



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS RUSSAS**  
**CURSO DE GRADUAÇÃO EM ENGENHARIA DE SOFTWARE**

**PEDRO ÍTALO CAMPOS DA SILVA**

**ANÁLISE DE SENTIMENTO EM REDES SOCIAIS: UMA REVISÃO SISTEMÁTICA  
DA LITERATURA**

**RUSSAS**

**2023**

PEDRO ÍTALO CAMPOS DA SILVA

ANÁLISE DE SENTIMENTO EM REDES SOCIAIS: UMA REVISÃO SISTEMÁTICA DA  
LITERATURA

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Engenharia de Software  
do Campus Russas da Universidade Federal do  
Ceará, como requisito parcial à obtenção do  
grau de bacharel em Engenharia de Software.

Orientador: Prof. Dr. Marcos Vinicius  
de Andrade Lima

RUSSAS

2023

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

S582t Silva, Pedro Ítalo Campos da.  
Técnicas utilizadas em análise de sentimento em redes sociais : uma revisão sistemática da literatura / Pedro Ítalo Campos da Silva. – 2023.  
50 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Russas, Curso de Engenharia de Software, Russas, 2023.  
Orientação: Prof. Dr. Marcos Vinicius de Andrade Lima.

1. Análise de sentimento. 2. Ferramentas. 3. Rede social. I. Título.

CDD 005.1

---

PEDRO ÍTALO CAMPOS DA SILVA

ANÁLISE DE SENTIMENTO EM REDES SOCIAIS: UMA REVISÃO SISTEMÁTICA DA  
LITERATURA

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Engenharia de Software  
do Campus Russas da Universidade Federal do  
Ceará, como requisito parcial à obtenção do  
grau de bacharel em Engenharia de Software.

Aprovada em: 05/12/2023

BANCA EXAMINADORA

---

Prof. Dr. Marcos Vinicius de Andrade  
Lima (Orientador)  
Universidade Federal do Ceará (UFC)

---

Profa. Dra. Rosineide Fernando da Paz  
Universidade Federal do Ceará (UFC)

---

Prof. Ms. Francisco Luciano Castro Martins Júnior  
Instituto Federal de Educação, Ciência e Tecnologia  
do Ceará (IFCE)

Às minhas duas famílias que me apoiaram tanto nesses últimos cinco anos de faculdade. O apoio de vocês foi fundamento para suportar cada momento que pensei em desistir e quis voltar para casa. Pai, obrigado por investir em mim.

## **AGRADECIMENTOS**

Ao amor da minha vida e minha futura noiva Ana Livia Nascimento Ramos por ter me apoiado muito enquanto escrevia minha monografia.

Aos meus pais, irmãos e sobrinhos, que nos momentos de minha ausência dedicados ao estudo superior, sempre fizeram entender que o futuro é feito a partir da constante dedicação no presente.

Ao Prof. Dr. Marcos Vinicus de Andrade Lima por me orientar na construção de minha monografia.

Ao meu amigo do curso de engenharia de software Bruno Holanda pelos últimos cinco de amizades que irei levar para fora da faculdade.

Agradeço a todos os professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender.

“Toda ação humana, quer se torne positiva ou negativa, precisa depender de motivação”

(Dalai Lama )

## RESUMO

Este trabalho explora a importância da análise de sentimentos em redes sociais, considerando o papel fundamental que essas plataformas desempenham na comunicação e no compartilhamento de informações. Dada a vasta quantidade de dados gerados diariamente, torna-se essencial compreender as opiniões e emoções expressas pelos usuários. O objetivo principal deste estudo é realizar uma revisão sistemática da literatura para identificar as ferramentas e algoritmos mais utilizados na análise de sentimentos em redes sociais. A metodologia proposta envolve a busca e análise de artigos científicos relevantes, destacando as abordagens mais eficazes e suas aplicações. Como resultado, espera-se fornecer uma base sólida para futuras pesquisas e melhorias nas técnicas de análise de sentimentos, acompanhando as tendências tecnológicas e o avanço desse campo.

**Palavras-chave:** análise de sentimento; redes sociais; revisão sistemática; métodos.

## **ABSTRACT**

This paper addresses the relevance of sentiment analysis in social networks, considering its crucial role in human interactions and information sharing in the contemporary era. Exploring the growing importance of online platforms as communication channels, the study emphasizes the need to comprehend the opinions and sentiments expressed by users amidst staggering volumes of data. Focusing particularly on message polarity, the paper proposes the evaluation of sentiment analysis tools in social networks, recognizing the constant evolution of this field as essential for understanding the emotions and opinions of online users.

**Keywords:** sentiment analysis; social network; tool.

## LISTA DE FIGURAS

Figura 1 – Fases de uma revisão sistemática . . . . .	22
Figura 2 – Gráfico da porcentagem da distribuição de trabalhos . . . . .	28
Figura 3 – Fluxograma da pesquisa . . . . .	30

## LISTA DE TABELAS

Tabela 1 – Total de trabalhos aceitos . . . . .	29
Tabela 2 – Situação dos trabalhos até a fase de triagem. . . . .	31
Tabela 3 – Situação dos trabalhos até a fase de triagem. . . . .	31
Tabela 4 – Sumarização dos dados por base resultado final . . . . .	32

## LISTA DE QUADROS

Quadro 1 – Protocolo da Revisão Sistemática da Literatura (RSL) (Parte 1) . . . . .	23
Quadro 2 – Protocolo da RSL (Parte 2) . . . . .	24
Quadro 3 – Estratégia para Seleção dos Estudos . . . . .	25
Quadro 4 – Expressões de busca em diferentes idiomas. . . . .	26
Quadro 5 – Critérios de Inclusão e Exclusão . . . . .	27

## LISTA DE ABREVIATURAS E SIGLAS

ACM	<i>Association Computing Machinery</i>
ANEW	<i>Affective Norms for English Words</i>
APIs	<i>Application Programming Interfaces</i>
AS	Análise de Sentimento
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CE	Critérios de Exclusão
CI	Critérios de Inclusão
CNN	Rede Neural Convolucional
FFNN	<i>Feedforward neural network</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
LIWC	<i>Linguistic Inquiry and Word Count</i>
LSTM	<i>Long Short Term Memory</i>
NLTK	<i>Natural Language Toolkit</i>
PLN	Processamento de Linguagem Natural
QP	Questões de Pesquisa
RS	Revisão Sistemática
RSL	Revisão Sistemática da Literatura
TCC	Trabalho de Conclusão de Curso
WWW	World Wide Web

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>15</b>
<b>1.1</b>	<b>Objetivo geral</b> . . . . .	<b>16</b>
<b>1.2</b>	<b>Objetivos específicos</b> . . . . .	<b>16</b>
<b>1.3</b>	<b>Estrutura do trabalho</b> . . . . .	<b>16</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	<b>17</b>
<b>2.1</b>	<b>A Web e as Redes Sociais Digitais</b> . . . . .	<b>17</b>
<b>2.2</b>	<b>Análise de Sentimento e o Avanço da Tecnologia</b> . . . . .	<b>18</b>
<b>2.3</b>	<b>Processamento de Linguagem Natural (Processamento de Linguagem Natural (PLN))</b> . . . . .	<b>19</b>
<b>3</b>	<b>PROCEDIMENTOS METODOLÓGICOS</b> . . . . .	<b>21</b>
<b>3.1</b>	<b>O que é revisão sistemática da literatura(RSL)</b> . . . . .	<b>21</b>
<b>3.2</b>	<b>Processo da revisão sistemática da literatura (RSL)</b> . . . . .	<b>21</b>
<b>3.3</b>	<b>Definição de protocolo</b> . . . . .	<b>22</b>
<b>3.3.1</b>	<i>Procedimentos da pesquisa baseada no protocolo da RSL</i> . . . . .	<b>22</b>
<b>3.3.2</b>	<i>Procolo da RSL</i> . . . . .	<b>23</b>
<b>3.3.3</b>	<i>Seleção dos estudos</i> . . . . .	<b>26</b>
<b>3.3.3.1</b>	<i>Critérios de exclusão e inclusão</i> . . . . .	<b>26</b>
<b>3.3.3.2</b>	<i>Avaliação da qualidade dos estudos</i> . . . . .	<b>28</b>
<b>3.3.4</b>	<i>Síntese dos dados e apresentação dos resultados</i> . . . . .	<b>29</b>
<b>3.3.4.1</b>	<i>Estratégia de divulgação da pesquisa e Publicação dos resultados</i> . . . . .	<b>29</b>
<b>4</b>	<b>RESULTADOS</b> . . . . .	<b>30</b>
<b>4.1</b>	<b>Identificação</b> . . . . .	<b>31</b>
<b>4.2</b>	<b>Triagem dos estudos</b> . . . . .	<b>31</b>
<b>4.3</b>	<b>Resposta para questões da pesquisa</b> . . . . .	<b>32</b>
<b>4.3.1</b>	<i>Para Questões de Pesquisa (QP) 1: Quais as ferramentas mais utilizadas no processo de análise de sentimento?</i> . . . . .	<b>32</b>
<b>4.3.2</b>	<i>Para a QP 2: Quais são as redes sociais mais estudadas por pesquisas de análise de sentimento?</i> . . . . .	<b>33</b>
<b>4.3.3</b>	<i>Para a QP 3: Quais algoritmos são mais utilizados no processo de análise de sentimento?</i> . . . . .	<b>33</b>

