

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
LINHA DE PESQUISA INFORMAÇÃO E TECNOLOGIA**

RONNIE SHIDA MARINHO

**APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL NO ESTUDO DA
ALTMETRIA: perspectivas na construção de uma abordagem qualitativa**

**Marília
2023**

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
LINHA DE PESQUISA INFORMAÇÃO E TECNOLOGIA**

RONNIE SHIDA MARINHO

**APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL NO ESTUDO DA
ALTMETRIA: perspectivas na construção de uma abordagem qualitativa**

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação, da Universidade Estadual Paulista – Campus de Marília, como requisito para a obtenção do título de Doutor em Ciência da Informação.

Área de Concentração: Informação, Tecnologia e Conhecimento.

Linha de Pesquisa: Informação e Tecnologia (Pós-graduação em Ciência da Informação - FFC/UNESP-Marília)

Orientador: Prof. Dr. José Eduardo Santarém Segundo

M338a Marinho, Ronnie Shida

Aplicação de técnicas de inteligência artificial no estudo da
altmetria : perspectivas na construção de uma abordagem qualitativa /
Ronnie Shida Marinho. -- Marília, 2023
154 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Filosofia e Ciências, Marília
Orientador: José Eduardo Santarem Segundo

1. Inteligência Artificial. 2. Altmtria. 3. Análise de sentimento. 4.
Web social. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de
Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

RONNIE SHIDA MARINHO

APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL NO ESTUDO DA

ALTMETRIA: perspectivas na construção de uma abordagem qualitativa

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista (UNESP), como requisito para a obtenção do título de Doutor em Ciência da Informação.

Área de Concentração: Informação, Tecnologia e Conhecimento.

Linha de Pesquisa: Informação e Tecnologia

Data da defesa: 17 de maio de 2023

Local: Faculdade de Filosofia e Ciências, UNESP - Campus de Marília

BANCA EXAMINADORA

Presidente e Orientador: Prof. Dr. José Eduardo Santarém Segundo UNESP

Professor no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP Campus de Marília. Professor da Universidade de São Paulo, USP. Campus de Ribeirão Preto.

Membro Titular: Prof. Dr. Leonardo Castro Botega

Professor no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP Campus de Marília

Membro Titular Externo: Prof. Dr. Ronaldo Ferreira de Araujo

Professor no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Alagoas, UFAL Campus de A.C Simões

Membro Titular: Prof. Dr. Edberto Ferneda

Professor no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP Campus de Marília

Membro Titular Externo: Prof. Dr. Marcia Regina da Silva

Professora no Departamento de Educação, Comunicação e Informação da Universidade de São Paulo, USP. Campus de Ribeirão Preto.

***Este trabalho é dedicado a todas as pessoas que direta,
ou indiretamente contribuíram na minha formação pessoal, acadêmica e
profissional.***

***Em especial, à minha família e aos professores da Pós Graduação em
Ciência da informação da Unesp de Marília.***

AGRADECIMENTOS

A conclusão de meu trabalho de doutorado representa um grande marco em minha vida.

Agradeço aqui a todos que de algum modo estão envolvidos na concretização deste trabalho.

Agradeço em primeiro lugar a Deus que iluminou o meu caminho durante essa caminhada.

À Universidade estadual paulista e o programa de pós graduação em ciência da informação, apresento meus agradecimentos pela oportunidade de desenvolvimento deste trabalho, assim como todo conhecimento transmitido ao longo do curso.

Ao professor José Eduardo Santarém Segundo, agradeço pelo seu encorajamento, interesse, estímulo e orientação durante este trabalho. Sou muito grato por todos os ensinamentos me passados, sendo parte essencial em minha trajetória profissional e acadêmica.

Aos meus pais, Nilton Donizete Marinho e Ilda Taiko Shida Marinho por sempre acreditarem na minha capacidade e sabedoria.

Ao meu avô já falecido, Milton Donizete Marinho por ter sido inspiração para que eu nunca desistisse dos meus sonhos, e sempre seguisse em frente.

À minha avó Érica dos Santos Marinho pelo amor, carinho, compreensão e por sua capacidade de me trazer paz.

Aos meus irmãos Renan Shida Marinho e Rangel Shida Marinho por toda paciência e otimismo.

À Tamires Passadori Martins que de forma carinhosa me deu força e coragem, me apoiando nos momentos de dificuldade.

Aos meus amigos do GOU (Grupo de oração universitário), pois foi nesse meio que eu aprendi o valor da minha fé. Aqui aprendi a ver a vida de um jeito diferente. Em especial, ao meu amigo, e irmão de consideração Rafael Montanari por ser chave fundamental na minha vida.

Aos meus amigos do núcleo de estudos em web semântica e dados abertos da USP (NEWSDA).

Por fim, agradeço a todas as pessoas que contribuíram direta ou indiretamente para a concretização deste trabalho.

RESUMO

Com o advento da Web 2.0, conhecida como web social, os usuários passaram da condição de visualizadores, para produtores de conteúdo online. Com base nas ferramentas e tecnologias criadas por esta nova geração da web, diversas plataformas de mídias sociais foram criadas para pesquisadores e leitores. Nesse sentido, constantemente pesquisadores são estimulados a compartilhar, por meio de publicações de revisões de textos, quais as suas impressões, comentários e avaliações acerca dos trabalhos científicos lidos. A Altmétrie surgiu neste contexto, preocupada em mensurar como os resultados de pesquisas são vistos e utilizados no ambiente online. Para tal, os estudos altmétricos são aferidos com base no número de compartilhamentos e curtidas em mídias sociais, comentários, menções, etiquetas, visualizações de páginas, downloads, notícias, reportagens, sites e o uso de resultados de pesquisa dentro e fora do âmbito acadêmico. Diante desse cenário, a relevância de um trabalho baseia-se, exclusivamente em quesitos quantitativos para construção de indicadores. Além disso, é comum que usuários analisem as opiniões de outros usuários no processo de seleção de trabalhos a serem lidos. Considerando que parte das menções possuem textos opinativos, e sob a visão de que as menções na web social são uma das principais características da Altmétrie, a aplicação de técnicas de Inteligência Artificial proposta neste trabalho, pode automatizar este processo, ao capturar o sentimento dos leitores expressos em suas menções para criar uma nova perspectiva de avaliação que auxilie o usuário na tomada de decisão de leitura de um determinado trabalho. Para isso, esta pesquisa apresenta como objetivo a proposição de uma abordagem qualitativa que redesenha o estudo da Altmétrie, a partir da aproximação com a Inteligência Artificial, utilizando técnicas de análise de sentimento. Para o desenvolvimento deste trabalho, utilizou-se o método Design Science Research, sendo um estudo exploratório e aplicado. Além disso, o modelo Altmétrico construído foi validado por meio de uma prova de conceito, a qual utilizou dados extraídos da plataforma Altmetric, *com o intuito de confirmar a hipótese levantada e responder às questões de pesquisa. Os resultados obtidos pelas técnicas de IA contrastam as pontuações de atenção dos trabalhos mais mencionados na plataforma altmetric, apresentando diferentes perspectivas na construção de uma abordagem qualitativa de acordo com cada técnica. Visando melhorar tal abordagem, apresentou uma nova técnica qualitativa, a qual foi intitulada IA Altmetric. Ainda, foi desenvolvido um protótipo que demonstra a validade e a aplicação da nova abordagem.* Conclui-se que a aplicação de técnicas de inteligência artificial no estudo da altmétrie viabiliza uma nova perspectiva de análise qualitativa. Além disso, aponta-se que o presente trabalho realiza um importante estudo entre a Ciência da Informação e a Inteligência Artificial, em especial no contexto da Altmétrie.

Palavras-chave: Altmétrie; Inteligência Artificial; Análise de sentimento; Web Social;

ABSTRACT

With the advent of Web 2.0, known as the social web, users have moved from viewers to online content producers. Based on the tools and technologies created by this new generation of the web, several social media platforms have been created for researchers and readers. In this sense, researchers are constantly encouraged to share, through publications of text reviews, what are their impressions, comments and evaluations about the scientific works read. Altmetrics emerged in this context, concerned with measuring how research results are seen and used in the online environment. To this end, altmetric studies are measured based on the number of shares and likes on social media, comments, mentions, tags, page views, downloads, news, reports, websites and the use of research results outside of academia. Given this scenario, the relevance of a work is based exclusively on quantitative requirements for the construction of indicators. In addition, it is common for users to analyze the opinions of other users in the process of selecting works to be read. Considering that part of the mentions have opinion texts, and under the view that mentions on the social web are one of the main characteristics of Altmetrics, the application of Artificial Intelligence techniques proposed in this work, can automate this process, by capturing the feeling of readers expressed in their mentions to create a new evaluation perspective that helps the user in making a decision to read a certain work. For this, this research aims to propose a qualitative approach that redesigns the study of Altmetrics, based on an approximation with Artificial Intelligence, using sentiment analysis techniques. For the development of this work, the DSR method was used, being an exploratory and applied study. In addition, the constructed Altmetric model was validated through a proof of concept, which used data extracted from the Altmetric platform, in order to confirm the hypothesis raised and answer the research questions. The results obtained by the AI techniques contrast the attention scores of the most mentioned works on the altmetric platform, presenting different perspectives in the construction of a qualitative approach according to each technique. Still, aiming to improve this approach, he presented a new qualitative technique, which was entitled IA Altmetric. Still, a prototype was developed that demonstrates the validity and application of the new approach. It is concluded that the application of artificial intelligence techniques in the study of altmetrics enables a new perspective of qualitative analysis. In addition, it is pointed out that the present work performs an important study between Information Science and Artificial Intelligence, especially in the context of Altmetrics.

Keywords: Altmetric; Artificial intelligence; Sentiment analysis; Social Web;

LISTA DE ILUSTRAÇÕES

Figura 1 - Ciclos em DSR (adaptação de Hevner e Chetterjee (2010, p.11) e Wieringa (2014, p. 14).....	14
Figura 2 - Pesquisa do termo <i>sentiment analysis</i> no Google Trends.....	22
Figura 3 - Léxico de sentimentos.	25
Figura 4 - Modelo de um conceito do SenticNet no formato RDF/XML.	30
Figura 5 - Representação adotada pelo SentiWordNet.....	33
Figura 6 - Exemplo de perfil de usuário no ImpacStory.	46
Figura 7 - Informações de artigo da PLOS ALM.....	47
Figura 8 - Detalhamento das fontes utilizadas para compor o Altmetric Score de um trabalho.	49
Figura 9 - Ciclo regulador de Wieringa.....	55
Figura 10 - Menções do twitter de cunho negativo para o trabalho "Few-shot adversarial learning of realist neural talking head models ".	56
Figura 11 - Etapas realizadas para solucionar o problema investigado.....	57
Figura 12 - Página de informações de um trabalho na plataforma Altmetric.....	58
Figura 13 - Processamento de uma menção extraída do twitter	59
Figura 14 - Processo de classificação da técnica combinada IA Altmetric.....	121
Figura 15 - Pseudocódigo da técnica IA Altmetric.	122
Figura 16 - Autorização do uso dos dados da Altmetric.	127
Figura 17 - Tela inicial do protótipo.....	128
Figura 18 - Tela inicial do protótipo 2.....	129
Figura 19 - Cadastro das informações do título a ser processado.	130
Figura 20 - Exemplo de preenchimento de cadastro do título.....	130
Figura 21 - Pesquisa de artigo por meio do DOI.....	131
Figura 22 - Apresentação das informações qualitativas para o DOI investigado (10.1140/epjds/s13688-017-0110-z).....	132
Figura 23 - Apresentação da quantidade de menções classificadas como positivas, neutras e negativas de acordo com a técnica.	132
Figura 24 - Apresentação da quantidade de menções classificadas como positivas, neutras e negativas de acordo com a técnica. (parte 2).....	133

LISTA DE ABREVIATURAS E SIGLAS

ALM	<i>Article-Level Metrics</i>
API	<i>Interface de programação de aplicação</i>
CSS	<i>Cascading Style Sheets</i>
DOI	<i>Digital Object Identifier</i>
DSR	<i>Design Science Research</i>
FI	<i>Fator de Impacto</i>
HTML	<i>HyperText Markup Language</i>
JCR	<i>Journal Citations Reports</i>
HTML	<i>HyperText Markup Language</i>
LIWC	<i>Linguistic Inquiry and Word Count</i>
IA	<i>Inteligência Artificial</i>
PDF	<i>Portable Document Format</i>
PHP	<i>Hypertext Preprocessor</i>
PLN	<i>Processamento de Linguagem Natural</i>
PLOS	<i>Public Library of Science</i>
PMID	<i>PubMed record ID</i>
SGBD	<i>Sistema de Gerenciamento de Banco de Dados</i>
NLTK	<i>Natural language Processing</i>
VADER	<i>Valence Aware Dictionary for sentiment Reasoning</i>
XML	<i>Extensible Markup Language</i>

LISTA DE TABELAS

Tabela 1 - Cálculo da pontuação para criação do ranque de sentimento.....	60
Tabela 2 - Número de menções coletadas dos 20 trabalhos com maiores pontuações de atenção do a plataforma Altmetric do ano de 2019.	61
Tabela 3 - Classificação de acordo com o método de sentimento Vader para a fonte de dados de Blogs.....	63
Tabela 4 - Classificação de acordo com o método de sentimento Vader para a fonte de dados de Notícias.	64
Tabela 5 - Classificação de acordo com o método de sentimento Vader para a fonte de dados de Twitter.....	65
Tabela 6 - Classificação de acordo com o método de sentimento Vader para a fonte de dados do Facebook.....	66
Tabela 7 - Classificação de acordo com a técnica de sentimento SenticNet para a fonte de dados Blogs.....	68
Tabela 8 – Classificação de acordo com a técnica de sentimento Senticnet para a fonte de dados de notícias	69
Tabela 9 - Classificação de acordo com a técnica de sentimento Senticnet para a fonte de dados Twitter.	70
Tabela 10 - Classificação de acordo com a técnica de sentimento Senticnet para a fonte de dados Facebook.....	71
Tabela 11 - Classificação de acordo com a técnica de sentimento SentiStrength para a fonte de dados Blogs.....	72
Tabela 12 - Classificação de acordo com a técnica de sentimento SentiStreght para a fonte de dados de notícias.....	73
Tabela 13 - Classificação de acordo com a técnica de sentimento SentiStrength para a fonte de dados Twitter.	74
Tabela 14 - Classificação de acordo com a técnica de sentimento SentiStrength para a fonte de dados Facebook.	75
Tabela 15 -Classificação de acordo com a técnica de sentimento Sentiwordnet par a fonte de dados Blogs.....	76
Tabela 16 - Classificação de acordo com a técnica de sentimento Sentiwordnet para a fonte de dados de notícias.....	77

Tabela 17 - Classificação de acordo com a técnica de sentimento Sentiwordnet para a fonte de dados Twitter.	78
Tabela 18 - Classificação de acordo com a técnica de sentimento Sentiwordnet para a fonte de dados Facebook.	80
Tabela 19 - Classificação de acordo com a técnica de sentimento Textblob par a fonte de dados Blogs.	81
Tabela 20 - Classificação de acordo com a técnica de sentimento Textblob para a fonte de dados de notícias.	82
Tabela 21 - Classificação de acordo com a técnica de sentimento Textblob para a fonte de dados do Twitter.	83
Tabela 22 - Classificação de acordo com a técnica de sentimento Textblob para a fonte de dados do Facebook.	84
Tabela 23 - Ranque de sentimento segundo o método Vader.	85
Tabela 24 - Ranque de sentimento segundo o método Senticnet.	86
Tabela 25 - Ranque de sentimento segundo o método Sentistreght.	87
Tabela 26 - Ranque de sentimento segundo o método Sentiwordnet.	88
Tabela 27 - Ranque de sentimento segundo o método Textblob.	89
Tabela 28 - Classificação de acordo com os títulos com maiores <i>altmetric score</i>	92
Tabela 29 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Vader.	94
Tabela 30 - Relação dos trabalhos que mais subiram posições no ranque de análise de sentimento Vader quando comparados com o critério de trabalhos com maior altmetric score.	95
Tabela 31 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Vader quando comparados com o critério de trabalhos com maior altmetric score.	95
Tabela 32 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Senticnet.	97
Tabela 33 - Relação dos trabalhos que mais ganharam posições no ranque de análise de sentimento Senticnet quando comparados com o critério de trabalhos com maior altmetric score.	98
Tabela 34 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Sentic quando comparados com o critério de trabalhos com maior altmetric score.	99

Tabela 35 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Sentistrength.	100
Tabela 36 - Relação dos trabalhos que mais ganharam posições no ranque de análise de sentimento Sentistrength quando comparados com o critério de trabalhos com maior altmetric score.	101
Tabela 37 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Sentistrength quando comparados com o critério de trabalhos com maior altmetric score.	102
Tabela 38 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Sentiwordnet.	103
Tabela 39 - Relação dos trabalhos que mais ganharam posições no ranque de análise de sentimento Sentiwordnet quando comparados com o critério de trabalhos com maior altmetric score.	104
Tabela 40 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Sentiwordnet quando comparados com o critério de trabalhos com maior altmetric score.	105
Tabela 41 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Textblob.	106
Tabela 42 - Relação dos trabalhos que mais ganharam posições no ranque de análise de sentimento Textblob quando comparados com o critério de trabalhos com maior altmetric score.	108
Tabela 43 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Textblob quando comparados com o critério de trabalhos com maior altmetric score.	108
Tabela 44 - Posições ganhas ou perdidas dos trabalhos de acordo com cada técnica de IA. .	110
Tabela 45 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Vader de acordo com a área Medical and Health Sciences.	114
Tabela 46 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Senticnet de acordo com a área Medical and Health Sciences	115
Tabela 47 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Sentistrength de acordo com a área Medical and Health Sciences	115

Tabela 48 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Sentiwordnet de acordo com a área Medical and Health Sciences	116
Tabela 49 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Textblob de acordo com a área Medical and Health Sciences	117
Tabela 50 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo IA Altmetric de acordo com a área Medical and Health Sciences	124

LISTA DE QUADROS

Quadro 1 - Termos de busca.....	15
Quadro 2 - Comparação de trabalhos relacionados.	17
Quadro 3 - Exemplo da aplicação da técnica Vader.....	27
Quadro 4 - Exemplo da aplicação da técnica SenticNet.....	29
Quadro 5 - Exemplo da aplicação da técnica SentiStrength.....	31
Quadro 6 - Estrutura da base de dados do léxico SentiWordNet.	34
Quadro 7 - Exemplo de aplicação da técnica Textblob.	36
Quadro 8 - Características das Técnicas de Inteligência Artificial.	38
Quadro 9 - Pontuação padrão para diferentes fontes.	50
Quadro 10 - Exemplos de cálculo com base no peso de cada fonte.....	52
Quadro 11 - Área de publicação dos top 20 trabalhos com maiores altmetric score.	112

SUMÁRIO

INTRODUÇÃO	9
1.1. CONTEXTUALIZAÇÃO	9
1.2. PROBLEMA	10
1.3. TESE.....	10
1.4. HIPÓTESE	10
1.5. PROPOSIÇÃO	11
1.6. OBJETIVOS.....	11
1.6.1. Geral.....	11
1.6.2. Específicos.....	11
1.7. JUSTIFICATIVA.....	12
1.8. METODOLOGIA.....	13
1.9. TRABALHOS RELACIONADOS	15
1.10. ORGANIZAÇÃO DA PROPOSTA	18
INTELIGÊNCIA ARTIFICIAL.....	19
2.1 CONCEITO DE INTELIGÊNCIA ARTIFICIAL	19
2.2 CONCEITO DE ANÁLISE DE SENTIMENTO.....	21
2.2.1 Abordagem baseada em aprendizado de máquina.....	23
2.2.2 Abordagem léxica	24
2.3 TÉCNICAS DE ANÁLISE DE SENTIMENTO.....	26
2.3.1 Vader	26
2.3.2 SenticNet.....	28
2.3.3 SentiStrength	30
2.3.4 SentiWordNet.....	32
2.3.5 TextBlob.....	34
2.4 ANÁLISE E COMPARAÇÃO DAS TÉCNICAS DE SENTIMENTO	37
2.5 CONSIDERAÇÕES FINAIS.....	40
MEDIÇÃO DA PRODUÇÃO CIENTÍFICA.....	41
3.1 INDICADORES TRADICIONAIS.....	41

3.2	INDICADORES ALTMÉTRICOS.....	42
3.3	FERRAMENTAS ALTMÉTRICAS	44
3.3.1	ImpactStory	45
3.3.2	PLOS ALM	46
3.3.3	Plum Analytics.....	47
3.3.4	Altmetric	48
3.4	CONSIDERAÇÕES FINAIS.....	53
	CICLO DE DESENVOLVIMENTO DO MODELO.....	54
4.1	METODOLOGIA DESIGN SCIENCE RESEARCH.....	54
4.2	INSTANCIÇÃO DO MODELO DSR	55
4.2.1	Investigação do problema.....	56
4.2.2	Projeto de soluções	56
4.2.3	Validação do Projeto.....	58
4.2.4	Implementação	60
4.3	CONSIDERAÇÕES FINAIS.....	90
	AValiação da Implementação	91
5.1	ESTUDO 1: ANÁLISE DA QUALIDADE DOS TOP 20 TRABALHOS COM MAIOR “ALTMETRIC SCORE” CONSIDERANDO O SENTIMENTO POSITIVO	91
5.2	ESTUDO 2: ANÁLISE QUALITATIVA DAS MENÇÕES DE ACORDO COM A ÁREA DE ESTUDO	112
5.3	ESTUDO 3: ANÁLISE DA COMBINAÇÃO DE TÉCNICAS DE IA NA ALTMETRIA: IA ALTMETRIC	120
5.4	CONSIDERAÇÕES FINAIS.....	125
	PROTÓTIPO	126
6.1	FERRAMENTAS AUXILIARES.....	126
6.2	RESULTADOS.....	128
6.3	CONSIDERAÇÕES FINAIS.....	133
	CONCLUSÃO	134
	REFERÊNCIAS.....	137

INTRODUÇÃO

O presente Capítulo se dedica a apresentar uma introdução sobre os temas que embasam o arcabouço científico e tecnológico deste trabalho, junto ao domínio estudado, assim como apresentar a motivação, problemática, objetivos e metodologia, resultados esperados e organização dos capítulos.

