

# AVICO CNF: UM SUPORTE FERRAMENTAL PARA A CLASSIFICAÇÃO DE NOTÍCIAS FALSAS SOBRE VACINAS DA COVID

Cassio Ceolin Júnior\*  
Fábio Paulo Basso\*\*  
Gabriel Machado Lunardi\*\*\*

## RESUMO

O crescimento imensurável das informações falsas nas mídias sociais e outras plataformas durante situações de pandemia como o COVID-19 pode causar danos significativos à estabilidade física e mental das pessoas. Para detectar essas informações falsas, pesquisadores têm aplicado várias técnicas com Mineração de Dados (*Web Mining*), Aprendizado de Máquina (*Machine Learning*) e Aprendizado Profundo (*Deep Learning*). O objetivo deste estudo foi revisar, avaliar e sintetizar sistematicamente artigos de pesquisa que utilizaram diferentes técnicas para detectar *fake news* sobre a COVID-19. Uma pesquisa de literatura foi realizada nos bancos de dados bibliográficos relevantes para garantir que a pesquisa fosse centrada apenas em pesquisas de alta integridade. Em particular, identificamos vários conjuntos de dados de *fake news* e revisamos diferentes técnicas de processamento de dados, extração de recursos e classificação para detectar as *fake news*. No final, discutimos os desafios e limitações encontrados na nossa revisão.

**Palavras-chaves:** machine learning, deep learning, natural language processing, fake news, literature survey, systematic literature review, systematic mapping study, systematic mapping review.

## ABSTRACT

The immeasurable growth of fake news on social media and other platforms during pandemic situations like COVID-19 can cause significant damage to people's physical and mental stability. To detect this false information, researchers have applied various techniques such as Web Mining, Machine Learning and Deep Learning. The aim of this study was to systematically review, evaluate and synthesize research articles that used different techniques to detect fake news about COVID-19. A literature search was performed on the relevant bibliographic databases to ensure that the search focused only on high integrity research. In particular, we identify several fake news datasets and review different data processing, feature extraction and classification techniques to detect fake news. At the end, we discuss the challenges and limitations found in our review.

**Keywords:** machine learning, deep learning, natural language processing, fake news, literature survey, systematic literature review, systematic mapping study, systematic mapping review.

\* Aluno do Curso de Engenharia de Software da Universidade Federal do Pampa, Alegrete, Rio Grande do Sul, Brasil  
E-mail: [cassioceolin.aluno@unipampa.edu.br](mailto:cassioceolin.aluno@unipampa.edu.br)

\*\* Orientador, Professor do Curso de Engenharia de Software da Universidade Federal do Pampa, Alegrete, Rio Grande do Sul, Brasil, E-mail: [fabio basso@unipampa.edu.br](mailto:fabio basso@unipampa.edu.br)

\*\*\* Co-Orientador, Professor do Departamento de Eletrônica e Computação da Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brasil, E-mail: [gabriel.lunardi@ufsm.br](mailto:gabriel.lunardi@ufsm.br)

## 1. INTRODUÇÃO

*Fake news* é um termo usado para descrever informações falsas ou enganosas que são deliberadamente disseminadas através de meios de comunicação, como sites de notícias, redes sociais e mensagens instantâneas, visando influenciar a opinião pública e enganar as pessoas. A disseminação de *fake news* pode ter consequências graves, incluindo a desinformação, o aumento da polarização política e até mesmo a instabilidade social. Por isso, é importante identificar e desacreditar notícias falsas para garantir a integridade da informação e proteger a sociedade (SYNTAX, 2021).

Tendo um impacto significativo na sociedade contemporânea. Elas não apenas distorcem a percepção pública e política, mas também podem influenciar decisões individuais e coletivas, especialmente em temas sensíveis como saúde pública e eleições. A disseminação rápida e ampla de informações falsas através de plataformas digitais pode minar a confiança nas fontes tradicionais de informação e aumentar a polarização social. Além disso, o desafio crescente de identificar e mitigar *fake news* destaca a necessidade urgente de desenvolver abordagens robustas e tecnologicamente avançadas para proteger a integridade da informação e promover um ambiente de comunicação mais seguro e confiável (MISINFORMATION, 2022).

Para isso a detecção de *fake news* envolve o uso de técnicas e ferramentas para identificar informações falsas ou enganosas e distingui-las de informações verdadeiras. Isso pode incluir a verificação de fontes, a análise de evidências e a comparação com informações confiáveis. Além disso, a inteligência artificial e técnicas de aprendizado de máquina estão sendo utilizados para ajudar na detecção de *fake news*, por meio da identificação de padrões e tendências nas informações enganosas. No entanto, é importante lembrar que a tecnologia não é uma solução definitiva e que a responsabilidade final de distinguir o verdadeiro do falso ainda é do público. Também é importante desenvolver habilidades críticas e a capacidade de avaliar a informação de forma independente (COSTA et al., 2020).

O presente trabalho visa desenvolver um suporte ferramental para a Associação de Vítimas de COVID-19 (AVICO) sendo esta uma organização comunitária com mobilização social que visa minimizar os diferentes impactos da pandemia no Brasil, lutando pela memória das vítimas da COVID-19, assim como pela garantia de acesso aos Direitos Humanos. Usando da utilizando de técnicas e métodos para a identificação automática de *fake news*, tendo em vista que a quantidade de informações disponíveis torna demorada e difícil a classificação manual. Assim, para o desenvolvimento do presente Trabalho de Conclusão de Curso, é necessário pesquisar sobre o tema proposto e utilizar algoritmos de classificação de texto para a AVICO <sup>1</sup>.

Para isso foi conduzido um estudo de mapeamento sistemático com a finalidade de identificar métodos, técnicas ou algoritmos utilizados na detecção de *Fake News*. Portanto, de cunho técnico-científico, o estudo foca na área de *machine learning* e *web mining*.

### 1.1. METODOLOGIA

Nesta seção, será demonstrada a metodologia pretendida a ser executada durante o TCC. Para isso, foi desenvolvida a Figura 5 em um modelo BPMN com a esquema-

---

<sup>1</sup><http://avicobrasil.com.br>

tização das atividades do trabalho. Neste modelo, é possível analisar os seguintes processos: Concepção, Fundamentação Teórica, TCC I e TCC II. A seguir, tem-se uma descrição do que foi realizado em cada etapa:

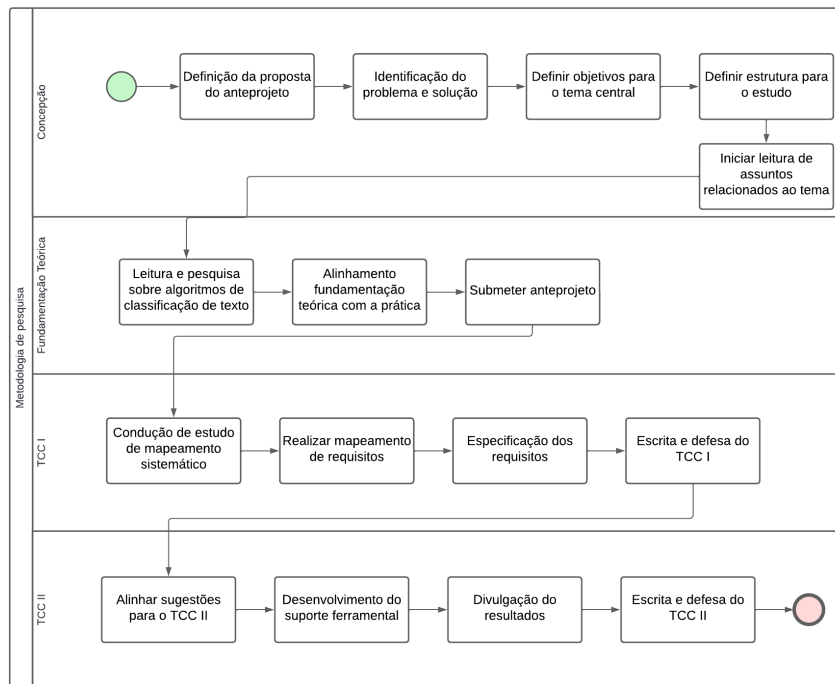


Figura 1 - Desenho da metodologia adotada. Fonte: Autor

- **Concepção:** Nesta etapa inicial, o objetivo principal foi a definição do tema do projeto, sua estruturação, bem como uma leitura inicial sobre as ferramentas relacionadas ao tema.
- **Fundamentação Teórica:** Nesta fase, serão realizadas pesquisas exploratórias em algoritmos de classificação de texto. Isso contribuiu para o conhecimento necessário para desenvolver o algoritmo que foi implementado no suporte ferramental.
- **TCC I:** Na etapa do TCC I, ocorreu a leitura dos artigos relacionados ao tema, no qual o TCC foi escrito e defendido.
- **TCC II:** Por fim, na etapa do TCC II, foram aplicados alguns ajustes referentes à escrita. Foi desenvolvido o suporte ferramental que contém o algoritmo para verificar se uma notícia é verdadeira ou não. Foi criada uma *API* que recebe como requisição o texto da notícia, o qual será fornecido pelos algoritmos responsáveis por realizar a mineração dos dados da notícia para assim verificar e retornar como resposta um texto indicando se é verdadeira ou falsa. Além disso, ocorreu a escrita e a defesa do TCC II.

