os principais conceitos relacionados ao tema para uma melhor compreensão do trabalho.

Foi desenvolvida uma abordagem com método classificação de texto e outra abordagem com método de *fact-checking* e *question answering* com o intuito de verificar e comparar seus resultados para responder à questão de pesquisa proposta.

A partir dos resultados falso-negativos do método de classificação, foram elaboradas 5 perguntas a serem respondidas pelo método de *fact-checking* e *question answering* para aferir se este método poderia extrair informações dos textos classificados equivocadamente, configurando um resultado complementar.

1.4 Estrutura do Texto

O texto está estruturado da seguinte forma. No capítulo 2 são tratados os conceitos para a compreensão e realização deste trabalho. No capítulo 3 são destacados estudos que compõem o estado da arte sobre o tema. O capítulo 4 apresenta a plataforma Web Fake que será utilizada para integrar os métodos propostos. O capítulo 5 apresenta o conjunto de materiais e métodos utilizados no desenvolvimento do trabalho. O capítulo 6 apresenta os resultados da pesquisa e por fim, o capítulo 7 apresenta a conclusão e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados conceitos utilizados para o desenvolvimento do trabalho. Inicialmente é apresentada uma caracterização do termo *Fake News*, o contexto no qual este tema ganhou grande relevância, métodos manuais, métodos automatizados, extração de informação e processamento de linguagem natural.

2.1 Fake News

Segundo o dicionário de Cambridge, *Fake News* podem ser definidas como “histórias falsas que aparentam serem notícias, espalhadas na internet ou usando outros meios de comunicação, geralmente criadas para influenciar visões políticas ou como uma piada” (FAKE NEWS, 2021).

O termo vem sendo amplamente estudado pois se mostrou impactante em diversos campos da sociedade no mundo todo, com destaque às eleições presidenciais de 2016 nos Estados Unidos (DELMAZO; VALENTE, 2018). O mesmo ocorreu nas eleições presidenciais de 2018 no Brasil, no qual os candidatos tiveram que lidar com uma massificação de notícias falsas ou até mesmo enviesadas em torno de suas campanhas (DE OLIVEIRA et al., 2021).

Outro exemplo mais recente, é a proliferação de *Fake News* em torno do combate à pandemia do Covid-19 que teve seu início nos primeiros meses de 2020. Uma pesquisa apontou que 73% dos brasileiros acreditam em ao menos um conteúdo desinformativo sobre a pandemia (AVAAZ, 2020). Esta falta de informação resulta negativamente no combate ao covid-19, podendo expor ainda mais a população ao vírus.

A partir da afirmação de Shu et al. (2017, p.1),

[...]notícias falsas intencionalmente persuadem os consumidores a aceitar crenças tendenciosas ou falsas. As notícias falsas geralmente são manipuladas por propagandistas para transmitir mensagens ou influência política.

Torna-se interessante pensarmos também para além do conteúdo e propagação, quais atores possivelmente podem obter benefícios a partir das *Fake News*, considerando que há interesses por trás da propagação destas.

2.2 Contexto

As redes sociais digitais estão consolidadas como meios de comunicação de nossa sociedade, tendo suas diferentes características influenciadas pelas possibilidades da internet como a alta velocidade e volume de informação. Ao mesmo tempo em que estes espaços digitais estabelecem benefícios como a acessibilidade e democratização da informação, como efeito colateral, tornam-se propícios à disseminação de *Fake News* (DELMAZO; VALENTE, 2018; DE OLIVEIRA et al., 2021; SHU et al., 2017).

Dentre as diferenças para com os meios de comunicação mais tradicionais como o rádio e a televisão, a internet possibilita o anonimato e facilita a propagação de informações falsas, tornando na maioria das vezes, impossível rastrear os autores das *Fake News* e eventualmente os responsabilizando (SHARMA et al. 2019).

A maneira com que consumimos notícias vem mudando ao longo dos anos e nota-se como nosso comportamento torna-se inerente à natureza das plataformas de mídias sociais (SHU et al., 2017), seja pela acessibilidade e custos para a leitura de notícias, como também, pela facilidade em compartilhar, comentar e discutir as notícias com pessoas da rede.

Nas redes sociais, tornou-se popular o uso de robôs para iniciar o disparo e compartilhamento de *Fake News* com o intuito de alcançar mais pessoas e até mesmo gerar afirmações de apoio ou rejeição a determinado assunto (SHARMA et al., 2019). Dentre as principais redes para a propagação de *Fake News* destacam-se o Whatsapp, Twitter e Facebook. Além destes, é possível encontrar sites que funcionam como portais de notícias que costumam propagar *Fake News* ou conteúdos tendenciosos.

As principais redes sociais realizaram algumas iniciativas a fim de controlar a propagação de *Fake News* em suas plataformas, nota-se um maior esforço em analisar as denúncias feitas pelos usuários e posteriormente excluir as *Fake News*, caso assim sejam identificadas, porém, não foram apresentados métodos automatizados que de fato assegurem a possível exposição dos usuários à *Fake News*.

2.3 Métodos manuais

Apurar fatos é uma tarefa desempenhada por jornalistas já há um bom tempo, o que também mostra como as notícias falsas ou informações tendenciosas não são um problema exclusivamente encontrado nas redes sociais digitais. O jornalismo investigativo desempenha um papel fundamental no processo de checagem de fatos manual, assunto que vem ganhando notoriedade e espaço até mesmo em grandes portais de notícias mais tradicionais.

O G1 por exemplo, um dos maiores portais de notícias do Brasil, adicionou em 2018 uma seção nomeada “Fato ou Fake”, na qual realizam a apuração de notícias que circulam em redes sociais sobre os mais variados assuntos. A checagem dos fatos para identificar a veracidade das informações se dá de maneira manual por uma equipe composta por jornalistas. O processo consiste em apurar fontes, checar contexto, consultar especialistas e outras possíveis tarefas que possam validar a apuração (G1, 2018).

Enquanto temos sites de checagem de fatos já consolidados mundialmente como o PolitiFact, fundado em 2007 com o propósito de apurar informações políticas, é interessante pensarmos como a mídia hegemônica começa a incorporar a identificação de *Fake News* também em seus portais, podemos entender isto como uma representação de como o assunto se tornou presente em nossas vidas e também que há interesse das pessoas em consumir este tipo de informação (CORDEIRO et al., 2019).

A partir disso, entende-se como os métodos manuais de verificação de fatos são custosos em tempo, mão de obra e conseqüentemente financeiro. Além disso, métodos manuais não conseguem lidar com a escala do alto volume de dados e a velocidade com que surgem novas informações (ZHOU et al., 2020).

2.4 Métodos automatizados

Diante dos grandes volumes de dados produzidos na *web*, identificar *Fake News* com métodos automatizados torna-se essencial para conter a propagação e realizar a verificação de fatos de maneira mais ágil. Para esta tarefa, costuma-se

utilizar recursos de inteligência artificial e aprendizado de máquina com diferentes abordagens (SHARMA et al., 2019).

Em uma das abordagens existentes, estes trabalhos são agrupados como métodos baseados em conteúdo, os quais consistem em realizar a classificação de *Fake News* com base nas informações encontradas no conteúdo da notícia (ZHOU et al., 2020). Isto possibilita identificar uma *Fake News* antes mesmo de sua propagação e possivelmente conter seus impactos. Porém, métodos baseados em conteúdo precisam de algoritmos bem treinados no tema do texto a ser classificado para que obtenham uma melhor precisão (BHATT et al., 2018; CUI et al., 2019; SHU et al., 2017; ZHOU et al., 2020).

O Processamento de Linguagem Natural (PLN) é amplamente utilizado em métodos baseados em conteúdo devido ao seu potencial em extrair informações do texto e processar estas informações para que gerem sentido na classificação destas. Com isto, é possível utilizar diversos recursos linguísticos e estatísticos baseados nas palavras do texto, como o uso da sintaxe e semântica (BHATT et al., 2018).

O outro grupo é composto por métodos baseados em propagação, os quais têm como objetivo utilizar as informações de rede da notícia, como por exemplo, quem são os atores que propagam a notícia e como os usuários reagem, podendo classificar as reações dos usuários a partir das informações geradas por interações como comentários e compartilhamentos (AJAO; BHOWMIK; ZARGARI, 2018; CONROY; RUBIN; CHEN, 2015; CORDEIRO et al., 2019; SHU et al., 2017). Isto possibilita identificar uma *Fake News* sem conhecimento prévio do tema, entretanto, métodos baseados em propagação necessitam de uma amostragem com um número significativo de dados, o que significa que a notícia deve já ter sido propagada para ser classificada com precisão.

O uso de Redes Neurais Artificiais (RNA) é presente em ambos os grupos de métodos devido ao seu potencial em processar uma grande quantidade de dados e realizar relações entre estes, inclusive viabilizando a hibridização entre diferentes abordagens (AJAO; BHOWMIK; ZARGARI, 2018). Como pode ser visto na seção de trabalhos relacionados, esta abordagem híbrida tem se mostrado promissora na identificação de *Fake News*.

A extração de conhecimento é uma etapa majoritariamente presente em abordagens de identificação de *Fake News*, principalmente no pré-processamento de dados e treinamento dos algoritmos, tendo grande impacto no resultado final da

abordagem. Enquanto isso, bases de conhecimento como ontologias, que podem contribuir bastante para a extração de conhecimento, ainda se mostram pouco exploradas no campo de *Fake News*. A partir da representação de conhecimento, as ontologias servem para realizar inferências entre palavras de um determinado domínio, podendo estabelecer relações, atributos e classes.

2.5 Extração de Informação

A tarefa de extração de informação (EI) serve para extrair determinadas informações de um texto de acordo com as necessidades da abordagem e preencher estas em um novo espaço no modelo definido (ZHU e CHANG, 2008; KUMAR, 2017). EI é uma tarefa que incorpora uma série de outras subtarefas para alcançar os objetivos de identificar, estruturar e armazenar as informações extraídas, passando por processos de reconhecimento de entidades nomeadas, extração de relações, extração de eventos e entre outros (KANG, 2020; YU et al., 2021).

Apesar da vasta quantidade de informações disponíveis na internet, os dados encontrados majoritariamente encontram-se não estruturados, isto é, textos brutos que não possuem entidades definidas ou dados que possam gerar inferências diretamente a partir de um simples processamento. Desta forma, processos de EI são de extrema importância para a construção de sistemas que necessitem processar estes dados para posterior utilização, visto a capacidade de EI de identificar e extrair informação a partir de fontes de dados não estruturados, semi-estruturados e até mesmo estruturados. (FEILMAYR, 2011).

A tarefa de extrair informações de textos simples encontrados em domínios abertos na internet é classificada como Extração de Informações Abertas (Open Information Extraction - OIE). Esta tarefa converte textos simples em representações semi-estruturadas a partir da extração de fatos organizados em triplas compostas por sujeito, predicado e objeto (YU et al., 2021).

EI pode ser aplicada e gerar benefícios até mesmo para bases de conhecimento como ontologias e gráficos que já possuem dados estruturados, pelo potencial de aprimorar estas fontes de conhecimento a partir da extração de novas informações e inferências de relações entre pares.

Métodos de extração de informação não supervisionados são utilizados para identificar similaridades ou diferenças entre objetos e gerar agrupamentos ou

clusterizações sem conhecimentos pré-definidos, baseando-se principalmente na relação entre frases. Métodos de EI supervisionados realizam tarefas a partir da identificação de pares de entidades para um conjunto de relações já estabelecidas a partir de documentos. Um desafio de EI é trabalhar com conteúdos previamente não vistos, podendo prejudicar o processo de identificar novas informações relevantes, ao mesmo tempo, trabalhar com entidades já conhecidas exige um grande conjunto de dados de treinamento para o aprendizado.

Tarefas de EI são amplamente implementadas com etapas de processamento de linguagem natural, principalmente quando é necessário extrair informações de fontes de dados não estruturados, devido a capacidade das técnicas de PLN lidarem com a interpretação de frases em diferentes níveis textuais, contribuindo extensamente em sistemas de busca e *question answering*.

2.6 Processamento de Linguagem Natural

O processamento de linguagem natural (PLN) exerce a função de decodificar a linguagem natural, língua falada e escrita pelos humanos, para processos computacionais, possibilitando a interação entre humanos e máquinas. Esta é uma área em intersecção com disciplinas de linguística, inteligência artificial e ciência da computação (GUPTA, 2014).

Tendo seu início com regras escritas manualmente, o processamento de linguagem natural teve grandes avanços a partir do aprendizado de máquina, integrando sistemas mais complexos e sendo utilizado para tarefas como extração de informação, sumarização de textos, mineração de dados, tradução automática, sistemas de *chatbots* e *question answering* (WU et al., 2021).