<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	<b>35</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>36</b>

## 1 INTRODUÇÃO

Segundo Araslanov *et al.* (2020), as redes sociais funcionam como plataformas para conexões entre pessoas e proporcionam um espaço para expressar opiniões e sentimentos.

Nesse cenário, observamos que as redes sociais se estabeleceram como plataformas para a interação humana e o compartilhamento de informações. Araújo *et al.* (2013) destacam que a quantidade de usuários ativos e o volume de dados criados diariamente nessas redes atingem números impressionantes. Por exemplo, uma rede social como o *Twitter*<sup>1</sup>, que possui mais de 200 milhões de usuários, compartilhando cerca de 400 milhões de *tweets*<sup>2</sup> por dia.

No universo das redes sociais, também podem ser encontrados outros participantes, como o *Facebook*<sup>3</sup>. O *Facebook* ganhou notoriedade nos anos 2000, tornando-se, nesse momento, a maior rede social do mundo, alcançando 2,3 bilhões de pessoas Valente (2020).

Em meio a essa volumetria de dados, surge a necessidade de entender como os usuários se comportam no ambiente digital. Existem estudos focados em compreender a polaridade das mensagens compartilhadas, partindo do pressuposto de que a maioria das mensagens trocadas nessas redes está relacionada ao humor e às emoções expressas pelos usuários Araújo *et al.* (2013).

Sendo assim, este trabalho busca conhecer e descrever as ferramentas de análise de sentimento em redes sociais, um campo em constante evolução que desempenha um papel fundamental na compreensão das emoções e opiniões expressas pelos usuários *online* Xavier *et al.* (2020).

Nair *et al.* (2021) define que o termo Análise de Sentimento (AS) consiste no uso de processamento de linguagem natural PLN e ferramentas de análise de texto para identificar e extrair os sentimentos de um autor a partir de seu texto, geralmente encontrado em forma de *tweets*. Esses sentimentos podem ser positivos, negativos ou neutros, refletindo a emoção do autor da mensagem.

A análise de sentimentos desempenha um papel significativo na determinação da opinião sobre o valor da marca de um produto e serviços, ajudando a entender as atitudes dos consumidores. Ela pode ser classificada em dois métodos: aprendizado de máquina supervisionado e aprendizado de máquina não supervisionado (NAIR *et al.*, 2021, p. 2).

<sup>1</sup> Disponível em <https://twitter.com>. Acesso em: 25/11/2023.

<sup>2</sup> Os *tweets* são mensagens com, no máximo, 140 caracteres, compartilhadas na rede social online *Twitter*.

<sup>3</sup> Disponível em <https://www.facebook.com>. Acesso em: 25/11/2023.

De acordo com Liu (2022), a análise de sentimentos, também chamada de mineração de opinião, é o campo de estudo que examina as opiniões, sentimentos, avaliações, apreciações, atitudes e emoções das pessoas em relação a entidades como produtos, serviços, organizações, indivíduos, questões, eventos, tópicos e seus aspectos.

Este trabalho de revisão busca responder à seguinte questão de pesquisa: Quais ferramentas são utilizadas na análise de sentimentos e quais redes sociais são mais comumente analisadas por meio dessas ferramentas? Para responder a essas questões, os objetivos geral e específicos são apresentados a seguir.

### **1.1 Objetivo geral**

Analisar a utilização de ferramentas e algoritmos de análise de sentimentos adotados nas redes sociais, com o intuito de identificar as técnicas de extração mais utilizadas no período de 2018 a 2023. Essa identificação é fundamental para compreender quais abordagens têm sido mais eficazes e amplamente aceitas pela comunidade científica e pelo mercado, facilitando a escolha de métodos mais apropriados para futuros estudos e aplicações práticas na área.

### **1.2 Objetivos específicos**

- Identificar as publicações científicas que realizaram análise de sentimentos em redes sociais.
- Listar as ferramentas, algoritmos e técnicas de extração utilizadas nas publicações científicas identificadas.
- Classificar as ferramentas, algoritmos e técnicas, apresentando suas características e indicações de uso.

### **1.3 Estrutura do trabalho**

Visando facilitar a compreensão das informações relatadas neste trabalho de conclusão de curso, sua organização foi estabelecida da seguinte forma: A Seção 1 apresenta as considerações introdutórias e os objetivos geral e específicos; na Seção 2 são abordadas as definições de alguns conceitos-chave para o entendimento da dissertação; a Seção 3 descreve os procedimentos metodológicos adotados, mostrando a metodologia do estudo da RSL; a Seção 4 apresenta os resultados; e, finalmente, na Seção 5 tecemos nossas considerações finais.

## 2 FUNDAMENTAÇÃO TEÓRICA

Esta seção expõe os principais conceitos necessários para a compreensão deste trabalho. Nela, explicaremos a origem da *web* e os avanços que possibilitaram o surgimento das redes sociais. Também abordaremos a análise de sentimento e suas ferramentas.

### 2.1 A *Web* e as Redes Sociais Digitais

Em 1994, surgiu uma nova tecnologia chamada *World Wide Web* (*WWW*), mais conhecida atualmente como *web*. Essa tecnologia foi desenvolvida por Berners-Lee, que buscava criar um repositório de conhecimento comum entre vários pesquisadores em qualquer lugar do mundo Berners-Lee *et al.* (1994). A ideia foi motivada por um projeto pessoal, cujo objetivo era manter o controle de informações pessoais.

A primeira versão da *web*, conhecida como 1.0, apresentava páginas estáticas e não permitia que os usuários interagissem com o conteúdo. Essa versão era conhecida por apenas permitir a leitura do conteúdo exibido na tela Aghaei *et al.* (2012). Com essa limitação, surgiu a necessidade de permitir que os usuários, além de ler, pudessem escrever em páginas, o que levou ao surgimento da versão *web* 2.0.

O termo *web* 2.0 foi definido em 2004 por Dale Dougherty, na época vice-presidente da *O'Reilly Media*, durante uma conferência de *brainstorming*. Ele descreveu a *web* 2.0 como uma "revolução empresarial na indústria de informática". Essa revolução trouxe as principais tecnologias e serviços que usamos atualmente Aghaei *et al.* (2012).

Dessa forma, a *web* 2.0 é amplamente conhecida como a *web* social, devido à sua abordagem colaborativa e de interação social. Como destaca Latorre (2018), o avanço da *web* 2.0 democratizou a internet e possibilitou a troca de informações entre pessoas ao redor do mundo.

Com essa democratização, surgiu um novo contexto de aplicabilidade chamado rede social. Esse termo está relacionado à caracterização de sites que consistem em comunidades virtuais, como o *Facebook*<sup>1</sup>, *Twitter*<sup>2</sup>, *Instagram*<sup>3</sup>, entre outros Kujur e Chhetri (2015). Portanto, é necessário entender como surgiu o termo rede social e quais são suas principais características Aghaei *et al.* (2012).

O termo "rede social" tornou-se sinônimo de tecnologia da informação e comunica-

<sup>1</sup> Disponível em <https://www.facebook.com>. Acesso em: 21/11/2023.

<sup>2</sup> Disponível em <https://twitter.com>. Acesso em: 21/11/2023.

<sup>3</sup> Disponível em <https://www.instagram.com>. Acesso em: 21/11/2023.