1.1. CONTEXTUALIZAÇÃO

Nos últimos anos, uma série de transformações têm ocorrido na maneira como os pesquisadores produzem e impulsionam o conhecimento, bem como nas formas de avaliações de trabalhos científicos e seus impactos acadêmicos.

Com o aumento do volume de conhecimento produzido diariamente, a seleção de um trabalho para colocar na lista de leituras tornou-se uma tarefa árdua e exaustiva para os pesquisadores (ARAÚJO, 2018). Em reflexo, emergiram métodos bibliométricos que auxiliam na análise e construção de indicadores para aferirem a importância de um trabalho.

Tradicionalmente, a bibliometria avalia as atividades de produção e comunicação do conhecimento por meio de publicações científicas. Para isso, os estudos bibliométricos são medidos, de maneira quantitativa, com base nas publicações científicas de um autor ou instituição. Assim, um trabalho reconhecido além do círculo em que foi produzido é tratado como referência em sua área (MATTOS, 2004). Embora tal método seja atualmente a principal forma de avaliação da produção acadêmica (PRITCHARD, 1969), uma crítica recorrente é a falta de informações para avaliar a produtividade dos pesquisadores, levando em consideração a repercussão online do trabalho. Outro ponto bastante questionado é em relação ao processo de publicação de artigos científicos ser lento, visto que o fator de impacto para um determinado ano é calculado utilizando as publicações citáveis dos dois últimos anos (CHANG; DESAI; GOSAIN, 2019; NOAH; PEARCE; MODGIL, 2020).

Em resposta a essas limitações, a Almetria surgiu como alternativa, incorporando novas técnicas que mensuram o espectro de interações na web social. Para tal, os estudos alométricos são aferidos com base no número compartilhamentos de artigos, menções em redes sociais, visualizações de páginas, downloads e o uso de resultados de pesquisa fora do âmbito acadêmico, e que atualmente passaram a ocorrer no ambiente web (NASCIMENTO, 2016).

Nesse cenário, trabalhos que utilizam tal métrica baseiam-se, majoritariamente, em quesitos quantitativos para a construção de indicadores.

1.2. PROBLEMA

A Altmtria surgiu como medida alternativa, apresentando diversas perspectivas relacionadas ao modo de medir o impacto da produção científica e descobrir indicativos da dimensão de uma pesquisa, tanto no contexto acadêmico, quanto fora dele (NASCIMENTO, 2016). Entretanto, tais indicativos se dão majoritariamente de forma quantitativa, com base nas menções em redes sociais, visualizações de páginas e compartilhamentos.

Como resultado, a medida retorna uma pontuação para os trabalhos científico que pode não refletir às reais necessidades dos pesquisadores, visto que parte das menções carregam textos opinativos.

Perante a concepção de que as menções na web social são uma das principais fontes alométricas (ARAÚJO, 2018), esta pesquisa apresenta como principal indagação: “A análise do conteúdo textual pode representar uma nova perspectiva, de cunho qualitativo, para o contexto da Altmtria?”.

1.3. TESE

A partir da problemática apresentada, este trabalho defende a tese de que o estudo da aplicação de técnicas de Inteligência Artificial na Altmtria possibilita gerar uma perspectiva na construção de uma abordagem de caráter qualitativo ao atendimento às necessidades informacionais dos pesquisadores, viabilizando, de modo direto a forma como os trabalhos são mensurados nas mídias sociais, além de proporcionar um avanço na Ciência da Informação no que tange ao uso de indicadores científicos.

1.4. HIPÓTESE

A hipótese levantada neste trabalho é a de que se houver a união entre as técnicas de Inteligência Artificial e o conceito de Altmtria, será possível estabelecer uma nova perspectiva por parte das ferramentas alométricas da importância da análise das menções da web social. Tal hipótese é formulada pois a Inteligência Artificial busca identificar opiniões perante as menções

produzidas pelos usuários, enquanto a Altmétrie tem como meta identificar menções existentes nas mídias sociais.

Complementarmente, a Inteligência Artificial estabelece uma forma automatizada de analisar as menções de acordo com o seu sentimento, permitindo que o cálculo Altmétrico ocorra com base em textos opinativos. Neste cenário, no âmbito da Ciência da Informação permite aperfeiçoar a Altmétrie, para que, assim o cálculo da produção científica seja melhor aferidos.

1.5. PROPOSIÇÃO

Este trabalho propôs-se a viabilizar um modelo Altmétrico qualitativo, centrado no uso de técnicas de Inteligência Artificial, tal como no sentimento expresso nas menções da web social. O modelo proposto está embasado nos conceitos de Altmétrie e da Inteligência Artificial.

1.6. OBJETIVOS

Levando em consideração a proposição apresentada, dividiu-se a explanação dos objetivos em duas partes, sendo: objetivo geral e objetivo específico.

1.6.1. Geral

A pesquisa relatada nesta tese tem como objetivo geral desenvolver um modelo Altmétrico que utiliza os conceitos e as técnicas de Inteligência Artificial para aprimorar a forma como é medida a produção científica no contexto das mídias sociais. A partir desse objetivo, buscou-se demonstrar uma nova perspectiva de análise altimétrica qualitativa.

1.6.2. Específicos

Como objetivos específicos buscou-se:

- Analisar o processo de construção dos indicadores na Altmétrie de forma geral;
- Propor a construção de um modelo qualitativo no contexto da Altmétrie;
- Relacionar como a Inteligência Artificial pode apoiar o estudo da Altmétrie;

- Viabilizar um novo parâmetro de avaliação que ajude os pesquisadores na tomada de decisão sobre a leitura de um trabalho científico;
- Apresentar os diferentes cenários gerados por meio da aplicação de algoritmos de Inteligência Artificial;
- Definir técnicas de Inteligência Artificial para construção do modelo;
- Validar o modelo proposto por meio de uma prova de conceito;
- Detalhar de forma prática a técnica combinada intitulada IA Altmetric;
- Implementar um protótipo utilizando o modelo Altmétrico.

1.7. JUSTIFICATIVA

Sabendo-se que as menções na web social são uma das principais fontes alométricas (ARAÚJO, 2018), esta pesquisa pode ser justificada ao propor um modelo alternativo realizar a construção de indicadores alométricos, sob a perspectiva na construção de uma abordagem qualitativa. Tal abordagem complementa os estudos métricos tradicionais, permitindo avaliar o impacto de uma pesquisa científica para além da quantidade de menções que recebem, e até mesmo para além do âmbito das comunidades científicas. A partir desse contexto, tem-se um novo cenário, uma vez que as menções na web social podem fornecer opiniões positivas e negativas, possibilitando aferir o comportamento dos usuários em relação aos trabalhos científicos que leram.

Por esse princípio, entende-se que os modelos computacionais, baseados em técnicas de Inteligência Artificial, podem se apresentar como elementos importantes na identificação e compreensão de padrões, mais especificamente, na análise qualitativa das menções em redes sociais sobre a produção científica. Vale destacar que a Inteligência Artificial é baseada na evolução dos modelos computacionais, avanço na infraestrutura computacional e nas técnicas para coletar, processar e analisar dados em grandes volumes, tornando viável e factível o uso de algoritmos de mineração de opiniões para análises da produção científica em larga escala. Assim, a utilização de IA poderá transformar o processo semântico como um todo, pois contribuirá na identificação das opiniões das menções produzidas pelos usuários.

Dentro desse contexto, a aplicação de técnicas de Inteligência Artificial no estudo da Altmétrie permitirá que pesquisas na Ciência da Informação avancem significativamente no processo de construção de indicadores que reflitam as necessidades informacionais dos usuários.

1.8. METODOLOGIA

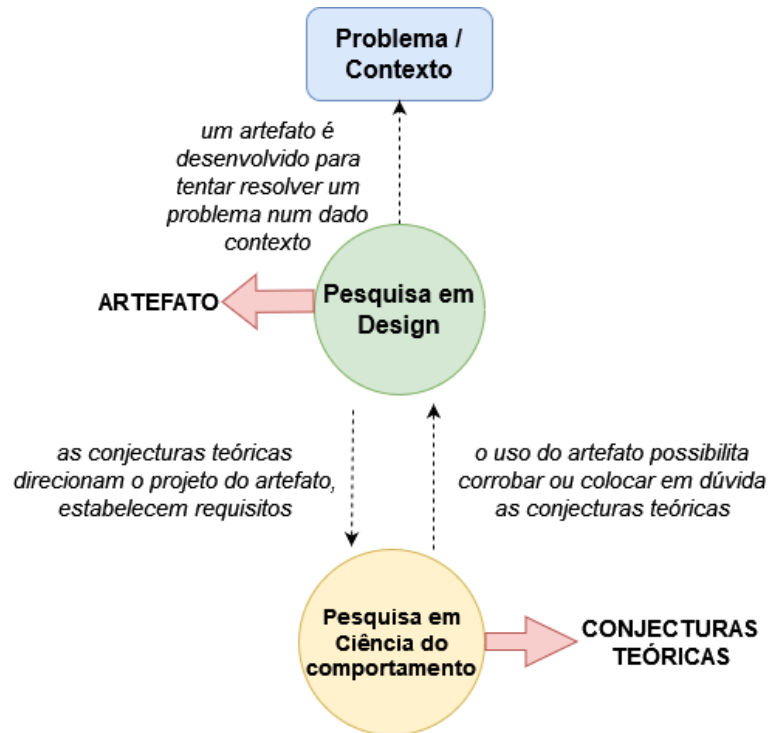
Esta pesquisa é de natureza aplicada, no qual buscou gerar conhecimentos para aplicação prática e dirigida à solução de problemas específicos (SILVA; MENEZES, 2005), uma vez que essa tese tem como finalidade elaborar um modelo altmétrico para auxiliar os pesquisadores na seleção de trabalhos científicos. Ainda, é classificada como exploratória, pois utiliza fontes bibliográficas, devido ser um tema bastante relevante no meio científico. Além disso, é uma pesquisa quantitativa e qualitativa, que utiliza a Inteligência Artificial, de modo a automatizar a identificação do grau de polaridade das opiniões contidas nas anotações dos usuários, o qual pode ser positivo, negativo ou neutro.

Para o desenvolvimento desta pesquisa foi utilizado o método DSR (*Design Science Research*), onde o pesquisador fica comprometido com dois objetivos: solucionar um problema prático num cenário específico através de um artefato, e gerar novo conhecimento científico (PIMENTEL; FILIPPO; SANTORO, 2019, p. 6).

No que diz respeito ao problema, pode-se citar a forma de aferir as menções de usuários utilizadas para compor a métrica Altmétrica. A possibilidade de auxiliar no processo de tomada de decisão de leitura de trabalhos, e a construção de uma nova perspectiva altmétrica qualitativa é a maior contribuição desta pesquisa.

A Figura 1 apresenta o ciclo em DSR, que propõe o desenvolvimento de um artefato para resolver um problema, e o uso deste permite avaliar as conjecturas teóricas elaboradas sobre o mesmo. Ao final do ciclo deve ser possível avaliar se o problema foi solucionado de modo satisfatório por meio do uso do artefato e se esse uso corroborou as conjecturas iniciais.

Figura 1 - Ciclos em DSR (adaptação de Hevner e Chetterjee (2010, p.11) e Wieringa (2014, p. 14))



Fonte: Pimentel; Filippo; Santoro (2019, p.7)

Recentemente, alguns autores inseriram essa metodologia nas pesquisas com foco em tecnologia e informação no cenário da Ciência da Informação. Ainda nesse contexto, Lacerda et al, (2013) afirma que:

O método DSR tem sido considerado como um método para a condução de pesquisas de caráter tecnológico, como tecnologias de gestão, constituindo-se em uma abordagem que, quando bem aplicada, produz rigor científico efetivo.

A condução da pesquisa pelo método DSR neste trabalho teve como base o ciclo de Wieringa (2009). Dessa forma, as ideias principais fizeram uso do ciclo regulador para auxiliar na resolução do problema de pesquisa.

1.9. TRABALHOS RELACIONADOS

Diversos esforços têm sido realizados com intuito de utilizar os comentários produzidos na web social no processo de extração de opiniões. Nesta subseção serão apresentados alguns trabalhos, que foram selecionados pela semelhança dos objetivos, pela relevância no cenário da Ciência da Informação, ou por conta do número de citações.

A pesquisa bibliográfica foi realizada como revisão de literatura nas duas grandes ciências que contemplam esta pesquisa, ciência da informação e ciência da computação, em bases de dados nacionais e internacionais, tais como: Google Scholar, BRAPCI e Portal de periódico da capes com os termos de pesquisa definidos no Quadro 1, e com um recorte temporal de 2010 a 2022 nos idiomas inglês e português.

Quadro 1 - Termos de busca.

INGLÊS	PORTUGUÊS
Altmetric	Altmetria
Artificial Intelligence	Inteligência Artificial
Sentimento analysis	Análise de sentimento
Qualitative	Análise Qualitativa
Vader	Vader
SenticNet	SenticNet
SentiStrength	SentiStrength
SentiWordNet	SentiWordNet
TextBlob	TextBlob

Fonte: Elaborado pelo autor.

Em relação aos termos de buscas utilizados, destaca-se o estudo de Paulino *et al.* (2018), que realizou uma análise sobre um tema de impacto abordado na mídia tradicional por meio de interações realizadas na rede social Twitter. Para isso, foram coletados dados de menções de dois casos relevantes na mídia brasileira no ano de 2018: a condenação do presidente Lula e a morte da vereadora Marielle. Os resultados da pesquisa levaram em consideração o sentimento dos comentários, assim como a identificação de clusters por meio da afinidade.

Um outro trabalho, desenvolvido por Piryani *et al.* (2018), apresenta um processo de avaliação de livros de Ciência da Computação por meio de menções extraídas da Amazon¹ e

¹ www.amazon.com

*Goodreads*². Para tanto, os autores aplicam a análise de sentimento nas resenhas e comparam os resultados com as métricas tradicionais de avaliação de impacto.

No que se refere às mídias sociais, Borba (2017) descrevem uma análise altmétrica do termo “Repositório Institucional” no Twitter no período de 2009 a 2015. O estudo utilizou a Altmétrie, buscando aferir as postagens relacionadas ao termo, assim como realizar uma análise de conteúdo das mesmas, com o intuito de identificar as relações estabelecidas entre o termo e assuntos associados. As conclusões indicaram que as primeiras publicações sobre o termo “Repositório Institucional” em língua portuguesa no Twitter aparecem em 2009, sendo que a maior quantidade de tweets sobre o tema aconteceu em 2013. Salientou-se a prevalência da categoria “Citações sociais” e “Divulgação de Repositório Institucional”, que, na maior parte das vezes, refere-se à inserção de repositórios. A hashtag mais encontrada foi #opendoar, que é o diretório oficial de repositórios acadêmicos de acesso aberto, desenvolvido pela Universidade de Nottingham, no Reino Unido.

Um outro recente trabalho é de Marinho *et al.* (2022), que propuseram um instrumento alternativo para tomada de decisão no processo de seleção de livros. Para tal, realizou-se a coleta e análise de opiniões contidas nas resenhas publicadas na plataforma Skoob.

Em uma abordagem diferente, Zhou *et al.* (2016) propõem medir o impacto de livros acadêmicos por meio de mineração de resenhas em plataformas online e da aplicação da técnica de análise de sentimento sobre as resenhas coletadas.

Cavalcanti *et al.* (2012) realizaram um estudo para aferir o impacto de uma publicação utilizando análise de sentimento positivo em artigos extraídos de uma revista internacional eletrônica, chamada *British Medical Journal*, que possui publicações relacionadas à medicina.

O trabalho descrito por Gontijo (2019) apresenta um processo de análise de artigos científicos de acesso aberto sobre Inteligência Artificial, mensurando o seu impacto acadêmico por meio dos indicadores bibliométricos e altmétricos. Assim, cerca de 5043 artigos publicados entre 2012 e 2018, foram analisados. Os dados acerca de citações foram coletados no site Dimensions, enquanto os de atenção on-line foram retirados da plataforma Altmetric. Os resultados demonstraram que o uso da Altmétrie, associada à bibliometria possibilita a expansão das análises relacionadas ao impacto acadêmico e social das produções científicas.

O trabalho proposto por Therwall *et al.* (2018) avaliou se as pontuações do Altmetric.com são indicadores confiáveis e, também, se podem refletir impactos fora do âmbito

² www.goodreads.com

acadêmico. Para tanto, realizou-se uma análise de regressão dos dados do altmetric de novembro de 2015 e a contagem de citações do Scopus de outubro de 2017.

Além da utilização de algoritmos de Inteligência Artificial, pode-se citar o trabalho de Vysakh *et al.* (2020), onde os autores apresentaram um estudo para medir a quantidade de artigos mencionados nas redes sociais sobre o assunto da Covid-19 durante o período de pandemia. Assim, foram analisados 145 artigos da revista *Nature*, tendo mostrado a influência dos mesmos na web social, revelando que o Twitter é o principal veículo de menções, seguido por portais de notícias. Além disso, o trabalho demonstrou a quantidade de menções a partir dos meses investigados.

O Quadro 2 apresenta uma síntese dos trabalhos relatados anteriormente. Para isso, os mesmos foram organizados em três tipos de assuntos, que refletem o aspecto principal de cada pesquisa. Os assuntos são:

1. Inteligência Artificial – trabalhos que apresentam um enfoque maior no uso de técnicas de IA;
2. Altmetria – tratam de trabalhos que exploram a métrica alternativa como fonte de impacto social e acadêmico;
3. Dados Qualitativo – trabalhos que utilizam dados de maneira qualitativa para um determinado propósito.

Quadro 2 - Comparação de trabalhos relacionados.