Com o avanço da área de *deep learning*, redes neurais estão sendo amplamente implementadas em abordagens de processamento de linguagem natural para lidar com o processamento de sequências de palavras. Dentre as mais variadas, destaca-se o uso de redes neurais de longa memória de curto prazo (*Long Short Term Memory - LSTM*) pela capacidade de interpretar dados em sequências próximas e longas, contribuindo para processar o contexto das frases contidas nos textos (AJAO; BHOWMIK; ZARGARI, 2018; CONROY; RUBIN; CHEN, 2015; CORDEIRO et al., 2019; SHU et al., 2017).

Frequentemente as sentenças em PLN são representadas em formatos sequenciais de tokens, porém, há iniciativas de realizar estas representações a partir de grafos com o intuito de aprimorar a visualização de problemas de PLN como a implementação de árvores ou análises sintáticas (WU et al., 2021).

Em abordagens de verificação de *Fake News*, técnicas de PLN podem ser utilizadas na etapa de mineração de dados e extração de informações de textos. O processo inicial passa pela etapa de *tokenização*, na qual as palavras são segmentadas e convertidas em *tokens*, tornando-se inteligíveis computacionalmente. A partir disto, é possível realizar etapas de pré-processamento de dados em diferentes níveis textuais como léxico, semântico e sintaxe. Devido a complexidade da linguagem natural, o PLN enfrenta desafios como eliminar ambiguidades e compreender os significados das palavras em contextos diferentes (DE OLIVEIRA et al., 2021).

3 TRABALHOS RELACIONADOS

Neste capítulo serão expostos alguns trabalhos que contribuem para a discussão sobre a detecção de *Fake News* com propostas de métodos automatizados utilizando recursos de *deep learning*. Apresentando etapas de extração e classificação de informação, destaca-se o potencial das redes neurais artificiais, assim como o emprego de recursos linguísticos para o alcance de melhores resultados.

Para o acesso a estes estudos, foram utilizadas as seguintes bases de dados: IEEE Xplore, Elsevier, Web of Science e ACM. Nas quais foram pesquisadas as seguintes palavras-chave: *Fake News*, *fact-checking* e *Fake News detection*. Os artigos foram selecionados baseando-se pela classificação de relevância das plataformas de busca, pela abrangência dos *datasets* utilizados, pela contemporaneidade dos estudos e pelo uso de técnicas de *deep learning*, *machine learning*, extração de informação e classificação de texto.

3.1 Fake News Identification on Twitter with Hybrid CNN and RNN Models

Com o intuito de identificar *Fake News* em posts no Twitter, Ajao, Bhowmik e Zargari (2018) apresentam um modelo baseado na hibridização entre dois modelos de redes neurais, a Rede Neural Convolutiva (*Convolutional Neural Network* - CNN) e a Rede Neural Recorrente (*Recurrent Neural Networks* - RNN). O destaque deste estudo está na classificação de *Fake News* sem conhecimento prévio do assunto, introduzindo heurísticas com recursos semânticos e linguísticos.

Dentre as redes sociais mais populares, talvez o Twitter seja a plataforma com o maior potencial de velocidade de propagação de informação, ao mesmo tempo em que não há um controle sobre a produção de conteúdo, tornando assim, um ambiente democrático no âmbito de publicações. Isto possibilita com que *Fake News* possam se difundir com mais facilidade, além de dificultar a classificação de texto, devido às novidades no conteúdo.

O estudo mostrou que não é necessário conhecimento prévio do conjunto de dados para que o algoritmo tenha bons resultados. Utilizando *Long Short Term Memory* - LSTM, foi obtido 82% de precisão na identificação de *Fake News* na amostra de 5,800 tweets. Ao mesmo tempo, os autores ressaltam que modelos de *deep learning* como CNN e RNN devem obter resultados ainda melhores em situações com

um maior conjunto de dados, devido à possibilidade de um melhor treinamento do algoritmo.

3.2 What is Real or Fake? - Machine Learning Approaches for Rumor Verification using Stance Classification

Com foco em reduzir o impacto das *Fake News* em sites de redes sociais, Cordeiro et al. (2019) propõem um sistema de recomendação de URL com *fact-checking* a partir de técnicas de *deep learning*. Baseado em duas amostras de dados de eventos ocorridos em 2017 e 2019, são criadas duas abordagens diferentes a fim de comparar com os resultados obtidos pelos melhores algoritmos de *deep learning* utilizados nas ocasiões.

A primeira abordagem utiliza *deep learning* com modelo LSTM (Long Short Term Memory) para classificação de postura. Enquanto a segunda, utiliza algoritmos de Regressão Logística e SVM (Support Vector Machine).

A técnica de Regressão Logística obteve o melhor resultado de classificação de postura, superando as técnicas de *deep learning* utilizadas. Ao mesmo tempo, adicionar classificação de postura ao processo de identificação de rumores resultou em uma redução da performance em todos os cenários.

3.3 Combining Neural, Statistical and External Features for Fake News Stance Identification

Bhatt et al. (2018) apresentam um trabalho que explora a detecção de postura tendo como objetivo determinar a relevância do título com o corpo da notícia. Para isso, eles combinam recursos neurais e estatísticos com heurísticas a fim de encontrar a melhor classificação. Sendo estes, em ordem: rede neural skip-thought, vetores TF-IDF n-gram e recursos criados manualmente.

O algoritmo teve dificuldades em lidar com a distribuição desigual de dados disponibilizados, além de ser necessário um maior número de amostras a fim de aprimorar o treinamento do algoritmo. Apesar disto, os experimentos mostraram que o modelo proposto superou todas as melhores técnicas de identificação de postura até então observadas pelos autores.

3.4 dEFEND: A System for Explainable Fake News Detection

Cui et al. (2019) propõem o sistema dEFEND, que além de identificar a autenticidade das notícias, busca também uma melhor usabilidade a fim de conquistar a confiança do usuário nos resultados da plataforma de fact-checking. Ao adicionar um módulo de explicação de detecção, o sistema retorna como chegou até os resultados exibidos.

O algoritmo faz uso de Redes Neurais Recorrentes (Recurrent Neural Networks - RNN) e Unidades Recorrentes Bloqueadas (Gated Recurrent Units - GRU) para calcular os pesos das palavras nas sentenças. Com isto, é possível classificar sentenças de notícias e aproveitar as informações dispostas pelos usuários nos comentários, operando com uma sub-rede de co-atenção relacionando os dados incorporados na notícia com os dados dos comentários.

Esta estrutura baseada em *deep learning* se mostra recente e apresenta grande eficácia, nos testes realizados com conjuntos de dados reais, o sistema proposto superou 7 métodos populares para a detecção de Fake News.

3.5 Fake News Early Detection: A Theory-driven Model

Interessados em identificar *Fake News* antes de serem propagadas, Zhou et al. (2020) apresentam um estudo interdisciplinar focado em explorar a identificação de *Fake News* a partir do conteúdo da notícia, o método investiga diferentes níveis textuais: léxico, semântico, sintaxe e discurso. O conhecimento para estes recursos é extraído de textos de dois datasets reais, PolitiFact e BuzzFeed.

Nos testes realizados, os recursos léxicos e de sintaxe obtiveram os melhores resultados separadamente em relação aos outros, porém, os melhores resultados são obtidos quando todos os níveis textuais são considerados. O método desenvolvido obteve em média 88% de acurácia e superou os trabalhos relacionados, dentre estes, métodos baseados em conteúdo, em propagação e híbridos.

O estudo também expõe a forte relação entre click baits e *Fake News*, pontuando que click baits se apresentam em maior volume em *Fake News* do que em notícias verdadeiras. Além disso, os autores identificam potenciais padrões no conteúdo de notícias falsas. Em ambos os datasets, *Fake News* compartilham maior

informalidade, diversidade e subjetividade, utilizam mais palavras emocionais, frases mais longas e palavras mais curtas em comparação às notícias verdadeiras.

3.6 SocialTruth Project Approach to Online Disinformation (Fake News) Detection and Mitigation

Com o intuito de descentralizar os serviços de verificação de informação, Choras et al. (2019) manifestam ideias de como desenvolver um ecossistema mais plural e distribuído, destacando desafios como a extração de informação a partir de diferentes fontes de dados e posteriormente sua clusterização.

Os autores também citam a possibilidade do uso de blockchain para armazenar e compartilhar dados relativos à reputação e credibilidade dos usuários baseados nos conteúdos compartilhados por estes. Além de ser uma grande contribuição para a verificação e identificação de notícias falsas, também possibilitaria um ambiente mais transparente em relação aos pontos de reputação de cada usuário.

3.7 Análise crítica e comparação entre os trabalhos

Os estudos focados em identificar *Fake News* exploram diferentes recursos de inteligência artificial, passando por algoritmos com uso de *deep learning* e *machine learning*, destacando-se processos de extração de informação e classificação de texto. A partir da leitura do estado da arte, percebe-se que a identificação da veracidade de informações está sujeita ao contexto ao qual está inserido, dispondo de características heterogêneas a ponto de necessitar de diferentes algoritmos para diferentes conteúdos e plataformas. Por este motivo, torna-se importante o conhecimento das características de cada ambiente e quais abordagens podem apresentar melhores resultados.

No trabalho proposto por Cordeiro et al. (2019), a abordagem de regressão logística superou os resultados da abordagem baseada uma rede neural LSTM. Apesar do algoritmo de *deep learning* ser promissor para a tarefa de classificação, estima-se que seria necessária uma amostra de dados maior para que a abordagem com *deep learning* pudesse obter resultados melhores.

Ao mesmo tempo, temos trabalhos como o de Ajao, Bhowmik e Zargari (2018) em que a implementação da rede neural LSTM obteve bons resultados mesmo com uma amostra de dados limitada, conforme expressado pelos autores.

Percebe-se como a construção de algoritmos híbridos que combinam recursos linguísticos e estatísticos, assim como redes neurais, tendem a obter bons resultados (AJAO; BHOWMIK; ZARGARI, 2018; BHATT et al., 2018; CHORAS et al., 2019; ZHOU et al., 2020). Isto se dá pela possibilidade de usufruir dos diferentes benefícios encontrados em cada técnica, como a capacidade de lidar com alto volume de dados e pesos das palavras providos por recursos estatísticos, tal qual, da melhor adequação semântica das palavras em seus contextos providos por recursos linguísticos.

Segundo Zhou et al. (2020), os métodos de verificação de *Fake News* podem ser pensados em dois grupos contendo elementos complementares: métodos baseados em conteúdo e métodos baseados em propagação. Quanto aos métodos baseados em propagação, existem outros termos tratados como *feedback-based* (SHARMA et al., 2019) e *network-based* (CONROY; RUBIN; CHEN, 2015), porém, possuem significados semelhantes e trataremos neste trabalho como métodos baseados em propagação.

Tabela 1 – Trabalhos relacionados

Trabalhos Relacionados	Recursos	Acurácia (%)	Métodos baseados em	
			Conteúdo	Propagação
Ajao, Bhowmik e Zargari (2018)	CNN+RNN (LSTM)	82%		X
Cordeiro et al. (2019)	RNN (LSTM) / LR / SVM	67%		X
Bhatt et al. (2018)	MLP (<i>Skip Thought</i> , TF e heurísticas)	83%	X	
Cui et al. (2019)	3x RNN (GRU)	85%		X
Zhou et al. (2020)	BOW, POS, CFG, CNN	88%	X	

Fonte: Jonathan Camargo

Na tabela 1, é possível visualizar de maneira resumida os trabalhos relacionados e os recursos utilizados em cada abordagem, além da acurácia e a categoria de método baseado em conteúdo ou propagação.

A partir da leitura do estado da arte, considera-se que não há abordagem que possa ser aplicada em todas as situações possíveis sem comprometer a eficácia dos algoritmos devido às delimitações de cada abordagem. É possível avaliar as complementaridades de cada abordagem, reforçando a importância de um ambiente que integre diferentes abordagens e seu potencial para obter bons resultados.

4 PLATAFORMA WEB FAKE

Ao desenvolver uma proposta que possa integrar mais de um método de verificação de *Fake News*, é necessário também planejar um ambiente capaz de comportar estes métodos.

Para isto, este trabalho conta com a contribuição de uma plataforma já existente denominada *Web Fake*, desenvolvida por CALDERINI (2021). A plataforma foi projetada de maneira modular, o que torna possível acoplar módulos futuros que complementem a plataforma. A plataforma já possui um módulo implementado com uma abordagem de classificação de *Fake News* utilizando vetores de palavras, desenvolvida pelo autor deste trabalho.