ção; seu uso transcendeu áreas e rompeu fronteiras, sendo amplamente adotado por diversos atores sociais Vermelho *et al.* (2015, p. 6). Com a explosão das redes sociais, elas se tornaram a principal ferramenta de comunicação entre os usuários da *web* Rabelo *et al.* (2012).

As redes sociais digitais possuem características distintas que as diferenciam de outros meios de comunicação. Entre as principais, destacam-se a conectividade global, que permite a interação entre indivíduos de diferentes partes do mundo em tempo real; a interatividade, na qual os usuários podem curtir, comentar, compartilhar e gerar conteúdo de forma ativa; e a criação de comunidades virtuais, que agregam pessoas com interesses comuns em torno de temas, marcas, ou ideologias.

Outro ponto importante é a rapidez na disseminação de informações. As redes sociais facilitam o compartilhamento instantâneo de notícias, tendências e ideias, transformando os usuários em agentes de propagação de conteúdo. Além disso, a personalização do conteúdo, com base em algoritmos que analisam as preferências e comportamentos dos usuários, torna o consumo de informações altamente direcionado e relevante para cada indivíduo.

Nos últimos anos, as redes sociais tornaram-se uma das principais razões para o uso da internet por milhões de pessoas. O número de usuários com contas em plataformas de mídia digital aumentou rapidamente, tornando-se um fenômeno cultural para milhões de pessoas. Kaya e Badwan (2020, p. 1).

Nessas redes, os usuários expressam suas opiniões sobre uma variedade de assuntos, desde política, viagens e pratos favoritos até a qualidade de produtos que pretendem comprar. Como resultado, o volume de dados gerados é imenso, criando uma quantidade significativa de informações Rabelo *et al.* (2012).

As opiniões compartilhadas nas redes sociais são extremamente valiosas para empresas e pesquisadores que buscam entender melhor o comportamento e as preferências dos usuários. A próxima seção abordará a análise de sentimento, uma técnica utilizada para avaliar o comportamento dos usuários na *web*.

## **2.2 Análise de Sentimento e o Avanço da Tecnologia**

O principal objetivo da Análise de Sentimento (AS) é definir técnicas automáticas capazes de extrair informações subjetivas de textos em linguagem natural, como opiniões e sentimentos, a fim de gerar conhecimento estruturado que possa ser utilizado por sistemas de apoio ou tomadores de decisão Benevenuto *et al.* (2015).

Para Miner (2012), a capacidade de compreender as percepções das pessoas a respeito de algo é um recurso informacional importante para auxiliar nos processos de tomada de decisão. Entender como os indivíduos compartilham suas opiniões com outras pessoas impacta diversos setores da sociedade Indurkha e Damerou (2010).

Miner (2012) também destaca que conhecer a opinião dos indivíduos é uma vantagem tanto para empresas quanto para consumidores decidindo o que comprar na internet. Maffazzolu (2021) complementa afirmando que as empresas têm grande interesse nesse tipo de análise para identificar a opinião do público-alvo.

Com a explosão das redes sociais globais, a AS passou a ter um valor social muito relevante França e Oliveira (2014). Além de seu valor social, as redes sociais tornaram-se uma verdadeira mina de ouro para quem busca entender como os usuários se expressam na *web*. Analisar essas opiniões é uma tarefa complexa, pois o volume de dados é massivo Xavier *et al.* (2020). Além disso, compreender o impacto de um comentário na *web* é desafiador. Para Pang e Lee (2008), a opinião pode ser medida em três viés: neutro, positivo e negativo. A polaridade neutra ocorre quando o texto traz opiniões positivas e negativas na mesma proporção. Opiniões positivas expressam felicidade e emoções agradáveis, enquanto as negativas estão associadas a tristeza, raiva, medo e ansiedade Lindner (2013). Assim, torna-se necessário desenvolver algoritmos e ferramentas que possam auxiliar pesquisadores nessa área.

Dada a complexidade da AS, dentro da área de PLN, algumas ferramentas ajudam nesse processo. Essas ferramentas são combinadas com algoritmos específicos para análise de sentimentos. O processo de análise de sentimento segue fases, incluindo a análise de comentários em redes sociais.

### **2.3 Processamento de Linguagem Natural (PLN)**

O Processamento de Linguagem Natural (PLN) é uma subárea da inteligência artificial que tem como objetivo permitir que máquinas compreendam, interpretem e processem a linguagem humana de maneira automatizada. Sua relevância tem crescido significativamente nos últimos anos, especialmente com o aumento do volume de dados textuais gerados em plataformas digitais, como redes sociais, blogs e fóruns online. Para análise de sentimento, o PLN desempenha um papel central, pois fornece as ferramentas e técnicas necessárias para extrair e processar informações subjetivas de textos em linguagem natural.

No contexto da análise de sentimento AS, o PLN permite que os sistemas identifi-

quem nuances semânticas, emoções e opiniões expressas em grandes volumes de texto. Técnicas de PLN são empregadas para pré-processamento de dados, incluindo tarefas como *tokenização*<sup>4</sup>, remoção de *stopwords*<sup>5</sup>, *lematização*<sup>6</sup> e *stemming*<sup>7</sup>, que visam transformar os textos em um formato que pode ser compreendido por algoritmos de *machine learning*<sup>8</sup> ou *deep learning*<sup>9</sup>. Após essa fase, os algoritmos podem classificar as opiniões de acordo com a polaridade (positiva, neutra ou negativa) e até mesmo identificar emoções mais complexas.

Após a compreensão dos conceitos abordados nesta seção, a próxima seção discutirá a metodologia utilizada neste trabalho.

---

<sup>4</sup> A tokenização é o processo de dividir um texto em unidades menores chamadas tokens.

<sup>5</sup> Stopwords são palavras comuns em um idioma que geralmente são filtradas ou removidas durante o processamento de texto, pois têm pouca relevância para a análise de conteúdo.

<sup>6</sup> Lematização é o processo de transformar uma palavra em sua forma base ou raiz, conhecida como lema.

<sup>7</sup> Stemming é o processo de reduzir palavras a sua forma base ou raiz (stem), removendo sufixos e prefixos, sem considerar o contexto ou a gramática.

<sup>8</sup> Machine Learning é uma subárea da inteligência artificial que permite que computadores aprendam e façam previsões ou decisões sem serem explicitamente programados para isso.

<sup>9</sup> Deep Learning é uma subcategoria de machine learning que utiliza redes neurais artificiais com várias camadas (também chamadas de redes neurais profundas).

### 3 PROCEDIMENTOS METODOLÓGICOS

Metodologia literalmente refere-se ao estudo sistemático e lógico dos métodos empregados nas ciências, seus fundamentos, sua validade e sua relação com as teorias científicas. No caso desse estudo, para alcançar os resultados esperados, fez uso da RSL, para avaliar a presença de trabalhos que abordem AS, a descrição detalhada da metodologia se encontra descrita na elaboração do protocolo da Revisão Sistemática (RS).

#### 3.1 O que é revisão sistemática da literatura(RSL)

A RSL visa identificar, analisar e interpretar as evidências disponíveis relacionadas com um particular tópico de pesquisa ou fenômeno de interesse Nakagawa *et al.* (2017). Dessa forma existem muitas razões para realizar uma revisão sistemática, os motivos mais comuns são: para resumir as evidências existentes relativas a um tratamento ou tecnologia, por exemplo, para resumir a evidência empírica dos benefícios e limitações de um determinado método ágil Kitchenham *et al.* (2009).

A maioria das pesquisas começa com algum tipo de revisão da literatura. No entanto, a menos que uma revisão da literatura seja realizada de forma justa, tem pouco valor científico. Dessa forma, esta é a principal razão para realizar revisões sistemáticas Kitchenham (2004). Uma revisão sistemática sintetiza o trabalho existente de uma maneira justa, por exemplo, as revisões sistemáticas devem ser realizadas de acordo uma estratégia de pesquisa predefinida para haver consistência nos resultados Kitchenham (2004).