## 1.2. OBJETIVO GERAL E ESPECIFICO

O objetivo geral deste trabalho é fornecer um suporte ferramental que verifica uma possível *fake news*. A partir disso, podem ser elencados os seguintes objetivos específicos:

- Estudar sobre técnicas de Inteligência Artificial aplicadas ao processamento de linguagem natural;
- Implementar o suporte ferramental para a classificação de notícias;
- Implantar e testar a aplicação de detecção na plataforma da AVICO;
- Realizar uma revisão sistemática dos estudos relacionados ao trabalho.

### **1.3. CONTRIBUIÇÕES**

As principais contribuições para este projeto de TCC estão ligadas ao desenvolvimento de um algoritmo para classificação de texto, o qual, conseqüentemente, visa ser integrado a um suporte ferramental que irá classificar a veracidade das notícias. Sendo assim, as contribuições esperadas para este trabalho de TCC são:

- Criação de um suporte ferramental integrado para a classificação de *Fake News*;
- Aplicação e consolidação do conhecimento técnico e científicos adquiridos ao longo do curso;

### **1.4. ORGANIZAÇÃO**

O presente artigo está estruturado da seguinte forma: A Seção 2 apresenta a fundamentação teórica e a motivação para o presente trabalho. A seção 3 introduz como foi realizada toda a etapa de planejamento do mapeamento sistemático. A Seção 4 apresenta a proposta de desenvolvimento, bem como a implementação e a demonstração conceitual. A seção 5 discute as ameaças à validade e, na seção 6, são realizadas algumas discussões acerca do tema. A seção 7 apresenta a conclusão do trabalho, debatendo sobre as considerações finais e trabalhos futuros.

## **2. FUNDAMENTAÇÃO TEÓRICA**

A disseminação de notícias falsas pode causar impactos negativos na sociedade e na política, a teoria por trás da detecção de *fake news* se baseia na utilização de métodos de processamento de linguagem natural, mineração de dados e aprendizado de máquina. Esses métodos procuram identificar diferenças entre notícias verdadeiras e falsas, utilizando técnicas de análise para alcançar resultados precisos.

A Mineração de Dados é um conjunto de técnicas e processos que são voltados para a busca de padrões, relações e exploração de grandes quantidades de informações em banco de dados ou em páginas web. Ocasionalmente, com o crescimento de informações geradas diariamente na web, a produção dessa grande quantidade de informação causa um problema de organização. Tal problema pode ser solucionado utilizando a técnica de mineração de dados, que visa estruturar e organizar essas informações. Um dos principais objetivos de quem utiliza a técnica de mineração de dados é o de analisar e explorar os dados em busca de tendências, relações entre as informações e com isso prever comportamentos futuros, onde seja possível classificar notícias em verdadeiras ou em *Fake News* (MARTINS, 2003).

A *machine learning* e *deep learning* são subcampos da Inteligência Artificial que se concentram em desenvolver sistemas de computação capazes de aprender sem que sejam explicitamente programadas para isso. Isso é alcançado através de algoritmos que são treinados com grandes quantidades de dados, a fim de identificar padrões e tomar decisões com base em novos *inputs*. *deep learning*, em particular, é uma abordagem avançada de *machine learning* que usa camadas profundas de redes neurais para modelar complexos relacionamentos entre os dados de treinamento. O objetivo é imitar a capacidade humana de aprendizado e desenvolvimento de conhecimento, tornando-o um dos caminhos mais promissores para a realização da inteligência artificial.

O algoritmo CNN Convolutional Neural Network é uma arquitetura de *deep learning* projetada para reconhecer padrões complexos e hierárquicos nos dados por meio de camadas convolucionais, que aplicam filtros para extrair características locais importantes e camadas de *pooling* que reduzem a dimensionalidade dos dados e aumentam a robustez contra variações (ANALYSIS, 2019).

As Redes Neurais, são um tipo de modelo de aprendizado de máquina inspirado na estrutura e funcionamento do cérebro humano, sendo composta por um grande número de unidades de processamento interconectadas, chamados de neurônios artificiais, que trabalham em conjunto para resolver problemas complexos. As redes neurais são capazes de aprender a partir de exemplos e podem ser usadas para tarefas como reconhecimento de padrões, classificação e previsão (ANALYSIS, 2019).

BERT (*Bidirectional Encoder Representations from Transformers*) é uma técnica de pré-treinamento de Programação Neurolinguística (PNL) desenvolvida pelo Google, é profundamente bidirecional porque aprende a representação de texto para ambas as direções para uma melhor compreensão do contexto e do relacionamento. Existem principalmente dois tipos de modelos: BERT Base e BERT Large, bem como alguns modelos baseados em idiomas como inglês, chinês e um modelo multilíngue (mBERT) que cobre 102 idiomas e é treinado na Wikipédia (MISINFORMATION, 2022).

O algoritmo C4.5 é utilizado para criar uma árvore de decisão, foi desenvolvido por Ross Quinlan. É uma extensão do algoritmo anterior de Quinlan's ID3 que visa combater o problema de *overfitting*. As árvores de decisão geradas pelo algoritmo C4.5 podem ser utilizadas para classificação e são portanto conhecidas como classificadores estatísticos, o algoritmo C4.5 escolhe o atributo dos dados que mais efetivamente particiona o seu conjunto de amostras em subconjuntos tendendo a uma categoria ou a outra (MISINFORMATION, 2022).

Os algoritmos SVM (*Support Vector Machine*) e RF (*Random Forest*) são dois algoritmos de aprendizado de máquina, sendo métodos usados em aprendizado de máquina molecular e previsão de propriedades de compostos. O algoritmo SVM é baseado nas propriedades geométricas dos dados e maximiza a margem entre diferentes pontos (SOUZA et al., 2020). Por outro lado, o RF funciona bem com uma mistura de características numéricas e categóricas e pode usar os dados como eles são, sendo intrinsecamente adequado para problemas multi classe, enquanto o SVM é intrinsecamente binário.

LR (Logistic Regression) é um algoritmo de classificação de aprendizagem supervisionada usado para prever a probabilidade de uma variável de destino. Esse tipo de modelo estatístico é frequentemente usado para classificação e análise preditiva. A

regressão logística estima a probabilidade de ocorrência de um evento, com base em um determinado conjunto de dados de variáveis independentes (MISINFORMATION, 2022).

A estrutura dEFEND é baseada na arquitetura HAN que é um modelo de aprendizado de máquina usado para classificação de texto. Ele usa redes neurais recorrentes empilhadas no nível da palavra, seguidas por um modelo de atenção para extrair palavras importantes para o significado da frase e agregar a representação dessas palavras informativas para formar um vetor de frase. Envolve o HAN no conteúdo do texto e um mecanismo de co-atenção entre o conteúdo do texto e os comentários do usuário para classificar a desinformação (MISINFORMATION, 2022).

C-LSTM (*Convolutional Long Short-Term Memory*) é uma arquitetura de rede neural utilizada para tarefas de processamento de linguagem natural, que combina dois componentes principais sendo eles uma camada de convolução e uma camada de LSTM (Long Short-Term Memory), sendo camada de convolução é responsável por extrair recursos locais e detectar padrões em sequências de palavras ou caracteres (MISINFORMATION, 2022).