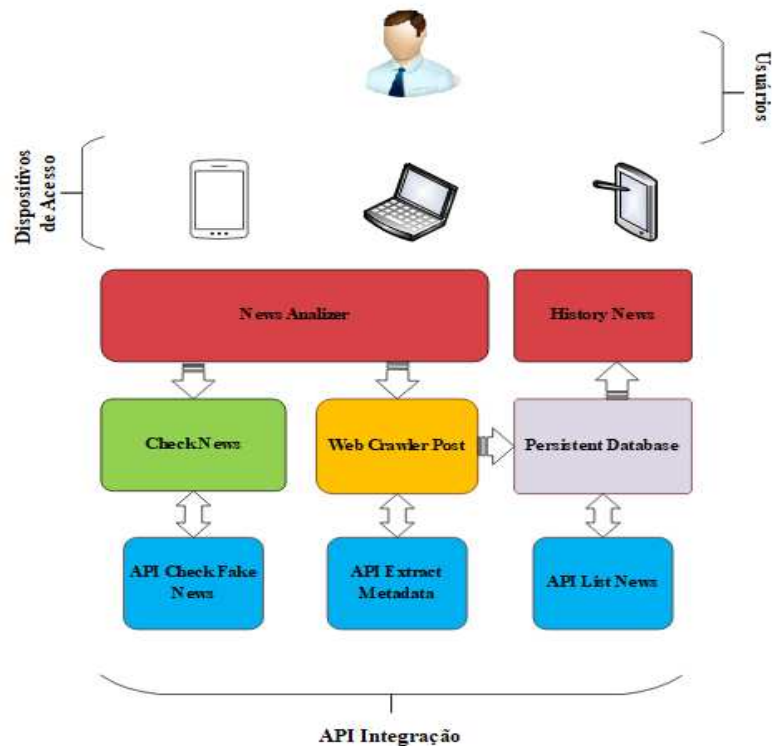
Este capítulo visa descrever a plataforma existente com detalhes sobre o modelo, módulos e serviços, além da usabilidade e funcionamento.

4.1 Modelo

O modelo proposto na plataforma desenvolvida por CALDERINI (2021) é dividido em camadas de Serviços, *APIs* de integração, Apresentação, Dados, Análise e *Scraper* de notícias. O modelo utiliza os conceitos de *API REST*, uma abordagem da arquitetura de *software* cujo principal papel é implementar as funcionalidades das aplicações em forma de serviços, que são conectados via interfaces *HTTP*. Toda a comunicação da *API* é feita utilizando o tráfego de entrada e saída no formato de dados *JSON*.

Há um protótipo baseado em serviços construído para gerar uma base de dados capaz de integrar diversos módulos de análise de *Fake News*. A plataforma foi projetada em uma arquitetura modular, conforme a Figura 1, permitindo acoplar novas funcionalidades quando necessário. Existem duas interfaces, sendo uma interface *web* para usuários interessados em validar artigos e uma *API REST* para o acoplamento de novos módulos.

Figura 1 - Modelo geral da aplicação



Fonte: CALDERINI (2021, p.3).

O sistema foi criado para ser flexível, possibilitando assim uma fácil manutenção e alteração de componentes, caso necessário, para atualizações futuras.

4.2 Módulos e serviços

O módulo *News Analiser* implementa a interface *web* do sistema, onde o sistema realiza o processamento da notícia e retorna o status de verdadeira ou fake. No módulo *History News* ficam registrados os últimos artigos analisados pelo sistema.

O módulo *Web Crawler Post* implementa o *Scraper* responsável por extrair os metadados de uma notícia, utilizando *Xpath* para extração das seguintes informações: autor, texto, data de publicação, título, descrição e imagem principal. Após este processo os dados são registrados em um banco de dados. O serviço Check News recebe os metadados do artigo e encaminha para os módulos de verificação de *Fake News*.

O serviço *Persistent Database* é responsável pelo armazenamento dos metadados extraídos das notícias, bem como as informações de análise dos diversos módulos de análise de *Fake News*. A camada *APIs* é responsável pela comunicação de serviços externos à plataforma desenvolvida.

A plataforma *Web Fake* possui alguns serviços que visam facilitar a análise de *Fake News* e também acessar a base de notícias já pesquisadas. Um destes serviços é voltado para a extração automática de metadados de notícias publicadas na web. Outro é responsável por retornar um conjunto de notícias já analisadas, podendo carregar notícias avaliadas como falsas ou não. Por fim, o serviço de análise permite que diversos modelos sejam conectados através da Interface *API REST*, que recebem os metadados da notícia em análise e executam o componente desejado.

4.3 Usabilidade e funcionamento

Para realizar uma verificação de *Fake News*, o usuário deve acessar a plataforma e indicar o endereço da notícia para a qual deseja verificar a validade que retornará avaliação como “Verdadeiro” ou “Fake”.

A figura 2 ilustra a página com o resultado de uma notícia classificada como *fake*. O sistema disponibiliza ao usuário também uma lista com as notícias já verificadas. Esta interface de uso simples e amigável para o usuário permite que este não se envolva com os detalhes dos métodos que serão utilizados para a verificação da notícia consultada. Ao mesmo tempo, o ambiente permite que os detalhes dos métodos e seus parâmetros sejam conhecidos e inspecionados pelo usuário.

Figura 2 - Resultado após uma busca

A Figura 3 apresenta um esquema de como os metadados são coletados. Estas informações ficam disponíveis em uma base de dados *Mysql*, para que não seja necessário fazer uma nova extração caso seja consultada uma notícia que já esteja na base de dados, o que possibilita que novo módulos que demandem um tempo maior de processamento possam apresentar seus resultados de forma assíncrona.

Desta forma, a plataforma trata destes dados obtidos, contendo, como apresenta a figura 3, os metadados das notícias processadas, com campos como: Título; descrição; autor; data de publicação; imagem de destaque; Texto da notícia.

5 MATERIAIS E MÉTODOS

Neste capítulo será apresentada a metodologia adotada, expondo os materiais utilizados e os métodos desenvolvidos. No item 5.1 é descrita a visão geral deste trabalho. Nos itens 5.2 e 5.3 são descritas as abordagens exploradas nos experimentos. O item 5.4 descreve os conjuntos de dados utilizados e o item 5.5 apresenta a metodologia desenvolvida para analisar os resultados.

5.1 Visão geral

O objetivo deste trabalho é avaliar a integração de métodos distintos de verificação de Fake News com o intuito de ampliar a abrangência de aplicação em contextos diferentes.

Para alcançar este objetivo, está previsto o uso de uma plataforma já implementada com arquitetura modular, capaz de receber e acoplar novos métodos de verificação, além de contar com serviços de extração de metadados de textos de notícias e base de dados que se atualiza conforme são realizadas novas pesquisas na plataforma. Esta plataforma foi desenvolvida por CALDERINI (2021) e está descrita em detalhes no capítulo 4.

A plataforma possibilita integrar novos métodos para trabalhos futuros de verificação de *Fake News* e está alinhada ao intuito de validar a hipótese proposta neste trabalho, avaliando a complementaridade entre diferentes abordagens. O desenvolvimento desta proposta tem potencial de contribuição para outras áreas, tais como a comunicação, além de servir de apoio à portais de notícias e outros interessados no tema.

Como parte inicial do trabalho, foi desenvolvido um experimento preliminar com classificação de texto utilizando o conjunto de dados *Sirene-News* com notícias reais extraídas de dois portais de notícias e utilizando o algoritmo *Random Forest* como classificador, obtendo 97,81% de acurácia. Este experimento foi validado com sua implementação e funcionamento na plataforma, integrando assim, o primeiro módulo disponível.

A partir deste experimento, foi implementado o mesmo método de classificação com um conjunto de dados diferente, o *Fake.Br* é um corpus com dados mais heterogêneos e com metodologia de anotação mais refinada, sendo extraídas

informações de 7 sites diferentes. Este segundo experimento obteve 95,33% de acurácia na classificação das notícias.

Visando a proposta de métodos complementares, foi realizado um experimento com método de *fact-checking* com *question answering* para analisar e comparar aos resultados do experimento de classificação.

A proposta considerou aferir a possibilidade de, a partir destes experimentos, demonstrar como a integração de métodos diferentes pode proporcionar resultados complementares e aumentar a abrangência da verificação de *Fake News* com aplicabilidade em diferentes contextos, visto que métodos específicos tendem a ter suas aplicações com escopo limitado, prejudicando o desempenho e resultados do algoritmo.

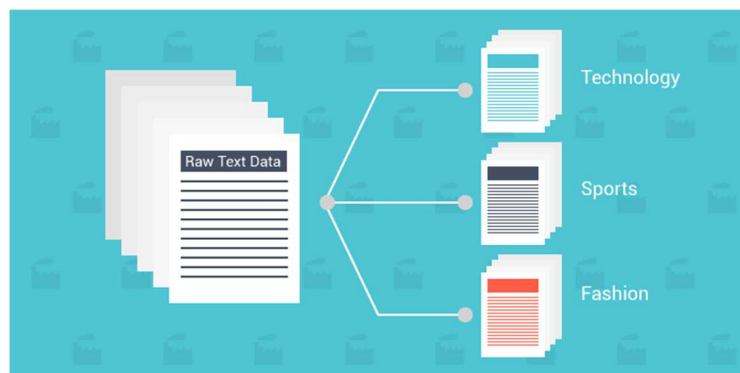
Para a análise dos resultados, foi definida uma metodologia de comparação, baseada na análise qualitativa de resultados integrados. A metodologia atua a partir da leitura das notícias dos resultados falso-negativos apurados pelo experimento de classificação de texto. Com base nestes resultados foram geradas perguntas a serem respondidas pelo experimento de *fact-checking* e *question answering*. Ao final do experimento os resultados são analisados qualitativamente para identificar o potencial de apoio na identificação correta das notícias.

5.2 Classificação de textos

Métodos de classificação de texto são capazes de organizar, estruturar e categorizar textos em diferentes formatos. A partir de recursos de processamento de linguagem natural, é possível extrair informações de textos com dados não estruturados e organizar estas informações de acordo com a aplicação. Esta característica é de extrema relevância pois a maior parte dos dados disponíveis em plataformas de mídias sociais digitais são dados não estruturados (SHARMA et al., 2019).

Na figura 4 é possível visualizar um esquema básico de classificação de texto no qual é possível fazer a leitura de dados brutos e classificar os textos por categorias, baseando-se em regras implementadas.

Figura 4 – Esquema de classificação de texto



Fonte: AGRAWAL, 2019.

Para um bom desempenho de classificação de texto, é necessário que o treinamento seja feito de acordo com o assunto a ser classificado, além de ser importante o tamanho do conjunto de dados. Entre os desafios encontrados, está a localização de textos similares que portam ideias distintas, além da dificuldade em lidar com problemas semânticos como ambiguidades (BHATT et al., 2018; CUI et al., 2019; ZHOU et al., 2020).

Métodos de *deep learning* como Redes Neurais Artificiais são capazes de aprimorar bastante os processos de classificação pela sua capacidade em lidar com memórias de longa e curta duração, possibilitando maior desempenho na leitura de conjuntos de palavras e sentenças, além da atribuição de *embeddings* para aprimorar o treinamento. Existem trabalhos sendo realizados com propostas híbridas entre recursos linguísticos e estatísticos e estão conquistando resultados interessantes (AJAO; BHOWMIK; ZARGARI, 2018).

5.3 Extração de informação e *Fact-checking*

Extração de informação pode ser uma tarefa difícil quando considerada a diversidade de expressões textuais devido à dependência da compreensão textual. Outro grande desafio é que por trabalhar com representações simbólicas, podem ocorrer problemas semânticos em lidar com polissemias e ambiguidades. O uso de *deep learning* com redes neurais profundas tem apresentado grandes avanços para estes desafios devido à possibilidade de usar representação distribuída ao invés de símbolos, utilizando operações através de vetores distribuídos (LIU, 2020).

Fact-checking, como a tradução sugere, é definido por verificação de fatos. Para além da verificação de *Fake News*, verificar um fato possui uma série de usos como refinar bases de conhecimento incompletas, prover evidências em detecção de erro em bases de conhecimento, melhorar a qualidade de pesquisas e integrar diferentes bases de conhecimento (LIN; SONG; WU, 2018).

Bases de conhecimento com dados reais geralmente são definidas por gráficos de conhecimentos e ontologias. As bases de conhecimento possuem dados estruturados e conectados, na qual os fatos são guardados em estrutura de gráficos. Os nós representam as entidades e as conexões entre qualquer nó correspondem às relações semânticas entre estes (LIU, 2020; LIN et al., 2018).