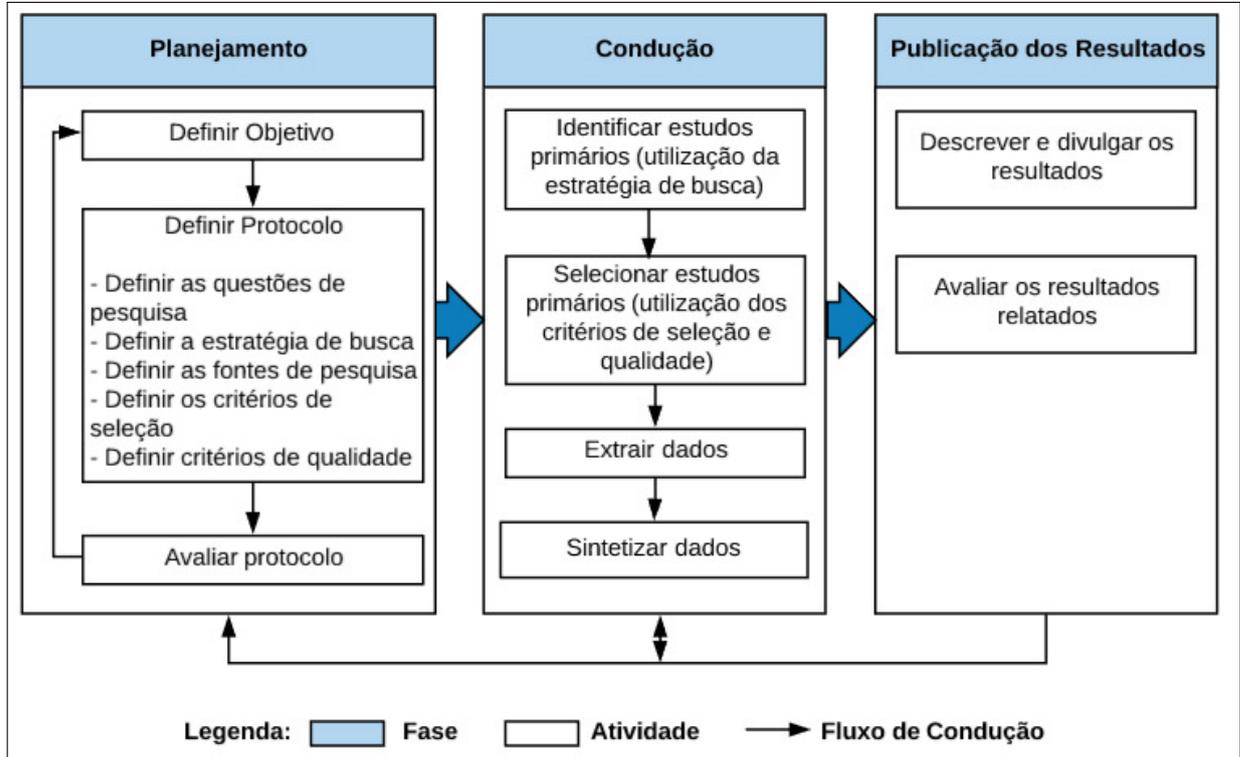
Uma revisão sistemática exige muito mais esforço que uma pesquisa tradicional, assim podemos destacar que a principal vantagem de uma pesquisa como essa, são os resultados que conseguimos coletar ao final da revisão, ou seja, se tivermos resultados consistentes a revisão apresenta um resultado robusto, se o estudo apresentar um fenômeno inconsistente conseguiremos estudar as fontes de variação e assim estudar como melhorar a pesquisa Kitchenham (2004).

#### 3.2 Processo da revisão sistemática da literatura (RSL)

O processo de uma RSL Figura 2 é composto por três fases: planejamento, condução e publicação dos resultados. A fase de planejamento visa identificar a real necessidade, ou seja, a motivação para a execução de uma RS. No entanto, antes de iniciar o planejamento da revisão, é fundamental identificar se existem estudos secundários no mesmo tema Nakagawa *et*

al. (2017, p. 19). O próximo passo após essa análise, temos que definir um protocolo e realizar uma validação do mesmo.

Figura 1 – Fases de uma revisão sistemática



Fonte: Nakagawa *et al.* (2017, p. 19)

### 3.3 Definição de protocolo

Ao se criar um protocolo, definimos a estrutura da nossa pesquisa, ou seja, quais métodos que serão usados para realizar um RS. Dessa forma, é necessário criar um protocolo pre-definido, para reduzir as possibilidades do pesquisador cometer equívocos ao selecionar trabalhos. O processo de desenvolvimento de um protocolo é composto por 5 sessões: (1) Informações Gerais, (2) Questões de pesquisa, (3) Identificação de estudos, (4) Seleção e avaliação de estudos e (5) Síntese dos dados e apresentação dos resultados Nakagawa *et al.* (2017).

#### 3.3.1 Procedimentos da pesquisa baseada no protocolo da RSL

Dessa forma, o protocolo dessa pesquisa utilizou as definições mencionadas no tópico anterior e a ferramenta *Parsifal*<sup>1</sup>, uma ferramenta online que auxilia no processo de RSL.

<sup>1</sup> Disponível em <https://parsif.al> Acesso em: 16/11/2023.

### 3.3.2 *Procolo da RSL*

Os quadros 1, 2 e 3 apresentam, de forma organizada, as informações que constituem o protocolo que desenvolvemos para a pesquisa.

Quadro 1 – Protocolo da RSL (Parte 1)

<b>Campo</b>	<b>Informação</b>
Título	Técnicas utilizadas em análise de sentimento em redes sociais: uma revisão sistemática da literatura.
Pesquisador(es)	Pedro Ítalo Campos da Silva – Pesquisador responsável Marcos Vinicius de Andrade Lima (UFC) – Supervisor
Descrição	Entender as principais ferramentas e algoritmos utilizados para análise de sentimento em redes e sumarizar esses dados.
Objetivos	Analisar o uso de ferramentas e algoritmos de análise de sentimentos nas redes sociais, identificando as técnicas de extração mais utilizadas de 2018 em diante.
QP	<ul style="list-style-type: none"> <li>• QP 1 Quais as ferramentas mais utilizadas no processo de análise de sentimento?</li> <li>• QP 2 Quais são as redes sociais mais estudadas por pesquisas de análise de sentimento?</li> <li>• QP 3 Quais algoritmos são mais utilizados no processo de análise de sentimento?</li> </ul>
Palavras-chave	Português: "análise de sentimento", "ferramentas", "redes sociais". Inglês: "sentiment analysis", "social network", "tool".

Fonte: Elaborado pelo autor, baseado em Nakagawa *et al.* (2017, p. 18-19).

Quadro 2 – Protocolo da RSL (Parte 2)

<b>Campo</b>	<b>Informação</b>
Critérios de Seleção das Fontes de Busca	Artigos completos, teses e dissertações
Lista das Fontes de Busca	<ul style="list-style-type: none"> <li>• <i>Institute of Electrical and Electronics Engineers (IEEE)</i><sup>2</sup></li> <li>• <i>Association Computing Machinery (ACM)</i><sup>3</sup></li> <li>• <i>Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)</i><sup>4</sup></li> </ul>
Estratégia de Busca	Busca automática abrangente nas bases selecionadas por meio das expressões de busca.
<b>Critérios de Inclusão (CI)</b>	<ul style="list-style-type: none"> <li>• <b>CI 1:</b> O estudo deve abordar análise de sentimentos;</li> <li>• <b>CI 2:</b> O estudo deve abordar aplicação de ferramenta(s);</li> <li>• <b>CI 3:</b> O estudo deve estar escrito em português ou inglês;</li> <li>• <b>CI 4:</b> O estudo deve focar nas redes sociais;</li> <li>• <b>CI 5:</b> O estudo deve ter data de publicação de 2018 a 2023;</li> <li>• <b>CI 6:</b> O estudo deve ser trabalho completo;</li> <li>• <b>CI 7:</b> Todos os critérios de inclusão foram atendidos.</li> </ul>
<b>Critérios de Exclusão (CE)</b>	<ul style="list-style-type: none"> <li>• <b>CE 1:</b> Algum critério de inclusão não foi atendido;</li> <li>• <b>CE 2:</b> Não foi possível acessar o resumo ou o texto completo;</li> <li>• <b>CE 3:</b> O estudo não foi publicado como artigo completo, tese ou dissertação;</li> <li>• <b>CE 4:</b> O estudo não possui resumo;</li> <li>• <b>CE 5:</b> Não é um estudo primário.</li> </ul>

Fonte: Elaborado pelo autor, baseado em Nakagawa *et al.* (2017, p. 18-19).