TRABALHO	INTELIGÊNCIA ARTIFICIAL	ALTMETRIA	DADOS QUALITATIVO
Paulino <i>et al.</i> (2018)	SIM	NÃO	SIM
Piryani <i>et al.</i> (2018)	SIM	NÃO	SIM
Borba <i>et al.</i> (2017)	NÃO	SIM	NÃO
Gontijo (2019)	NÃO	SIM	NÃO
Cavalcanti <i>et al.</i> (2012)	SIM	NÃO	SIM
Therwall <i>et al.</i> (2018)	NÃO	SIM	NÃO
Zhou <i>et al.</i> (2016)	SIM	NÃO	SIM
Vysakh (2020)	NÃO	SIM	NÃO

Fonte: Elaborado pelo autor.

Todos os trabalhos mencionados trazem grandes contribuições para as áreas de Altmétrie e Inteligência Artificial, porém nenhuma das pesquisas realizam uma análise qualitativa das menções nas redes sociais. Neste sentido, o presente trabalho buscou inserir a Inteligência Artificial no cálculo Altmétrico para a construção de uma nova perspectiva.

O uso da IA poderá transformar o processo semântico como um todo, pois trabalhará na identificação das opiniões contidas nas menções. As principais diferenças deste trabalho, com relação aos outros apresentados, estão primeiramente no uso de Inteligência Artificial juntamente com a Altmétrie e, principalmente, pela utilização dos resultados como uma nova forma de mensurar os dados da Altmétrie.

No âmbito da Ciência da Informação, os trabalhos mencionados não trazem a conceituação de uma perspectiva de análise qualitativa alométrica, enquanto a tese defendida demonstra com a definição do modelo proposto, como a Inteligência Artificial pode contribuir para aprimorar a maneira como as menções são aferidas e, assim, criar uma nova visão no contexto da Altmétrie.

1.10. ORGANIZAÇÃO DA PROPOSTA

Este trabalho está organizado em capítulos, por meio dos quais apresentam-se os principais conceitos e técnicas associadas à pesquisa realizada.

No Capítulo 2 são apresentadas informações básicas a respeito da área de Inteligência Artificial, bem como uma visão geral sobre esta área de pesquisa. Nele, são apresentados os principais subcampos, além de técnicas de análise de sentimento para avaliação de textos não estruturados. No Capítulo 3 são apresentados alguns conceitos relacionados à medição da produção científica. Inicialmente, faz-se um panorama geral dos indicadores tradicionais bibliométricos e, em seguida, apresenta-se o conceito de métricas alternativas, assim como as ferramentas e a forma como é calculada a pontuação de atenção na plataforma altmetric.com. No Capítulo 4, apresenta-se o ciclo de desenvolvimento do modelo Altmétrico proposto, a instanciação do método DSR, além de cada uma das etapas do ciclo regular, bem como a aplicação de cada técnica de Inteligência Artificial utilizada. No Capítulo 5 são expostos e discutidos os resultados parciais obtidos em cada experimento realizado. Por fim, no Capítulo 6, são apresentadas as considerações parciais produzidas por esta pesquisa até o presente momento, destacando-se os possíveis próximos passos.

INTELIGÊNCIA ARTIFICIAL

O desenvolvimento deste trabalho está embasado em um conceito-chave, que é a Inteligência Artificial. Nesse sentido, este capítulo destaca e apresenta a área da Inteligência Artificial, além de mostrar como a Análise de sentimento pode auxiliar nesse processo.

As investigações sobre Ciência da Informação e Inteligência Artificial iniciaram há vários anos, mas a evolução desta segunda possibilitou diversos impactos nos últimos anos, desde questões teóricas, aplicadas e de investigação científica.

Dessa forma, esta seção discute uma nova aproximação nesse sentido, produzindo campos de estudos e perspectivas da Ciência da Informação, aproximando-se da área da Inteligência Artificial e, em específico, da área de Análise de Sentimento.

2.1 CONCEITO DE INTELIGÊNCIA ARTIFICIAL

Nos últimos anos, o estudo da Inteligência Artificial vem sendo utilizada nas mais diversas áreas do conhecimento. Tal estudo desponta em conjunto com a área de Ciência da Computação, motivado pelo crescimento tecnológico das últimas décadas, o que acarretou à criação de diversas tecnologias para o dia a dia das pessoas.

De modo geral, podemos definir a Inteligência Artificial em: sistemas que agem como humanos, sistemas que pensam como humano, sistemas que pensam de forma racional e sistemas que agem de forma racional. A abordagem centrada nos humanos apresenta um caráter empírico, já a abordagem racional preocupa-se com questões matemáticas relacionadas à engenharia (RUSSEL; NORVIG,2016).

Kaufman (2019) diz que:

A Inteligência Artificial refere-se a um campo de conhecimento associado à linguagem e à inteligência, ao raciocínio, à aprendizagem e à resolução de problemas. A IA propicia a simbiose entre o humano e a máquina ao acoplar sistemas inteligentes artificiais ao corpo humano e a interação entre o homem e a máquina como duas espécies distintas conectadas.

Uma outra definição realizada por Rich (1994) afirma que:

Inteligência Artificial é o estudo de como fazer com que a máquina possa realizar tarefas que hoje são melhor realizadas pelo homem.

As definições apresentadas correspondem como a IA foi elaborada por alguns pensadores possuindo um caráter mais filosófico, onde as máquinas assemelham-se ao aspecto da cognição humana, até uma visão pragmática, com uma perspectiva voltada na resolução de problemas (WEBER, 1996).

Em decorrência dessas visões distintas a IA influencia diversas pesquisas no âmbito acadêmico que visam produzir serviços e produtos, como por exemplo podemos citar os *chatbots*, onde tal serviço visa agir como um humano, por meio de interações em conversas. Outro exemplo é a visão computacional, onde busca-se reproduzir a capacidade humana de analisar e identificar imagens. Consoante aos exemplos apresentados pode-se identificar objetivos extremamente distintos aderentes da área de Inteligência Artificial.

Tendo em vista os diferentes objetivos da IA surgiram algumas abordagens que evoluíram de maneira significativa e possibilitaram a criação de subcampos que propõem a solucionar inúmeros problemas.

Russel e Norvig (2016), apresentam uma abordagem que faz uso de redes neurais, simulando o pensamento do cérebro humano, para refletir em como sucede o aprendizado de máquinas. As redes neurais são empregadas em diferentes áreas da IA, visto que ao usar uma abordagem probabilística, é possível identificar padrões e correlações existentes entre os dados.

Além da utilização de redes neurais, pode-se citar outra abordagem importante da IA, os agentes inteligentes que são uma representação de algo que pode perceber o ambiente através de sensores. Dessa forma, os agentes usam componentes computacionais como: sensores e câmeras para atuarem no mundo (RUSSEL; NORVIG, 2016). A característica de um agente inteligente baseia-se na possibilidade de ele ser treinado e ir aperfeiçoando-se com o tempo.

Diferentemente das abordagens supracitadas, a linguagem natural faz uso de símbolos e estatísticas, possibilitando uma compreensão de textos e falas, de maneira que ao desenvolver sistemas, encontre uma grande semelhança entre a linguagem computacional e a linguagem natural. Dentro do campo de processamento de linguagem natural, encontra-se o conceito de análise de sentimento, que tem como foco identificar, de modo qualitativo, a emoção expressa em um texto.

Com o aumento da quantidade de informações e participações de usuários no conteúdo da web social, pesquisadores apresentam grande interesse em explorar as vantagens advindas de tal conceito. Nesse contexto, diversas aplicações surgiram com a intenção de extrair opiniões, bem como inferir o sentimento público utilizando técnicas de Inteligência Artificial.

2.2 CONCEITO DE ANÁLISE DE SENTIMENTO

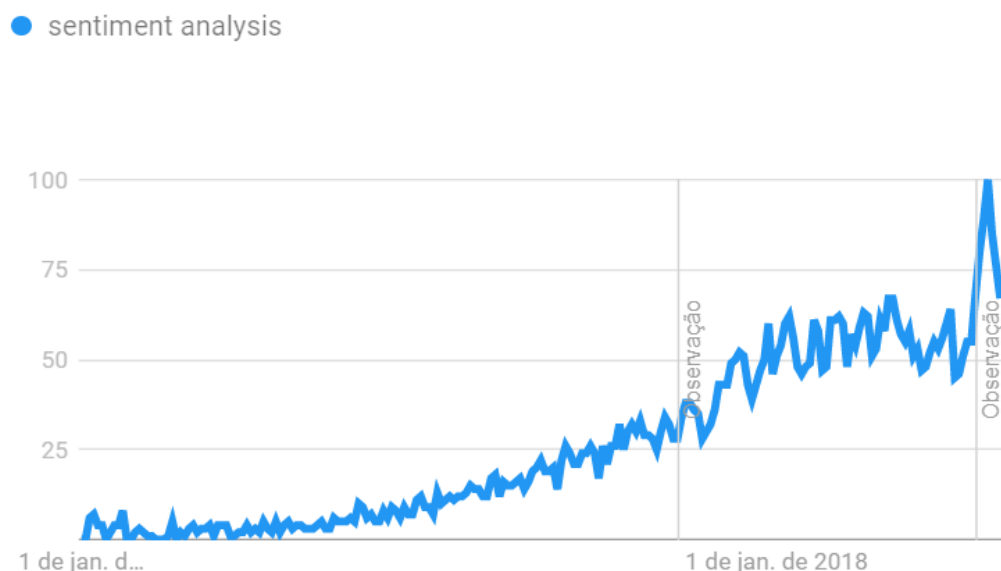
A análise de sentimento é um conceito estudado e aplicada por muitas áreas, como por exemplo: a antropologia, a filosofia, a psicologia, a ciência da informação, a ciência da computação, entre outras. Tal conceito tem por objetivo identificar e extrair, de maneira automática, opiniões, sentimentos e emoções expressados em um texto, podendo revelar, por exemplo, como as pessoas se sentem em relação a um determinado assunto ou alvo (BING LIU, 2012).

Em especial, uma opinião é composta por dois elementos: um alvo e um sentimento sobre esse alvo, que pode ser um assunto, um produto, uma pessoa, entre outras coisas. Assim, pode-se caracterizar um sentimento como sendo uma emoção ou atitude de uma pessoa em relação ao seu alvo.

Segundo Cambria, Livingstone e Hussain (2012), emoções são estados de sentimento que resultam em reações físicas e psicológicas, pensamento e comportamento, e é um dos assuntos mais complexos da psicologia, principalmente pelo aspecto ambíguo da língua. O estudo das emoções e opiniões tenta compreender de que modo os sentimentos são expressos a partir de opiniões positivas e negativas acerca de um assunto. Para isso, os métodos empregados compreendem desde questionários até pesquisas de campo (BECKER; TUMITAN, 2013). No entanto, este tipo de abordagem geralmente apresenta custos elevados, limitando-se a pequenas amostras.

Com o crescimento das redes sociais de uso global como o Twitter em 2006, a análise de sentimento começou a ter um valor social significativo. A Figura 2 mostra tal cenário ao apresentar a evolução na quantidade de buscas pelo termo “*sentiment analysis*” no motor de busca do Google à partir de 2004 até nos dias atuais. A simplicidade para difusão de informações disponibilizada pelas redes sociais reflete uma das novidades mais interessantes desses últimos anos, captando o interesse de jornalistas, pesquisadores, empresas, entre outros. Assim, as redes sociais são a criação de uma revolução digital, possibilitando a difusão das emoções e opiniões por meio da rede (PANG; LEE, 2008).

Figura 2 - Pesquisa do termo *sentiment analysis* no Google Trends.



Fonte: Retirada do Google Trends.

A crescente quantidade de dados publicados diariamente na Web criou diversos obstáculos para o estudo da análise de sentimento, exigindo a construção e o domínio de ferramentas que auxiliam no tratamento de grandes bases de dados. Aferir o sentimento requer uma análise aprofundada do contexto na qual a mensagem se encontra (BECKER; TUMITAN, 2013). Além disso, exige o reconhecimento de regras semânticas e sintáticas para apontar o entendimento da linguagem, tanto implícita quanto explícita.

Para tornar os dados textuais qualitativos, a quantificação de “o que as pessoas pensam” torna-se uma etapa primordial e de relevância de diferentes comunidades científicas. Para isso, inúmeras abordagens têm sido utilizadas para aferir o sentimento em conteúdos publicados na Web, dentre as quais podemos citar: a abordagem baseada em aprendizado de máquina e a abordagem de dicionários léxicos.

A abordagem baseada em aprendizado de máquina necessita de bases de dados rotuladas para treinar classificadores, tendo sua eficiência prejudicada devido ao alto custo na obtenção de dados rotulados.

Em contrapartida, a abordagem de dicionários léxicos conta com uma lista e dicionários de palavras associadas a sentimentos específicos. Assim, sua eficiência está diretamente associada a generalização do vocabulário usado, para os diferentes cenários existentes (PANG et al, 2002).

2.2.1 Abordagem baseada em aprendizado de máquina

De acordo com Benevuto, Ribeiro e Araújo (2015) a abordagem baseada em aprendizado de máquina é composta por técnicas supervisionadas, utilizando a palavra supervisionado em razão de necessitar de uma etapa de treinamento de um modelo com amostras previamente classificadas. Para isso, tal estratégia necessita de quatro fases importantes:

- Obtenção de dados rotulados que serão utilizados para treino e testes;
- Seleção das características que possibilitam a distinção entre os dados;
- Treinamento de um modelo computacional com algoritmo de aprendizagem;
- Aplicação do modelo.

Na fase dos dados rotulados, o intuito é adquirir uma entrada com o seu devido rótulo. No cenário de análise de sentimento, seria uma sentença seguida de sua polaridade.

Em seguida, na fase de definição das características, o foco é classificar os dados com características, intituladas *features*. Tais características devem ser atributos que reconheça cada dado em um conjunto de dados a serem rotulados, possibilitando desta forma uma boa distinção entre o conjunto de dados.

Posteriormente, na fase de treinamento do modelo com algoritmo de aprendizagem, a finalidade é proporcionar um modelo de classificação de palavras segundo a etapa de definições das características.

Por fim, na fase de aplicação do modelo, o objetivo é considerar uma sentença de entrada, rotular de acordo com o modelo treinado e receber como saída uma classificação. No cenário de análise de sentimentos, seria apontado no texto o sentimento de uma sentença.

O aprendizado de máquina apresenta-se de modo eficaz na utilização de um conjunto de dados para qual foi treinado, possuindo uma considerável aplicabilidade de seu modelo, entretanto é limitado ao conjunto de dados. Tal estratégia possui um custo computacional alto em relação de processamento da CPU e memória, características que podem limitar a habilidade de avaliar um sentimento em dados de streaming (BENEVUTO; RIBEIRO; ARAÚJO, 2015).

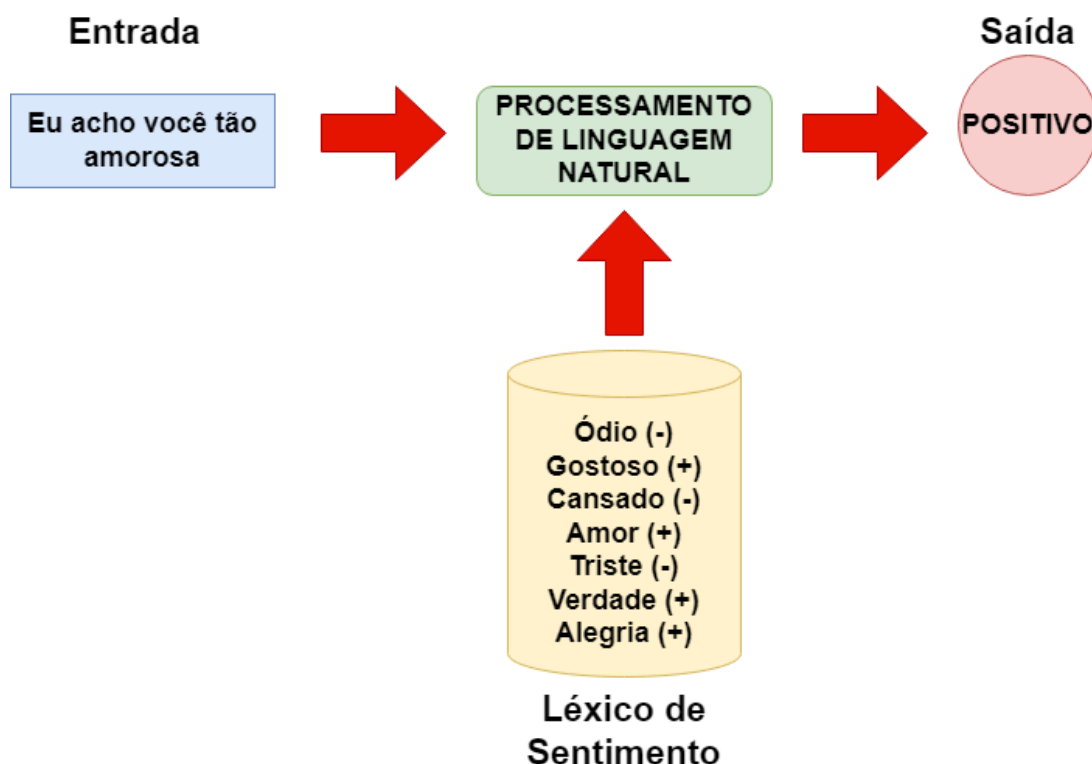
2.2.2 Abordagem léxica

Diferentemente da abordagem baseada em aprendizado de máquina, a abordagem léxica não necessita de sentenças previamente rotuladas e treinos para a criação de um modelo. Está é uma das suas principais vantagens visto que desta maneira não mantém aplicação restrita ao contexto para o qual foram treinados. Nesta abordagem utiliza-se o método de detecção de polaridade, que tem como objetivo a classificação de uma sentença como sendo positiva, neutra ou negativa.

A detecção de polaridade utiliza uma espécie de dicionário de palavras que ao invés de possuir como conteúdo o significado de cada termo, possui em seu lugar um significado quantitativo, por exemplo, pode ser um número entre -1 a 1, onde -1 [e o valor sentimental mais negativo e 1 o valor mais positivo ou mesmo valor qualitativo, como exemplo, positivo ou negativo, feliz ou triste. Tal técnica assume que palavras individuais apresentam o que é chamada de polaridade prévia, que é, uma orientação semântica independente de contexto e que pode ser expressa com um valor numérico ou classe (TABOADA et al, 2011).

Atualmente, a análise de sentimento com base em abordagens léxicas é a estratégia mais poderosa, seja na capacidade de predição, ou então na utilização de recursos computacionais. Para isso, tal abordagem baseia-se na aplicação de um dicionário de palavras, onde cada palavra está relacionada a um sentimento a Figura 3 demonstra a execução de um método de análise de sentimento léxico.

Figura 3 - Léxico de sentimentos.



Fonte: Elaborado pelo autor.

A classificação começa quando o método recebe uma sentença de entrada, em seguida é realizado um processamento de linguagem natural, bem como uma busca no léxico das palavras que compõem esta mensagem. Por fim, na etapa final do processo o método é capaz de concluir qual é a polaridade ou sentimento implícito na sentença de entrada.

Existem diversos dicionários clássicos na literatura como: SentiWordNet, SenticNet, LIWC, Opinion Lexicon, entre outros. Tais dicionários estão organizados na forma como foram classificados originalmente, sendo de forma binária, ou seja positivo ou negativo, ou a partir da intensidade do sentimento, como por exemplo: '-1' negativo, '0' neutro e '+1' positivo. Tal diferença ocorre visto que muitas aplicações se beneficiam caso forem capazes de não somente apontar a polaridade binária ou ternária (positiva, negativa ou neutral), mas também a força de um sentimento expressado no texto (HUTTO and GILBERT, 2014). A detecção da polaridade possibilita identificar alterações na intensidade do sentimento ao longo do tempo, entre outras aplicações.

A próxima seção apresenta as principais técnicas de análise de sentimento que fazem uso de dicionários léxico.