O experimento proposto neste trabalho busca explorar o método de *fact-checking* com modelo *question answering* fazendo uso do *Haystack*¹. O *Haystack* consiste em um *framework* de código aberto para a criação de sistemas de buscas em documentos que conta uma série de serviços e possui compatibilidade com modelos de linguagem como *transformers*, possibilitando a importação direto da plataforma de código aberto *hugging face*², que busca democratizar o acesso à inteligência artificial e conta com uma série de recursos de compartilhamento de códigos disponibilizados pela própria comunidade. Foi utilizado o *Elasticsearch*³ para a indexação e busca dos documentos, enquanto para ler e interpretar as informações dos textos foi importado um modelo de *deep learning* em *BERT-large*⁴, treinado no *dataset Squad v1.1* no idioma português.

É importante destacar o interesse em manter o escopo dos experimentos em trabalhar com o idioma português do Brasil, contribuindo cientificamente para o cenário nacional.

5.4 Conjunto de dados e recursos

No experimento preliminar, foi utilizado o conjunto de dados SIRENE-news⁵, composto por 2.371 notícias classificadas como verdadeiras, extraídas do site G1

¹ Haystack. 2020. Disponível em: <https://haystack.deepset.ai/overview/intro>. Acesso em: 29 abr. 2022.

² Hugging Face. Disponível em: <https://huggingface.co/>. Acesso em: 29 abr. 2022.

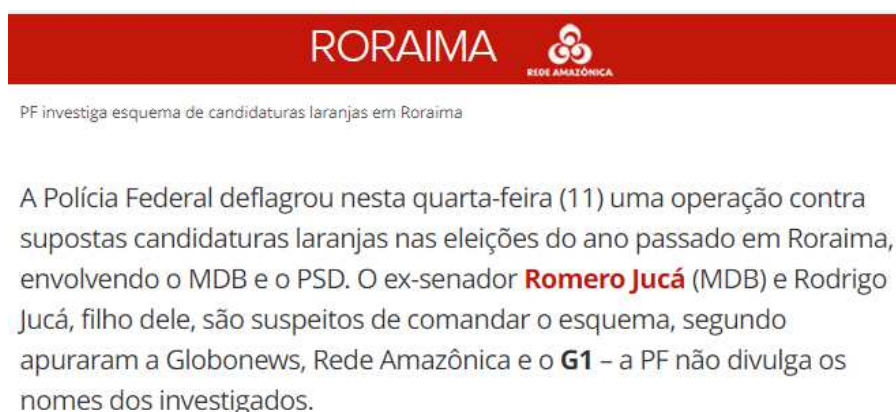
³ Elasticsearch. Disponível em: <https://www.elastic.co/pt/what-is/elasticsearch>. Acesso em: 30 abr. 2022.

⁴ BERT-large. Disponível em: <https://huggingface.co/bert-large-uncased>. Acesso em em: 02 mai. 2022.

⁵ SIRENE-News. 2019. Disponível em: <https://github.com/ViniciusNunes0/SIRENE-news>. Acesso em: 20 jun. 2021.

(<https://g1.globo.com/>) e 2.371 notícias classificadas como falsas, extraídas do site Sensacionalista (<http://sensacionalista.com.br/>). Na figura 5, é possível visualizar um exemplo de texto classificado como verdadeiro encontrado neste corpus.

Figura 5 – Texto do corpus SIRENE-News



Fonte: G1 (2019)

Nos experimentos posteriores de classificação de texto e *fact-checking*, foi utilizado o conjunto de dados do corpus *Fake.Br* (MONTEIRO et al., 2018), composto por 3.600 notícias classificadas como verdadeiras e 3.600 falsas, totalizando 7.200 notícias extraídas de 7 sites diferentes. Após o processo de mineração de dados e a seleção das notícias, estas foram manualmente rotuladas em 6 categorias, tendo as amostras distribuídas nas seguintes categorias: 4.180 em política, 1.544 em TV e celebridades, 1.276 em sociedade e notícias diárias, 112 em ciência e tecnologia, 44 em economia e 44 em religião. Na figura 6 é possível visualizar um exemplo de texto classificado como falso encontrado neste corpus. É válido destacar a dificuldade em encontrar os textos falsos na internet, por já terem sido excluídos dos sites citados no corpus.

Figura 6 – Texto do corpus Fake.br

Bolsonaro vai militarizar todas as escolas do Brasil em 4 anos. Bolsonaro vai militarizar todas as escolas do Brasil em 4 anos. Para o desespero dos professores comunistas que terão que procurar urgentemente um novo emprego, ou ir vender miçangas em praça pública.

Não dobrar short ou camiseta da educação física para diminuir seu tamanho nem usar óculos com lentes ou armações de cores esdrúxulas.

Fonte: Estadão (2016)

5.5 Metodologia e métricas de avaliação

Considerando os experimentos que dispõem de conjuntos de dados já rotulados, é interessante identificar o número correto de Verdadeiro-Positivos (VP), Falso-Positivos (FP), Verdadeiro-Negativos (VN) e Falso-Negativos (FN). Foi gerada uma matriz de confusão para a visualização destes resultados encontrados no experimento de classificação de texto.

Também foram implementadas as métricas de acurácia e *F1 Score*. A acurácia é definida pela razão do total de amostras classificadas corretamente pelo número total de amostras e o *F1 Score* é composto pela relação harmônica entre precisão e recall, quanto maior o valor de *F1 Score*, melhor é a classificação do algoritmo (LIU, 2020; DE OLIVEIRA et al., 2021). Outras métricas tradicionais como precisão e *recall* não foram consideradas relevantes para a proposta experimental do trabalho, visto que para avaliar a complementaridade entre os métodos desenvolvidos, não é possível utilizar as mesmas métricas.

A partir dos resultados FN do experimento de classificação, foram elaboradas 5 perguntas para serem respondidas no experimento de *fact-checking* com modelo *question answering*. As respostas foram lidas e interpretadas manualmente a partir das informações encontradas nos textos utilizados para treinamento.

6 EXPERIMENTOS E RESULTADOS

Neste capítulo serão expostos os resultados obtidos através da pesquisa deste trabalho. No item 6.1 é descrito em detalhes o experimento realizado com método de classificação de texto. No item 6.2 é descrito em detalhes o experimento com método de *fact-checking* com *question answering*. No item 6.3 é descrita a plataforma de integração e seu potencial para outros trabalhos. No item 6.4 é apresentada em detalhes a metodologia de comparação e a análise dos resultados.

6.1 Experimento de classificação de texto

Foi realizado um experimento com classificação de texto a fim de promover uma aproximação com o tema de maneira prática. O modelo proposto foi desenvolvido em linguagem *python* e foi utilizado o recurso *TF-IDF* para calcular a frequência dos termos e realizar o treinamento e classificação com o algoritmo *random forest*. Em um experimento preliminar, foi utilizada a base de dados SIRENE-News para incorporar o treinamento e teste de classificação, atingindo 97,81% de acurácia. O experimento final utiliza a base de dados *Fake.Br*, pois esta possui um corpus mais desenvolvido e com processos de seleção de notícias mais sofisticados. No experimento final, a acurácia obtida foi de 95,33%. Nos próximos itens são apresentadas as etapas com mais detalhes.

6.1.1 Conjunto de dados

Inicialmente, para a realização de testes preliminares, foi utilizada a base de dados “SIRENE-news”, composta por 2.371 notícias classificadas como verdadeiras, extraídas do site G1 (<https://g1.globo.com/>) e 2.371 notícias classificadas como falsas, extraídas do site Sensacionalista (<http://sensacionalista.com.br/>). As palavras do texto passaram pelo processo de tokenização e remoção de pontuações. Este processo é importante para o pré processamento de dados com *Term Frequency* (TF), pois a partir desta etapa as palavras serão tratadas como tokens, que são unidades individuais.

Figura 7 - Notícia verdadeira no site G1



Fonte: G1 (2019)

A biblioteca de software Pandas foi utilizada para a leitura do conjunto de dados presentes em um arquivo com extensão CSV e a criação de um dataframe. Após carregar o arquivo, as notícias são adicionadas ao dataframe onde os dados binários de cada notícia são etiquetados como 0 (verdadeiro) e 1 (falso).

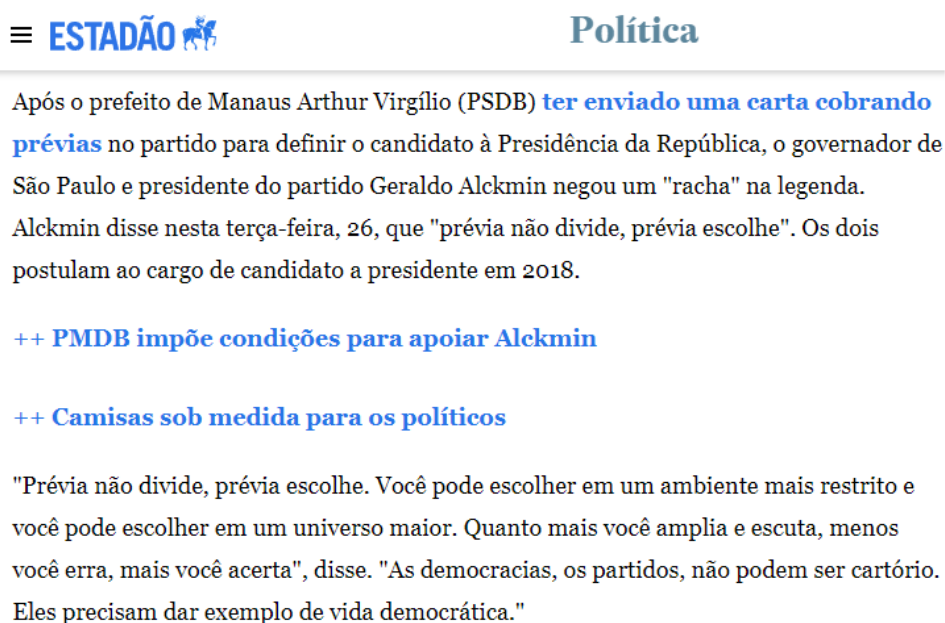
O *dataframe* do Pandas é composto por uma estrutura de dados tabulares bidimensionais. Para o desenvolvimento desta aplicação, o *dataframe* gerado conta com a disposição de três colunas (ID, notícia e classificação) e 4.738 linhas, estas são compostas pelo número total de notícias inseridas no conjunto de dados selecionado.


A base de dados utilizada no experimento final foi o corpus *Fake.Br* (MONTEIRO et al., 2018), que é composta por 7.200 notícias distribuídas igualmente entre verdadeiras e falsas, somando 3.600 notícias classificadas como verdadeiras e 3.600 classificadas como falsas.

O processo de seleção das notícias falsas foi realizado manualmente. Foram feitas buscas dentro do período de janeiro de 2016 à janeiro de 2018 em quatro sites diferentes: Diário do Brasil, A Folha do Brasil, The Jornal Brasil e Top Five TV, sendo escolhidas apenas notícias consideradas totalmente falsas.

As notícias verdadeiras foram selecionadas a partir de um processo semiautomático, o qual realizou uma busca dentro do mesmo período citado nos sites: G1, Folha de São Paulo e Estadão. A busca foi baseada nas palavras-chave das *Fake News*, coletando 40.000 notícias verdadeiras no primeiro momento. Para filtrar e selecionar as notícias verdadeiras com conteúdo similar às falsas, foram utilizadas medidas de similaridade léxica. Após isto, foi realizado um processo manual para garantir que as notícias tenham de fato assuntos relacionados.

Figura 8 - Notícia verdadeira no site Estadão



≡ **ESTADÃO**  **Política**

Após o prefeito de Manaus Arthur Virgílio (PSDB) **ter enviado uma carta cobrando prévias** no partido para definir o candidato à Presidência da República, o governador de São Paulo e presidente do partido Geraldo Alckmin negou um "racha" na legenda. Alckmin disse nesta terça-feira, 26, que "prévia não divide, prévia escolhe". Os dois postulam ao cargo de candidato a presidente em 2018.

++ PMDB impõe condições para apoiar Alckmin

++ Camisas sob medida para os políticos

"Prévia não divide, prévia escolhe. Você pode escolher em um ambiente mais restrito e você pode escolher em um universo maior. Quanto mais você amplia e escuta, menos você erra, mais você acerta", disse. "As democracias, os partidos, não podem ser cartório. Eles precisam dar exemplo de vida democrática."

Fonte: Estadão (2017)

As notícias que compõem o corpus foram manualmente rotuladas em 6 categorias, tendo as amostras distribuídas em: 4.180 notícias de política, 1.544 notícias de TV e celebridades, 1.276 notícias de sociedade e notícias diárias, 112 notícias de ciência e tecnologia, 44 notícias de economia e 44 notícias de religião.