Quadro 3 – Estratégia para Seleção dos Estudos

Campo	Informação
Estratégia para Seleção dos Estudos	<ul style="list-style-type: none"> <li>• <b>Etapa 1:</b> Após a execução da “Busca automática”, será feita a seleção inicial dos estudos por meio da leitura do título, resumo e palavras-chave, aplicando-se os critérios de inclusão e exclusão definidos neste protocolo.</li> <li>• <b>Etapa 2:</b> Os estudos selecionados na primeira etapa serão lidos na íntegra para responder às questões de pesquisa e extração de metadados. Nessa etapa, os estudos passam por reanálise da seleção (utilizando-se dos critérios de inclusão e exclusão) e, se confirmado seu atendimento a esses critérios, serão adicionados na lista final desta RSL.</li> <li>• <b>Etapa 3:</b> Após a conclusão da etapa de leitura completa, será realizada a sumarização dos dados e publicações dos resultados.</li> </ul>
Estratégia de Extração de Dados	Os dados serão extraídos pelo pesquisador e armazenados utilizando-se a ferramenta <i>Parcifal</i> .
Estratégia de Sumarização dos Dados	Utilização do método descritivo.

Fonte: Elaborado pelo autor, baseado em Nakagawa *et al.* (2017, p. 18-19).

Vale destacar que, no início do planejamento e da elaboração do protocolo da RSL, foi definido que seriam considerados, na busca, apenas os estudos escritos em português ou inglês. Desse modo, as palavras-chave e as expressões de busca que constam no protocolo possuem duas versões, como já foi descrito nos Quadro 1. Na fase de planejamento foram estabelecidos os critérios para os estudos que iriam compor a RSL: artigos completos (publicados em periódicos), teses e dissertações. A pesquisa descartou livros, resumos.

Para encontrar os estudos, foram selecionadas as seguintes bases que indexam as principais revistas na área de tecnologia: *IEEE*<sup>5</sup>, *ACM*<sup>6</sup> e *CAPES*<sup>7</sup>. Para escolha dessa base foi levado em conta a relevância que as mesma tem para área tecnológica.

Além de definimos as fontes de dados necessárias para esse estudo, definimos no

<sup>5</sup> Disponível em <https://ieeexplore.ieee.org/Xplore/home.jsp> Acesso em: 17/11/2023.

<sup>6</sup> Disponível em <https://dl.acm.org/> Acesso em: 12/11/2023.

<sup>7</sup> Disponível em <https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/> Acesso em: 12/11/2023.

protocolo da revisão a "Estratégia para Seleção dos Estudos". A estratégia de seleção consiste em estabelecer as etapas da seleção (seleção inicial, seleção final e revisão da seleção), e a maneira como os critérios de inclusão e exclusão deverão ser aplicados durante essa etapa, ou seja, é preciso definir se para aceitar um estudo basta que ele satisfaça um ou todos os critérios estabelecidos Nakagawa *et al.* (2017).

Dessa forma, no quadro 3, do nosso protocolo estão previstas três etapas: na primeira a (iniciada após o processo de busca automática), os títulos e os resumos dos estudos são lidos pelos pesquisadores, de forma que eles possam identificar quais são as investigações relevantes, com base nos CI e CE da RSL; na segunda etapa, é iniciado o processo de leitura completa dos estudos selecionados na etapa anterior. Essa leitura é fundamental para responder às questões de pesquisas registradas no protocolo e para proceder com a extração dos dados. Após a leitura na íntegra de cada estudo, os critérios de inclusão e seleção são aplicados novamente para definir quais trabalhos serão inclusos no relatório final.

### 3.3.3 Seleção dos estudos

Durante a primeira etapa de seleção, é importante poder contar com algum mecanismo de exportação de resultados da busca automática, buscando facilitar o trabalho de documentação por parte dos pesquisadores. As bases de dados selecionadas possuem um mecanismo que auxilia quem busca realizar esse tipo de pesquisa, esse recurso é chamado pesquisa avança.

A pesquisa avança permitir que o pesquisador insira sua *String* de busca, para essa revisão a *String* está definida no protocolo do quadro 4. As expressões possuem um elemento lógicos conhecidos como (*AND e OR*). Nessa revisão, a *String* adotada se encontra na quadro 4.

Quadro 4 – Expressões de busca em diferentes idiomas.

<b>Idioma</b>	<b>Expressão de Busca</b>
Português	"análise de sentimento" <i>AND</i> "rede social" <i>AND</i> "ferramenta"
Inglês	"sentiment analysis" <i>AND</i> "social network" <i>AND</i> "tool"

Fonte: Elaborado pelo autor (2023)

#### 3.3.3.1 Critérios de exclusão e inclusão

Critérios de exclusão e inclusão são definidos pelo pesquisador, e funcionam com um guia para selecionar os estudos primários que são relevantes para o RS. Dessa forma,

os critérios de inclusão servem para indicar por qual ou quais critérios um estudo foi incluído na revisão, já o critérios de exclusão servem para indicar por quais critérios o estudo não foi selecionado Nakagawa *et al.* (2017). Assim, para essa pesquisa, os CI e CE estão destacados na Quadro 2.

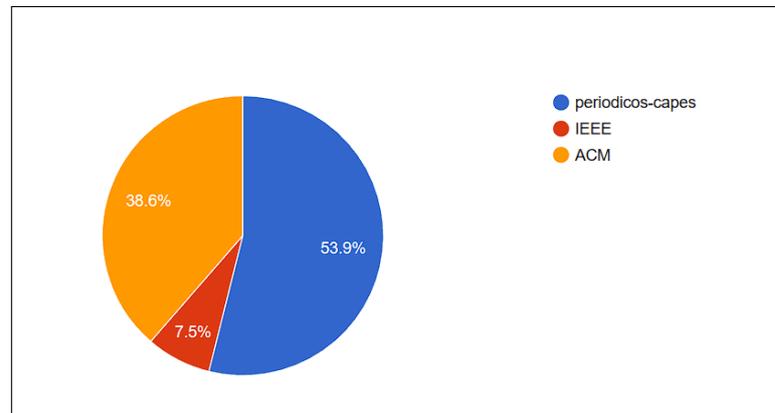
Quadro 5 – Critérios de Inclusão e Exclusão

<b>CI</b>	<ul style="list-style-type: none"> <li>• <b>CI 1:</b> O estudo deve abordar análise de sentimentos;</li> <li>• <b>CI 2:</b> O estudo deve abordar aplicação de ferramenta(s);</li> <li>• <b>CI 3:</b> O estudo deve estar escrito em português ou inglês;</li> <li>• <b>CI 4:</b> O estudo deve ter como foco as redes sociais;</li> <li>• <b>CI 5:</b> O estudo deve ter data de publicação de 2018 a 2023;</li> <li>• <b>CI 6:</b> O estudo deve ter sido publicado como trabalho completo;</li> <li>• <b>CI 7:</b> Todos os critérios de inclusão foram atendidos.</li> </ul>
<b>CE</b>	<ul style="list-style-type: none"> <li>• <b>CE 1:</b> Algum dos critérios de inclusão não foi atendido;</li> <li>• <b>CE 2:</b> Não foi possível ter acesso ao resumo ou ao texto completo;</li> <li>• <b>CE 3:</b> O estudo não foi publicado como artigo completo, tese ou dissertação;</li> <li>• <b>CE 4:</b> O estudo não possui resumo;</li> <li>• <b>CE 5:</b> O trabalho não é um estudo primário.</li> </ul>

Fonte: Elaborado pelo autor.

Durante a primeira etapa de seleção dos estudos, esses critérios ajudaram o pesquisador a selecionar os principais trabalhos relacionados com o tema, sem esses critérios o processo de análise seria bastante complicado, pela abundância de trabalhos relacionados. A imagem 2 mostra o gráfico com o percentual de trabalhos encontrados durante a primeira fase de seleção.

Figura 2 – Gráfico da porcentagem da distribuição de trabalhos



Fonte: Elaborado pelo autor

### 3.3.3.2 Avaliação da qualidade dos estudos

O objetivo principal ao utilizar critérios de qualidade é avaliar aspectos metodológicos dos estudos Nakagawa *et al.* (2017). Ao fazer isso estamos garantindo a confiabilidade da RSBreerton *et al.* (2007).

Desse modo, após uma análise primária dos estudos na etapa 1, temos que realizar a validação da qualidade dos estudos que passaram na primeira fase. Dando início aí etapa 2, que consiste em realizar uma revisão secundária desses estudos.