O MCNNet é um sistema de detecção generalizada de notícias falsas que utiliza uma arquitetura de CNN (*Convolutional Neural Network*) multicanal, essa arquitetura usa diferentes tamanhos de filtros e *kernels* em diferentes redes CNN paralelas. Ela concatena características de canais diferentes em um único vetor e utiliza algumas camadas de *dropout* para fornecer capacidade de generalização na classificação de notícias falsas (MISINFORMATION, 2022).

Cross-SEAN é um modelo de atenção neural semi-supervisionado baseado em *cross-stitch*, esse modelo ajuda a reduzir a dependência dos dados rotulados, uma vez que utiliza dados não rotulados. Esse modelo é utilizado principalmente para tarefas de processamento de linguagem natural, como análise de sentimentos em textos curtos, como por exemplo *tweets* onde é feito o uso de diferentes tipos de informações, incluindo o texto do *tweet*, metadados do usuário e do *tweet*, bem como conhecimento externo, como bases de dados ou recursos adicionais relacionados ao tema dos *tweets* (MISINFORMATION, 2022).

LM-BFGS (*Limited-Memory Broyden-Fletcher-Goldfarb-Shanno*) é um algoritmo de otimização utilizado para problemas de otimização não lineares, é uma variante do algoritmo BFGS, que é um método quase Newton para otimização. O LM-BFGS é especialmente adequado para problemas em que é caro computar a matriz Hessiana completa, sendo uma abordagem que utiliza memória limitada para aproximar essa matriz e realizar atualizações iterativas (COMPUTATIONAL; SECURITY, 2021).

A LSTM (*Long Short-Term Memory*) foi projetada para superar o problema de desvanecimento do gradiente em redes neurais recorrentes tradicionais, permitindo que informações relevantes de longo prazo sejam capturadas e preservadas ao longo do tempo. A LSTM tem sido amplamente utilizada em uma variedade de aplicações, incluindo tradução automática, geração de texto, análise de sentimento, reconhecimento de voz e muitas outras tarefas que envolvem sequências temporais. Sua capacidade de capturar dependências de longo prazo a torna uma escolha popular em muitos cenários de processamento de linguagem natural (COMPUTATIONAL; SECURITY, 2021).

XGBoost é uma biblioteca de código aberto para aprendizado de máquina base-

ado em árvores de decisão, que é amplamente utilizado para problemas de regressão e classificação, o nome XGBoost é uma abreviação de *Extreme Gradient Boosting*, o que indica que é uma técnica que aprimora o algoritmo de *gradient boosting* com algumas melhorias adicionais. O XGBoost é conhecido por sua eficácia e desempenho em uma ampla gama de problemas de aprendizado de máquina. Ele foi desenvolvido para lidar com conjuntos de dados grandes e complexos, com muitas variáveis características e amostras (REVIEW, 2019).

O algoritmo TM, abreviação de "*DeepWalk*", é um algoritmo de aprendizado de representação de grafos utilizado para aprender representações latentes de nós em um grafo não direcionado, o *DeepWalk* é inspirado no conceito de representações distribuídas e utiliza técnicas de aprendizado não supervisionado para capturar informações estruturais do grafo, sendo uma abordagem poderosa para aprender representações de grafos e tem sido amplamente utilizado em diferentes domínios, desde redes sociais até redes de colaboração científica, para extrair informações significativas dos dados de grafos complexos (REVIEW, 2019).

Outro meio que pode ser utilizado para detecção de *fake news* é a análise de sentimentos, como visto no artigo (ANALYSIS, 2022) para identificar a percepção do público sobre a COVID-19. A utilização ou combinação desses métodos citados anteriormente permite a criação de sistemas automatizados capazes de identificar notícias falsas. No entanto, é importante destacar que a detecção de *fake news* ainda é um desafio em constante evolução.

Com base nisso surge a nossa motivação de realizar um mapeamento sistemático na área de detecção automática de *fake news*. Uma revisão pode ajudar a consolidar o conhecimento existente sobre o tema, identificando as principais abordagens e técnicas usadas para a detecção de *fake news*, outro benefício de uma revisão é identificar as lacunas de conhecimento e áreas de tendências. Por fim uma revisão pode fornecer orientações valiosas para ajudar a desenvolver abordagens eficazes e precisas na identificação de desinformação.

### **3. EMBASAMENTO TEÓRICO E REVISÃO DE LITERATURA**

#### **3.1. PROTOCOLO**

##### **3.1.1. ESCOPO E OBJETIVOS**

Neste trabalho realizamos um mapeamento sistemático para entender quais são as metodologias utilizadas na área de aprendizado de máquina. A pesquisa focou nos estudos que tratam da detecção de notícias falsas. Conseguiu-se mapear as técnicas e algoritmos utilizados pelos estudos e relacionar quais deles apresentam uma melhor desempenho dentro do contexto de detecção automática de notícias falsas.

##### **3.1.2. CRITÉRIOS DE INCLUSÃO E EXCLUSÃO**

Uma etapa fundamental durante o planejamento de um mapeamento sistemático é a definição dos Critérios de Inclusão e Critérios de Exclusão. Esses critérios auxiliam na seleção dos artigos apropriados e são utilizados para reduzir o número de artigos retornados pelos mecanismos de busca. Os critérios levantados para esse estudo foram:

Tabela 1 - Bases de Busca de Artigos e Respectivas Adaptações das *Strings* de Busca

Base de Busca
IEEE
SCOPUS
ACM
Science Direct
SPRINGER

### 3.1.3. INCLUSÃO

- **INC1:** Estudos secundários tratando de detecção automática de *fake news* utilizando termos como *Machine Learning*, ou *Deep Learning* ou *Artificial Intelligence*.
- **INC2:** Estudos secundários tratando de detecção automática de *Fake News* utilizando termos como *Web Scraping*, *Web Crawling*, *Web Mining*.

### 3.1.4. EXCLUSÃO

- **EXC1:** Possuir menos de 6 páginas.
- **EXC2:** Estudo publicado antes de 2014.
- **EXC3:** Não ser uma revisão da literatura.
- **EXC4:** Artigos não escritos em inglês.
- **EXC5:** Artigos duplicados.
- **EXC6:** Não estar disponível para download pelo portal de periódicos da CAPES.
- **EXC7:** Não se aplica na busca de estudos secundários tratando de detecção automática de *Fake News* utilizando dos termos definidos.

### 3.1.5. STRING E BASES DE BUSCA

Para a presente pesquisa foram utilizadas bases de dados conhecidas e que possuem um sistema de pesquisa por palavras-chave. Foram utilizadas as bases demonstradas na tabela 1. A tabela 2 ilustra a *string* adotada para examinar as bases de dados.

```
("machine learning"OR "deep learning"OR "natural language processing") AND "fake news"AND ("literature survey"OR "systematic literature review"OR "systematic mapping study"OR "systematic mapping review")
```

Tabela 2 - String de busca

Para definição da *string* foram utilizados os termos presentes na Tabela 2, separados por operadores *"OR"* para selecionar termos e sinônimos alternativos e *"AND"* para selecionar termos para o tipo de estudo requerido. No entanto, a *string* criada não



pôde ser usada como planejado, pois algumas das bases possuem algumas formas específicas de lidar com as *strings* de pesquisa. Nesses casos, foi necessário adotar estratégias específicas para construir uma ou mais *strings* equivalentes à *string* de pesquisa apresentada na tabela 2.

### 3.1.6. QUESTÕES DE PESQUISA

Definimos as seguintes questões de pesquisa:

- **RQ1** - Quais das propostas identificadas são utilizadas na detecção de *fake news* relacionadas a COVID-19?
- **RQ2** - Quais algoritmos/técnicas são utilizadas pelas abordagens. Quais delas apresentam melhores indicadores de acurácia?
- **RQ3** - Quais são as principais ameaças à validade encontradas em estudos de detecção automática de *fake news*?
- **RQ4** - Quais são as principais tendências de pesquisa de detecção de *fake news*?