6.1.2 Pré-processamento

Na etapa de pré-processamento de dados, foi utilizado Processamento de Linguagem Natural e importados recursos da biblioteca NLTK (*Natural Language Toolkit*) para aprimorar o treinamento do conjunto de dados. Foi realizada a importação de um conjunto de palavras da biblioteca de stop-words encontrados na NLTK para filtrar o texto removendo palavras que possam aparecer com regularidade,

ao mesmo tempo que possuem menor peso devido a sua irrelevância ao contexto das orações.

Foi utilizada a biblioteca *Scikit-learn* para transformar o texto em vetores de frequência de palavras com o recurso *TfidfVectorizer*, sendo que esta abordagem estatística é usada para encontrar a frequência com que termos aparecem nas sentenças. O recurso de Frequência Inversa de Documentos (*Term Frequency - Inverse Document Frequency*) foi utilizado para reduzir o peso das palavras na lista de *Stopwords*. Isto faz com que palavras sem relevância tenham um peso menor para o algoritmo. Também foram definidos valores para *n-gram* considerando $n=2$, para que o comando seja aplicado separadamente em cada token. *N-gram* é utilizado em probabilidade computacional para determinar sequências de tokens adjacentes em uma amostra, no qual N determina o valor da sequência.

6.1.3 Classificação

O tamanho do conjunto de dados escolhido para treinamento foi de 75% do total dos dados dispostos, enquanto os outros 25% foram utilizados para a classificação. Para realizar a classificação, foi utilizado o algoritmo *Random Forest*, que se adequa bem ao cenário do experimento considerando o tamanho limitado e a baixa complexidade do conjunto de dados. Este método de aprendizagem supervisionada opera gerando uma série de árvores de decisão com o intuito de classificar resultados, apresentando bom desempenho mesmo com *datasets* menores.

6.1.4 Resultados

A partir do experimento realizado, foi possível explorar alguns recursos de processamento de linguagem natural e identificar como alguns parâmetros podem impactar a aplicação. O modelo proposto obteve a acurácia de 95,33% e *F1 score* de 95,48% utilizando o classificador *Random Forest*.

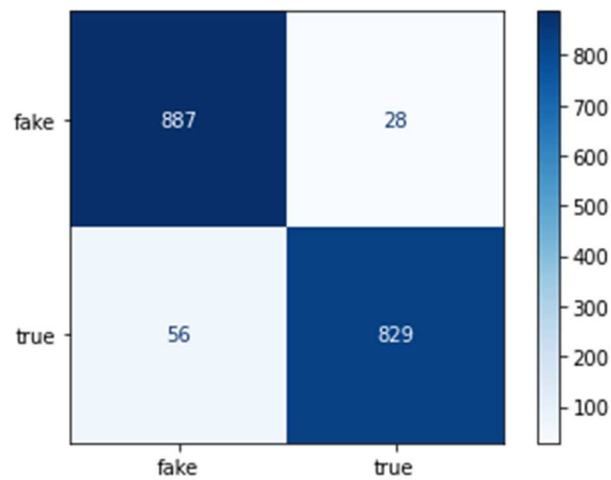
As métricas foram implementadas a partir da biblioteca *scikit-learn*⁶, uma biblioteca de código aberto em Python que possui diversas ferramentas de *machine learning*. A acurácia representa a razão entre o número de acertos divididos pelo

⁶ Scikit-learn. Disponível em: <https://scikit-learn.org/stable/index.html>. Acesso em: 20 fev. 2022.

número total da amostra, ou seja, $Acurácia = (VP+VN / Total)$. O *F1 score* representa a harmonia entre precisão e *recall*, sua equação é $F1\ Score = 2 * (Precisão * Recall) / (Precisão + Recall)$.

A figura 9 apresenta a matriz de confusão dos resultados, foram classificados 887 verdadeiro-positivos, 829 verdadeiro-negativos, 28 falso-negativos e 56 falso-positivos.

Figura 9 – Matriz de confusão



Fonte: Jonathan Camargo

Além das métricas tradicionais amplamente utilizadas em métodos de predição, foi criado um *dataframe* com as classificações originais e as predições e exportado em formato “*csv*”. Esta etapa é importante pois possibilita a leitura dos 28 textos classificados como falso-negativos que posteriormente serão utilizados para a formulação de perguntas para o experimento de *fact-checking* e *question answering*.

Figura 10 – Resultados parciais do experimento de classificação de textos

Texto	Original	Predição
8	fake	fake
14	fake	fake
15	fake	fake
17	fake	fake
19	fake	true
23	fake	true
26	fake	true
31	fake	fake
33	fake	fake
37	fake	true
41	fake	true

Fonte: Jonathan Camargo

A figura 10 apresenta um trecho de como os resultados são exibidos em uma tabela com 3 colunas e 1.800 linhas. A coluna “texto” refere-se ao número do documento de cada notícia, a coluna “original” refere-se ao valor original da classificação da notícia e a coluna “predição” refere-se ao resultado da predição do algoritmo de classificação. As 1.800 linhas são compostas pelo número do documento de cada notícia, estas são as 1.800 notícias selecionadas aleatoriamente para o teste com o *Random Forest*, representando 25% do *dataset*.

Além disso, este experimento possibilitou a implementação de um módulo na plataforma *Web Fake*, sendo então a primeira abordagem de verificação de fatos implementada para testes na plataforma.

6.2 Experimento de *fact-checking* com *Question Answering*

Para complementar o experimento de classificação de texto e validar a metodologia proposta, foi realizado um experimento de *fact-checking* com um modelo de *question answering* baseado no modelo *BERT Large*, que desde o seu lançamento em 2018 pelo *Google*, se tornou referência em tarefas de processamento de linguagem natural. O modelo de *question answering* é capaz de imprimir respostas para perguntas a partir da extração de informações de textos, identificando o contexto e selecionando a sequência de palavras que constituem uma resposta.

A partir dos resultados falso-negativos (FN) encontrados no experimento de classificação, foi realizada a leitura dos textos, analisados o seu conteúdo, e geradas

questões a serem respondidas pelo experimento de *fact-checking* com o intuito de apresentar como um método pode complementar os resultados de outro método diferente.

As perguntas foram elaboradas com a intenção de atingir as informações centrais de cada texto, a partir da leitura e interpretação de cada notícia. Também foram pensadas perguntas que pudessem imprimir resultados interessantes para o experimento, contribuindo para a análise exposta no capítulo 6.4.

O algoritmo foi desenvolvido no *Haystack*, um *framework* de código aberto que possibilita a criação de sistemas de buscas em documentos, o que agiliza a criação de sistemas com processamento de linguagem natural por possibilitar um ambiente de desenvolvimento com técnicas já pré-carregadas, além da importação de modelos e bases de dados.

Como base de conhecimento, foram utilizadas as 3.600 notícias verdadeiras do corpus Fake.Br, as quais são indexadas à busca do Elasticsearch e posteriormente processadas por um modelo *BERT*⁷.

A partir da leitura das sequências de palavras impressas nas respostas, é possível validar que as perguntas testadas foram respondidas adequadamente pelo algoritmo e os resultados são considerados satisfatórios, visto que o propósito principal é contrastar com os resultados falso negativos impressos pelo método de classificação.

6.3.1 Conjunto de dados

A base de dados utilizada neste experimento foi previamente descrita no item 6.1.1, o corpus Fake.Br, composto por 7.200 notícias distribuídas proporcionalmente entre verdadeiras e falsas.

Devido à metodologia proposta, neste experimento foram utilizadas apenas as notícias verdadeiras para comporem o conjunto de dados, sendo 3.600 textos completos, sem normalização. O conjunto de dados foi hospedado na plataforma *GitHub* em formato *zip* com os 3.600 arquivos de texto, estes são carregados no algoritmo e passam por um processo de limpeza de exclusão de espaços em branco

⁷ Disponível em: <https://huggingface.co/pierreguillou/bert-large-cased-squad-v1.1-portuguese>. Acesso em: 20 fev. 2022.

e divisão de parágrafos, para então serem indexados em formato de documento no armazenamento do *Elasticsearch*.

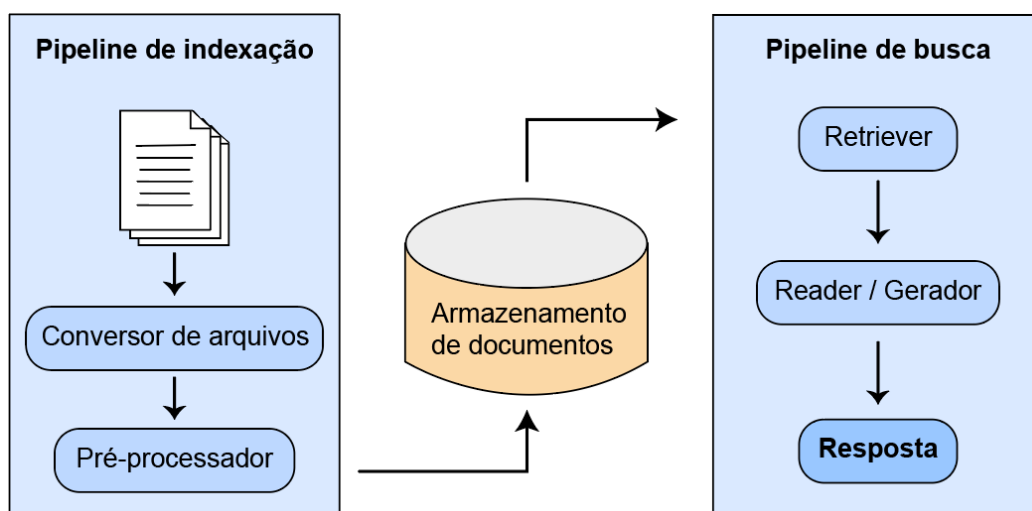
É importante destacar que houve uma busca extensa por bases de conhecimento e ontologias em português, porém, não foram encontradas bases compatíveis com a proposta do experimento. Ao mesmo tempo, utilizar uma base de dados com corpus real de notícias para validar questões sobre os mesmos assuntos permite com que sejam alcançados bons resultados.

6.3.2 *Haystack*

O *Haystack* é um *framework* de código aberto utilizado para a construção de sistemas de busca baseados em documentos. O *framework* é compatível com uma série de algoritmos de processamento de linguagem natural e agiliza o desenvolvimento de métodos de PLN.

O *Haystack* possibilita o uso de *transformers*, modelos de *deep learning* que lidam com a entrada de dados sequenciais, porém, não necessariamente processam os dados em ordem devido aos mecanismos de auto atenção. Além disso, no *Haystack* é possível carregar diferentes bases de dados, criar escalabilidade para lidar com milhões de documentos e implementar *API REST*, além de contar com uma série de ferramentas de anotação, coleta de *feedback* dos usuários, avaliação dos componentes e aprimoramento de modelos. Estas características possibilitam que o método desenvolvido seja aprimorado no futuro sem grandes dificuldades de implementação.

A figura 11 representa um esquema básico de um sistema de busca em documentos. No *pipeline* de indexação, são carregados arquivos de texto em diferentes formatos, estes passam por um conversor de arquivos e um pré-processador que realiza funções de limpeza de texto, como excluir espaços em branco e dividir os textos em fragmentos menores, para então serem armazenados em forma de documentos. No *pipeline* de busca, o *retriever* faz um filtro e seleciona os documentos mais relevantes a serem enviados ao *reader*, este é composto por um modelo de *deep learning* mais robusto capaz de extrair informações dos textos.

Figura 11 - Visão geral do *Haystack*

Fonte: *HAYSTACK* (2020).

Após o *reader* realizar a extração de informações dos textos selecionados, são identificadas as sequências de palavras que mais se adequam as respostas, portando também um *score* de confiança que representa a expectativa da resposta estar correta.

6.3.3 Descrição do método

O experimento foi desenvolvido em linguagem *Python* no *Google Colab*, uma plataforma gratuita com serviço em nuvem que possibilita o desenvolvimento de códigos em *Python 3* sem a necessidade de configuração, além de disponibilizar o acesso à *GPUs* e facilitar o compartilhamento do código.

O *framework Haystack* foi utilizado para a construção do sistema de *question answering* com extração de informações em textos. O experimento segue a estrutura básica de um sistema de buscas em documentos conforme pode ser visto na figura 11.