Na etapa 2, os estudos selecionados são lidos na íntegra, essa leitura busca confirmar se esses trabalhos de fato estão seguindo os critérios de inclusão e exclusão que foram definidos no protocolo dessa RS. A Tabela 1 destaca a quantidade de trabalhos que após seleção foram aceitos, através dessa tabela observamos a importância de utilização dos critérios de inclusão e exclusão.

Notamos também que apesar de ter mais de 1000 trabalhos selecionados, o número final de aceitos é bem menor após realizada a revisão secundária. Vale salientar que após essa seleção devemos sintetizar esses dados e avaliar se os mesmos respondem nossas questões de pesquisa. Na subseção seguintes destacamos esses processo.

Tabela 1 – Total de trabalhos aceitos

Base	Selecionado	Aceito
Periodico-Capes	973	62
IEEE	135	2
ACM	698	12
Total	1806	76

Fonte: Elaborado pelo autor.

### 3.3.4 Síntese dos dados e apresentação dos resultados

Nessa subseção será mostrado a estratégia para mostra os resultados obtidos durante as fases da RS. Desse modo apresentamos também a importância de exibir os resultados adquiridos nessa trabalho para toda a comunidade acadêmica.

Para brereton2007lessons a sumarização dos dados pode ser tanto qualitativas (ou descritiva) ou quantitativa, por meio da aplicação de métodos estatístico. nakagawa2017revisão destaca também que nessa fase a sumarização dos dados já podem auxiliar a responder às perguntas levantadas para dá início a pesquisa. Dessa forma a estratégia adotada foi a utilização dos dados gerados pela plataforma *Parfisal* após a análise dos estudos.

#### 3.3.4.1 Estratégia de divulgação da pesquisa e Publicação dos resultados

A estratégia de divulgação adotada para publicação desse trabalho o desenvolvimento de um Trabalho de Conclusão de Curso (TCC), onde após publicado ficará disponível para toda a comunidade acadêmica. Após o desempenho de todas as etapas propostas devemos começar a sumarização dos dados.

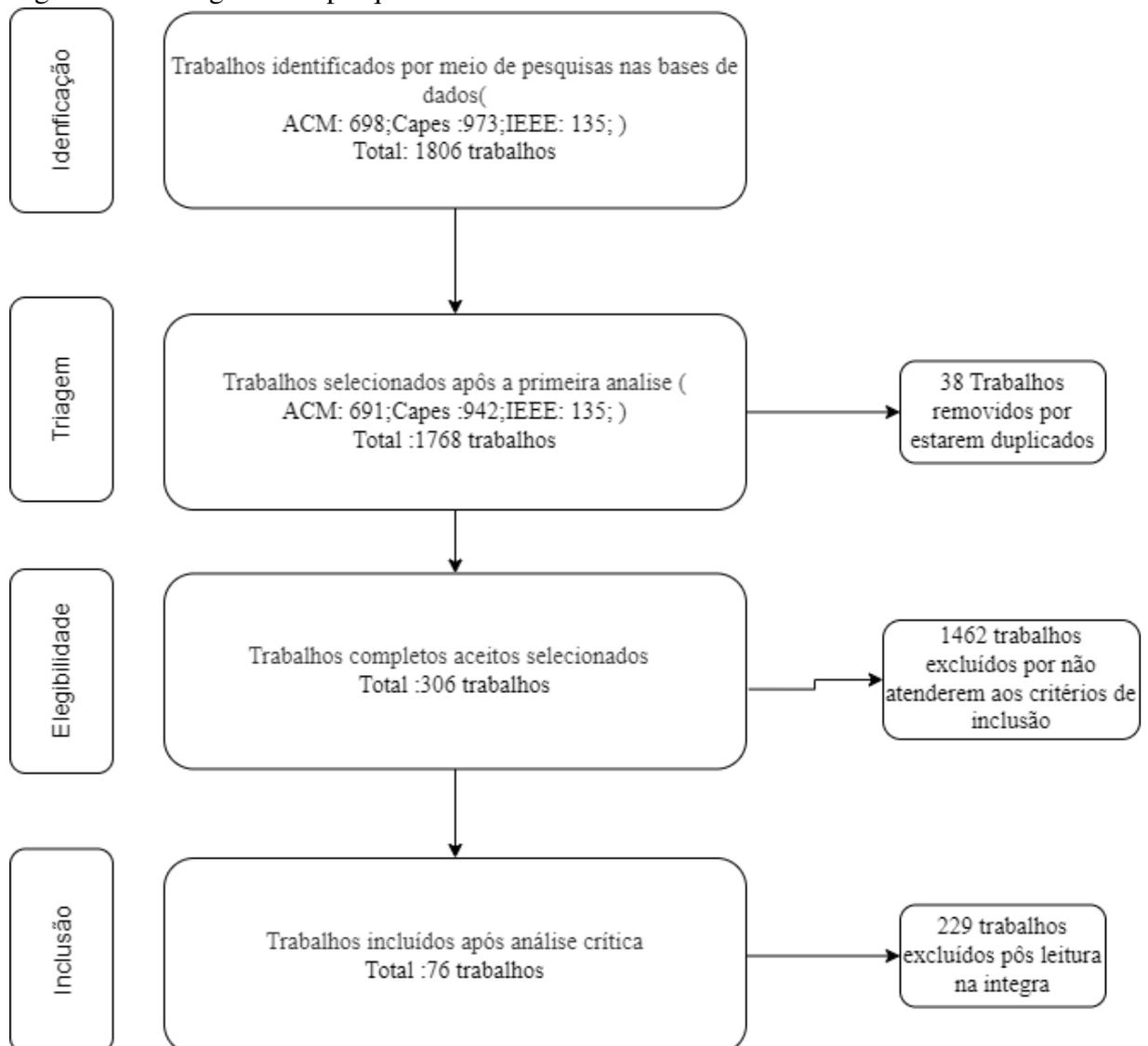
Após realizar a sumarização dos dados, o pesquisador deve se concentrar em juntar os dados conforme as perguntas que definiu no começou da pesquisa. nakagawa2017revisão sugere que esses resultados sejam divulgados no meio acadêmico, conferências, revistas científicas, relatórios técnicos, pôsteres, paginas, *web* e livros. Deste modo, a etapa de divulgação dos resultados é a etapa final da RS, assim aproxima sessão ira aborda o resultado obtidos na nossa revisão.

## 4 RESULTADOS

Nessa sessão será mostrado todos os resultados obtidos durante a RSL, dessa forma essa sessão está organizada da seguinte forma: identificação (5.1), triagem dos estudos, (5.2), (5.3) Elegibilidade e (5.4) Inclusão.

Na figura 3 destacamos um fluxograma das etapas dessa pesquisa, assim como a quantidade de estudos selecionados durante cada etapa da nossa pesquisa.

Figura 3 – Fluxograma da pesquisa



Fonte: Elaborado pelo autor.

#### 4.1 Identificação

No processo de identificação, realizamos a sumarização dos estudos que vão ser analisados, na Figura 2 destacamos o total de trabalhos encontrados na fase de identificação. Durante essa fase encontramos um total de 1806 trabalhos relacionados com o nosso tema.

Na Tabela ?? conseguimos ver melhor esses dados, notasse que ACM é o portal de periódicos da CAPES, foram as bases de dados que mais tiveram trabalhos relacionados, a base da IEEE apresentou o menor número quando comparada as demais bases.

Tabela 2 – Situação dos trabalhos até a fase de triagem.

Total de trabalhos aceitos	Total de trabalhos excluídos	Total duplicados
306	1462	38

Fonte: Elaborado pelo autor.

#### 4.2 Triagem dos estudos

Após a sumarização dos trabalhos, realizamos a importação dos estudos para a plataforma *parsifal* que indicou a duplicidade de 38 trabalhos. Assim ficamos com o total de 1768 trabalhos para analisar. Durante essa etapa, foi realizado também a leitura dos resumos, títulos e palavras-chave. Durante essa fase os critérios de inclusão e exclusão foram utilizados para triagem dos estudos.

Vale destacar que nessa fase tivemos 1462 trabalhos excluídos pelos CE e CI tivemos cerca de 306 estudos que estavam conforme os critérios estabelecidos no protocolo. Na etapa dois da RS, ocorreu a leitura na íntegra de todos os trabalhos selecionados, durante essa fase o pesquisador precisa ser bem criterioso, e seguir o protocolo, devido à possibilidade de haver uma seleção de estudos baseada somente em aspectos subjetivos. Na tabela 3 destacamos esses números para melhor entendimento.