### 3.1.7. CRITÉRIOS DE QUALIDADE

O objetivo dos critérios de qualidade é avaliar os estudos como forma de medir sua relevância em relação aos outros para a resposta das questões de pesquisa. Esses critérios de avaliação de qualidade podem ser baseados em várias questões. Os critérios de qualidade foram realizados individualmente por dois pesquisadores, também utilizando do uso de uma terceira pessoa caso houvesse discordância entre as notas.

- **CQ01:** Foi abordada questão sobre uma técnica, ferramenta, método, ou processo para *Machine Learning*, *Deep Learning*, *Artificial Intelligence*.
- **CQ02:** Foi abordada questão sobre uma técnica, ferramenta, método ou processo para *Web Scraping*, *Web Crawling*, *Web Mining*.
- **CQ03:** Os métodos de busca usados para identificar estudos relevantes estão claramente descritos.
- **CQ04:** O estudo relata o número de estudos identificados por meio das pesquisas, além dos números excluídos e fornece os motivos apropriados para as exclusões, com base em critérios explícitos de inclusão e exclusão.

A pontuação geral dos critérios de qualidade podem ser vista na Tabela 3, onde foi decidido que a aprovação mínima deveria ser 2.6 a 3.75 (Boa Qualidade), precisando que no mínimo todas os critérios fossem parciais como pode ser visto na Tabela 4.

## 3.2. CONDUÇÃO

Nesta seção, o processo de mapeamento é demonstrado desde a pesquisa nas bases até a seleção de trabalhos para extração e análise dos resultados dos estudos primários, até chegar às respostas das perguntas de pesquisa definidas na Seção 3.

Tabela 3 - Pontuação geral usada para o trabalho.

<b>Começo do intervalo</b>	<b>Final do Intervalo</b>	<b>Descrição</b>
0	1,25	Baixa Qualidade
1,26	2,5	Média Qualidade
2,6	3,75	Boa Qualidade
3,76	5	Ótima Qualidade

Tabela 4 - Questões de qualidade usadas no trabalho.

<b>ID</b>	<b>Regras</b>	<b>Peso</b>	<b>Mínimo para Aprovar</b>
CQ01	N	1,25	P
	P		
	S		
CQ02	N	1,25	P
	P		
	S		
CQ03	N	1,25	P
	P		
	S		
CQ04	N	1,25	P
	P		
	S		

Além de todo o passo-a-passo mostrado na Figura 2, os critérios de avaliação da qualidade foram aplicados nesta revisão com o objetivo de avaliar a confiabilidade dos estudos primários. A Tabela 5 fornece informações sobre os estudos de qualidade incluídos no mapeamento sistemático. Cada estudo pode ser identificado através de sua referência, assim como a pontuação geral atingido por aquele artigo e seu status.

Tabela 5 - Avaliação de qualidade

<b>Referência</b>	<b>CQ01</b>	<b>CQ02</b>	<b>CQ03</b>	<b>CQ04</b>	<b>Pontuação Geral</b>
(MISINFORMATION, 2022)	S	P	S	S	Ótima Qualidade
(MEDIA, 2022)	S	P	S	S	Ótima Qualidade
(ANALYSIS, 2022)	S	P	S	S	Ótima Qualidade
(MODELING, 2021)	P	P	S	P	Boa Qualidade
(SYNTAX, 2021)	S	P	S	S	Ótima Qualidade
(COMPUTATIONAL; SECURITY, 2021)	S	P	S	S	Ótima Qualidade
(ASSESSMENT, 2020)	S	P	S	S	Ótima Qualidade
(REVIEW, 2019)	S	P	S	S	Ótima Qualidade
(SILVA; ALVES; GARCIA, 2019)	S	P	S	P	Boa Qualidade
(ANALYSIS, 2019)	S	P	S	S	Ótima Qualidade
(ALKHAMEES et al., 2021)	S	P	S	S	Ótima Qualidade
(SOUZA et al., 2020)	S	P	S	S	Ótima Qualidade
(T; MATHEW, 2022)	P	P	S	S	Boa Qualidade

Todos os artefatos de estudo gerados durante a revisão foram devidamente guardados no *figshare*<sup>2</sup>, vale ressaltar também o uso da ferramenta Toth<sup>3</sup> para conduzir

<sup>2</sup><[https://figshare.com/articles/online\\_resource/Detec\\_o\\_Autom\\_tica\\_de\\_Fake\\_News/26181050](https://figshare.com/articles/online_resource/Detec_o_Autom_tica_de_Fake_News/26181050)>

<sup>3</sup><http://200.132.136.13/Thoth>

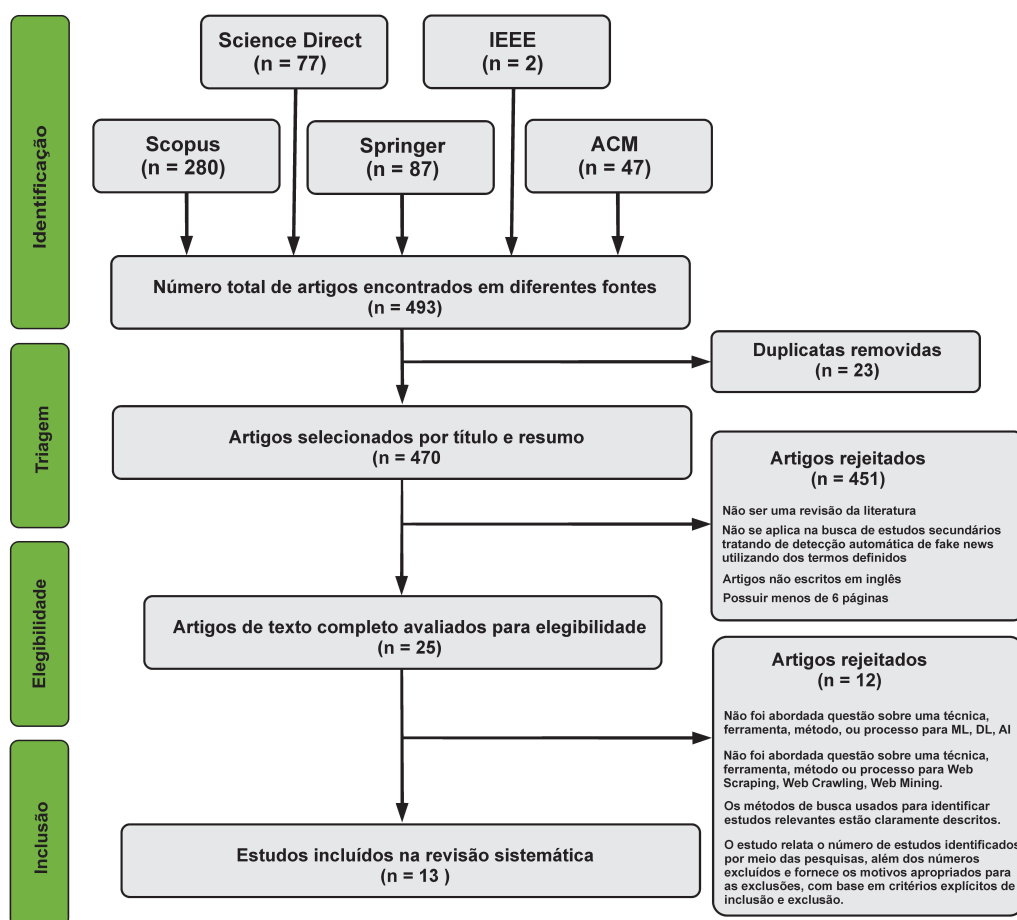


Figura 2 - Fluxograma das atividades para a seleção e avaliação sistemática dos artigos.

a revisão, aumentando o desempenho na hora de realizar a exportação dos materiais selecionados e na condução de todo o passo-a-passo.

### 3.2.1. ANÁLISE DOS RESULTADOS

Nesta seção são respondidas as questões de pesquisa levantadas para o estudo.

#### **RQ01: Quais das propostas identificadas são utilizadas na detecção de *fake news* relacionadas a COVID-19?**

A partir da análise realizada entre as 13 publicações, foi revelado que apenas 23,1% se concentram na identificação de *fake news* sobre o COVID-19. Nesse grupo, há temas como a percepção pública sobre a vacinação e a desinformação sobre o vírus. A maioria dos estudos, por outro lado, aborda questões como combate a boatos maliciosos, análise de conteúdo falso, imagens e vídeos falsos e combate à desinformação nas redes sociais.