Após realizado o carregamento dos serviços necessários para o ambiente de desenvolvimento, é realizado o *download* da base de dados composta pelas 3.600 notícias verdadeiras encontradas no corpus *Fake.Br*. A base possui os textos de cada notícia em sua versão completa divididos entre arquivos com extensão de texto. Os textos passam por um processo simples de limpeza excluindo espaços em branco e dividindo os parágrafos, para então serem indexados ao armazenamento do *Elasticsearch*.

O *Elasticsearch* foi utilizado para realizar a primeira busca e filtrar os documentos mais relevantes a serem lidos. Este é um componente de busca e análise de dados distribuídos com ótimo desempenho e escalabilidade devido a sua estrutura de dados de índice invertido.

Após a seleção dos documentos relevantes, os textos são lidos por um modelo em *BERT Large* e treinado no idioma português do Brasil a partir da base de dados *SQUAD v1.1 (Stanford Question Answering Dataset)*, uma base de *question answering* com dados extraídos de artigos da wikipédia amplamente utilizada em métodos de PLN, esta versão da base de dados conta com mais de 100.000 pares de perguntas e respostas (RAJPURKAR et al., 2016).

O modelo é carregado através da biblioteca do site *Hugging Face*, uma plataforma colaborativa com modelos e bases de dados hospedadas pelos usuários. A plataforma promove serviços de rápido acesso e conta com modelos com performance próximas ao estado da arte. É interessante pontuar que no período da pesquisa, em fevereiro de 2022, foram encontrados apenas 4 modelos de *question answering* no idioma português, em contraste ao inglês por exemplo, que possui 115 modelos disponíveis.

Antes de criar uma pergunta, é definida uma *pipeline* de extração de *quest answering* para combinar o *retriever* e o *reader* (modelo BERT) na busca e seleção das respostas. Por último, são realizadas as perguntas e a impressão das respostas.

6.3.4 Resultados

O método de *fact-checking* com *quest answering* obteve resultados satisfatórios ao ser capaz de ler as perguntas e imprimir as respostas baseadas no conjunto de dados escolhido. As perguntas e respostas serão totalmente expostas no item 6.4, no qual explica detalhadamente a metodologia de comparação e os resultados obtidos.

As perguntas realizadas foram elaboradas a partir da leitura dos textos classificados como falso-negativos pelo experimento de classificação, para isto, foi realizada uma análise qualitativa das informações centrais dos textos e informações auxiliares. O intuito foi verificar se o método de *fact-checking* com *quest answering* poderia extrair informações destes textos e validar respostas a partir dos contextos, constituindo uma abordagem capaz de complementar os resultados de um método diferente.

Figura 12 – Resposta para a pergunta “Qual é a crítica de Zico ao Flamengo?”

```
Resposta: 'montagem do elenco',  
Contexto: 'no estádio do Maracanã, no Rio de Janeiro, Zico foi '  
           'questionado sobre o ano do Flamengo. E o ídolo do clube '  
           'não poupou críticas à montagem do elenco.',  
Score: 0.8862317204475403
```

Fonte: Jonathan Camargo

Na figura 12 é possível visualizar um exemplo de resposta com o melhor *score* para a pergunta “Qual é a crítica de Zico ao Flamengo?”. O primeiro campo exibe uma pequena sequência de palavras definidas como a resposta, já o segundo, exibe o contexto sobre qual a resposta foi extraída, compondo uma sequência maior de palavras. O *score* representa a probabilidade de a resposta estar adequada à pergunta, variando de 0 a 1, sendo que quanto mais próximo de 1, maiores são as chances de estar correta.

6.3 Plataforma de integração

Dada a intenção de integrar diferentes abordagens em um único ambiente, foi incorporada a possibilidade de utilizar uma plataforma em desenvolvimento com objetivos em comum. Esta já está estruturada para receber algoritmos de verificação de *Fake News* e combinar as abordagens de maneira modular.

Descrita no capítulo 4 deste trabalho, a plataforma *Web Fake* foi desenvolvida com arquitetura de software baseada em APIs, o que possibilita implementar as funcionalidades das aplicações em forma de serviços, criando um sistema bem estruturado e ao mesmo tempo maleável devido às conexões dos módulos, o que é o ambiente ideal para a integração de novos módulos com diferentes abordagens.

Além das APIs responsáveis pela extração e armazenamento de metadados, destaca-se o serviço denominado *check news*, responsável por receber metadados dos artigos e encaminhar para os módulos de verificação de *Fake News*, este é o serviço que faz conexão com o experimento de classificação de texto descrito no capítulo 6 e também pode ser explorado nas conexões dos métodos propostos neste trabalho.

A plataforma teve seu funcionamento validado a partir de testes do sistema que permitiram identificar a operação adequada da plataforma no que tange ao acesso,

extração e cadastro de metadados de notícias de sites, bem como permitiram verificar a adequada funcionalidade da inclusão e utilização de um módulo inicial de classificação de *Fake News*.

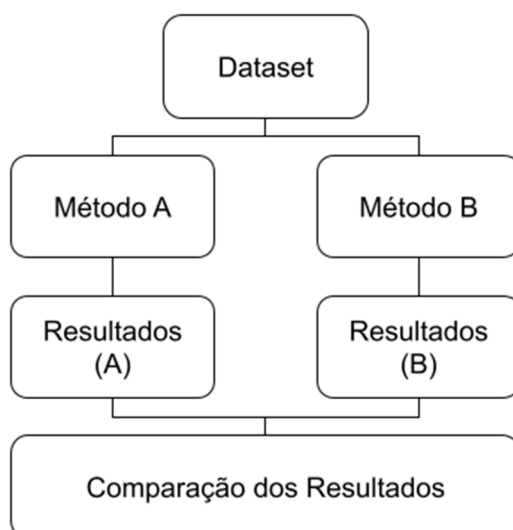
Estes testes funcionais foram complementados com o uso do modelo de classificação desenvolvido e descrito no item 6.1, sendo que este material foi utilizado em uma submissão de artigo preliminar.

Atualmente, a plataforma não integra o experimento de *fact-checking* e *question answering* citado no item 6.2, porém, o *Haystack* permite implementar o método *API REST* e conectar aos serviços da plataforma, viabilizando a proposta de integração de métodos na plataforma e podendo ser implementado em trabalhos futuros.

6.4 Metodologia de comparação

Para avaliar a integração de métodos distintos de verificação de *Fake News*, foi proposta uma metodologia a fim de comparar os resultados apresentados nos experimentos realizados. A figura 13 apresenta o esquema básico da metodologia proposta, esta estrutura pode ser ampliada para comportar mais *datasets* e mais métodos em trabalhos futuros.

Figura 13 – Metodologia de comparação



Fonte: Jonathan Camargo

O *dataset* utilizado em ambos os experimentos foi o corpus Fake.Br, composto por 3.600 notícias classificadas como verdadeiras e 3.600 classificadas como falsas. A escolha deste corpus se justifica pela pesquisa e leitura do estado da arte em trabalhos de processamento de linguagem natural com textos de notícias no idioma português. A metodologia de criação deste *dataset* passa por processos de definição dos sites, mineração de dados, filtro de similaridade léxica e análises manuais para garantir a qualidade do corpus.

O primeiro experimento realizado consiste no desenvolvimento de um método de classificação de texto, no qual foi implementado *TF-IDF* para calcular a frequência dos termos e realizar o treinamento e classificação com *Random Forest*. Métodos de classificação de texto podem enfrentar dificuldades com problemas semânticos como ambiguidades, além da possível similaridade de textos com ideias diferentes. Outra característica é a necessidade de treinamento no assunto a ser classificado, tendo dependência dos conjuntos de palavras e sentenças.

Após a classificação do algoritmo, os resultados foram exportados em um arquivo com extensão “*csv*” junto com os valores originais para possibilitar a leitura e identificação dos falso-negativos, que foram 28.

Figura 14 – Resultados do método de classificação de texto

Texto	Original	Predição			
8	fake	fake	3601	true	fake
14	fake	fake	3605	true	fake
15	fake	fake	3608	true	true
17	fake	fake	3612	true	fake
19	fake	true	3616	true	true
23	fake	true	3626	true	true
26	fake	true	3641	true	true
31	fake	fake	3647	true	fake
33	fake	fake	3653	true	fake
37	fake	true	3656	true	fake
41	fake	true	3658	true	fake
44	fake	fake	3661	true	true
45	fake	fake	3662	true	true
50	fake	fake	3667	true	true
51	fake	fake	3680	true	true
57	fake	fake	3682	true	true
59	fake	true	3684	true	true
61	fake	true	3685	true	true
63	fake	fake	3686	true	true
			3689	true	true

Fonte: Jonathan Camargo

Conforme a figura 14, os resultados podem ser visualizados em uma tabela com 3 colunas e 1.800 linhas, estas são compostas pelo número do documento de cada notícia que foram selecionadas aleatoriamente para o teste com o *Random Forest*, representando 25% do *dataset*.

A partir dos resultados do método de classificação de texto, foram identificados os resultados falso-negativos e realizada a leitura e análise dos textos das notícias. Não foi possível identificar um padrão sobre o conteúdo dos textos, porém, considerando que foi utilizado o método *TF-IDF*, estima-se que estes textos possuam palavras com frequências similares em textos classificados falsos.

Figura 15 – Exemplo de texto verdadeiro classificado como falso

O deputado Wladimir Costa (SD-PA), que ganhou notoriedade ao [tatuar o nome](#) do presidente Michel Temer no ombro, divulgou nesta terça (8), via WhatsApp, imagem de caráter homofóbico, cujo alvo é o jornalista Ricardo Boechat, apresentador do grupo Bandeirantes.

A reportagem da **Folha** recebeu o meme do deputado: uma montagem em que o microfone do jornalista foi substituído por um órgão sexual masculino.

O título da mensagem é "Ricardo Bichat e o microfone de nervo" e balões de texto trazem mensagem de cunho homofóbico.

A **Folha** procurou o deputado nesta manhã. Ele disse que sofreu críticas pesadas de Boechat e que "apenas compartilhou" um dos "tantos" memes que já recebeu sobre o jornalista.

"São de pessoas que também não concordam com o comportamento dele, produzem e compartilham na mídia social. E compartilhei com você", disse.

Boechat não atendeu as ligações da **Folha** na tarde desta terça e a assessoria de imprensa do grupo Bandeirantes informou que não vai comentar o gesto do deputado.

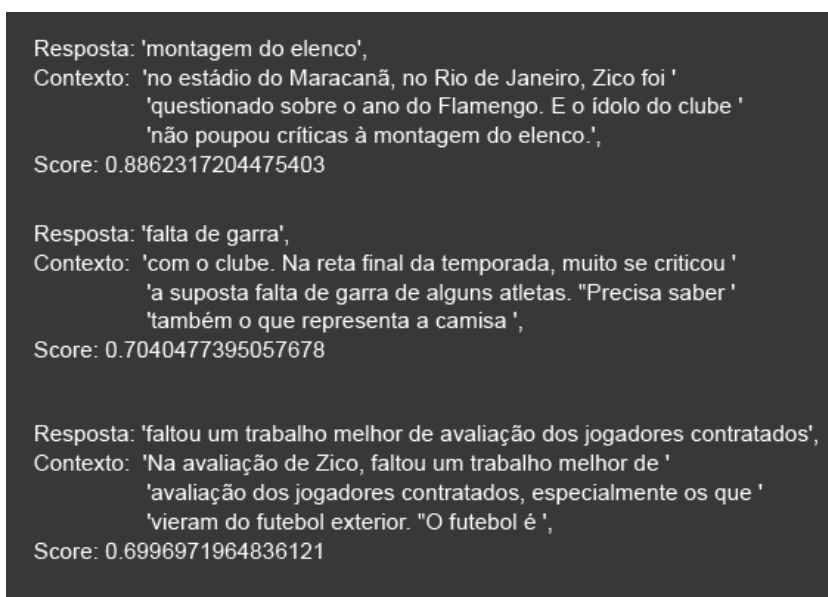
Fonte: Folha de São Paulo

A partir da leitura dos textos classificados como falso-negativos, foram elaboradas 5 perguntas com o intuito de testar se o método de *fact-checking* e *question answering* seria capaz de formular respostas com as informações corretas, configurando uma abordagem complementar à classificação de texto.

O segundo experimento é baseado em um método de *fact-checking* com modelo de *question answering*, o *dataset* é composto pelas 3.600 notícias verdadeiras do corpus *Fake.Br*, funcionando como uma base de conhecimento para a busca das informações. Métodos de *fact-checking* realizam consultas em bases de

conhecimento e dados estruturados, possibilitando fornecer uma explicação para a predição, além de promover mais confiabilidade em razão das fontes. Ao mesmo tempo, isto significa uma necessidade de conhecimento prévio e no caso de bases que não se atualizam, um conhecimento limitado.