Tabela 3 – Situação dos trabalhos até a fase de triagem.

Total de trabalhos aceitos	Total de trabalhos excluídos	Total duplicados
306	1462	38

Fonte: Elaborado pelo autor.

Após a leitura na íntegra dos trabalhos selecionados, o pesquisador identificou apenas 76 trabalhos relacionados com o tema dessa pesquisa. Na Tabela ?? abaixo, realizamos

a sumarização dos dados com bases na seleção dos estudos. O portal de periódicos CAPES apresentou um total de 62 trabalhos aceitos, 880 rejeitados, e um total de 31 duplicados. Já a ACM apresentou apenas 12 trabalhos dos que haviam sido selecionados, deixando essa base em segundo lugar em número de artigo aceitos. O IEEE foi a base de dados que apresentou o menor número de trabalhos aceitos, praticamente todos foram rejeitados nessa fase, deixando apenas 2 para a fase de análise e extração de dados. Na Tabela 4 destacamos esses números para melhor entendimento do leitor.

Tabela 4 – Sumarização dos dados por base resultado final

Bases	Total de trabalhos aceitos	Total de trabalhos rejeitados	Total de trabalhos duplicados
CAPES	62	880	31
ACM	12	679	7
IEEE	2	134	0
Total	76	1693	38

Fonte: Elaborado pelo autor (2023).

### 4.3 Resposta para questões da pesquisa

Esta subseção se dedicará a resolução da questão de pesquisa. Para atingir esse objetivo, uma abordagem abrangente foi adotada, começando pela análise das questões secundárias propostas anteriormente.

#### 4.3.1 Para QP 1: Quais as ferramentas mais utilizadas no processo de análise de sentimento?

Dentro da seleção de trabalhos aprovados, que se propuseram a análise de sentimento em redes sociais, mencionam-se as ferramentas *Vader*, *SentiWordnet*, *Bert*.

A ferramenta *Vader* foi desenvolvida por Hutto e Gilbert (2014), em 2014, em sua construção foram utilizados alguns dicionários léxicos conhecidos como *Linguistic Inquiry and Word Count* (LIWC) e *Affective Norms for English Words* (ANEW) ela possui regras próprias em sua avaliação, com, por exemplo, letras maiúscula, palavras no aumentativo e pontuações. Dentro dos estudos a validados essa ressa ferramenta apresentou um percentual de 23,2% de utilização, tornando-se a ferramenta mais utilizadas.

A segunda ferramenta mais utilizada, foi a *SentiWordnet*, essa ferramenta foi desenvolvida por Esuli e Sebastiani (2006) em 2006, sua principal característica está no seu dicionário

de palavras, este em comparação com a primeira ferramenta mencionada é bem mais robusto com aproximadamente 147307 palavras Benevenuto *et al.* (2015).

A terceira ferramenta mais utilizada segundo nossa pesquisa é a *Bert*, ela esteve presente em aproximadamente 6,1% dos trabalhos analisados, desenvolvida por Devlin *et al.* (2019) em 2019, é o primeiro modelo de representação baseado em ajustes finos a níveis de frases Devlin *et al.* (2019).

Além dessas três mencionadas anteriormente, identificamos outras 28 ferramentas, dentre essas podemos destacar as ferramentas *TEXTBLOB*<sup>1</sup> e *Natural Language Toolkit (NLTK)*<sup>2</sup>, estas ferramentas apresentaram um percentual de 4,4%. Elas são bibliotecas criadas para *Python*. No processo de pesquisa identificamos que apesar serem apenas 4.4% elas são utilizadas em conjunto com as outras ferramentas, como relata alguns atores da nossa pesquisa.

A ferramenta *Sentiwordnet*, desenvolvida por Esuli e Sebastiani (2006), que possui um dicionário bastante robusto com aproximadamente 147307 palavras Benevenuto *et al.* (2015).

#### **4.3.2 Para a QP 2: Quais são as redes sociais mais estudadas por pesquisas de análise de sentimento?**

Durante a nossa pesquisa notamos vários trabalhos abordavam várias redes sócias, mas a que se destacava era o *Twitter*, essa rede apresentou um total de 79% ocorrência entre os trabalhos. Os trabalhos apresentavam essa rede com uma das mais fáceis de serem estudada pela facilidade de extraírem os dados que estão dentro dessa rede social. O *Facebook*, apresentou um total de 6% de ocorrência. Outro fator que se destaca é que essas mídias sócias estão desde dos anos 2000 em atividade, isso faz com que elas possuam uma quantidade maior de informações.

#### **4.3.3 Para a QP 3: Quais algoritmos são mais utilizados no processo de análise de sentimento?**

Ao examinarmos a sumarização dos dados sobre os algoritmos mais utilizados, observamos que existe uma variabilidade significativa na escolha dos mesmo. Deste modo destaca-se, notavelmente, o algoritmo de *Feedforward neural network* (FFNN), que se destacou

<sup>1</sup> TextBlob é uma biblioteca Python (2 e 3) para processamento de dados textuais. Ele fornece uma API simples para mergulhar em tarefas comuns de processamento de linguagem natural (PNL), como marcação de classe gramatical, extração de sintagmas nominais, análise de sentimento, classificação, tradução e muito mais.

<sup>2</sup> NLTK é uma plataforma líder para a construção de programas Python para trabalhar com dados de linguagem humana.

com uma representação significativa de 21,8%. Tivemos também outros dois algoritmos com mesmo percentual de utilização, sendo eles Rede Neural Convolutacional (CNN) e *Long Short Term Memory* (LSTM) ambos com 14,9%.

O algoritmo CNN é uma rede neural especializada em processar dados bidimensionais, embora também possa ser utilizada para dados uni ou tridimensionais. Sua principal característica é a utilização de uma operação matemática linear chamada “convolução”, a qual confere o nome da rede, em pelo menos uma de suas camadas Goodfellow *et al.* (2016). LSTM, também é um algoritmo que utiliza redes neurais, sua principal funcionalidade é resolver um problema chamado de dissipação de gradiente que afeta as redes neurais. FFNN também é uma rede neural<sup>3</sup>, sua principal característica na direção do fluxo de informações entre suas camadas Hemeida *et al.* (2020)

Depois de fornecer as respostas para nossas perguntas de pesquisa, conseguimos identificar que a rede social mais utilizada no processo de AS é o *Twitter*, isso ocorre pelo fato dessa rede possuir uma variada de usuários de diferentes países e conter um enorme número de postagens em formato de texto Mehta *et al.* (2012). Além disso, a *Application Programming Interfaces* (APIs) do *Twitter* permitir que pesquisadores manipulem os dados e analisem de maneira bem eficiente.

Deste modo, também percebemos que pode haver uma relação entre as ferramentas utilizadas e os algoritmos mencionados na QP 3, dado que ambos buscam o mesmo propósito de análise. No entanto, não podemos afirmar com certeza que o uso de técnicas que utilizam redes neurais como CNN, LSTM e FFNN são tão eficientes quanto as ferramentas que utilizam análise léxica baseada em dicionários, pois os resultados dos trabalhos avaliados não foram comparados diretamente. Mais estudos e uma análise comparativa detalhada entre essas abordagens seriam necessários para confirmar essa hipótese.

---

<sup>3</sup> Rede neural artificial consiste em um conjunto de unidades que contêm funções matemáticas, unidas por pesos. As redes são capazes de aprender, mediante modificação dos pesos sinápticos, e generalizar o aprendizado para outros arquivos desconhecidos. O projeto de redes neurais é composto por três etapas: pré-processamento, processamento e, por fim, pós-processamento dos dados.

## 5 CONSIDERAÇÕES FINAIS

Esta monografia apresentou os resultados de uma RSL, que investigou as principais técnicas e ferramentas utilizadas no processo de análise de sentimento. Durante a análise, 1806 trabalhos foram submetidos para avaliação, dos quais 76 foram selecionados com base nos critérios de inclusão. A partir das informações obtidas, conseguimos responder às principais perguntas que orientaram esta pesquisa.