Concluindo essa questão, podemos notar que apenas uma pequena parcela dos estudos finais aprovados tem foco em COVID-19, o que já era esperado pois a *string* utilizada para buscar os trabalhos nas bases foi realizada de maneira genérica para que pudessemos identificar de maneira geral quais era os algoritmos utilizados em

diversas abordagens.

**RQ02: Quais algoritmos/técnicas são utilizadas pelas abordagens? Quais delas apresentam melhores indicadores de acurácia?**

Para responder esta questão foi necessário fazer o mapeamento proposto na Tabela 6. Ela demonstra os algoritmos e a acurácia obtida por cada um: os resultados de acurácia são valores que vão de 0 a 100%. Como existem trabalhos que possuem mais de um resultado de acurácia, são apresentados apenas os algoritmos que obtiveram o melhor resultado. Fazendo a comparação dos algoritmos na tabela, percebe-se que a Rede neural convolucional *CNN* obteve a maior média entre todos, seguido de *NN* utilizando da abordagem *TF-IDF* e do *BERT*.

Também nota-se que as variações de acurácia podem ocorrer pelo fato da quantidade de dados fornecidos para o treinamento do algoritmo ou da base de dados não possuir uma boa integridade. Outro motivo da possível variação de acurácia entre os artigos pode se dar pela escolha da técnica de processamento de dados utilizada, onde se pode destacar as mais utilizadas como o *TF-IDF* e *BoW*.

A razão disso acontecer é porque o *BoW* cria uma matriz de contagem de palavras, onde cada documento é representado por um vetor com a contagem de cada palavra, sem considerar a importância relativa das palavras em diferentes documentos, isso pode resultar na influência de palavras muito frequentes, como artigos e preposições, que podem não ser úteis para distinguir entre diferentes classes de documentos. Já o *TF-IDF* por sua vez além de contar as palavras, pondera cada termo com base em sua frequência no documento e sua frequência inversa em todos os documentos, conferindo mais peso a termos que são mais exclusivos para um documento específico e menos peso a termos comuns, essa ponderação ajuda a destacar palavras mais relevantes e distintivas, o que pode levar a uma melhoria na acurácia do modelo.

**RQ03: Quais são as principais ameaças à validade encontradas em estudos de detecção automática de *fake news*?**

Assim como qualquer trabalho que utiliza o mapeamento sistemático, existem falhas provenientes de vícios de análise. Por este motivo, nesta pergunta, foi proposta a realização de uma busca de ameaças à validade apresentadas nos trabalhos selecionados. Ou seja, buscou-se abstrair as possíveis limitações encontradas em trabalhos de detecção automática de *fake news*.

Ao realizar a análise nos estudos, pode-se perceber ameaças como a falta de dados confiáveis e equilibrados para treinar os modelos de aprendizado de máquina, o uso inadequado ou não apropriado de técnicas de processamento de linguagem natural, a falta de transparência e interpretabilidade dos modelos de aprendizado de máquina, a falta de atualização dos modelos para refletir mudanças na linguagem e no contexto da mídia social e a dependência da qualidade dos dados de entrada para a precisão dos resultados da detecção de *fake news* e falta de consenso quanto ao que constitui uma *fake news*, que também foi levantado como uma ameaça à validade que impede a replicação do estudo em ambiente controlado.

Também pode-se observar que os estudos encontrados apresentam a subjetividade na avaliação dos resultados como ameaça importante. Uma vez que julgamento humano pode ser influenciado por preconceitos ou opiniões prévias, e que isso pode resultar em uma avaliação enviesada dos resultados, estudos enviesados por esta

Tabela 6 - Acurácia dos algoritmos.

Algoritmos	Acurácia	ID do artigo
CNN	99,99%	(MISINFORMATION, 2022)
NN	99,68%	(MISINFORMATION, 2022)
BERT	99,20%	(MISINFORMATION, 2022)
BERT-large	98,13%	(MISINFORMATION, 2022)
BERT-base (DNN)	87,5%	(MISINFORMATION, 2022)
SVM/Gaussian (ICA)	81,2%	(MISINFORMATION, 2022)
C4.5	95,11%	(MISINFORMATION, 2022)
SVM+RF	97,8%	(MISINFORMATION, 2022)
LR(Multi Class)	96,30%	(MISINFORMATION, 2022)
dEFEND w.news (Binary)	98%	(MISINFORMATION, 2022)
Ro-CT-BERT	99,02%	(MISINFORMATION, 2022)
C-LSTM	98,62%	(MISINFORMATION, 2022)
MCNNNet	98,20%	(MISINFORMATION, 2022)
Cross-SEAN	95,40%	(MISINFORMATION, 2022)
SVM	96,77%	(REVIEW, 2019)
SVM+RF (Ensemble-model)	97,8%	(MISINFORMATION, 2022)
HBL	99,40%	(COMPUTATIONAL; SECURITY, 2021)
LM-BFGS	99,36%	(COMPUTATIONAL; SECURITY, 2021)
LSTM	95,30%	(COMPUTATIONAL; SECURITY, 2021)
XGBoost	91,24%	(REVIEW, 2019)
TM (Deepwalk)	91,24%	(REVIEW, 2019)
TM (LSTM-RNN)	91,24%	(REVIEW, 2019)
TM (LINE)	91,24%	(REVIEW, 2019)

ameaça podem levar a conclusões incorretas sobre a efetividade dos sistemas de detecção de *fake news*. Portanto, é importante que trabalhos futuros na área considerem métodos objetivos e rigorosos de avaliação dos resultados, tais como métricas de avaliação de precisão para garantir a sua validade.

#### **RQ04: Quais são as principais tendências de pesquisa de detecção de *Fake News*?**

Ao analisar os estudos, pode-se observar oportunidades para futuras pesquisas nas áreas de *Big Data* envolvendo *fake news* e maior aprofundamento nas técnicas de aprendizado de máquina como classificação, aprendizado profundo, análise de sentimentos, mineração de texto e modelagem de tópicos.

As pesquisas (ANALYSIS, 2022) e (MISINFORMATION, 2022) também têm explorado o papel da inteligência artificial na análise de notícias falsas relacionadas ao COVID-19. No caso da mineração, os artigos identificados pelos autores apontam que os dados são coletados de mídias sociais, portanto não estruturados. Os autores também destacam as principais abordagens e algoritmos utilizados na área.

## **4. PROPOSTA E DESENVOLVIMENTO**

No início do desenvolvimento do suporte ferramental para classificar se uma determinada notícia é verdadeira ou falsa, será realizada a escolha do algoritmo que

melhor se adapte às nossas necessidades. Para complementar, será utilizada uma base de dados que contenha informações falsas para o treinamento desse algoritmo, assim melhorando sua acurácia e precisão na detecção de notícias falsas.

Por fim, será criada uma *API* cuja funcionalidade é verificar a veracidade de uma notícia fornecida a ele. Essa *API* será implementada como uma funcionalidade em um *chatbot* disponível atualmente para qualquer visitante do site da AVICO onde será possível verificar se uma determinada notícia é verdadeira ou falsa. Tudo sendo comunicado via *API*, ajudando assim aqueles que desejam realizar verificações de *fake news*. Cabe ressaltar que o desenvolvimento dessa ferramenta não anula completamente a necessidade de uma avaliação humana e serve apenas como apoio, pois a detecção de *fake news* é um desafio em constante evolução e nenhum sistema é totalmente infalível.

#### **4.1. REQUISITOS**

Os seguintes requisitos são necessários para o desenvolvimento dessa ferramenta:

- O suporte ferramental deve ser capaz de importar e utilizar conjuntos de dados rotulados para treinamento do modelo.
- O suporte ferramental deve ser capaz de classificar automaticamente as notícias como verdadeiras ou falsas com base nas características extraídas.
- A interface deve incluir funcionalidades para inserir notícias para análise e visualizar os resultados de detecção de *fake news* de forma clara e compreensível.
- O usuário deve acessar o suporte ferramental por uma interface intuitiva que permita interagir com o sistema facilmente.
- O suporte ferramental deve classificar automaticamente as notícias como verdadeiras ou falsas com base nas características extraídas.