Figura 16 - “Qual é a crítica de Zico ao Flamengo?”



Fonte: Jonathan Camargo

Na figura 16 é possível visualizar as três melhores respostas para a pergunta “Qual é a crítica de Zico ao Flamengo?”. A pergunta foi formulada a partir da leitura de um texto classificado como falso-negativo e o método de *quest answering* foi capaz de formular as respostas baseadas nas informações do documento com este mesmo texto. Nesta amostra é possível observar a importância da explicação da predição característica dos métodos de *fact-checking*, pois a apresentação do contexto torna a resposta mais rica e melhora a confiabilidade.

É importante ressaltar que, segundo o *Haystack*, o *score* é uma métrica que estima a probabilidade de o resultado estar correto baseado na relação entre resposta e pergunta, porém, em alguns momentos esta métrica não terá total conformidade com a resposta correta. Isto significa que em algum momento, uma resposta com score alto pode estar incorreta, enquanto o contrário também é possível, reforçando a importância do contexto e da leitura.

Figura 17 - “O que o deputado Wladimir Costa divulgou no *whatsapp*?”

Resposta: 'imagem de caráter homofóbico',
 Contexto: 'Michel Temer no ombro, divulgou nesta terça (8), via 'WhatsApp, imagem de caráter homofóbico, cujo alvo é o 'jornalista Ricardo Boechat, apresentador do g',
 Score: 0.9590848982334137

Resposta: 'uma montagem',
 Contexto: 'o aplicativo Whastapp – do qual fazem parte deputados e 'assessores – uma montagem com o propósito de “atacar a 'condição de mulher, mãe e parlamentar”’,
 Score: 0.8906220197677612

Resposta: 'fotos íntimas',
 Contexto: '-PA). Em uma das representações, ele foi acusado pelo PT 'de divulgar fotos íntimas de uma filha da deputada Maria 'do Rosário (RS). "O Conselho de Étíc',
 Score: 0.7061140984296799

Fonte: Jonathan Camargo

Na figura 17 são exibidas as três melhores respostas para a pergunta “O que o deputado Wladimir Costa divulgou no *whatsapp*?”. Apesar da pergunta ter sido formulada a partir de um texto em específico, estas três respostas são baseadas em três documentos de texto diferentes. Apesar de o contexto não apresentar com clareza a explicação para o resultado, ao acessar os textos é possível identificar a relação entre resposta e pergunta.

A figura 18 apresenta o maior conjunto de respostas corretas para as perguntas selecionadas, isto acontece pela alta quantidade de informações relacionadas ao assunto disponíveis no *dataset* utilizado, conforme a maior parte do corpus é composta por notícias de política, esta é a categoria com mais informações disponíveis.

Figura 18 - “Quem é Jair Bolsonaro?”

Resposta: 'pré-candidato à Presidência',
Contexto: 'nto é o deputado federal Eduardo Bolsonaro (PSC-SP), filho ' 'do pré-candidato à Presidência Jair Bolsonaro, que repete ' 'a exaustão que "bandido só respei',
Score: 0.9899130165576935
Resposta: 'deputado federal e militar da reserva',
Contexto: 'militares podem voltar ao poder por meio do voto –é o que o ' 'deputado federal e militar da reserva Jair Bolsonaro, em ' 'segundo lugar nas pesquisas, afirmou',
Score: 0.9754301905632019
Resposta: 'PSC-RJ',
Contexto: "parlamentar afirmou em entrevista que cuspiria de novo em ' 'Jair Bolsonaro (PSC-RJ). No ano passado, quando a Câmara ' 'analísava o processo de impeachment d',
Score: 0.948394387960434
Resposta: 'pai de Eduardo Bolsonaro',
Contexto: 'Dilma Rousseff, Jean Wyllys cuspiu em Jair Bolsonaro ' '(PSC-RJ), pai de Eduardo Bolsonaro, logo após votar contra ' 'o processo. À época, Wyllys disse ter ',
Score: 0.927505612373352
Resposta: 'um nacional-desenvolvimentista tipo Geisel',
Contexto: 'a com a candidatura do deputado Jair Bolsonaro, que é um ' 'nacional-desenvolvimentista tipo Geisel, e apoiou a ' 'candidatura da Marie Le Pen na França e d',
Score: 0.8501009047031403

Fonte: Jonathan Camargo

Os resultados desta pergunta tornam evidente a necessidade da atualização da base de conhecimento. Devido ao período de buscas para a anotação do corpus, o conhecimento disponível nos textos é de janeiro de 2016 à janeiro de 2018. Apesar das três melhores respostas estarem corretas dentro do período de análise das notícias, ao considerarmos a data de março de 2022, é possível definir que estas três respostas estão incorretas, pois estas informações sofreram mudanças.

Este exemplo elucida a importância de bases de dados abertos e estruturados que se atualizam frequentemente para que as informações corretas possam ser extraídas e transformadas em conhecimento.

Figura 19 - “Qual foi a captação da poupança em 2017?”

Resposta: '6,61%',
 Contexto: 'Em 2017, a remuneração dos depósitos de poupança foi de '
 '6,61%, superando largamente a inflação estimada em menos '
 'de 3%. Em 2016, as cadernetas já haviam',
 Score: 0.9804734885692596

Resposta: 'R\$ 14,8 bilhões',
 Contexto: 'a de poupança se recuperaram em 2017, com captação líquida '
 'anual de R\$ 14,8 bilhões no Sistema Brasileiro de Poupança '
 'e Empréstimo (SBPE). Passou, ass',
 Score: 0.927505612373352

Fonte: Jonathan Camargo

A figura 19 apresenta um exemplo de uma resposta equivocada que obteve um *score* alto. Ambas as respostas foram extraídas do mesmo documento de texto, porém, a resposta com o melhor *score* está incorreta. A partir da leitura do contexto é possível interpretar que a informação contida na resposta não está devidamente correta ao relacionar com a pergunta, pois tratam de dados distintos. Neste exemplo, a resposta correta obteve *score* de 0.927, enquanto a incorreta obteve *score* de 0.980. Apesar da similaridade das palavras em ambas as sequências, um especialista seria capaz de interpretar com facilidade as diferenças entre os termos.

Figura 20 - “Qual foi a inflação em 2017?”

Resposta: '2,95%',
 Contexto: 'A inflação encerrou 2017 em 2,95%, divulgou o IBGE nesta '
 'quarta-feira (10), abaixo do piso da meta do governo, de '
 '3%.',
 Score: 0.9960058033466339

Fonte: Jonathan Camargo

A figura 20 apresenta uma resposta com um *score* de 0.996. Este resultado é alcançado pela alta similaridade entre as informações contidas no documento de texto e na pergunta, facilitando a atuação do modelo de *question answering* utilizado.

Os resultados apresentados pelo método de *fact-checking* e *question answering* ilustram como é possível propor métodos complementares para a abrangência de mais contextos. Ainda que tenha sido utilizado o mesmo *dataset*, textos que foram classificados como falso-negativos podem ter suas informações extraídas corretamente a partir do método de *fact-checking*. A análise destes

resultados proporcionou a resposta para a questão de pesquisa, representando como a integração de abordagens distintas pode sim, ampliar a abrangência de resultados de verificação de Fake News em contextos distintos.

Além disso, os experimentos foram importantes para a explorar as características de cada método e visualizar na prática com resultados em um *dataset* real. Além disso, pensar um ambiente para a integração destes métodos também potencializa a maneira com que estas informações são consumidas.

É possível ainda pressupor resultados mais sólidos com algoritmos mais complexos de *deep learning* com uso de redes neurais artificiais para métodos de classificação de texto, além de ontologias e bases de conhecimento com dados abertos e estruturados para a consulta de métodos de *fact-checking*.

7 CONCLUSÃO

Este trabalho descreveu a dinâmica pela qual as *Fake News* são impulsionadas nas plataformas digitais dadas as características inerentes aos meios de comunicação na internet, que geram impactos em nossa sociedade por conta da crença das pessoas em informações falsas, ou ainda pela descrença em informações verdadeiras. A emergência de conter a proliferação de informações falsas tem recebido grande contribuição da área da computação a partir dos estudos de métodos automatizados de verificação de *Fake News*.

As *Fake News* se tornaram um problema emergente a partir das plataformas de mídias sociais digitais e ao que parece, tendem a seguir crescendo no âmbito político e social. A desinformação a partir de informações falsas ou tendenciosas costumam beneficiar apenas um grupo seleto de atores, enquanto a sociedade como um todo tende a herdar impactos negativos.

Ao realizar a leitura do estado da arte, observa-se que as abordagens propostas possuem delimitações de aplicabilidade em contextos específicos e não há um método capaz de lidar com diferentes contextos sem comprometer seus resultados. Com base nisto, este trabalho propôs a integração de métodos distintos com o intuito de ampliar a cobertura de aplicabilidade em diferentes contextos. Foram desenvolvidos um método de classificação de texto e um método de *fact-checking* com modelo *question answering* para validar a hipótese de pesquisa.

A metodologia de análise dos resultados proporcionou a resposta para a questão de pesquisa, que representa como a integração de abordagens distintas pode sim, ampliar a abrangência de resultados de verificação de *Fake News* em contextos distintos.

A escolha dos métodos propostos se justifica pelas características que os diferenciam, uma vez que métodos de classificação de texto podem enfrentar dificuldades semânticas como ambiguidades, similaridade entre textos com ideias distintas e terem seu desempenho dependente de treinamento no assunto abordado. Ao mesmo tempo, métodos de *fact-checking* são capazes de extrair informações de bases de conhecimento para a validação dos resultados e possibilitam explicação da predição, porém, são dependentes de conhecimento prévio.

O *dataset* utilizado em ambos os experimentos contribuiu para a avaliação dos resultados de forma mais clara, visto que os experimentos compartilham das mesmas

informações a serem validadas. Também foi importante utilizar um corpus com dados reais, extraídos de sites de notícias que são amplamente acessados, pois se aproxima mais da linguagem atribuída e consumida pelos leitores em mídias digitais.

7.1 Contribuições

A metodologia de comparação adotada proporcionou realizar uma análise dos resultados de ambos os experimentos e validar o objetivo de avaliar a integração de métodos distintos de verificação de *Fake News*, além da possibilidade de ampliar a metodologia para outros experimentos e outros *datasets*.

A partir dos resultados apresentados, foi possível atestar que os resultados falso-negativos do método de classificação de texto puderam ser interpretados corretamente pelo método de *fact-checking* proposto, caracterizando a complementaridade entre estes dois métodos distintos e aumentando a abrangência de aplicação.

Este trabalho promoveu o desenvolvimento de dois experimentos com métodos distintos de verificação de *Fake News* e a análise dos resultados alcançados. Além disso, proporcionou a contribuição na submissão de um artigo ao Simpósio Brasileiro de Sistemas de Informação (SBSI 2021), além da implementação do método de classificação na plataforma *Web Fake*.

Também foi realizada uma apresentação no Seminário Desinformação e Pandemia promovido pelo Programa de Pós Graduação em Comunicação pela Unisinos, reforçando a relação deste trabalho para com a área de comunicação, visto que o estudo a partir de uma visão interdisciplinar sobre o tema *Fake News* potencializa os resultados dos trabalhos.

7.2 Limitações do trabalho

Durante a pesquisa ficou evidente a dificuldade em encontrar bases de dados estruturados no idioma português do Brasil, a ontologia *DBpedia* em português não possui as triplas *RDF* completas e está desatualizada. Ao mesmo tempo em que métodos com processamento de linguagem natural estão evoluindo no campo da

computação, o investimento na ciência no país tem o potencial de impulsionar a produção de conhecimento científico na área.

A proposta inicial previa a um conjunto de experimentos utilizando de forma integrada a plataforma Web-Fake, envolvendo na medida do possível uma grande quantidade de usuários e volumes maiores de itens em teste. A partir de um conjunto de experimentos ampliado e com uso de maior quantidade de recursos seria possível evidenciar de forma mais assertiva os resultados da metodologia proposta. Esta avaliação ampliada não foi implementada tal como prevista, porém está alinhada com atividades futuras sugeridas a seguir.

7.3 Trabalhos futuros

Devido à natureza deste trabalho, há uma gama de potenciais melhorias para trabalhos futuros. Ao pensar uma plataforma para integrar os métodos complementares, pode-se propor um ambiente não só capaz de integrar os métodos de verificação de *Fake News*, mas também pensar a experiência do usuário em diferentes níveis para promover um serviço inteligível e ao mesmo tempo tecnicamente satisfatório, ampliando o alcance para mais pessoas.