2.3 TÉCNICAS DE ANÁLISE DE SENTIMENTO

Devido ao volume exorbitante de conteúdo textual produzido diariamente na web como, publicações de cunho jornalísticos, reviews sobre produtos, opiniões em blogs e menções em redes sociais, torna-se difícil aferir o sentimento de um texto visto a diversidade de vocabulário e semântica existentes. O uso de gírias, ironias, sarcasmos, erros ortográficos, presença de emoticons e vocabulário informal torna ainda mais desafiadora a tarefa de classificação de um sentimento.

Existem diversas técnicas de análise de sentimento na literatura que se diferenciam por suas regras, estratégias para aferir a polaridade de uma frase e técnicas de predição (LIU, 2012). Segundo Biswas (2020) a escolha da técnica de sentimento depende do conjunto de dados e da abordagem de raciocínio utilizada. Isso indica que, a depender da técnica utilizada, é possível apresentar cenários distintos como resultados.

Consoante a este contexto, a seguir são apresentadas cinco técnicas de análise de sentimento, sendo essas as mais difundidas no meio acadêmico, tendo inclusive inúmeros trabalhos relacionados, a saber: Vader, SenticNet, SentiStrength, SentiWordNet e TextBlob.

2.3.1 Vader

A técnica de Vader³ (*Valence Aware Dictionary for sentiment Reasoning*) é um classificador de sentimentos que tem como base regras de classificação simbólica. Em especial, tal técnica foi construída para ser utilizadas em redes sociais, onde se tem um contexto vago, ou seja, textos curtos, emoticons e gírias (HUTTO and GILBERT, 2014).

A atribuição de um sentimento é realizada por meio da separação da frase em palavras, dado que para cada palavra da frase é concedida uma pontuação de intensidade em uma escala de -4 até +4, sendo o primeiro um sentimento negativo, e o segundo positivo.

O quadro abaixo apresenta um pequeno exemplo da aplicação do classificador Vader em uma frase.

³ <http://comp.social.gatech.edu/papers/>.

Quadro 3 - Exemplo da aplicação da técnica Vader.

FRASE	<i>This place is great</i>			
PALAVRA	<i>This</i>	<i>Place</i>	<i>Is</i>	<i>great</i>
INTENSIDADE	0	0	0	3.1

Fonte: Elaborado pelo autor.

Analisando o Quadro 3, observa-se que a palavra *great* tem intensidade de 3.1, tal pontuação foi concedida por meio de um dicionário construído através do método *wisdom of the crowd*, em que um conjunto de pessoas foi responsável por designar as intensidades para cada palavra. Por outro lado, as palavras *this*, *place* e *is*, apresentam intensidade 0, pelo fato de não expressarem sentimento, e, consequentemente, não existirem no dicionário. Além disso, são empregadas diversas regras, as quais utilizam valores empiricamente derivados para inferir a intensidade do sentimento (HUTTO and GILBERT, 2014).

Entre as regras, verifica-se que quando uma palavra possui sentimento é escrita em letras maiúsculas, pois neste cenário é aumentada ou subtraída 0,733 na intensidade do sentimento, sem interferir na orientação semântica. Outra regra utilizada pelo Vader diz respeito à averiguação se alguma das três palavras anteriores é um adverbio intensificador, sendo acrescida ou diminuída em 0,293.

Verifica-se, também, se a frase possui a palavra *but*, pois esta aponta uma inversão do sentimento da frase, visto que o texto seguinte expressa um sentimento dominante. Para este cenário, o método multiplica a intensidade dos sentimentos expressos até a palavra *but* por 0.5, e os sentimentos expressos após a palavra, por 1.5. A Vader analisa se a frase apresenta pontos de exclamação, pois este tipo de pontuação acresce a intensidade, sem interferir na orientação semântica. Assim, é acrescentado 0,292 a cada ponto de exclamação, considerando-se um limite máximo de 4 pontos.

Outra regra da técnica Vader, consiste na investigação das três palavras anteriores a do sentimento, verificando a existência de uma negação que inverta a polaridade da frase. Dessa forma, quando é encontrada uma negação na frase, a intensidade de cada palavra de sentimento é multiplicada por -0,74.

Depois de calcular a intensidade, realiza-se a normalização da pontuação, por meio da equação 2.1. Assim, 15 é um valor fixo para aproximar o resultado final do valor máximo esperado, que é -1 para palavras negativas, e +1 para palavras positivas.

$$Pontuação Normalizada = \frac{Pontuação}{\sqrt{Pontuação^2 + 15}} \quad (2.1)$$

Utilizando-se a equação para a frase *This place is great*, será concedida uma pontuação final de 0,6249, a equação 2.2 ilustra tal instanciação.

$$Pontuação Normalizada = \frac{3.1}{\sqrt{3.1^2 + 15}} = 0,649 \quad (2.2)$$

Nos cenários em que a pontuação da frase for menor que -1 ou maior que 1, essa seria limitada aos valores de -1 ou 1, respectivamente. Assim, nesta técnica são consideradas frases positivas aquelas que possuem uma pontuação de 0.5 até 1; frases neutras, aquelas que apresentam uma pontuação -0.5 até 0.5; e frases negativas que consideram a pontuação de -1 até -0.5. Em vista disso, a frase *This place is great* seria atribuída como positiva.

Segundo o estudo conduzido por Hutto (HUTTO and GILBERT, 2014), Vader apresentou resultados superiores à técnica de *Naive Bayes*⁴, tendo uma assertividade de 96%, 63% e 55% na análise de tweets, avaliações de produtos da *Amazon* e editoriais do *New York Times*, respectivamente, contra 56%, 49% e 44% da técnica *Naive Bayes*.

2.3.2 SenticNet

O SenticNet⁵ é um recurso semântico e afetivo, disponível para o conceito de mineração de opinião e análise de sentimentos. Tal técnica foi construída por meio de Sentic Computing, um paradigma que explora Inteligência Artificial além de métodos de Web Semântica, a fim de identificar e processar opiniões de PLN por meio da web (CAMBRIA et al, 2010),

Criado em 2010, a partir de uma iniciativa concebida pelo *MIT Media Laboratory*, tal técnica tem sido utilizada em projetos de reconhecimento de emoção em áreas que vão desde a mineração de dados, até interação homem-máquina. Assim, o principal objetivo é fazer com que as informações conceituais e afetivas transmitidas por linguagem natural estejam mais facilmente acessíveis às máquinas. Para isso, busca inferir a polaridade das palavras associada a conceitos de senso comum, e representá-la em um formato semântico. Tais conceitos podem

⁴ Algoritmo que se baseia nas descobertas de Thomas Bayes para realizar previsões em aprendizagem de máquina.

⁵ <http://sentic.net/>.

ser formados por uma ou várias palavras associadas, como exemplo: o conceito *afraid* e *afraid death*.

SenticNet possui aproximadamente 14.000 conceitos e concede uma pontuação para a polaridade na escala entre -1(negativo) a 1(positivo), com valores neutros igual a zero. Desta forma, conceitos com a polaridade menor que zero, indicam que o sentimento apresentado é de cunho negativo. Em contrapartida, conceitos com o valor da polaridade maior que zero apresentam sentimento positivo. O Quadro 4 mostra um exemplo da aplicação da técnica para a frase “*Boring, its’s Monday Morning*”.

Quadro 4 - Exemplo da aplicação da técnica SenticNet.

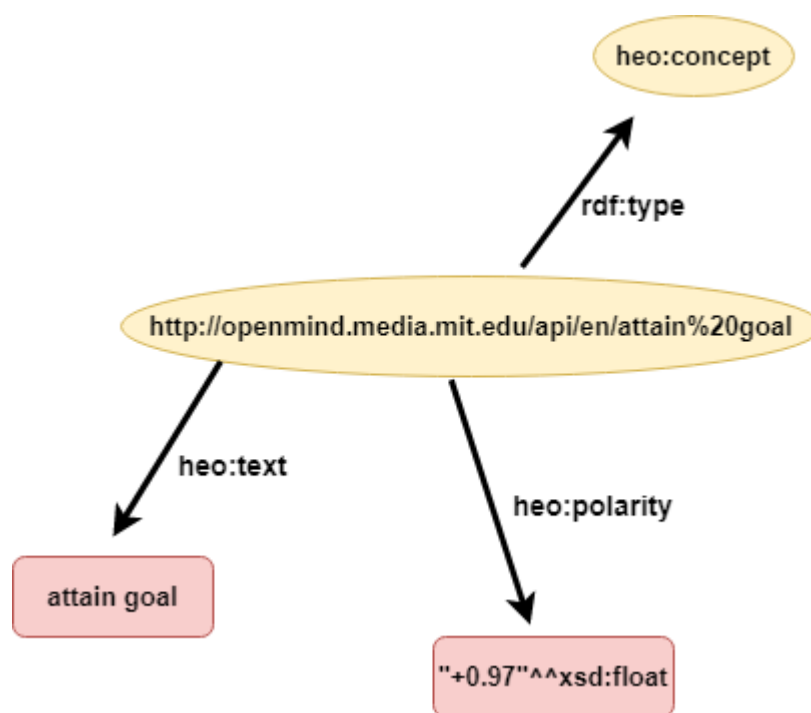
FRASE	<i>Boring, its’s Monday Morning</i>	
CONCEITOS	<i>Boring</i>	<i>Monday Morning</i>
POLARIDADE	-0,383	+0,228
POLARIDADE FINAL	-0,077	

Fonte: Elaborado pelo autor.

Inicialmente, a técnica tenta reconhecer conceitos, que neste exemplo seriam *boring* e *Monday morning*, por conseguinte, computa a polaridade para cada conceito, no caso -0,383 para *boring*, e +0,228 para *Monday morning*. A polaridade final consiste na média dos valores encontrados para cada conceito, resultando em -0,077.

O acesso ao SenticNet é por meio do formato RDF/XML. A Figura 4 ilustra um conceito, modelado como um arquivo RDF.

Figura 4 - Modelo de um conceito do SenticNet no formato RDF/XML.



Fonte: 1 Elaborado pelo autor.

Analisa-se que, para um determinado conceito, têm-se o texto, que apresenta o nome do conceito e a polaridade, que demonstra o quanto ele é negativo ou positivo. O SenticNet associa pontuações de polaridade a um subconjunto de conceitos. Assim, a técnica SenticNet está estruturada, de modo que cada conceito possui uma tag RDF, que possui atributos para selecionar o tipo de recurso, texto e sua polaridade.

A validação da técnica foi realizada por meio da medição de níveis de polaridade em opiniões de pacientes sobre o *National Health Service*, na Inglaterra, e em uma base de dados coletadas do *Live Journal*, em que mensagens foram rotuladas em 130 estados de humor pelos próprios usuários, e convertidas em negativo e positivo (CAMBRIA et al, 2010).

2.3.3 SentiStrength

A técnica SentiStrength⁶ utiliza um dicionário, anotado por humanos, e aperfeiçoado com a execução de aprendizado de máquina. Para isso, ele concede pontuações às palavras de um dicionário, onde *emoicons* também fazem parte. Valores entre 1 e 5 são atribuídas para

⁶ <http://sentistrength.wlv.ac.uk/>.

emoções de cunho positivo, em contrapartida valores entre -5 e -1 são concedidos para as palavras de cunho negativo.

Para computar o sentimento, o SentiStrength separa a frase em trechos de uma ou mais sentenças e concede valores negativos ou positivos para cada uma, indicando a pontuação mínima ou máxima entre todas as palavras de uma sentença. A classificação final para a frase é a força máxima positiva e negativa de cada sentimento (THELWALL, 2013). O Quadro 5 apresenta a aplicação do SentiStrength,

Quadro 5 - Exemplo da aplicação da técnica SentiStrength.

FRASE	<i>Ronnie is horrible and nasty but I am lovely.</i>
SENTENÇA 1	<i>Ronnie is horrible [-4]</i>
SENTENÇA 2	<i>and nasty [-3]</i>
SENTENÇA 3	<i>but I am lovely [2]</i>
SENTIMENTO	Negativo

Fonte: Elaborado pelo autor.

Observa-se que, após encontrar o máximo valor negativo (-4) e positivo (2) de sentimento nas sentenças, a classificação final para a frase é atribuída por meio da comparação do maior grau de sentimento, ou seja, neste caso -4, resultando em um sentimento negativo.

O SentiStrength utiliza para classificação uma versão expandida do dicionário *do Linguistic Inquiry and Word Count* (LIWC) (TAUSCZIK AND PENNEBAKER, 2010), com a adição de características para o contexto de redes sociais. Essas características estão expressas em quatro arquivos de texto, que contêm um conjunto extra de palavras, emoticons, palavras de negação, gírias e palavras de intensidade, juntamente com seus respectivos rótulos de classificação, são os arquivos:

- *EmoticonLookupTable.txt* – retrata os emoticons com as suas respectivas emoções, por exemplo, o emoticon “=(” apresenta sentimento de cunho negativo, pois trata de uma carinha triste;
- *EmotionLookupTable* – apresenta como intuito relacionar uma palavra com um peso positivo ou negativo, sendo que tais pesos ficam em uma escala de -5 e 1 para palavras negativas, e entre 1 e 5 para palavras de cunho positivo, o 0 indica neutralidade;

- *NegatingWordList.txt* – tem como foco identificar palavras de negação;
- *BoosterWordList.txt* – apresenta como propósito intensificar as emoções, por exemplo, “*very beautiful*”, a palavra “*very*” presente na frase estaria intensificando a emoção.

Segundo Thelwall, a técnica SentiStrength faz uso de técnicas de pré-processamento, como: uso de algoritmos de correção ortográfica, que reconhece a grafia padrão de palavras que foram digitadas de modo incorreto pela adição de letras repetidas, ou ainda uma lista de palavras de reforço, que contêm palavras que aumentam ou reduzem a emoção das palavras subsequentes. Além disso, embora exista dificuldade em detectar sarcasmos e ironias, a técnica apresentou resultados significativos e uma alta versatilidade de aplicações em diversos contextos.

Thelwall *et. al.* (2010) apresenta a proposta do classificador, o qual realiza testes comparativos com outras técnicas de classificação, sendo que o SentiStrength apresentou resultados melhores quando comparados às técnicas de aprendizado de máquina.

2.3.4 SentiWordNet

Outra técnica existente, desenvolvida por Esuli (2008) para colaborar em aplicações para classificação de opinião, é o SentiWordNet⁷. Tal técnica, possui 117.374 entradas resultantes de anotações automatizadas em todos os *synsets*⁸ da base do WordNet⁹ (MILLER, 1995), com scores de sentimentos. Tal base de dados agrupa adjetivos, verbos e outras classes gramaticais.

No SentiWordNet, cada *synset* é relacionado a três valores numéricos, Pos(s), Neg(s) e Obj(s) que apontam o quanto é positivo, negativo ou neutro. Cada um dos três valores varia no intervalo [00, 1.0], e a soma deles é 1.0. A equação 2.3 apresenta o cálculo realizado para cada *synset*:

$$Obs(s) + Pos(s) + Neg(s) = 1 \quad (2.3)$$

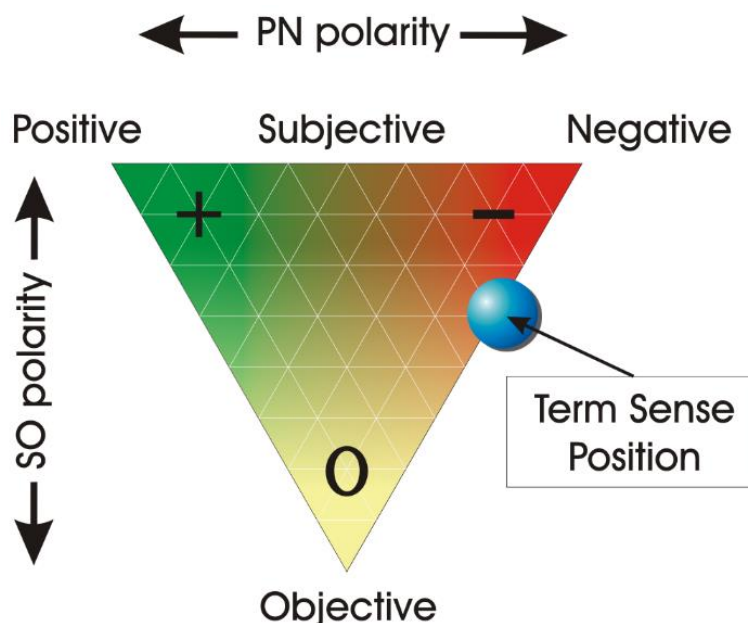
⁷ <http://sentiwordnet.isti.cnr.it/>

⁸ conjunto de sinônimos representando um conceito

⁹ grande banco de dados lexical

Ilustra-se, na Figura 5, a representação gráfica utilizada pelo SentiWordNet, demonstrando as propriedades relacionadas ao synset. As bordas do triângulo caracterizam uma das três classificações, e um ponteiro (*term sense position*) indica para a classificação de maior valor.

Figura 5 - Representação adotada pelo SentiWordNet.



Fonte: Retirado de <http://sentiwordnet.isti.cnr.it/>.

O método utilizado na construção do SentiWordNet tem como base a análise quantitativa de termos relacionados a *synset* e o uso de representação vetorial de termos resultantes para classificação semi-supervisionada de *synsets*. Assim, o método depende da formação de um conjunto de classificadores ternários, cada um deles indicando se um *synset* é positivo, objetivo ou negativo. Cada classificador ternário produz diferentes categorias dos *synsets* do WordNet, sendo que os scores relacionados às opiniões de cada *synset* são definidos pela proporção dos classificadores ternários que tenham dado o correspondente rótulo a ele. Dessa forma, se todos os classificadores ternários apresentarem o mesmo rótulo a um *synset*, esse terá o maior score, caso contrário, cada rótulo terá um score proporcional aos classificadores (ESULI and SEBASTIANI, 2006).

A versão do SentiWordNet 3.0¹⁰ conta com uma versão web para consulta online, além de disponibilizar a base de dados do léxico em arquivo de dados .txt. O Quadro 6 apresenta a estrutura dos dados do léxico SWN, onde o *POS-ID* identifica o *synset*, e os campos *PosScore*

¹⁰ <https://www.nltk.org/howto/sentiwordnet.html>

e *NegScore* dizem respeito à positividade e negatividade, respectivamente. O cálculo da objetividade é dado pela equação 2.4:

$$Obj(s) = 1 - (Pos(s) + Neg(s)) \quad (2.4)$$

O campo *SynseTerms* refere-se a termos separados por espaço, pertencente ao synset, com a classe gramatical e número equivalente ao sentido. Por fim, o campo *Gloss* apresenta o sentido do termo.

Quadro 6 - Estrutura da base de dados do léxico SentiWordNet.

POS	ID	POSSCORE	NEGSCORE	SYNSETERMS	GLOSS
A	00016756	0	0.25	<i>scarce#1</i>	<i>deficient in quantity or number compared with the demand;</i>
A	00851744	0.5	0	<i>eligible#1</i>	<i>qualified for or allowed or worthy of being chosen;</i>
A	00852101	0.375	0	<i>bailable#2</i>	<i>eligible for bail;</i>

Fonte: Retirado da base de dados do léxico SentiWordNet.

2.3.5 TextBlob

TextBlob¹¹ é uma biblioteca desenvolvida na linguagem de programação *Python*, que disponibiliza dois modelos de classificação treinados, com conjuntos de dados na língua Inglesa fornecidos pela biblioteca NLTK (LORIA, 2018). O classificador padrão disponibilizado por este sistema pertence à biblioteca *Pattern* (SMEDT AND DAELEMANS, 2012), que recorre a um léxico de adjetivos frequentes em resenhas de filmes.