#### **4.2. IMPLEMENTAÇÃO E DEMONSTRAÇÃO CONCEITUAL**

##### **4.2.1. CONFIGURAÇÃO DO AMBIENTE**

Inicialmente, como ambiente para testar o algoritmo *SVM*, foi utilizado o Google Colab. No entanto, uma das desvantagens dessa ferramenta é a limitação de recursos computacionais, que pode resultar em erros de estouro de memória ao utilizar uma grande base de dados para o treinamento do algoritmo.

Por isso, optou-se pela utilização da configuração de um ambiente local, onde foi utilizado a linguagem *Python 3* e o editor de código *Visual Studio Code*, em um computador com processador Intel Core i7-11800H @ 2.30GHz e 16GB de memória RAM.

##### **4.2.2. ESCOLHA DA BASE DADOS**

Para o treinamento do algoritmo, foi utilizada a base de dados *Fake.br*, uma das poucas bases de dados em português disponíveis. O conjunto de dados total que

ela possui é de 7.200 notícias, contendo 3.600 notícias verdadeiras e 3.600 notícias falsas. As notícias são divididas em seis categorias: sendo 4.180 sobre “política”, 1.544 em “TV e celebridades”, 1.276 em “sociedade e notícias diárias”, 112 sobre “ciência e tecnologia”, 44 sobre “economia” e 44 notícias sobre “religião”. Todas elas são do período de janeiro de 2016 a janeiro de 2018, com formatação de texto livre. A disponibilização de um grande número de notícias, artigos e dados anotados na base de dados *Fake.br* foi impactante para a decisão de usá-la em vez de criar um novo banco de dados. Além disso, o conjunto de dados *Fake.br* é diversificado, contendo uma variedade de artigos de notícias falsas, cobrindo diferentes tópicos e provenientes de diferentes fontes. Essa diversidade é essencial para garantir que o algoritmo seja testado em uma variedade de tipos de notícias falsas e seja robusto o suficiente para identificar notícias falsas em diferentes contextos. A escolha de não utilizar uma base de dados criada manualmente foi porque o desenvolvimento de uma nova base de dados exigiria, além de coletar um grande número de notícias, a anotação manual de cada notícia como verdadeira ou falsa. Esta tarefa, portanto, exigiria uma avaliação humana para garantir sua validação para ser utilizada no treinamento do algoritmo, o que levaria muito tempo para projetar, desenvolver e avaliar esta nova base.

#### 4.2.3. IMPLEMENTAÇÃO DO ALGORITMO

Algoritmo	SVM
Iteração 1	96.39%
Iteração 2	96.39%
Iteração 3	95.14%
Iteração 4	96.25%
Iteração 5	95.83%
Iteração 6	96.81%
Iteração 7	97.64%
Iteração 8	96.67%
Iteração 9	96.11%
Iteração 10	96.53%
<b>Média</b>	<b>96.37%</b>

Figura 3 - BoW Cross Validation Resultado

Algoritmo	SVM
Iteração 1	97.36%
Iteração 2	96.94%
Iteração 3	95.28%
Iteração 4	97.64%
Iteração 5	96.81%
Iteração 6	96.39%
Iteração 7	96.53%
Iteração 8	95.56%
Iteração 9	95.69%
Iteração 10	96.81%
<b>Média</b>	<b>96.50%</b>

Figura 4 - TF-IDF Cross Validation Resultado

Acurácia	Precisão		Recall		Medida-F	
	Verdadeiro	Falso	Verdadeiro	Falso	Verdadeiro	Falso
96,37%	97,42%	95,38%	95,28%	97,47%	96,33%	96,41%

Tabela 7 - Resultado utilizando o BoW

Acurácia	Precisão		Recall		Medida-F	
	Verdadeiro	Falso	Verdadeiro	Falso	Verdadeiro	Falso
96,50%	96,89%	96,12%	96,08%	96,92%	96,49%	96,51%

Tabela 8 - Resultado utilizando o TF-IDF

A não escolha do algoritmo BERT como candidato se deu devido ao problema de *overfitting*, que ocorre quando um modelo se ajusta muito bem aos dados de treina-

mento, porém não generaliza bem para novos e diferentes conjuntos de dados. Em outras palavras, o modelo aprende padrões específicos dos dados de treinamento que podem não ser relevantes ou aplicáveis a outros conjuntos de dados.

Após a conclusão dessa etapa de verificação do SVM, foi desenvolvida uma *API* para permitir que o suporte ferramental recebesse requisições. Para isso, utilizou-se a biblioteca *pickle* do Python para salvar o algoritmo e reutilizá-lo sem a necessidade de treinamento novamente. Além disso, o *framework Flask* simplificou a criação dessa *API*, onde uma rota recebe uma requisição *POST* contendo um texto a ser classificado pelo algoritmo, fornecendo como retorno "Verdadeira" ou "Falsa". Como ambiente de implantação para essa *API*, optou-se pelo site Vercel, pois oferece uma maneira fácil de realizar a implantação, além de um plano gratuito para hospedagem da mesma.

A utilização do *DialogFlow* foi escolhida, pois no site da AVICO já existe um chatbot atualmente que os usuários podem consultar para tirar algumas dúvidas que, antigamente, eram respondidas pela presidente. Para a interação com o chatbot, o usuário pode clicar no símbolo do chatbot no qual irá responder com as possibilidades de escolha de como classificar a notícia. Pode-se escolher a classificação onde o usuário informa uma entrada em formato de texto ou um link.

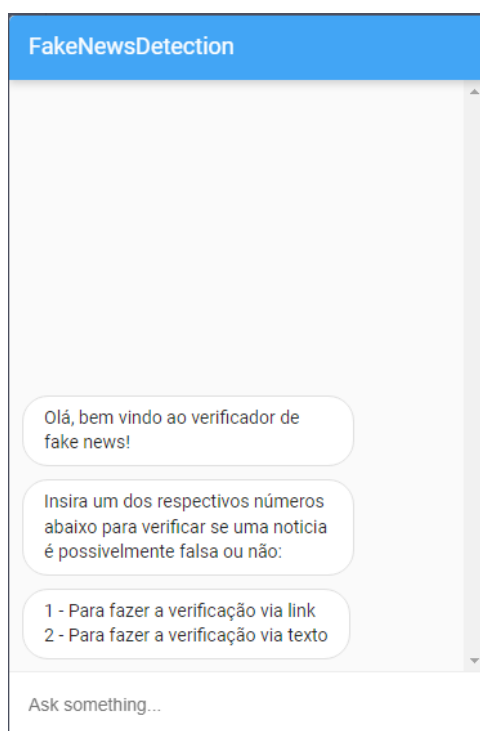


Figura 5 - Interação inicial do chatbot

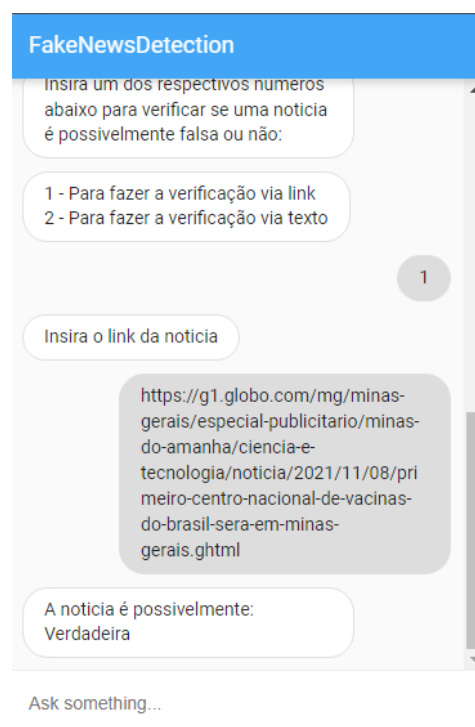


Figura 6 - Resposta retornada pelo chatbot

Caso o usuário informe a entrada por link, o chatbot irá se comunicar enviando esse *input* para uma *API* que chamará um recurso de mineração, o qual irá extrair o texto do link. Após isso, será realizado um pré-processamento no texto, no qual serão retiradas as *stopwords*, que são palavras muito comuns em um idioma e geralmente são removidas durante o pré-processamento de textos em PLN (Processamento de Linguagem Natural), devido ao seu alto índice de ocorrência e baixo valor informativo. Após isso, ele será formatado para um modelo de dados que o algoritmo de classifi-



cação consiga ler e prever se a notícia é verdadeira ou não. O retorno será fornecido conforme demonstrado na imagem 6.