A integração de métodos distintos pode tornar os resultados confusos e para isto, elaborar um serviço de verificação cruzada entre os métodos poderia aprimorar os resultados e organizá-los de acordo com o contexto aplicado.

As redes neurais artificiais são um recurso que tem apresentado resultados compatíveis com o estado da arte em métodos de classificação e seria importante o desenvolvimento e inclusão de um método com RNA na plataforma desenvolvida. Além disso, a construção de bases de conhecimento com dados abertos e estruturados como ontologias no idioma português teria grande contribuição para o desenvolvimento de métodos de *fact-checking*.

REFERÊNCIAS

- AGRAWAL, R. Analyzing Text Classification Techniques on Youtube Data. 2019. Disponível em: <https://towardsdatascience.com/analyzing-text-classification-techniques-on-youtube-data-7af578449f58>. Acesso em: 20 out. 2021
- AJAO, O.; BHOWMIK, D.; ZARGARI, S. **Fake News Identification on Twitter with Hybrid CNN and RNN Models**. Proceedings of the 9th International Conference on Social Media and Society. **Anais**. In: Smsociety '18: International Conference on Social Media and Society. Copenhagen Denmark: ACM, 18 jul. 2018. Disponível em: <https://dl.acm.org/doi/10.1145/3217804.3217917>. Acesso em: 10 mar. 2021.
- AVAAZ. O Brasil está sofrendo uma infodemia de Covid-19. 4 mai. 2020. Disponível em: https://secure.avaaz.org/campaign/po/brasil_infodemia_coronavirus/. Acesso em: 17 mar. 2021.
- BHATT, G.; SHARMA, A.; SHIVAM, S.; NAGPAL, A.; RAMAN, B.; MITTAL, A. **Combining Neural, Statistical and External Features for Fake News Stance Identification**. Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18. **Anais**. In: Companion Of The The Web Conference 2018. Lyon, France: ACM Press, 2018. Disponível em: <http://dl.acm.org/citation.cfm?doid=3184558.3191577>. Acesso em: 20 mar. 2021.
- CALDERINI, R. W. HUB de Integração para Métodos de Análise de Fake News. 2020. Acesso em: 20 mai.2021
- CAMBRIDGE Dictionaries Online. Fake News. Disponível em: <http://dictionary.cambridge.org/>. Acesso em: 17 mar. 2021.
- CHORAS, M.; PAWLICKI, M.; KOZIK, R.; DEMESTICHAS, K.; KOSMIDES, P.; GUPTA, M. **SocialTruth Project Approach to Online Disinformation (Fake News) Detection and Mitigation**. Proceedings of the 14th International Conference on Availability, Reliability and Security. **Anais**... In: Ares '19: 14th International Conference On Availability, Reliability And Security. Canterbury CA United Kingdom: ACM, 26 ago. 2019. Disponível em: <https://dl.acm.org/doi/10.1145/3339252.3341497>. Acesso em: 28 mar. 2021.
- CONROY, N. K.; RUBIN, V. L.; CHEN, Y. Automatic deception detection: Methods for finding fake news. **Proceedings of the Association for Information Science and Technology**, v. 52, n. 1, p. 1–4, jan. 2015. Disponível em: https://www.researchgate.net/publication/281818865_Automatic_Deception_Detection_Methods_for_Finding_Fake_News. Acesso em: 10 abr. 2021.
- CORDEIRO, P.; PINHEIRO, V.; MOREIRA, R.; CARVALHO, C.; FREIRE, L. **What is Real or Fake? - Machine Learning Approaches for Rumor Verification using Stance Classification**. IEEE/WIC/ACM International Conference on Web Intelligence. **Anais**... In: WI '19: IEEE/WIC/ACM International Conference On Web Intelligence. Thessaloniki Greece: ACM, 14 out. 2019. Disponível em: <https://dl.acm.org/doi/10.1145/3350546.3352562>. Acesso em: 10 abr. 2021.

CUI, L.; SHU, K.; WANG, S.; LEE, D.; LIU, H. **dEFEND: A System for Explainable Fake News Detection**. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. **Anais...** In: CIKM '19: The 28th ACM International Conference On Information And Knowledge Management. Beijing China: ACM, 3 nov. 2019. Disponível em: <https://dl.acm.org/doi/10.1145/3357384.3357862>. Acesso em: 10 abr. 2021.

DE OLIVEIRA, N.R.; PISA, P.S.; LOPEZ, M.A.; DE MEDEIROS, D.S.; MATTOS, D.M. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. **Information**, v. 12, n. 1, p. 38, 18 jan. 2021. Disponível em: <https://doi.org/10.3390/info12010038>. Acesso em: 20 jun. 2021

DELMAZO, C.; VALENTE, J. C. Fake news nas redes sociais online: propagação e reações à desinformação em busca de cliques. **Media & Jornalismo**, v. 18, n. 32, p. 155–169, 18 maio 2018. Disponível em: https://doi.org/10.14195/2183-5462_32_11. Acesso em: 10 abr. 2021.

DIÓGENES, J. Alckmin diz que 'prévia não divide, prévia escolhe'. **Estadão**, dez, 2017. Disponível em: <https://politica.estadao.com.br/noticias/geral,alckmin-diz-que-previa-nao-divide-previa-escolhe,70002131259>. Acesso em: 10 dez. 2022.

FEILMAYR, C. **Text Mining-Supported Information Extraction: An Extended Methodology for Developing Information Extraction Systems**. 2011 22nd International Workshop on Database and Expert Systems Applications. **Anais...** In: 2011 22nd International Conference On Database And Expert Systems Applications (DEXA). Toulouse, France: IEEE, ago. 2011. Disponível em: <http://ieeexplore.ieee.org/document/6059820/>. Acesso em: 19 jun. 2021.

G1. G1 lança Fato ou Fake, novo serviço de checagem de conteúdos suspeitos. 2018. Disponível em: <https://g1.globo.com/fato-ou-fake/noticia/2018/07/30/g1-lanca-fato-ou-fake-novo-servico-de-checagem-de-conteudos-suspeitos.ghtml>. Acesso em: 10 mar. 2021.

GUILLOU, Pierre. Bert large cased squad v1.1 Portuguese. 2021. Disponível em: <https://huggingface.co/pierreguillou/bert-large-cased-squad-v1.1-portuguese>. Acesso em: 10 fev. 2022.

GUPTA, V. A Survey of Natural Language Processing Techniques. **Engineering Technology**, v. 5, n. 01, p. 3, 2014. Disponível em: <https://doi.org/10.1155/2015/674296>. Acesso em: 20 jun. 2021.

Haystack. Overview. 2020. Disponível em: <https://Haystack.deepset.ai/overview/intro>. Acesso em: 23 fev. 2022.

HISAYASU, A.; CAMARGO, I.; COSTA, E.; OLIVEIRA, V. PF investiga ex-senador Romero Jucá e filho em suposto esquema de candidaturas laranjas em Roraima. **Globonews**, dez, 2019. Disponível em: <https://g1.globo.com/rr/roraima/noticia/2019/12/11/pf-faz-buscas-em-investigacao-sobre-supostas-candidaturas-laranjas-em-roraima.ghtml>. Acesso em: 12 fev. 2022.

KANG, L. A survey on neural relation extraction. 2020. China. Disponível em: <https://doi.org/10.1007/s11431-020-1673-6>. Acesso em: 3 fev. 2022.

KUMAR, S. A Survey of Deep Learning Methods for Relation Extraction. 2017. Disponível em: <https://arxiv.org/pdf/1705.03645.pdf>. Acesso em: 22 jan. 2022.

LIN, P.; SONG, Q.; WU, Y. Fact Checking in Knowledge Graphs with Ontological Subgraph Patterns. **Data Science and Engineering**, v. 3, n. 4, p. 341–358, dez. 2018. Disponível em: <https://doi.org/10.1007/s41019-018-0082-4>. Acesso em: 20 jun. 2021.

LIU, K. A survey on neural relation extraction. **Science China Technological Sciences**, v. 63, n. 10, p. 1971–1989, out. 2020. Disponível em: <https://doi.org/10.1007/s11431-020-1673-6>. Acesso em: 20 jun. 2021.

MONTEIRO, R. A.; SANTOS, R. L.; PARDO, T. A.; DE ALMEIRA, T. A.; RUIZ, E. E.; VALE, O. A. Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. In: VILLAVICENCIO, A. et al. (Eds.). **Computational Processing of the Portuguese Language**. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018. v. 11122p. 324–334. Disponível em: https://doi.org/10.1007/978-3-319-99722-3_33. Acesso em: 30 mai. 2021.

OLIVEIRA, E. Manchas de óleo chegam a mais de 800 pontos após mudança de metodologia na contagem do Ibama. **G1**, nov, 2019. Disponível em: <https://g1.globo.com/natureza/desastre-ambiental-petroleo-praias/noticia/2019/11/28/manchas-de-oleo-chegam-a-mais-de-800-pontos-apos-mudanca-de-metodologia-na-contagem-do-ibama.ghtml>. Acesso em: 13 fev. 2022.

OLIVEIRA, J. Bolsonaro vai militarizar todas as escolas do Brasil em 4 anos. **Folha Brasil**, 2017. Disponível em: <https://afolhabrasil.com.br/politica/bolsonaro-vai-militarizar-todas-as-escolas-do-brasil-em-4-anos/>. Acesso em: 20 fev. 2022.

RAJPURKAR, P.; ZHANG, J.; LOPYREV, K.; LIANG, P. SQuAD: 100,000+ Questions for Machine Comprehension of Text. In: **Conference on Empirical Methods in Natural Language Processing**, p. 2383–2392, Austin, Texas, 11 out. 2016. Disponível em: <https://arxiv.org/abs/1606.05250>. Acesso em: 20 fev. 2022.

SHARMA, K.; QIAN, F.; JIANG, H.; RUCHANSKY, N.; ZHANG, M.; LIU, Y. Combating Fake News: A Survey on Identification and Mitigation Techniques. **ACM Transactions on Intelligent Systems and Technology**, v. 10, n. 3, p. 1–42, 31 maio 2019. Disponível em: <https://doi.org/10.1145/3305260>. Acesso em: 10 abr. 2021.

SHU, K.; SILVA, A.; WANG, S.; TANG, J.; LIU, H. Fake News Detection on Social Media: A Data Mining Perspective. **ACM SIGKDD Explorations Newsletter**, v. 19, p. 22, jun. 2017. Disponível em: <https://dl.acm.org/doi/10.1145/3137597.3137600>. Acesso em: 19 abr. 2021.

TANDOC, E. C.; LIM, Z. W.; LING, R. Defining “Fake News”: A typology of scholarly definitions. **Digital Journalism**, v. 6, n. 2, p. 137–153, 7 fev. 2018. Disponível em: <https://doi.org/10.1080/21670811.2017.1360143>. Acesso em: 19 abr. 2021.

WARDLE, C.; DERAKHSHAN, H. Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report, v. 27, 2017.

Disponível em:

https://www.researchgate.net/publication/339031969_INFORMATION_DISORDER_Toward_an_interdisciplinary_framework_for_research_and_policy_making_Informati_on_Disorder_Toward_an_interdisciplinary_framework_for_research_and_policymaki ng. Acesso em: 10 jun. 2021

WAZLAWICK, R. S. Metodologia de Pesquisa para Ciência da Computação. Elsevier Editora Ltda. Rio de Janeiro, v. 6, 2009. Disponível em: https://tsxvpsbr.dyndns.org/arquivos/UFFS/Metodologia%20De%20Pesquisa%20Cie nciaDaComputacao%20_%20TCC1.pdf. Acesso em: 19 mai. 2021.

YU, B. et al. Semi-Open Information Extraction. In Proceedings of the **Web Conference 2021**, Abr. 2021. Disponível em: <https://doi.org/10.1145/3442381.3450029>. Acesso em: 10 fev. 2022.

ZHOU, X.; JAIN, A.; PHOHA, V. V.; ZAFARANI, R. Fake News Early Detection: A Theory-driven Model. **Digital Threats: Research and Practice**, v. 1, n. 2, p. 1–25, 9 jul. 2020. Disponível em: <https://dl.acm.org/doi/10.1145/3377478>. Acesso em: 20 abr. 2021.

ZHU, Q.; CHENG, X. **The Opportunities and Challenges of Information Extraction**. 2008 International Symposium on Intelligent Information Technology Application Workshops. **Anais...** In: 2008 INTERNATIONAL SYMPOSIUM ON INTELLIGENT INFORMATION TECHNOLOGY APPLICATION WORKSHOPS (IITAW). Shanghai, China: IEEE, dez. 2008. Disponível em: <http://ieeexplore.ieee.org/document/4732009/>. Acesso em: 19 mai. 2021.