A técnica *Pattern* é organizada em vários módulos para mineração de dados, aprendizado de máquina, visualização e análise de sentimento. No módulo de análise de

¹¹ <https://pypi.org/project/textblob/>.

sentimento, a pontuação da polaridade é um float dentro do intervalo $[(-1,0), (1,0)]$. A subjetividade é um float dentro do intervalo $[(0,0), (1,0)]$, de modo que 1,0 é bastante subjetivo, e 0,0 é bastante objetivo. Assim, a técnica apresenta três classes de polaridade classificadas como: positivo $[(0,01), (1,0)]$, neutra $(0,0)$, e negativo $[(-0,01), (-1,0)]$. Desta maneira, para calcular o valor da subjetividade ($\bar{S}$) e polaridade ($\bar{P}$) de um único termo, a técnica utiliza o conceito de média aritmética simples, de modo que a soma dos valores da polaridade do termo $\sum_{i=1}^n X_i$ é dividido pelo número de vezes (n) que este termo aparece dentro do conjunto de dados. A equação 2.5 e 2.6 mostram o cálculo da polaridade e subjetividade de uma palavra do conjunto de dados:

$$\overline{P_{(x)}} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.5)$$

$$\overline{S_{(x)}} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.6)$$

A título de exemplo, a palavra *great* surge em quatro ocasiões no conjunto de dados léxico, tendo como polaridades $[(1,0), (1,0), (0,4) \text{ e } (0,8)]$ e subjetividades $[(1,0), (1,0), (0,2) \text{ e } (0,8)]$ distintas, mas com intensidades equivalentes a 1,0. As equações 2.7 e 2.8 demonstram o cálculo da polaridade e subjetividade, resultando em 0,8 e 0,75, respectivamente.

$$\overline{P_{(x)}} = \frac{1,0 + 1,0 + 0,4 + 0,8}{4} = 0,8 \quad (2.7)$$

$$\overline{S_{(x)}} = \frac{1,0 + 1,0 + 0,2 + 0,8}{4} = 0,75 \quad (2.8)$$

Em relação ao cálculo da polaridade ($\overline{PF}$) e subjetividade ($\overline{SF}$) em frases, a técnica realiza a média aritmética simples com o conjunto de valores da polaridade ($\bar{P}$) ou subjetividade ($\bar{S}$) dos termos na frase. Deste modo, obtêm-se o valor final da frase calculando e dividindo-se a soma dos valores dos termos que apresentam sentimento na frase pela quantidade de tais palavras.

A seguir, a equação 2.9 e 2.10 demonstram a fórmula para calcular a polaridade e a subjetividade de uma frase que possui n termos no conjunto de dados léxico.

$$\overline{PF_{(P)}} = \frac{1}{n} \sum_{i=1}^n \overline{P_i} \quad (2.9)$$

$$\overline{SF_{(S)}} = \frac{1}{n} \sum_{i=1}^n \overline{S_i} \quad (2.10)$$

A título de exemplo, o Quadro 7 apresenta a aplicação do algoritmo para a frase “*The world is beautiful and great*”. Observa-se que existem palavras que não estão no conjunto de dados do léxico. Por outro lado, a palavra “*beautiful*” e “*great*” apresentam duas e quatro entradas respectivamente.

Quadro 7 - Exemplo de aplicação da técnica Textblob.

FRASE	POLARIDADE $\overline{P_{(x)}} = \frac{1}{n} \sum_{i=1}^n x_i$	SUBJETIVIDADE $\overline{S_{(x)}} = \frac{1}{n} \sum_{i=1}^n x_i$
<i>The</i>		
<i>world</i>		
<i>is</i>		
<i>beautiful</i>	$\overline{P_{(x0)}} = \frac{0,7 + 1,0}{2} = 0,85$	$\overline{S_{(x0)}} = \frac{1,0 + 1,0}{2} = 1,0$
<i>and</i>		
<i>great</i>	$\overline{P_{(x1)}} = \frac{1,0 + 1,0 + 0,4 + 0,8}{4} = 0,8$	$\overline{S_{(x1)}} = \frac{1,0 + 1,0 + 0,2 + 0,8}{4} = 0,75$

Fonte: Elaborado pelo autor.

Finalmente, as equações 2.11 e 2.12 exibem a polaridade e subjetividade final da frase investigada.

$$\overline{PF_{(P)}} = \frac{0,85 + 0,8}{2} = 0,825 \quad (2.11)$$

$$\overline{SF_{(s)}} = \frac{1,0 + 0,75}{2} = 0,875 \quad (2.12)$$

O Pattern apresenta, regras no processo de cálculo do sentimento em frases: com palavras de negação, advérbios, negação e advérbios e frases com emoticons.

2.4 ANÁLISE E COMPARAÇÃO DAS TÉCNICAS DE SENTIMENTO

Muitas técnicas estão sendo aplicadas no desenvolvimento de sistemas sem um entendimento concreto da sua aplicabilidade em diferentes contextos, suas limitações e vantagens comparadas às demais técnicas.

Nesse cenário, o Quadro 8 apresenta uma descrição a respeito de cada uma das técnicas investigadas nesta pesquisa, trazendo informações como: *datasets* utilizados para validá-los, as saídas da técnica, como por exemplo, -1 (negativo), 0 (neutro) e +1(positivo), trabalhos utilizados como baselines, tamanho do dicionário léxico, além de apontar o uso ou não de algumas estratégias na classificação, tais como: web semântica, emotions, gírias, palavras de intensidade, capitalização, correção ortográfica e identificação de sarcasmo e ironia.

Quadro 8 - Características das Técnicas de Inteligência Artificial.

CARACTERÍSTICAS	TÉCNICAS DE SENTIMENTO				
	VADER	SENTICNET	SENTISTRENGTH	SENTIWORDNET	TEXTBLOB
Validação	Twitter, Reviews de filmes, Reviews técnicos de produtos, opiniões de usuários do NYT	Opiniões de pacientes	Twitter, Youtube, Digg, Myspace, BBC Forums and Runners Wordls	Baseado no dicionário WordNet	Reviews de produtos
Comparado	GI (STONE et.al, 1967), LIWC,(TAUSCZIK AND PENNEBAKER, 2010), SentiWordNet (ESULI and SEBASTIANI, 2006), SenticNet (CAMBRIA et al, 2010) e algumas técnicas de aprendizado de máquina	SentiStrength (THELWALL, 2013)	Nove melhores técnicas de Aprendizado de Máquina para cada teste	General Inquirer (STONE et.al, 1967)	Não informado
Saída	-1, 0, +1	Negativo, Positivo e Neutro	-5 ,0, +5	Negativo, Positivo e Neutro	-1, 0, +1

Tamanho do dicionário léxico	7.517	14.000	2.698	117.658	2973
Web Semântica	Não	Sim	Não	Não	Não
Emoticons	Sim	Sim	Sim	Não	Sim
Gírias	Sim	Não	Não	Não	Não
Palavras de Intensidade	Sim	Não	Sim	Não	Sim
Capitalização	Sim	Não	Sim	Não	Não
Semântico	Não	Sim	Não	Sim	Não
Sintático	Sim	Não	Sim	Não	Sim
Correção ortográfica	Não	Não	Sim	Não	Sim
Sarcasmo e Ironia	Sim	Não	Sim	Não	Não

Fonte: Elaborado pelo autor.

Aplicar técnicas de Inteligência Artificial no contexto da Altmatria, especialmente em comentários da web social, possibilita analisá-los e classificá-los de acordo com a polaridade dos mesmos, ou seja, positivo, negativo ou neutro, possibilitando inclusive o incremento de classes intermediárias que indicam variações de avaliações positivas ou negativas, e consequentemente, auxiliar na construção de uma perspectiva qualitativa aos já conhecidos indicadores quantitativos.

2.5 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os principais conceitos envolvidos na área de Inteligência Artificial. Em seguida foi descrito o conceito de Análise de sentimento, bem como as duas principais abordagens: baseada em aprendizado de máquina e baseada em dicionários léxicos.

Foram detalhadas cinco técnicas de análise de sentimento com o objetivo de investigar seu funcionamento, características e estratégias utilizadas empregadas no processo de detecção de sentimento. Tal investigação se faz importante uma vez que podem gerar resultados distintos, em especial no cenário da Altmatria.

No próximo capítulo, serão apresentados fundamentos relacionados à Medição da produção científica. Para tanto, é apresentado o conceito dos indicadores tradicionais e alométricos. Além disso, é explanado o conceito de Altmatria, suas ferramentas e o cálculo utilizado para aferir trabalhos na plataforma *Altmetric.com*.

MEDIÇÃO DA PRODUÇÃO CIENTÍFICA

Tradicionalmente, a medição da produção científica tem sido efetuada com o intuito de proporcionar subsídios para o estabelecimento e acompanhamento de políticas nacionais e internacionais de ensino e pesquisa, assim como apoiar nas decisões estratégicas dos gestores de políticas tecnológicas e científicas. Isso ocorre visto que essa medição possibilita um diagnóstico das reais competências de certos grupos e/ou instituições de pesquisa, o monitoramento de distintas áreas do conhecimento e a identificação dos projetos mais promissores (OLIVEIRA et al., 1992; SANTOS, 2003). Nesse cenário, tornou-se essencial a estruturação de sistemas de informação tecnológico e científico que viabilizassem o desenvolvimento de metodologias, indicadores e ferramentas para essa finalidade (SANTOS, 2003).

Os indicadores representam um índice ou uma medida que possibilita avaliar o desempenho de um fenômeno, do seu estado e de sua evolução, tratando-se da avaliação da ciência, permitem aferir as ações sistemáticas relacionadas com a produção e aplicação dos conhecimentos científicos tecnológicos, permitindo a padronização dos procedimentos e servindo de parâmetro para avaliações distintas (VANTI, 2011). Desta forma, nesta seção, serão abordados os principais indicadores utilizados para mensurar a produção científica.

3.1 INDICADORES TRADICIONAIS

Os indicadores bibliométricos tradicionais medem o impacto da pesquisa por meio de índices quantitativos baseados em dados de citações, como por exemplo o fator de impacto do periódico e o índice h (WAHEED, 2018).

Proposto por Eugene Garfield, em 1955, o Fator de Impacto (FI) começou a ser utilizado como ferramenta de avaliação da qualidade das publicações em meados dos anos 1960, tendo sido utilizado como critério de seleção dos periódicos a serem indexados pelo *Science Citation Index* (SCI). A partir deste momento, o Fator de Impacto estabeleceu-se como meio de avaliação dos periódicos nas inúmeras instâncias.

Anualmente é calculada pelo *Institute for Scientific Information / Thompson Scientific Reuters* para as revistas indexadas em sua base de dados e é publicado pelo *Journal Citations Reports* (JCR).

Frequentemente, autores consideram o valor do FI para selecionar o periódico que possa proporcionar maior visibilidade ao trabalho, os bibliotecários consideram o FI como um parâmetro de seleção dos títulos de maior interesse científico e que, por conseguinte, deverão formar o acervo das instituições.

De modo paralelo, os editores assistem atentos os FI dos periódicos, cientes da relevância desse índice como fator de influência na captação de recursos financeiros e na atração de bons trabalhos a serem publicados.

Para o cálculo do FI de um certo ano, são considerados o número de citações recebidas naquele ano pelos artigos publicados pelo periódico nos dois anos precedentes, dividido pelo número de artigos publicados pelo periódico no mesmo período. Dessa forma, o FI é eficaz em avaliar a qualidade de um periódico, não sendo porém, útil na análise da qualidade científica de um trabalho científico isolado, de um pesquisador (GISBERT, 2009).

Outras medidas surgiram em resposta ao FI como o Índice-H proposto por Jorge E. Hirsch para classificar e posicionar autores segundo suas distribuições de citação (HIRSCH, 2005). Com isso, esse índice, que é baseado no número de publicações e citações, pode indicar quais são os *experts* de um determinado campo científico ou instituição, medir o desempenho de um pesquisador no decorrer de sua carreira e identificar os pesquisadores mais citados.

O cálculo do índice h leva em consideração o equilíbrio de dois fatores quantitativos, que são: quantidade de citações e quantidade de publicações.

Com a chegada da internet e a Web, houve a demanda da criação de índices de impacto para a avaliação da produção científica, assim, com o surgimento e crescimento da Webometria, a partir da análise de links de sítios, que se criou o Fator de Impacto Web (FIW) (INGWERSEN, 1998) para aferir a influência e o nível de atratividade desses novos recursos de informação na Web (VANTI, 2014).

Com a evolução da Web 2.0, a web social é que têm melhor refletido o impacto e a influência das atividades acadêmicas, ainda, motores de busca como o Google, também impulsionaram diversas medidas de impacto aplicáveis à análise de redes sociais por meio da citação nas redes (BOLLEN, 2009).

3.2 INDICADORES ALTMÉTRICOS

Métricas alternativas ou Altmatria (do Inglês, *Altmetrics*) são indicadores da comunicação científica, produzidos por meio de dados de interações em ambientes e

ferramentas na internet (NASCIMENTO, 2016). Por meio deles, é possível medir e quantificar o impacto dos trabalhos científicos proporcionando apoio na utilização de plataformas que facilitam a produção e comunicação do conhecimento.

O surgimento da Altmatria originou-se acerca da influência de determinados princípios que vêm se formando nas instituições e redes sociais, como: o nascimento de novas ferramentas sociais que contribuíram para a evolução nas formas de comunicação, a carência de filtros para preferir a informação mais significativa dentro da ciência, o descontentamento com as abordagens tradicionais de medição de acordo com o impacto científico, e, por fim, o movimento denominado *open access*, que popularizou a divulgação e o alcance a qualquer tipo de informações no meio científico (PRIEM, 2010).

O texto que marca a origem de uma nova abordagem para aferir a produção científica nos diversos recursos informacionais pelos meios tecnológicos foi escrito por Jason Priem em 2010 e intitulado como “Altmatria: um manifesto” (PRIEM, 2010). Tal texto, discute a proposta de contabilizar as citações e a análise do impacto acadêmico através dos inúmeros recursos das redes sociais disponíveis, considerando a realidade atual de velocidade e o uso de tecnologias pela comunidade acadêmica. A ideia consiste na existência de uma grande quantidade de artefatos acadêmicos sendo avaliados na web social, de modo que possibilitam a disseminação dos conteúdos e analisam o impacto acadêmico fora da academia.

Em “*The Collection Altmetrics*” (PRIEM, 2012), a Altmatria é definida como:

O estudo e uso de medidas de impacto com base nas atividades acadêmicas e em ferramentas e ambientes online. O termo também tem sido utilizado para descrever métricas próprias, que podem ser expressas no plural “um conjunto de novas Altmetrics”. Altmatria é, na maioria dos casos, fruto da intersecção de ambas métricas, a Cientometria e a Webometria, sendo um subconjunto desta última, que se concentra mais especificamente na influência acadêmica, medida por meio de ferramentas e ambientes em linha, mais do que na Web em geral.

De maneira mais minuciosa, Taylor (Taylor, 2013) justifica que:

Altmatria é o termo coletivo para os dados de uso acadêmico que vão além de contagem de citações formais. Os dados altmétricos vêm de plataformas especializadas e ferramentas de pesquisa, porém, também incluem dados de aplicações gerais e plataformas técnicas. Além disso, o termo abrange referências da mídia de massa e dados de editoras, como visualizações de páginas web e downloads.

Por sua vez, para Alperin (ALPERIN, 2013) tais métricas concentram-se, exclusivamente, em aferir o impacto acadêmico na web. Já Thelwall (THELWALL, 2013) apresenta que a Altmétrie se refere às contagens de citações ou menções nas mídias sociais.

Certamente, o aparecimento das métricas alternativas tornou possível medir e importar práticas que são comuns na comunidade acadêmica, para o ambiente online, como: compartilhamento de artigos, discussões informais e o uso de resultados de pesquisas fora do âmbito acadêmico. Além disso, a Altmétrie possibilita que pesquisadores observem onde e por quem seu trabalho está sendo visualizado, recomendado, discutido e compartilhado (NASCIMENTO, 2014).

3.3 FERRAMENTAS ALTMÉTRICAS

Nos últimos anos ferramentas alométricas foram desenvolvidas com o intuito de mensurar e coletar os indicadores alométricos, o que possibilitou a construção de diversos estudos de revisão e aplicação.

Gracio (2020) apresenta um estudo que mesmo mensurando as mesmas fontes “[...] há variações de scores, dependendo de um agregador ou outro ser utilizado, de modo que tanto o uso de uma fonte quando de uma metodologia, podem fazer a diferença.”

Existem algumas ferramentas que proporcionam, o monitoramento de dados alométricos referentes a um ou vários documentos de forma integrada em plataformas online, como: ImpactStory, o PLOS ALM, Plum Analytics e o Altmetric. Sendo, este último, a principal empresa deste grupo (NASCIMENTO, 2016). Tais ferramentas preocupam-se de modo exclusivo na produção de métricas alternativas, entretanto cada abordagem tem buscado diferenciar-se dos demais, dedicando-se a atender a um determinado segmento acadêmico.

A ImpactStory tem como foco pesquisadores individuais, em contrapartida a PLOS ALM responde às demandas de editores científicos. Já a Plum Analytics desenvolve ferramentas para o mercado institucional acadêmico. Por fim, a Altmetric tem um caráter mais diversificado, apresentando ferramentas direcionadas para pesquisadores, editores e bibliotecários.

A seguir são apresentadas cada uma das ferramentas mencionadas.

3.3.1 ImpactStory

O ImpactStory¹² é uma ferramenta direcionada para usuários individuais, que possibilita aos pesquisadores criar um perfil, adicionar itens de sua produção acadêmica, como por exemplo: conjuntos de dados, apresentações, artigos, e assim por diante, sendo possível monitorar as menções recebidas. Tal ferramenta foi lançada no ano de 2011 e era intitulada Total Impact.

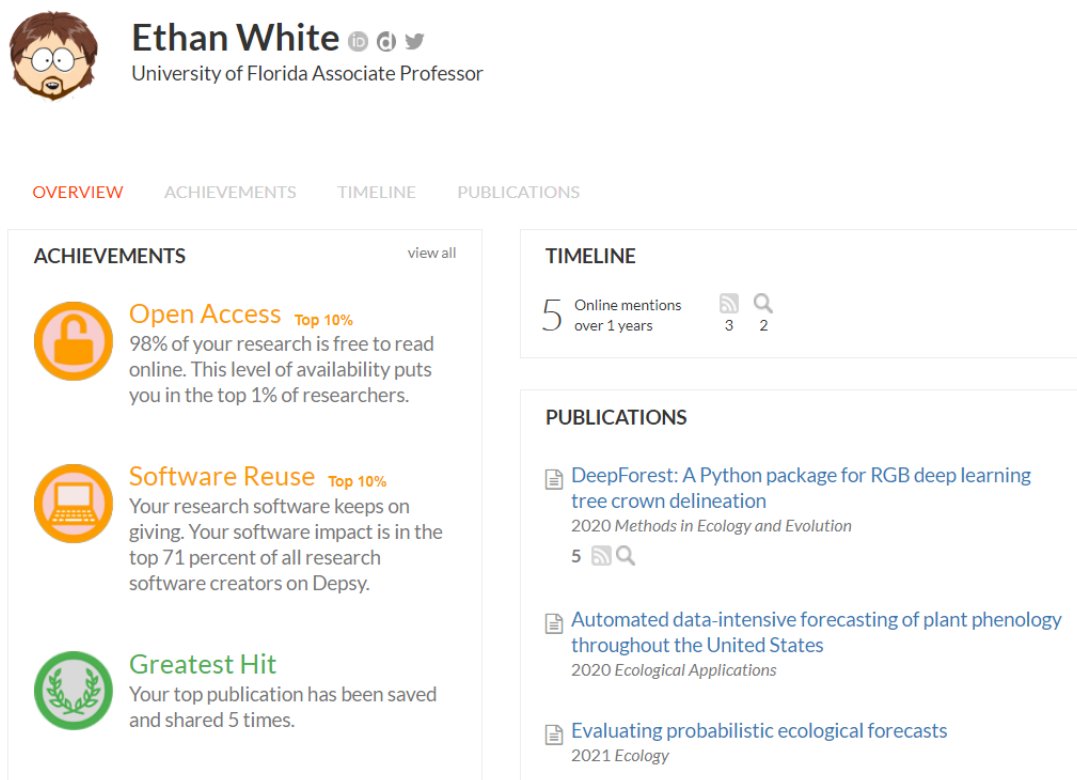
Em 2014 o ImpactStory começou a cobrança de uma assinatura anual para os usuários, porém voltou a ser concedida de forma gratuita no início do ano de 2016.