Caso o usuário opte por um texto, o processo será mais simples, pois não será necessário passar por uma extração. Somente será realizado um pré-processamento e classificação do conteúdo informado. Tendo potencial para verificar textos divulgados via *WhatsApp* ou outro meio de comunicação. Na imagem 7 podemos ver como o fluxo de informação é passado entre o algoritmo de classificação e o de mineração.

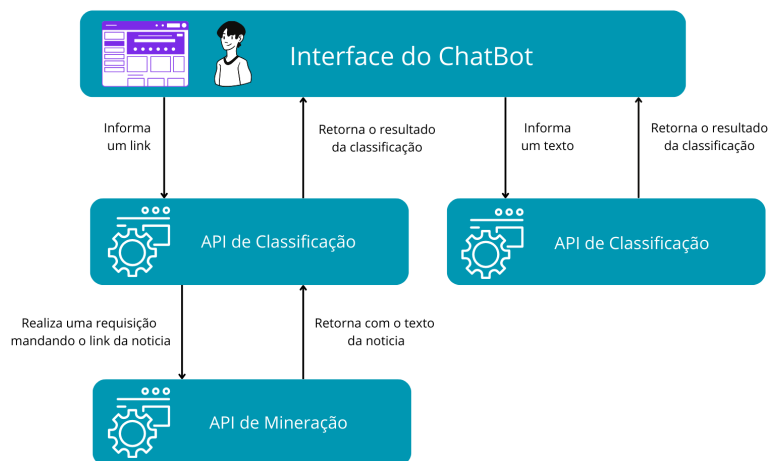


Figura 7 - Fluxo de requisições

## 5. AVALIAÇÃO E LIMITAÇÕES

Ameaças à validade podem limitar a capacidade de descrever ou interpretar resultados dos dados obtidos. Portanto, não há como desconsiderar as seguintes ameaças encontradas neste estudo:

### 5.1. VALIDADE INTERNA

#### 5.1.1. EXTRAÇÃO DE DADOS

Os pesquisadores foram responsáveis por extrair e classificar os estudos, viés ou extração de dados que poderiam ameaçar a validade da caracterização dos dados.

#### 5.1.2. VIÉS DE SELEÇÃO

Alguns artigos podem ter sido categorizados incorretamente, pois foram incluídos ou excluídos no mapeamento sistemático de acordo com o julgamento dos pesquisadores. Para diminuir essas ameaças, foram feitas análises de seleção e extração pelos pesquisadores envolvidos, com uma votação final dos professores sobre divergências.

#### 5.1.3. VIÉS DE CLASSIFICAÇÃO

Alguns artigos selecionados não deixaram clara a metodologia em detalhes, ou seja, a estratégia de pesquisa, avaliação ou validação. Para diminuir esse viés, esses

artigos foram lidos completamente pelos pesquisadores, a fim de encontrar características que se encaixassem em um tipo de pesquisa.

## **5.2. VALIDADE EXTERNA**

Embora a pesquisa tenha sido realizada nas principais bases digitais, é impossível dizer que os resultados desse mapeamento sistemático cobriram todos os trabalhos sobre o assunto. No entanto, este estudo apresentou evidências das principais técnicas utilizadas e lacunas a serem exploradas, servindo como um guia para trabalhos futuros nesta linha.

## **6. DISCUSSÕES**

Esta seção discute as possibilidades e lacunas percebidas durante a revisão sobre o uso do aprendizado de máquina na detecção de desinformação e notícias falsas, bem como na análise das percepções públicas e sentimentos em relação à pandemia da COVID-19 e à vacinação, oferecendo uma visão abrangente da área.

A primeira possível direção seria o mapeamento sistemático da literatura sobre a detecção de notícias falsas no meio online, incluindo as abordagens de aprendizado de máquina mais eficazes e as ferramentas mais utilizadas. Os estudos poderiam investigar a eficácia de novos algoritmos de aprendizado de máquina e o uso de métodos híbridos que combinam várias técnicas para melhorar a precisão na detecção de notícias falsas.

Outra possibilidade seria a análise da percepção pública sobre a vacinação COVID-19 como foi realizado no estudo (ANALYSIS, 2022), mas com base no processamento de texto e na detecção de sentimento em plataformas de mídia social. Os pesquisadores poderiam explorar as emoções predominantes em relação à vacinação em diferentes países e culturas, a fim de identificar as preocupações mais comuns e melhorar as estratégias de comunicação para aumentar a aceitação da vacina.

Outro tema interessante seria como a detecção de desinformação e notícias falsas pode ser usada para proteger a saúde pública durante crises globais de saúde, como foi o caso da pandemia do COVID-19. Os pesquisadores poderiam explorar o uso de algoritmos de aprendizado de máquina para identificar rapidamente informações enganosas ou imprecisas e fornecer informações verdadeiras e atualizadas para o público em geral.

Além disso, técnicas avançadas de aprendizado de máquina, como redes neurais profundas, podem ser utilizadas para melhorar a precisão da detecção de desinformação, em comparação com abordagens baseadas em métodos de aprendizado de máquina mais tradicionais.

A análise de sentimento e o processamento de linguagem natural também podem ser usados para detectar desinformação em plataformas de mídia social. Os estudos revisados (ANALYSIS, 2022) e (MODELING, 2021) sugerem que a análise de sentimentos pode ser útil na detecção de notícias falsas em plataformas de mídia social.

Outra possibilidade seria a aplicação de métodos de aprendizado de máquina para avaliação da credibilidade de fontes de informação. Alguns estudos sugerem que é possível usar técnicas de aprendizado de máquina para avaliar a credibilidade das fontes de informação assim ajudando a identificar notícias falsas.

Por fim existe uma necessidade de pesquisas futuras sobre a eficácia de diferentes abordagens para combater a disseminação de notícias falsas. Embora a maioria dos estudos revisados tenha se concentrado na detecção de desinformação é crucial examinar a eficácia de diferentes abordagens de combate à desinformação.

Em resumo, os estudos revisados neste texto destacam o potencial do aprendizado de máquina para ajudar a combater a desinformação e aumentar a conscientização pública sobre a saúde pública durante as crises globais de saúde. As direções e estratégias de pesquisa mencionadas acima podem fornecer ideias valiosas para pesquisas futuras nesta área. Estas são apenas algumas possíveis direções de pesquisa que emergem dos artigos revisados, existem muitas outras áreas de pesquisa relacionadas à detecção de desinformação e ao uso de aprendizado de máquina para combater a disseminação de notícias falsas, e o campo continua a evoluir à medida que novas técnicas e abordagens são desenvolvidas.

## 7. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

A disseminação de informações falsas e enganosas tem se tornado cada vez mais comum na era da internet e das redes sociais, gerando impactos negativos na sociedade e na política. A detecção automática de *fake news* é uma área de pesquisa em ascensão, com o objetivo de desenvolver algoritmos e técnicas para identificar e combater a disseminação de informações falsas na internet. No entanto, apesar do grande interesse e investimento nessa área, ainda há uma lacuna na compreensão do estado da arte e das tendências de pesquisa em detecção automática de *fake news*. É nesse contexto que surge a necessidade de um estudo terciário para o mapeamento sistemático da literatura, que visa sintetizar, analisar e avaliar criticamente a literatura existente sobre detecção automática de *fake news*. Esse tipo de estudo é essencial para fornecer *insights* e orientações para pesquisadores, desenvolvedores e tomadores de decisão, e para promover o avanço da área de detecção automática de *fake news* de forma consistente e baseada em evidências.