Após a avaliação dos resultados, conclui-se que o processo de análise de sentimento é uma tarefa árdua e exige do pesquisador uma boa compreensão e definição do objetivo que se deseja alcançar. Através da sumarização dos resultados, identificamos que a rede social mais utilizada para AS foi o *Twitter*. Durante a leitura dos trabalhos, observamos que essa rede social permite ao pesquisador manipular melhor os dados que está estudando, tornando-a uma ferramenta bastante versátil.

Outro dado importante a ser ressaltado é que a qualidade das ferramentas utilizadas está sujeita ao dicionário de palavras que elas possuem, visto que o maior problema da análise de sentimento é reconhecer padrões de escrita. Ter uma boa base de expressões é fundamental. Uma ferramenta bastante mencionada nos estudos foi a *Valder*, com um percentual de 23,2%. Isso ocorre pelo fato de estar disponível há mais tempo e possuir um dicionário léxico robusto.

Um fato que nos chamou bastante atenção relacionado aos algoritmos é que os três mencionados estão todos ligados a redes neurais, o que ocorre devido à volumetria de dados que o pesquisador precisa analisar. Dessa forma, o pesquisador utiliza técnicas de aprendizado de máquina para automatizar sua pesquisa. O algoritmo mais utilizado foi o *FFNN*, que apresentou 22% dos resultados.

Observou-se que, durante a consolidação dos resultados, os pesquisadores combinam algoritmos e ferramentas, buscando aumentar o percentual de acerto, já que alguns desses instrumentos apresentam dicionários léxicos pouco eficientes.

Ao conduzir nossa pesquisa, constatamos que uma amostra de dados extensa, associada a um intervalo de tempo superior a cinco anos, torna o processo de pesquisa mais complexo, dificultando o processo de revisão e não produzindo os resultados esperados.

Como sugestão para trabalhos futuros, recomenda-se que os autores que desejam realizar uma RSL considerem o percentual de acerto de cada algoritmo utilizado na análise de sentimento. Isso ajudará a entender por que certos algoritmos precisam ser combinados com outros para obter resultados melhores.

## REFERÊNCIAS

- AGHAEI, S.; NEMATBAKHSI, M. A.; FARSANI, H. K. Evolution of the world wide web: From web 1.0 to web 4.0. **International Journal of Web & Semantic Technology**, Academy & Industry Research Collaboration Center (AIRCC), v. 3, n. 1, p. 1–10, 2012.
- ARASLANOV, E.; KOMOTSKIY, E.; AGBOZO, E. Assessing the impact of text preprocessing in sentiment analysis of short social network messages in the russian language. In: **2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)**. [S.l.: s.n.], 2020. p. 1–4.
- ARAÚJO, M.; GONÇALVES, P.; BENEVENUTO, F.; CHA, M. Métodos para análise de sentimentos no twitter. In: SN. **Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)**. [S.l.], 2013. p. 19.
- BENEVENUTO, F.; RIBEIRO, F.; ARAÚJO, M. Métodos para análise de sentimentos em mídias sociais. **Sociedade Brasileira de Computação**, 2015.
- BERNERS-LEE, T.; CAILLIAU, R.; LUOTONEN, A.; NIELSEN, H. F.; SECRET, A. The world-wide web. **Communications of the ACM**, ACM New York, NY, USA, v. 37, n. 8, p. 76–82, 1994.
- BRERETON, P.; KITCHENHAM, B. A.; BUDGEN, D.; TURNER, M.; KHALIL, M. Lessons from applying the systematic literature review process within the software engineering domain. **Journal of systems and software**, Elsevier, v. 80, n. 4, p. 571–583, 2007.
- DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. 2019.
- ESULI, A.; SEBASTIANI, F. SENTIWORDNET: A publicly available lexical resource for opinion mining. In: CALZOLARI, N.; CHOUKRI, K.; GANGEMI, A.; MAEGAARD, B.; MARIANI, J.; ODIJK, J.; TAPIAS, D. (Ed.). **Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)**. Genoa, Italy: European Language Resources Association (ELRA), 2006.
- FRANÇA, T. C. de; OLIVEIRA, J. Análise de sentimento de tweets relacionados aos protestos que ocorreram no brasil entre junho e agosto de 2013. In: SBC. **Anais do III Brazilian Workshop on Social Network Analysis and Mining**. [S.l.], 2014. p. 128–139.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT press, 2016.
- HEMEIDA, A. M.; HASSAN, S. A.; MOHAMED, A.-A. A.; ALKHALAF, S.; MAHMOUD, M. M.; SENJYU, T.; EL-DIN, A. B. Nature-inspired algorithms for feed-forward neural network classifiers: A survey of one decade of research. **Ain Shams Engineering Journal**, v. 11, n. 3, p. 659–675, 2020. ISSN 2090-4479.
- HUTTO, C.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. **Proceedings of the International AAAI Conference on Web and Social Media**, v. 8, n. 1, p. 216–225, May 2014.
- INDURKHIA, N.; DAMERAU, F. J. **Handbook of natural language processing**. [S.l.]: CRC Press, 2010. v. 2.

- KAYA, G. A.; BADWAN, A. Effects of reputation systems and group decision making systems usage on online social networks. In: **2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS)**. [S.l.: s.n.], 2020. p. 1–7.
- KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele Univ.**, v. 33, 08 2004.
- KITCHENHAM, B.; BRERETON, P.; BUDGEN, D.; TURNER, M.; BAILEY, J.; LINKMAN, S. Systematic literature reviews in software engineering—a systematic literature review. **Information and Software Technology**, v. 51, p. 7–15, 01 2009.
- KUJUR, P.; CHHETRI, B. Evolution of world wide web: Journey from web 1.0 to web 4.0. **www.ijcst.com**, v. 6, 03 2015.
- LATORRE, M. Historia de las web, 1.0, 2.0, 3.0 y 4.0. **Universidad Marcelino Champagnat**, v. 1, 2018.
- LINDNER, E. G. O que são emoções? **Revista Brasileira de Sociologia das Emoções, RBSE**, v. 12, n. 36, p. 854–846, 2013.
- LIU, B. **Sentiment analysis and opinion mining**. [S.l.]: Springer Nature, 2022.
- MAFFAZZOLU, W. **UNIVERSIDADE FEDERAL DO PARANÁ WESLEY MAFFAZZOLLI APLICAÇÃO DE REDES LSTM, BILSTM E CNN-BILSTM PARA ANÁLISE DE SENTIMENTO EM AVALIAÇÕES DO GLASSDOOR CURITIBA 2021**. 2021. 1- 31 p.
- MEHTA, R.; MEHTA, D.; CHHEDA, D.; SHAH, C.; CHAWAN, P. M. Sentiment analysis and influence tracking using twitter. **International Journal of Advanced Research in Computer Science and Electronics Engineering**, v. 1, n. 2, p. 72–79, 2012.
- MINER, G. **Practical text mining and statistical analysis for non-structured text data applications**. [S.l.]: Academic Press, 2012.
- NAIR, A. J.; G, V.; VINAYAK, A. Comparative study of twitter sentiment on covid - 19 tweets. In: **2021 5th International Conference on Computing Methodologies and Communication (ICCMC)**. [S.l.: s.n.], 2021. p. 1773–1778.
- NAKAGAWA, E.; SCANNAVINO, K.; FABBRI, S.; FERRARI, F. **Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática**. [S.l.]: Elsevier Brasil, 2017. 1-141 p. ISBN 9788535285970.
- PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends in Information Retrieval**, v. 2, p. 1–135, 01 2008.
- RABELO, J. C. B.; PRUDÊNCIO, R. C. B.; BARROS, F. A. Leveraging relationships in social networks for sentiment analysis. In: **Anais do XVIII Simpósio Brasileiro de Sistemas Multimídia e Web**. Porto Alegre, RS, Brasil: SBC, 2012. p. 181–188.
- VALENTE, J. C. L. **Tecnologia, informação e poder: das plataformas online aos monopólios digitais**. 2020.
- VERMELHO, S. C.; VELHO, A. P. M.; BERTONCELLO, V. Sobre o conceito de redes sociais e seus pesquisadores. **Educação e Pesquisa**, v. 41, n. 4, p. 1– 18, dez. 2015.

XAVIER, F.; OLENSCKI, J. R. W.; ACOSTA, A. L.; SALLUM, M. A. M.; SARAIVA, A. Análise de redes sociais como estratégia de apoio à vigilância em saúde durante a covid-19. **Estudos avançados**, SciELO Brasil, v. 34, p. 261–282, 2020.