Em relação ao monitoramento da ferramenta, pode-se citar diversas redes sociais e coleta de dados como: visualizações de vídeos no Youtube e Vimeo, contagem de citações na base de dados Scopus, número de downloads no SlideShare, Figshare e Dryad, menções, e discussões nas redes sociais como Facebook, Twitter, Wikipedia, Mendeley, além de postagens em blogs.

A Figura 6 ilustra um exemplo de um perfil de usuário do site ImpactStory em companhia com as suas principais publicações acadêmicas. O usuário cadastrado nesta plataforma recebe emails semanais indicando possíveis novas ocorrências de acessos, visualizações e compartilhamento de seus artefatos de pesquisa.

¹² <https://profiles.impactstory.org/>

Figura 6 - Exemplo de perfil de usuário no ImpactStory.



Fonte: Retirada de <https://profiles.impactstory.org/u/0000-0001-6728-7745>.

3.3.2 PLOS ALM

Um dos pioneiros em se preocupar com as métricas em nível de artigo (*Article-Level Metrics*), foi o PLOS (*Public Library of Science*), com a ferramenta intitulada como PLOS ALM criada em 2009.

O megaperiódico foi elaborado com a intenção de não precisar de ferramentas de terceiros para gerar suas métricas, tendo como base um conjunto de fontes como: download e citações em HTML, XML e PDF no Pubmed Central, citações em bases de dados como Scopus, Google Acadêmico, CrossRef e Web of Science, entre outros serviços oferecidos (NASCIMENTO, 2016).

O monitoramento dos artigos a curto e longo prazo com indicadores de estatística de uso, bookmarks e citações podem ser visualizados na página do artigo. A Figura 7 ilustra como as informações das métricas do artigo são apresentadas no PLOS ALM¹³.

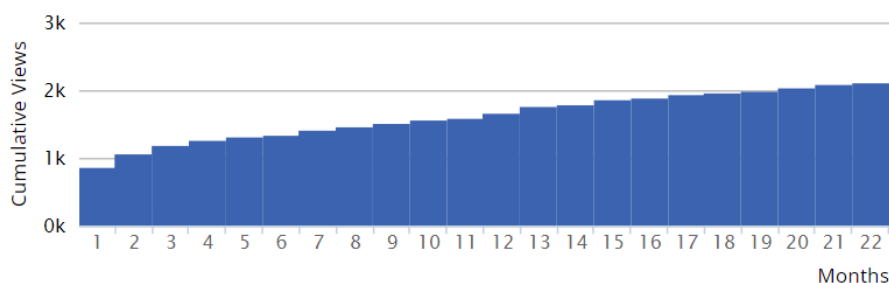
¹³ <https://plos.org/publish/metrics/>

A princípio a PLOS ALM utilizava apenas os artigos publicados em seus periódicos, porém com o passar do tempo começou a se utilizar por outras publicações, sendo que nos dias de hoje diversos periódicos também publicam métricas de seus artigos aderindo a PLOS ALM.

As informações do megaperiódico podem ser consultadas através de relatórios personalizados produzidos pelo aplicativo criado pela PLOS intitulado Lagotto, que retrata informações das fontes onde os usuários estão discutindo, mencionando, compartilhando ou recomendando trabalhos.

Figura 7 - Informações de artigo da PLOS ALM.

Total Article Views	HTML Page Views	PDF Downloads	XML Downloads	Total
2.110	777	680	653	2.110
87,516 % of article views led to PDF downloads				
Nov 11, 2020 (publication date) through Aug 22, 2022 *				



*Although we update our data on a daily basis, there may be a 48-hour delay before the most recent numbers are available.

Fonte: Extraído de <https://journals.plos.org/ploscompbiol/article/metrics?id=10.1371/journal.pcbi.1008227>.

3.3.3 Plum Analytics

A Plum Analytics¹⁴ foi criada em 2012 e tem como intuito as instituições acadêmicas e de pesquisas sendo o PlumX seu principal produto.

O PlumX é responsável pela coleta e organização de métricas alternativas relacionadas a mais de vinte diferentes tipos de artefatos, contendo livros, vídeos, artigos de periódicos, trabalhos de conferência, código de programação entre outros. Sendo utilizado por institutos de pesquisas como uma forma de mensurar a produtividade de pesquisadores, departamentos e grupos de pesquisa (NASCIMENTO, 2016). Entretanto, é uma ferramenta que funciona com

¹⁴ <https://plumanalytics.com/>

base ao pagamento de uma assinatura, deste modo, nenhuma API aberta (interface de programação de aplicação) pode utilizá-la.

Dentre as informações utilizadas pela ferramenta podemos citar mais de trinta fontes de dados externas, como: Pubmed, Plos e Youtube, sendo organizadas em cinco categorias:

1. Uso – responsável por medir o número de vezes que o produto foi visto;
2. Captura – quantas vezes foi marcado como favorito;
3. Menção – Quantidade de comentários e post;
4. Mídia social – Compartilhamentos e curtidas;
5. Citações - número de citações recebidas na Pubmed e outras fontes

Consoante a essas categorias, o resumo das métricas podem ser aplicadas ao repositório institucional ou em gráficos individuais por tipo de pesquisador ou unidade organizacional (NASCIMENTO, 2016).

3.3.4 Altmetric

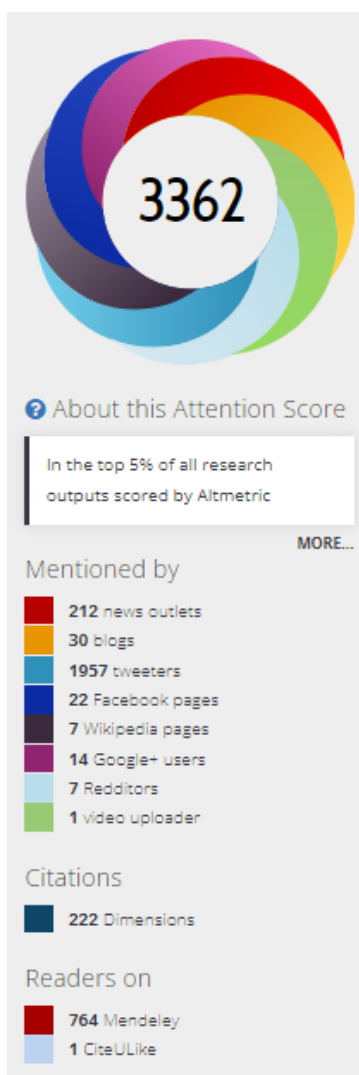
A Altmetric, foi fundada em 2011, em Londres. Tal companhia possui suas métricas na base de dados da Scopus, nas coleções Scielo, no site da revista Nature e em muitas outras bases de dados e periódicos. Sua atividade baseia-se na coleta metadados de citações, menções, comentários e downloads de artigos acadêmicos por meio de diversas fontes disponíveis na rede, como o gestor bibliográfico Mendeley, sites de notícias, revistas comerciais e jornais, blogs acadêmicos e mídias sociais, como Twitter, Google+, Facebook, Reddit, Youtube, dentre outros.

Ainda, passou a recolher menções de artigos acadêmicos em documentos de políticas públicas, evidenciando uma nova categoria da relevância da produção acadêmica, diferente das formas tradicionais de citações em periódicos científicos. Com o intuito de que os metadados de citação de um artigo sejam coletados e armazenados, é necessário que o trabalho acadêmico tenha, no mínimo, uma classe de identificador único, como um DOI (Digital Object Identifier) ou PMID (PubMed record ID), que possibilitam a distinção exclusiva e a condução de citações de maneira inequívoca.

É importante ressaltar que essa plataforma oferece relatórios que dizem respeito aos dados quantitativos dos conjuntos de trabalhos. Neste sentido, é essencial analisar como alguns

trabalhos apresentam-se em relação a outros, de modo que a pontuação medida pode formar o conceito da atenção do trabalho na web social. Essa pontuação é representada por uma contagem ponderada, considerando as diversas fontes que mencionam os trabalhos científicos, sendo possível uma reflexão quando à importância relativa de cada uma delas, visto que é concedida maior pontuação a uma menção feita em um sistema de promoção de notícias, do que em uma rede social. A título de exemplo, a Figura 8 ilustra as fontes que são utilizadas para compor a pontuação do altmetric score para o trabalho intitulado “*Instagram fotos reveal predictive markers of depression*”.

Figura 8 - Detalhamento das fontes utilizadas para compor o Altmetric Score de um trabalho.



Fonte: Extraída do site Altmetric.com.

Nota-se por meio da análise da figura as diferentes fontes de dados coletadas e o que cada cor representa na construção do *donut*. Para o cálculo altmétrico, a pontuação é derivada de um algoritmo automatizado, que utiliza o Quadro 9 de pesos abaixo.

Fang e Costas (2021) apresentaram que não são todos os indicadores que se acumulam na mesma velocidade, de modo que existem fontes rápidas e lentas. Com base nisso, apontaram que juntamente com a velocidade da acumulação, os diversos tipos de dados altmétricos também estão associados e modificam de acordo com o documento, assunto e tópico de pesquisa.

Quadro 9 - Pontuação padrão para diferentes fontes.

FONTES	PESO
Notícia	8
Blog	5
Documento de política	3
Patente	3
Wikipedia	3
Twitter (tweets e retweets)	1
Weibo	1
Avaliação por pares (Publons, Pubpeer)	1
Syllabi (Open Syllabus)	1
Google+	1
LinkedIn	0,5
Reddit	0,25
Facebook	0,25
Pinterest	0,25
Q&A (Stack Overflow)	0,25
Youtube	0,25
Quantidade de leitores no Mendeley	0
Quantidade de dimensões e citações da <i>Web of Science</i>	0

Fonte: Adaptação da tabela retirada do site www.altmetric.com

O modelo observado na Quadro 9 é uma indicação de como o algoritmo calcula a pontuação de um trabalho. Deste modo, a pontuação de atenção final (score) para cada trabalho

deve ser um número inteiro. Isso significa que as menções que contribuem com pontuações menores de 1, às vezes, são arredondadas para 1. Portanto, caso haja uma menção no Facebook para um trabalho, a pontuação aumentará para 1, entretanto, se tivermos 3 menções no Facebook para o mesmo trabalho, a pontuação continuará idêntica.

Outro fator importante, é que recentemente a Altmetric anunciou que algumas fontes como: Pinterest, Google+, Weibo e LinkedIn não são mais rastreadas, já que é necessário fazer o login nas páginas, impedindo Altmetric de acessar tais dados.

A título de exemplo, o Quadro 10 apresenta o cálculo da pontuação de atenção para o trabalho intitulado “Multimessenger Astronomy”, onde temos, uma menção (mentioned by) em site de notícia, duas em blogs, dezesseis no twitter, duas no facebook e uma na wikipedia.

Em relação a equação matemática utilizada pelo algoritmo da Altmetric, assume-se que $|I|$ é o número total de fontes e que uma menção m aparece em $|i : t \in i|$ fontes. Assim, assume-se também que $N_{m,i}$ é a frequência da menção m nas fontes i . Utiliza-se a pontuação P_x , onde x é a pontuação utilizada pelo algoritmo da Altmetric e apresentada na Tabela 1.

A atenção de uma fonte $AF_{m,i}$, pode ser definida como:

$$AF_{m,i} = P_x \times N_{m,i} \quad (3.1)$$

Por fim, a pontuação de atenção (*Altmetric score*) PA , é computada pela soma das atenções dadas para cada fonte, como:

$$PA = AF_0 + AF_1 \dots + AF_i \quad (3.2)$$

Quadro 10 - Exemplos de cálculo com base no peso de cada fonte

TRABALHO	FONTES	PONTUAÇÃO P_x	MENÇÕES $N_{M,I}$	CÁLCULO $AF_{m,i} = P_x \times N_{m,i}$	PONTUAÇÃO DE ATENÇÃO (ALTMETRIC SCORE) $PA = AF_0 + AF_1 \dots + AF_i$
Multimessenger Astronomy	Notícia	8	1	$8 \times 1 = 8$	$8 + 10 + 16 + 0,5 + 3 = 37,5$
	Blog	5	2	$5 \times 14 = 70$	
	Twitter	1	16	$1 \times 16 = 16$	
	Facebook	0,25	2	$0,25 \times 2 = 0,5$	
	Wikipedia	03	1	$3 \times 1 = 3$	
	Quantidade de leitores no Mendeley	0	33	$0 \times 33 = 0$	
	Quantidade de dimensões e citações da Web of Science	0	5	$0 \times 5 = 0$	

Fonte: Elaborado pelo autor.

Realizando o cálculo para o exemplo do Quadro 10, encontra-se o valor de 37,5. Entretanto, o resultado final para a pontuação de atenção (Altmetric score) é igual a 38 fato que é justificado pelo algoritmo da empresa que arredonda para cima a pontuação para um número inteiro.

O levantamento da visibilidade que a plataforma *Altmetric* têm dado aos trabalhos, apresenta novas maneiras de aferir a produção científica, de modo que tal ferramenta identifica instantaneamente quem está discutindo, salvando e compartilhando o trabalho em questão. Além do mais, possibilita reunir informações de como seu trabalho foi recebido por um público, identificando tendências, sendo útil principalmente nas áreas de Ciências humanas e sociais, onde no processo de avaliação inexistem indicadores equivalentes aos adotados para os periódicos (WILLIAMS, 2017).

3.4 CONSIDERAÇÕES FINAIS

Conforme visto neste capítulo, foram discutidos assuntos relacionados à medição da produção científica. Foram retratados princípios diretamente relacionados com a importância da relevância de um trabalho acadêmico. Também discutiu-se como a evolução da web social favoreceu para a criação de métricas alternativas, como a Altmetria.

Apesar da Altmetria ser considerada um campo em ascensão na área da Ciência da Informação, tal campo baseiam-se, majoritariamente, em quesitos quantitativos para a construção de indicadores, o que negligencia uma fonte de informação detalhada que são as menções de usuários. Dessa forma, algumas pesquisas estudam meios de analisar conteúdo das mídias sociais de modo qualitativo.

No capítulo seguinte, será descrita a construção do modelo proposto neste trabalho de doutorado, detalhando-se a aplicação da Inteligência Artificial na Altmetria.

CICLO DE DESENVOLVIMENTO DO MODELO

No desenvolvimento dessa tese foi crucial a elaboração de um modelo para ser utilizado no desenvolvimento de uma métrica de produção científica qualitativa no contexto da Altmetria, pois nas pesquisas sobre trabalhos relacionados não foram encontrados artefatos que atendessem às necessidades dos pesquisadores. Para alcançar tal objetivo, foi proposto um ciclo de desenvolvimento do modelo utilizando o método DSR, a fim de solucionar o problema, bem como validar e identificar melhorias.

4.1 METODOLOGIA DESIGN SCIENCE RESEARCH

Através da observação e utilização de métodos para validação das hipóteses, o conhecimento científico busca explicar o mundo, podendo ser classificado em: ciências naturais, que envolve o estudo dos fenômenos naturais; ciências formais, que compreende o estudo das relações abstratas e simbólicas; e ciências sociais, que abrange o estudo dos fenômenos humanos e sociais (ROCHADEL, 2016).

O avanço da tecnologia acarretou várias transformações na sociedade, motivando o modo de se fazer ciência. Cada vez mais o ser humano vem desenvolvendo artefatos para a solução de problemas. Segundo Simon (1980), o conceito de Design Science (DSR) é direcionado às pesquisas com foco na solução de problemas e, por conseguinte, ao projeto de artefatos (SIMON, 1980).

Em virtude destas características, as ciências tradicionais manifestam certas limitações, pois além de explicar fenômenos, há a necessidade de produzir sistemas para o avanço da ciência (DRESCH, 2015).

Assim, com o intuito de legitimar a produção de conhecimento científico por meio da construção de artefatos, foi desenvolvido o Design Science como método de pesquisa para o avanço da ciência e tecnologia (ROCHADEL, 2016) (DRESCH, 2015) (PEFFERS, 2007). Tal método vem sendo aplicado nas áreas de sistemas de informação, engenharia de produção, computação aplicada e ciência da informação (BARBOSA, 2017).

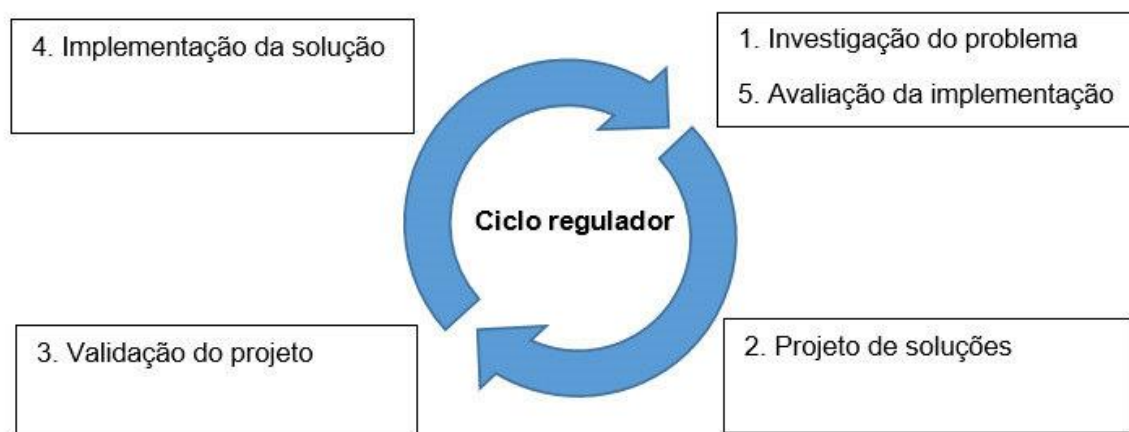
Os artefatos gerados nessa abordagem equivalem às criações humanas, como: frameworks, métodos, princípios de design, instanciações, métodos e modelos. Em razão disso, considera-se que a DSR apresenta como base a pesquisa em design e a pesquisa em ciência do

comportamento (DRESCH, 2015). Tais pesquisas detêm uma relação cíclica, da qual caracteriza a DSR como uma metodologia que se realiza em ciclos.

Wieringa (2009) propõe o ciclo regulador, ilustrado na Figura 9, que consiste em uma estrutura lógica para solucionar problemas. Dessa forma, cada uma das cinco etapas do ciclo é útil para coordenar a parte prática das pesquisas, por meio de um problema, ou resulta em conhecimentos por meio de respostas às questões de conhecimento.

A primeira etapa do ciclo regulador inicia-se com a descrição e investigação de um problema, sendo que este pode ser inicial ou um subproblema, resultante de uma interação anterior do ciclo. Por conseguinte, tem-se a etapa de projeto de soluções, onde há um problema prático, dado que esse projeto normalmente tem o intuito de melhorar o mundo de alguma maneira. A terceira etapa, denominada “Validação do Projeto”, visa verificar se o projeto irá atender aos objetivos desta tese. Posteriormente, há a etapa de implementação, que trata da execução do que foi planejado nas etapas anteriores. Ao final, o resultado é avaliado podendo ser o início de um novo ciclo.

Figura 9 - Ciclo regulador de Wieringa.



Fonte: Os autores, baseado em Wieringa (2009).

4.2 INSTANCIAÇÃO DO MODELO DSR

As etapas do ciclo regular são o processo utilizado no desenvolvimento do modelo, com o intuito de resolver um problema. Assim, realizou-se a instanciação do modelo DSR utilizando o ciclo regular.