A pesquisa apresenta um mapeamento sistemático das investigações existentes sobre a classificação de *fake news*, destacando abordagens e métodos utilizados no pré-processamento, extração de recursos e desempenho da classificação. Identificaram-se lacunas significativas, como escassez de conjuntos de dados adequados e precisão variável dos modelos de classificação, que requerem investigação adicional. O artigo visa fornecer *insights* sobre a detecção de *fake news*, orientando futuras pesquisas para melhorar estratégias de combate à disseminação de desinformação na web. Observou-se também que métodos de *machine learning* tradicionais demonstraram bom desempenho geral, apesar de limitações ocasionais, enquanto abordagens baseadas em *deep learning* mostraram eficácia particular na classificação de desinformação relacionada ao COVID-19.

A detecção de *fake news* é uma área emergente com diferentes subproblemas (detecção de rumor, classificação de veracidade, extração de reivindicações, assim por diante), e apenas conjuntos de dados de acesso aberto maiores e mais amplos estão disponíveis. Esses conjuntos de dados são muito distintos no número e tipo de informação (recursos artesanais, conteúdo da mensagem, imagem ou informações do usuário), curadoria (revisada por especialistas ou sem informações de destino), tópicos gerais de domínio (ou tópicos muito específicos) ou classes de saldo (contendo um número comparativamente baixo de instâncias). Portanto, uma comparação justa

entre as abordagens é desafiadora.

Vários conjuntos de dados disponíveis foram identificados por este estudo. A maioria dos conjuntos de dados disponíveis foi criada nos últimos anos e eles ainda são pouco utilizados. Identificou-se que a maioria dos trabalhos avalia a proposta usando dados indisponíveis para pesquisas adicionais.

Ao comparar os resultados dos artigos, diferenças metodológicas foram notadas, mesmo em estudos que usavam o mesmo conjunto de dados. Por exemplo, ao comparar o uso do conjunto de dados para o problema de classificação, alguns autores realizaram experimentos no conjunto de dados completos, enquanto outros usaram subconjuntos diferentes. Além disso, poucos trabalhos apresentaram tabelas com resultados detalhados, e a escolha de autores para diferentes métricas de avaliação torna a comparação entre resultados inconclusiva.

A falta de avaliação comparativa ainda é um problema em aberto nessa área, e os conjuntos de dados devem ser estruturados para cobrir diferentes subproblemas e instâncias com uma ampla gama de complexidade. É necessária uma análise estatística mais aprofundada dos resultados para entender melhor os recursos relevantes na detecção de informações erradas e como os métodos propostos se comportam em instâncias mais simples e complexas.

A implementação de uma *API* para verificação de notícias, integrada a um *chatbot* acessível pelo site da AVICO, oferece uma ferramenta prática e acessível ao público interessado em validar a veracidade de informações fornecidas a ela. Esta abordagem não substitui a avaliação humana, mas serve como suporte valioso em um campo complexo e em constante evolução.

Para futuros trabalhos, é recomendado explorar técnicas avançadas de processamento de linguagem natural, visando lidar com linguagem informal e contextos variados de informação, o que pode ampliar o alcance e a utilidade do sistema desenvolvido. Além disso, a criação e validação de uma base de dados manualmente curada poderia ser explorada para determinar se ela melhora a acurácia alcançada pelo modelo treinado com a base original, explorando tanto a qualidade dos dados quanto os benefícios potenciais para a detecção de *fake news*.

## REFERÊNCIAS

ALKHAMEES, M. et al. User trustworthiness in online social networks: A systematic review. **Applied Soft Computing**, 2021, v. 103, p. 107159, 2021. ISSN 1568-4946. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S156849462100082X>>.

ANALYSIS, C.-. vaccine; literature review; machine learning; public perception; sentiment. Role of artificial intelligence for analysis of covid-19 vaccination-related tweets: Opportunities, challenges, and future trends. **Mathematics**, 2022, v. 10, 2022. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85137805826&doi=10.3390/2fmath10173199&partnerID=40&md5=b2a1e5359eccc3820011941ce499aa>>.

ANALYSIS, D. learning; Rumor detection; Systematic review; T. Deep learning-based rumor detection on microblogging platforms: A systematic review. **IEEE Access**, 2019,

v. 7, p. 152788–152812, 2019. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85077808229&doi=10.1109%2fACCESS.2019.2947855&partnerID=40&md5=3b6ff7c2d25ac9a6c0a40d8765cb5678>>.

ASSESSMENT, C. D. quality; Fake news; Online data; Social media; V. Veracity assessment of online data. **Decision Support Systems**, 2020, v. 129, 2020. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85076227196&doi=10.1016%2fj.dss.2019.113132&partnerID=40&md5=46ae94465b8accf8997e7f4504ce4c0c>>.

COMPUTATIONAL; SECURITY, A. I. fake news; machine learning; public. Intelligent fake news detection: A systematic mapping. **Journal of Applied Security Research**, 2021, v. 16, p. 168–189, 2021. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085047943&doi=10.1080%2f19361610.2020.1761224&partnerID=40&md5=84ad8f9a2550cf1605c3cf8ed9b962f5>>.

COSTA, A. V. G. et al. Classificador de fake news utilizando um modelo de aprendizado de máquina com técnicas de processamento de linguagem natural. Universidade Federal Rural do Semi-Árido, 2020.

MARTINS, C. A. **Uma abordagem para pré-processamento de dados textuais em algoritmos de aprendizado**. Tese (Doutorado) — Universidade de São Paulo, 2003.

MEDIA deep learning; ensemble; fake news detection; machine learning; meta-analysis; misinformation; social. A systematic literature review and meta-analysis of studies on online fake news detection. **Information (Switzerland)**, 2022, v. 13, 2022. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85141756564&doi=10.3390%2finfo13110527&partnerID=40&md5=e08949f56bff3231ae462895fba13800>>.

MISINFORMATION, C. C.-. D. learning; Machine learning;. Applications of machine learning for covid-19 misinformation: a systematic review. **Social Network Analysis and Mining**, 2022, v. 12, 2022. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85135183706&doi=10.1007%2fs13278-022-00921-9&partnerID=40&md5=1f98c943b3ecc560d967dd3caf20da49>>.

MODELING, L. review; Marketing; Sentiment analysis; Services management; Social media; Text mining; T. Applications of text mining in services management: A systematic literature review. **International Journal of Information Management Data Insights**, 2021, v. 1, 2021. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85125385866&doi=10.1016%2fj.ijime.2021.100008&partnerID=40&md5=5e496b8d67ccc0af2a1fe89ad07aca14>>.

REVIEW, D. making; Deep learning; False information; Machine learning; Rumor detection; Social media; S. False information detection in online content and its role in decision making: a systematic literature review. **Social Network Analysis and Mining**, 2019, v. 9, 2019. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073055142&doi=10.1007%2fs13278-019-0595-5&partnerID=40&md5=eb2cee64486d9e601fd91617c1bd018c>>.

SILVA, F. da; ALVES, R. da C.; GARCIA, A. Can machines learn to detect fake news?' a survey focused on social media. In: . 2019: [s.n.], 2019. v. 2019-January,

p. 2763–2770. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083954801&partnerID=40&md5=4c5d420107a23ab01be5924bc76d7260>>.

SOUZA, J. b. V. de et al. A systematic mapping on automatic classification of fake news in social media. **Social Network Analysis and Mining**, 2020, v. 10, p. 48, 2020. ISSN 1869-5450, 1869-5469. Disponível em: <<https://link.springer.com/10.1007/s13278-020-00659-2>>.

SYNTAX, A. D. tools; Fake news; Linguistics; Machine learning; Semantics; Social media;. Approaches to identify fake news: A systematic literature review. **Lecture Notes in Networks and Systems**, 2021, v. 136, p. 13–22, 2021. Disponível em: <[https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087556918&doi=10.1007%2f978-3-030-49264-9\\_2&partnerID=40&md5=05e066d15e29fd28fea3499aa438e13a](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087556918&doi=10.1007%2f978-3-030-49264-9_2&partnerID=40&md5=05e066d15e29fd28fea3499aa438e13a)>.

T, S. M.; MATHEW, S. K. The disaster of misinformation: a review of research in social media. **International Journal of Data Science and Analytics**, 2022, v. 13, p. 271–285, 2022. ISSN 2364-415X, 2364-4168. Disponível em: <<https://link.springer.com/10.1007/s41060-022-00311-6>>.