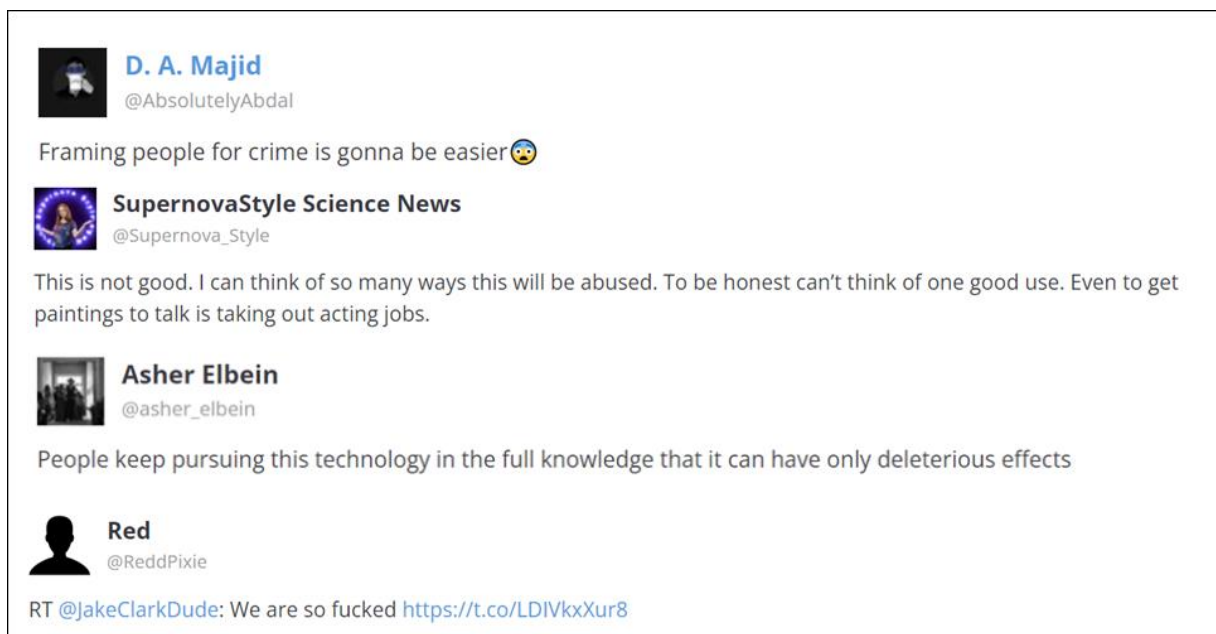
4.2.1 Investigação do problema

Primeiramente, a etapa de investigação do problema consistiu em um levantamento de informações necessárias para fundamentar a construção do modelo Altmétrico qualitativo. As menções da web social foram analisadas de modo a verificar a existência de sentimento, uma vez que tais menções são insumos que influenciam diretamente na composição da pontuação de atenção de um determinado trabalho.

A título de exemplo, a Figura 10 apresenta algumas menções de cunho negativo retiradas do Twitter, que compõem a pontuação de atenção para o trabalho “*Few-shot adversarial learning of realist neural talking head models*”.

Neste cenário, opiniões de cunho negativo, negligenciam a medida altmétrica quantitativa, onde são consideradas todas as menções, independentemente do sentimento expresso.

Figura 10 - Menções do twitter de cunho negativo para o trabalho "Few-shot adversarial learning of realist neural talking head models".



Fonte: Elaborado pelo autor.

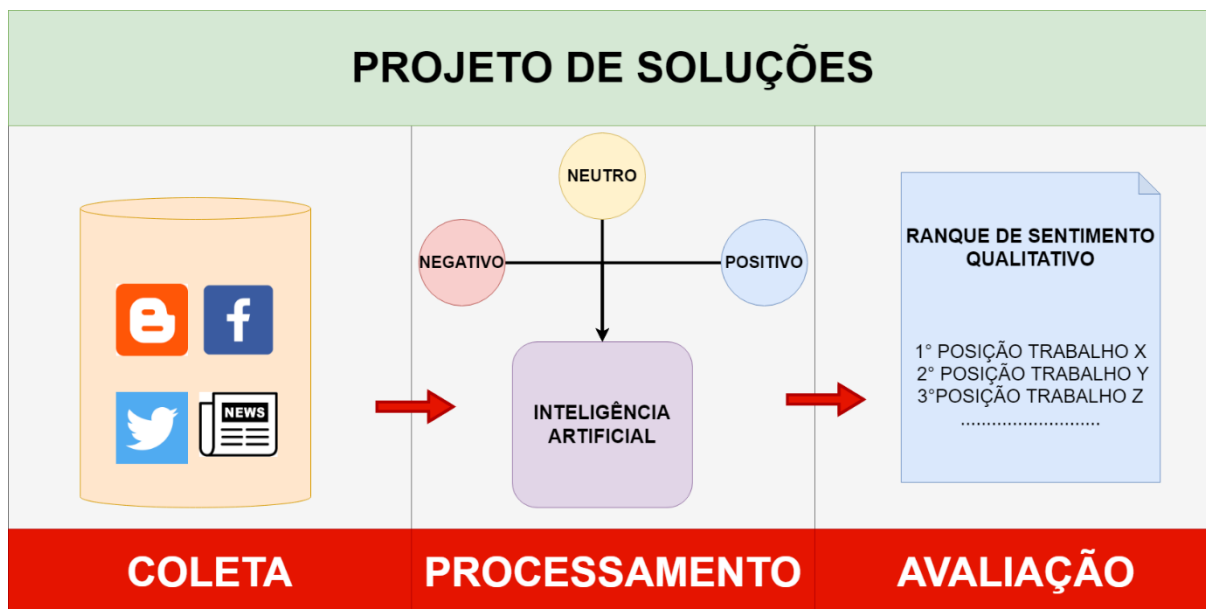
4.2.2 Projeto de soluções

A etapa de projeto de soluções focou em projetar uma possível maneira de aferir a qualidade da menção de acordo com a opinião do leitor, para, assim, solucionar o problema

investigado. Para tanto, organizou-se o projeto de solução em três fases essenciais, nas quais foram realizadas tarefas específicas, como: coleta, processamento e avaliação de textos.

A Figura 11 apresenta, em maiores detalhes, as fases do modelo proposto, bem como o funcionamento de cada módulo.

Figura 11 - Etapas realizadas para solucionar o problema investigado.



Fonte: Elaborado pelo autor.

A primeira fase do estudo compreendeu na coleta de menções de trabalhos científicos que compõem a pontuação de atenção, com o intuito de constituir o corpus da análise que foi investigado pela fase de processamento.

A fase de processamento foi responsável pela classificação e detecção da polaridade das menções na web social, focando a aferição qualitativa dos textos, sendo uma forma automatizada de compreensão utilizando Inteligência Artificial.

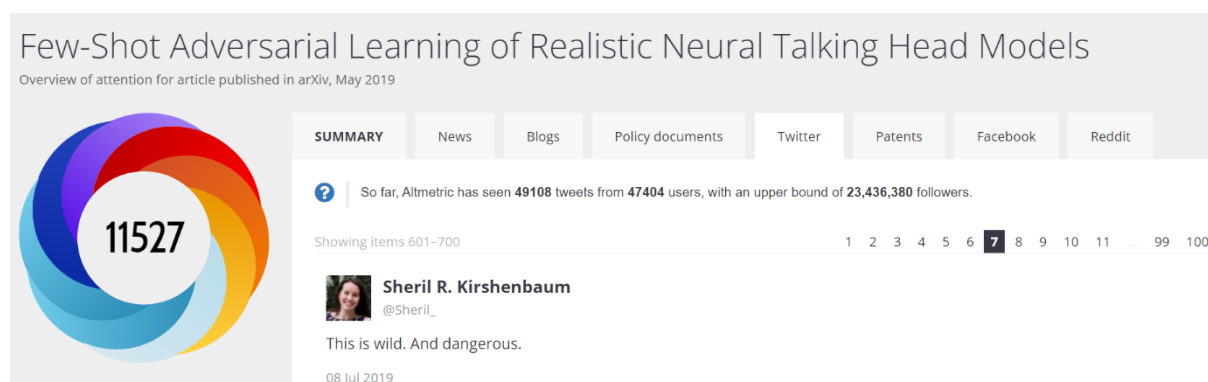
Uma vez realizadas as classificações, seguiu para a fase de avaliação, onde foram construídos ranques qualitativos, que utilizam informações do sentimento detectado das menções, visando não somente valores quantitativos.

4.2.3 Validação do Projeto

Já a etapa de validação do projeto, analisou a viabilidade da construção das fases mencionadas na etapa de proposta de solução. Desta forma, foram explorados recursos disponíveis da Altmetric e Inteligência Artificial.

Primeiramente, para validar a fase de coleta foi acessada a plataforma *Altmetric* e verificada a existência / disponibilidade de menções que compõem a pontuação de atenção para um determinado trabalho. A Figura 12 exibe a página do trabalho “*Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*”. Nesta página, é possível visualizar as menções advindas de diferentes fontes de dados, tais como: notícias, blogs, twitter, facebook, entre outras.

Figura 12 - Página de informações de um trabalho na plataforma Altmetric.



Fonte: Retirado de <https://www.altmetric.com/details/60733304/twitter/>

Posteriormente, na fase de processamento, foram aplicadas técnicas de Inteligência Artificial, com a finalidade de identificar a polaridade de sentimento contida em cada frase dos textos coletados nas menções. Nesta fase, por meio do processamento das menções pelas técnicas de sentimentos Vader, SenticNet, SentiStrength, Sentiwordnet e TextBlob, foi gerado um conjunto de resultados com as classificações obtidas de acordo com as respectivas polaridades dos sentimentos (positivo, negativo ou neutro). A seleção das técnicas justifica-se, pois, segundo a literatura, apresentam resultados superiores para o contexto de redes sociais, além de serem favoráveis em tarefas que envolvem grande quantidade de dados (HUTTO; GILBERT, 2014; PÅLSSON; SZERSZEN, 2016).

Para ilustrar esta fase, a Figura 13 apresenta, a título de exemplo, um esquema de seu funcionamento, que parte de uma menção sobre o trabalho “*Few-Shot Adversarial Learning of*

Realistic Neural Talking Head Models”, extraída da plataforma *Altmetric*, onde após ser processada pela técnica de sentimento, é então gerado um resultado.

Figura 13 - Processamento de uma menção extraída do twitter



Fonte: Elaborado pelo autor.

Por fim, na fase de avaliação do projeto de soluções, foi possível analisar diferentes cenários gerados pelas técnicas de IA, por meio da construção de ranques qualitativos de sentimento. Para isso, foi utilizada a tabela de pontuação das fontes, retirada do *Altmetric.com*, e apresentada na Seção 3.3.4. Além disso, considerou-se apenas as menções de cunho positivo para a construção do ranque qualitativo, devido ao fato destas indicarem tendências de leitura, e pelo potencial para engajarem novos leitores atraídos por indicações positivas (CAVALCANTI et. al, 2012).

O processo de construção de um ranque para um trabalho pode ser analisado na Tabela 1, onde é apresentada a quantidade de menções classificadas como positivas, de acordo com as fontes e técnicas de sentimentos, utilizando o trabalho ‘*Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*’ como exemplo. Além disso, demonstra-se o cálculo da pontuação de atenção utilizando os pesos das fontes selecionadas.

Tabela 1-Cálculo da pontuação para criação do ranque de sentimento.

Método de sentimento	Quantidade de menções positivas				Cálculo utilizando peso das fontes	Pontuação de atenção
	Blogs	News	Twitter	Facebook		
Vader	6	31	8391	2	$6 \times 5 + 31 \times 8 + 8391 \times 1 + 2 \times 0,25$	8670
SentiShenght	1	13	8288	2	$1 \times 5 + 13 \times 8 + 8288 \times 1 + 2 \times 0,25$	8398
SenticNet	8	68	7334	2	$8 \times 5 + 68 \times 8 + 7334 \times 1 + 2 \times 0,25$	7919
SentiWordNet	8	51	8725	4	$8 \times 5 + 51 \times 8 + 8725 \times 1 + 4 \times 0,25$	9174
TextBlob	6	36	1902	5	$6 \times 5 + 36 \times 8 + 1902 \times 1 + 5 \times 0,25$	2222

Fonte: Elaborado pelo autor.

4.2.4 Implementação

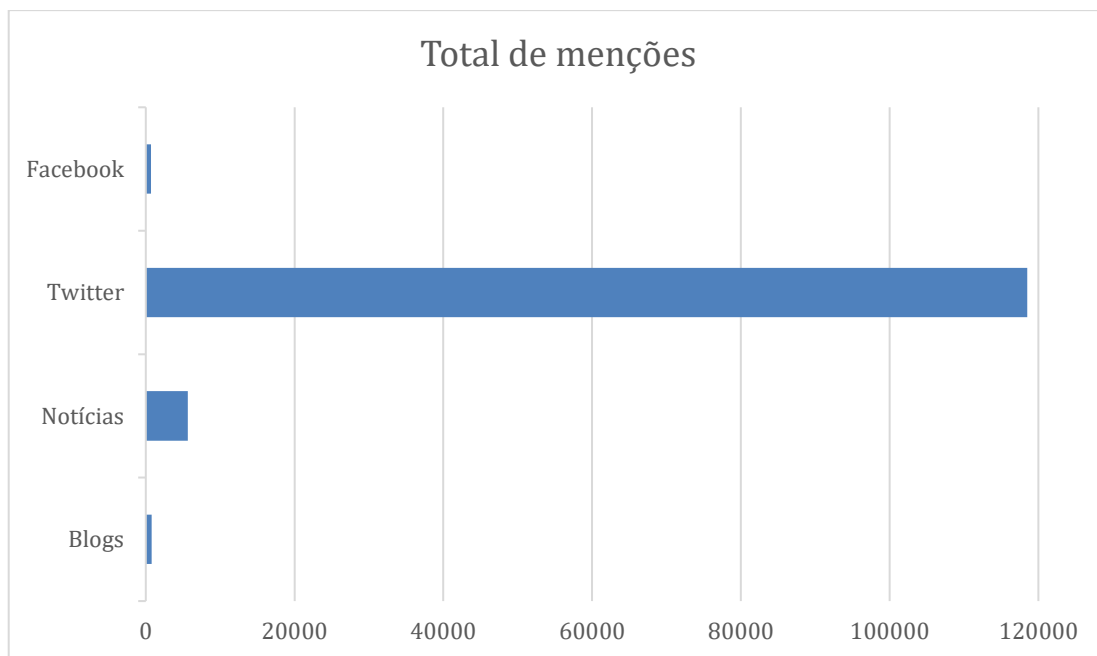
A etapa de implementação gerou o artefato desta pesquisa, sendo este: o modelo proposto e o protótipo. Para tal, seguiu as fases apresentadas na etapa de projeto de soluções, sendo validadas pelas ferramentas descritas na validação da solução.

A princípio, na fase de coleta de dados, foram extraídos entre os dias 2 e 10 de fevereiro de 2020, textos de menções de trabalhos publicados na plataforma *Altmetric*. Para as extrações de textos das menções selecionadas, utilizou-se a técnica computacional chamada *web scraping*, construída na linguagem de programação *Python*. Essa técnica permite extrair e analisar informações importantes de determinado site. Além disso, optou-se por manter apenas textos do idioma inglês, devido os métodos de sentimento não possuírem suporte multilíngue. Além disso, estudos demonstram que a maior parte das menções do *Altmetric* encontram-se no idioma inglês (WU, 2021). Assim, utilizou-se a biblioteca para linguagem Python chamada *Langdetect*.

Foram selecionadas, no total, 125.663 menções referentes às fontes twitter, páginas de notícias, blogs e facebook, para os 20 trabalhos com maiores pontuações de atenção da plataforma do ano de 2019, onde, a mídia social twitter contou com cerca de 94,3% do total de menções, enquanto blogs, notícias e facebook contaram com 0,6%, 4,5% e 0,5%, respectivamente. Portanto, a maioria das menções foram especialmente do twitter (118507). A análise demográfica do total de menções de acordo com as fontes blogs, notícias, twitter e facebook podem ser visualizadas no Gráfico 1. Já a Tabela 2, apresenta o número de menções

coletadas dos 20 trabalhos com maiores pontuações de atenção do site *altmetric.com* do ano de 2019.

Gráfico 1 - Total de menções investigadas de acordo com as fontes de dados.



Fonte: Elaborado pelo autor.

Tabela 2 - Número de menções coletadas dos 20 trabalhos com maiores pontuações de atenção da plataforma Altmetric do ano de 2019.

TÍTULOS DOS TRABALHOS	FONTES				QUANTIDADE TOTAL DE MENÇÕES
	BLOG	NEW	TWITTER	FACEBOOK	
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	14	86	9358	8	9466
Scientists rise up against statistical significance	44	41	8008	234	8327
Measles, Mumps, Rubella Vaccination and Autism	41	338	5531	55	5965
World Scientists' Warning of a Climate Emergency	103	639	7596	47	8385
A Style-Based Generator Architecture for Generative Adversarial Networks	29	90	6875	1	6995

Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	27	325	7602	11	7965
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	33	333	9585	15	9966
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	23	66	7708	40	7837
The global tree restoration potential	93	508	5607	12	6220
Civic honesty around the globe	15	260	3980	6	4261
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	25	470	3829	75	4399
Quantum supremacy using a programmable superconducting processor	46	386	2358	12	2802
Worldwide decline of the entomofauna: A review of its drivers	112	448	2792	40	3392
Ecological memory modifies the cumulative impact of recurrent climate extremes	11	63	9871	7	9952
Restoration of brain circulation and cellular functions hours post-mortem	39	352	1201	18	1610
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	19	115	9165	9	9308
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	7	21	5633	32	5693
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	59	487	2551	14	3111
Cast of the Right Bronchial Tree	3	186	4129	46	4364
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	40	455	5128	22	5645
Total	783	5669	118507	704	125663

Fonte: Elaborado pelo autor.

Nota-se que alguns trabalhos, mesmo estando em posições inferiores no ranque dos top 20, apresentaram um maior número de menções coletadas. Esse fato pode ser explicado pelo

peso aplicado a cada fonte nas menções, além do presente estudo considerar apenas quatro fontes da web social (Blogs, Notícias, Twitter e Facebook), desconsiderando as demais fontes que compõem o cálculo altmétrico.

Posteriormente, na fase de processamento, quando um texto é processado, um resultado é emitido de acordo com a classificação obtida a partir da polaridade do sentimento nela expresso, que pode ser positivo, negativo ou neutro. Desta maneira, foram aplicadas as técnicas de análise de sentimento mencionadas na seção 2.3, implementadas na linguagem de programação *Python*, com a finalidade de identificar a polaridade de sentimento contida em cada frase dos textos coletados nas menções. Este processo foi aplicado a todas as fontes e menções investigadas de um trabalho, as quais foram coletadas na plataforma.

As Tabelas 3, 4, 5 e 6 apresentam a classificação conforme a técnica Vader das menções coletadas das fontes de dados blogs, notícias, twitter e facebook, respectivamente.

Tabela 3 - Classificação de acordo com o método de sentimento Vader para a fonte de dados de Blogs.

BLOGS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	6	5	3
Scientists rise up against statistical significance	20	6	18
Measles, Mumps, Rubella Vaccination and Autism	14	10	17
World Scientists' Warning of a Climate Emergency	41	10	52
A Style-Based Generator Architecture for Generative Adversarial Networks	14	8	7
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	12	7	8
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	8	8	17
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	6	6	11
The global tree restoration potential	56	21	16
Civic honesty around the globe	10	1	4
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	9	4	12

Quantum supremacy using a programmable superconducting processor	29	9	8
Worldwide decline of the entomofauna: A review of its drivers	36	35	41
Ecological memory modifies the cumulative impact of recurrent climate extremes	8	0	3
Restoration of brain circulation and cellular functions hours post-mortem	8	4	27
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	10	4	5
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	3	0	4
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	28	21	10
Cast of the Right Bronchial Tree	3	0	0
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	12	4	24

Fonte: Elaborado pelo autor.

Tabela 4-Classificação de acordo com o método de sentimento Vader para a fonte de dados de Notícias.

NOTÍCIAS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	31	34	21
Scientists rise up against statistical significance	11	5	25
Measles, Mumps, Rubella Vaccination and Autism	111	100	127
World Scientists' Warning of a Climate Emergency	91	157	391
A Style-Based Generator Architecture for Generative Adversarial Networks	41	23	26
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	97	111	117
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	50	65	218
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	7	8	51
The global tree restoration potential	267	105	136

Civic honesty around the globe	93	81	86
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	60	119	291
Quantum supremacy using a programmable superconducting processor	229	77	80
Worldwide decline of the entomofauna: A review of its drivers	124	132	192
Ecological memory modifies the cumulative impact of recurrent climate extremes	47	6	10
Restoration of brain circulation and cellular functions hours post-mortem	49	115	188
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	51	17	47
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	9	4	8
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	244	135	108
Cast of the Right Bronchial Tree	60	91	35
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	207	123	125

Fonte: Elaborado pelo autor.

Tabela 5-Classificação de acordo com o método de sentimento Vader para a fonte de dados de Twitter.

TWITTER			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8391	279	688
Scientists rise up against statistical significance	5127	1128	1753
Measles, Mumps, Rubella Vaccination and Autism	1157	3020	1354
World Scientists' Warning of a Climate Emergency	2808	1822	2966
A Style-Based Generator Architecture for Generative Adversarial Networks	2458	3525	892
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	984	1130	5488
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	413	8147	1025

Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	3806	1120	2782
The global tree restoration potential	2638	1002	1967
Civic honesty around the globe	1851	451	1678
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	913	2525	391
Quantum supremacy using a programmable superconducting processor	1440	691	227
Worldwide decline of the entomofauna: A review of its drivers	900	923	969
Ecological memory modifies the cumulative impact of recurrent climate extremes	2610	1472	5789
Restoration of brain circulation and cellular functions hours post-mortem	244	395	562
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	7626	675	864
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	282	965	4386
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	1700	367	483
Cast of the Right Bronchial Tree	3148	698	283
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	2534	1395	1199

Fonte: Elaborado pelo autor.

Tabela 6-Classificação de acordo com o método de sentimento Vader para a fonte de dados do Facebook.

FACEBOOK			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	2	3	3
Scientists rise up against statistical significance	49	151	34
Measles, Mumps, Rubella Vaccination and Autism	21	20	14
World Scientists' Warning of a Climate Emergency	8	15	24
A Style-Based Generator Architecture for Generative Adversarial Networks	0	1	0

Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	2	7	2
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	6	3	6
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	13	6	21
The global tree restoration potential	9	2	1
Civic honesty around the globe	3	1	2
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	23	9	43
Quantum supremacy using a programmable superconducting processor	10	2	0
Worldwide decline of the entomofauna: A review of its drivers	8	12	20
Ecological memory modifies the cumulative impact of recurrent climate extremes	3	4	0
Restoration of brain circulation and cellular functions hours post-mortem	5	5	8
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	6	2	1
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	5	18	9
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	11	1	2
Cast of the Right Bronchial Tree	35	8	3
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	14	6	2

Fonte: Elaborado pelo autor.

Entre as técnicas utilizadas neste trabalho, está a classificação gerada pela Senticnet, que combina raciocínio de senso-comum e psicologia, além de explorar técnicas de Web Semântica para melhor reconhecer e processar opiniões de linguagem natural por meio da web. As classificações de acordo com as fontes Blogs, Notícias, Twitter e Facebook são apresentadas nas Tabelas 7, 8, 9 e 10.

Tabela 7 - Classificação de acordo com a técnica de sentimento SenticNet para a fonte de dados Blogs.

BLOGS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8	4	2
Scientists rise up against statistical significance	26	12	6
Measles, Mumps, Rubella Vaccination and Autism	27	7	7
World Scientists' Warning of a Climate Emergency	53	30	20
A Style-Based Generator Architecture for Generative Adversarial Networks	20	8	1
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	20	3	4
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	22	8	3
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	14	6	3
The global tree restoration potential	64	19	10
Civic honesty around the globe	9	5	1
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	13	3	9
Quantum supremacy using a programmable superconducting processor	29	12	5
Worldwide decline of the entomofauna: A review of its drivers	57	34	21
Ecological memory modifies the cumulative impact of recurrent climate extremes	9	1	1
Restoration of brain circulation and cellular functions hours post-mortem	25	9	5
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	10	4	5
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	7	0	0
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	33	14	12
Cast of the Right Bronchial Tree	1	1	1
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	23	10	7

Fonte: Elaborado pelo autor.

Tabela 8 – Classificação de acordo com a técnica de sentimento Senticnet para a fonte de dados de notícias

NOTÍCIAS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	68	10	8
Scientists rise up against statistical significance	29	6	6
Measles, Mumps, Rubella Vaccination and Autism	192	49	97
World Scientists' Warning of a Climate Emergency	446	99	94
A Style-Based Generator Architecture for Generative Adversarial Networks	51	22	17
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	256	60	9
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	193	48	92
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	58	6	2
The global tree restoration potential	336	80	92
Civic honesty around the globe	185	23	52
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	225	95	150
Quantum supremacy using a programmable superconducting processor	322	34	30
Worldwide decline of the entomofauna: A review of its drivers	288	70	90
Ecological memory modifies the cumulative impact of recurrent climate extremes	41	10	12
Restoration of brain circulation and cellular functions hours post-mortem	308	14	30
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	84	20	11
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	16	3	2
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	354	52	81
Cast of the Right Bronchial Tree	115	26	45
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	315	66	74

Fonte: Elaborado pelo autor.

Tabela 9 - Classificação de acordo com a técnica de sentimento Senticnet para a fonte de dados Twitter.

TWITTER			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	7334	1594	430
Scientists rise up against statistical significance	3626	1503	2879
Measles, Mumps, Rubella Vaccination and Autism	2638	2377	516
World Scientists' Warning of a Climate Emergency	3642	2643	1311
A Style-Based Generator Architecture for Generative Adversarial Networks	5575	889	411
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	6424	628	550
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	5294	1063	3228
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	3166	2025	2517
The global tree restoration potential	3999	403	1205
Civic honesty around the globe	2496	823	661
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	2537	566	726
Quantum supremacy using a programmable superconducting processor	1522	566	270
Worldwide decline of the entomofauna: A review of its drivers	1576	542	674
Ecological memory modifies the cumulative impact of recurrent climate extremes	4857	1862	3152
Restoration of brain circulation and cellular functions hours post-mortem	872	218	111
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	4425	400	4340
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	4149	1077	407
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	1028	562	960
Cast of the Right Bronchial Tree	726	3136	267
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	2612	1050	1466

Fonte: Elaborado pelo autor.

Tabela 10 - Classificação de acordo com a técnica de sentimento Senticnet para a fonte de dados Facebook.

FACEBOOK			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	2	5	1
Scientists rise up against statistical significance	110	98	26
Measles, Mumps, Rubella Vaccination and Autism	35	17	3
World Scientists' Warning of a Climate Emergency	21	23	3
A Style-Based Generator Architecture for Generative Adversarial Networks	0	1	0
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	7	3	1
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	15	0	0
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	18	6	16
The global tree restoration potential	9	1	2
Civic honesty around the globe	5	1	0
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	36	27	12
Quantum supremacy using a programmable superconducting processor	6	6	0
Worldwide decline of the entomofauna: A review of its drivers	19	12	9
Ecological memory modifies the cumulative impact of recurrent climate extremes	2	2	3
Restoration of brain circulation and cellular functions hours post-mortem	15	2	1
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	3	2	4
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	29	2	1
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	8	4	2
Cast of the Right Bronchial Tree	34	7	5
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	17	4	1

Fonte: Elaborado pelo autor.

A terceira técnica, intitulada SentiStrength, estima a força do sentimento positivo e negativo em textos curtos, até mesmo quando se emprega linguagem informal. Desta forma, nesta classificação foi padronizado o grau do sentimento em apenas positivo, neutro ou negativo. Os resultados para as diferentes fontes investigadas podem ser observados nas Tabelas 11, 12, 13 e 14.

Tabela 11 - Classificação de acordo com a técnica de sentimento SentiStrength para a fonte de dados Blogs.

BLOGS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	1	10	3
Scientists rise up against statistical significance	7	27	10
Measles, Mumps, Rubella Vaccination and Autism	4	18	19
World Scientists' Warning of a Climate Emergency	10	54	39
A Style-Based Generator Architecture for Generative Adversarial Networks	5	18	6
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	4	14	9
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	1	13	19
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	6	10	7
The global tree restoration potential	20	50	23
Civic honesty around the globe	2	12	1
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	2	8	15
Quantum supremacy using a programmable superconducting processor	26	15	5
Worldwide decline of the entomofauna: A review of its drivers	7	42	63
Ecological memory modifies the cumulative impact of recurrent climate extremes	2	4	5
Restoration of brain circulation and cellular functions hours post-mortem	0	12	27
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	3	8	8

US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	1	2	4
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	10	35	14
Cast of the Right Bronchial Tree	0	2	1
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	5	16	19

Fonte: Elaborado pelo autor.

Tabela 12 - Classificação de acordo com a técnica de sentimento SentiStreght para a fonte de dados de notícias.

NOTÍCIAS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	13	44	29
Scientists rise up against statistical significance	12	15	14
Measles, Mumps, Rubella Vaccination and Autism	41	138	159
World Scientists' Warning of a Climate Emergency	28	259	352
A Style-Based Generator Architecture for Generative Adversarial Networks	23	33	34
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	15	163	147
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	24	107	202
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	5	52	9
The global tree restoration potential	93	222	193
Civic honesty around the globe	47	199	14
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	22	152	296
Quantum supremacy using a programmable superconducting processor	198	96	92
Worldwide decline of the entomofauna: A review of its drivers	44	158	246
Ecological memory modifies the cumulative impact of recurrent climate extremes	15	29	19

Restoration of brain circulation and cellular functions hours post-mortem	4	135	213
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	7	55	53
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	6	9	6
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	154	208	125
Cast of the Right Bronchial Tree	37	113	36
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	50	179	226

Fonte: Elaborado pelo autor.

Tabela 13 Classificação de acordo com a técnica de sentimento SentiStrength para a fonte de dados Twitter.

TWITTER			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8288	447	623
Scientists rise up against statistical significance	2110	4042	1856
Measles, Mumps, Rubella Vaccination and Autism	747	3753	1031
World Scientists' Warning of a Climate Emergency	1722	3958	1916
A Style-Based Generator Architecture for Generative Adversarial Networks	2385	3723	767
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	616	3919	3067
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	622	4012	4951
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	2080	1822	3806
The global tree restoration potential	2042	2576	989
Civic honesty around the globe	1414	1161	1405
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	361	632	2836
Quantum supremacy using a programmable superconducting processor	1266	909	183

Worldwide decline of the entomofauna: A review of its drivers	134	477	2181
Ecological memory modifies the cumulative impact of recurrent climate extremes	3738	2809	3324
Restoration of brain circulation and cellular functions hours post-mortem	161	527	513
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	3175	1440	4550
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	148	1263	4222
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	440	1761	349
Cast of the Right Bronchial Tree	407	1115	2607
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	1042	2215	1871

Fonte: Elaborado pelo autor.

Tabela 14 - Classificação de acordo com a técnica de sentimento SentiStrength para a fonte de dados Facebook.

FACEBOOK			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	2	5	1
Scientists rise up against statistical significance	23	152	59
Measles, Mumps, Rubella Vaccination and Autism	10	34	11
World Scientists' Warning of a Climate Emergency	7	15	25
A Style-Based Generator Architecture for Generative Adversarial Networks	0	1	0
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	0	10	1
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	1	7	7
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	11	7	22
The global tree restoration potential	2	5	5
Civic honesty around the globe	4	2	0

Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	6	26	43
Quantum supremacy using a programmable superconducting processor	10	2	0
Worldwide decline of the entomofauna: A review of its drivers	3	7	30
Ecological memory modifies the cumulative impact of recurrent climate extremes	2	4	1
Restoration of brain circulation and cellular functions hours post-mortem	2	6	10
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	0	7	2
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	2	20	10
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	3	7	4
Cast of the Right Bronchial Tree	8	15	23
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	7	12	3

Fonte: Elaborado pelo autor.

Além das técnicas mencionadas anteriormente, utilizou-se neste trabalho a técnica SentiWordNet. Tal técnica tem como base o famoso dicionário de sentidos WordNet. As classificações obtidas podem ser observadas nas Tabelas 15, 16, 17 e 18..

Tabela 15 -Classificação de acordo com a técnica de sentimento Sentiwordnet par a fonte de dados Blogs.

BLOGS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8	1	5
Scientists rise up against statistical significance	27	0	17
Measles, Mumps, Rubella Vaccination and Autism	20	3	18
World Scientists' Warning of a Climate Emergency	60	3	40
A Style-Based Generator Architecture for Generative Adversarial Networks	14	0	15

Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	20	1	6
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	30	1	2
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	17	1	5
The global tree restoration potential	65	2	26
Civic honesty around the globe	10	0	5
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	11	1	13
Quantum supremacy using a programmable superconducting processor	30	0	16
Worldwide decline of the entomofauna: A review of its drivers	45	2	65
Ecological memory modifies the cumulative impact of recurrent climate extremes	9	0	2
Restoration of brain circulation and cellular functions hours post-mortem	18	1	20
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	10	0	9
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	6	1	0
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	42	2	15
Cast of the Right Bronchial Tree	1	0	2
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	23	5	12

Fonte: Elaborado pelo autor.

Tabela 16 - Classificação de acordo com a técnica de sentimento Sentiwordnet para a fonte de dados de notícias.

NOTÍCIAS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	51	3	32
Scientists rise up against statistical significance	31	0	10
Measles, Mumps, Rubella Vaccination and Autism	202	10	126

World Scientists' Warning of a Climate Emergency	344	59	236
A Style-Based Generator Architecture for Generative Adversarial Networks	46	1	43
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	233	4	88
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	246	7	80
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	60	1	5
The global tree restoration potential	391	13	104
Civic honesty around the globe	103	1	156
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	194	0	276
Quantum supremacy using a programmable superconducting processor	297	10	79
Worldwide decline of the entomofauna: A review of its drivers	221	9	218
Ecological memory modifies the cumulative impact of recurrent climate extremes	39	3	21
Restoration of brain circulation and cellular functions hours post-mortem	132	4	216
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	80	4	31
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	17	0	4
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	362	3	122
Cast of the Right Bronchial Tree	157	2	27
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	363	7	85

Fonte: Elaborado pelo autor.

Tabela 17 - Classificação de acordo com a técnica de sentimento Sentiwordnet para a fonte de dados Twitter.

TWITTER			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8725	172	461

Scientists rise up against statistical significance	6383	347	1278
Measles, Mumps, Rubella Vaccination and Autism	2658	1724	1149
World Scientists' Warning of a Climate Emergency	4261	736	2599
A Style-Based Generator Architecture for Generative Adversarial Networks	1583	251	5041
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	4292	236	3074
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	4872	4285	428
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	5898	503	1307
The global tree restoration potential	4555	564	488
Civic honesty around the globe	2950	165	865
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	1458	182	2189
Quantum supremacy using a programmable superconducting processor	1614	472	272
Worldwide decline of the entomofauna: A review of its drivers	1174	169	1449
Ecological memory modifies the cumulative impact of recurrent climate extremes	4933	276	4662
Restoration of brain circulation and cellular functions hours post-mortem	699	74	428
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	8143	203	819
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	3502	1472	659
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	1986	69	496
Cast of the Right Bronchial Tree	3268	356	505
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	3188	615	1325

Fonte: Elaborado pelo autor.

Tabela 18 - Classificação de acordo com a técnica de sentimento Sentiwordnet para a fonte de dados Facebook.

FACEBOOK			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	4	1	3
Scientists rise up against statistical significance	88	77	69
Measles, Mumps, Rubella Vaccination and Autism	24	2	29
World Scientists' Warning of a Climate Emergency	25	2	20
A Style-Based Generator Architecture for Generative Adversarial Networks	1	0	0
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	7	0	4
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	11	1	3
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	26	0	14
The global tree restoration potential	8	2	2
Civic honesty around the globe	5	0	1
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	42	4	29
Quantum supremacy using a programmable superconducting processor	10	2	0
Worldwide decline of the entomofauna: A review of its drivers	15	1	24
Ecological memory modifies the cumulative impact of recurrent climate extremes	6	0	1
Restoration of brain circulation and cellular functions hours post-mortem	7	0	11
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	5	0	4
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	10	4	18
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	11	1	2
Cast of the Right Bronchial Tree	28	1	17
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	17	0	5

Fonte: Elaborado pelo autor.

Por fim, a última técnica investigada nesse documento de tese é a TextBlob, desenvolvida a partir do *Natural Language Toolkit* (NLTK) e tem como classificador padrão o *Pattern Analyzer*. Os resultados podem ser visualizados nas Tabelas 19, 20, 21 e 22.

Tabela 19 -Classificação de acordo com a técnica de sentimento Textblob par a fonte de dados Blogs.

BLOGS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	6	4	4
Scientists rise up against statistical significance	19	18	7
Measles, Mumps, Rubella Vaccination and Autism	17	18	6
World Scientists' Warning of a Climate Emergency	46	38	19
A Style-Based Generator Architecture for Generative Adversarial Networks	8	9	12
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	22	2	3
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	19	8	6
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	8	13	2
The global tree restoration potential	55	26	12
Civic honesty around the globe	11	3	1
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	13	6	6
Quantum supremacy using a programmable superconducting processor	24	17	5
Worldwide decline of the entomofauna: A review of its drivers	45	34	33
Ecological memory modifies the cumulative impact of recurrent climate extremes	5	4	2
Restoration of brain circulation and cellular functions hours post-mortem	14	13	12
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	10	4	5
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	5	0	2

Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	28	27	4
Cast of the Right Bronchial Tree	1	1	1
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	24	9	7

Fonte: Elaborado pelo autor.

Tabela 20 - Classificação de acordo com a técnica de sentimento Textblob para a fonte de dados de notícias.

NOTÍCIAS			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	36	31	19
Scientists rise up against statistical significance	21	17	3
Measles, Mumps, Rubella Vaccination and Autism	193	121	24
World Scientists' Warning of a Climate Emergency	336	241	62
A Style-Based Generator Architecture for Generative Adversarial Networks	20	26	44
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	205	110	10
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	160	89	84
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	55	4	7
The global tree restoration potential	267	181	60
Civic honesty around the globe	129	126	5
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	103	179	188
Quantum supremacy using a programmable superconducting processor	202	123	61
Worldwide decline of the entomofauna: A review of its drivers	155	190	103
Ecological memory modifies the cumulative impact of recurrent climate extremes	44	16	3
Restoration of brain circulation and cellular functions hours post-mortem	57	171	124

Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	58	37	20
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	10	8	3
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	210	235	42
Cast of the Right Bronchial Tree	60	107	19
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	289	112	54

Fonte: Elaborado pelo autor.

Tabela 21 - Classificação de acordo com a técnica de sentimento Textblob para a fonte de dados do Twitter.

TWITTER			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	1902	6940	516
Scientists rise up against statistical significance	3730	3499	779
Measles, Mumps, Rubella Vaccination and Autism	2229	2890	412
World Scientists' Warning of a Climate Emergency	2831	4315	450
A Style-Based Generator Architecture for Generative Adversarial Networks	3850	2174	851
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	4541	785	2276
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	1383	3791	4411
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	5521	1791	396
The global tree restoration potential	2032	3417	158
Civic honesty around the globe	1802	881	1297
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	1684	1007	1138
Quantum supremacy using a programmable superconducting processor	615	1500	243
Worldwide decline of the entomofauna: A review of its drivers	1128	1341	323

Ecological memory modifies the cumulative impact of recurrent climate extremes	3356	3385	3130
Restoration of brain circulation and cellular functions hours post-mortem	306	771	124
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	3928	4711	526
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	2563	783	2287
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	961	665	924
Cast of the Right Bronchial Tree	3401	534	194
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	2658	1727	743

Fonte: Elaborado pelo autor.

Tabela 22 - Classificação de acordo com a técnica de sentimento Textblob para a fonte de dados do Facebook.

FACEBOOK			
TÍTULO	POSITIVO	NEUTRO	NEGATIVO
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	5	2	1
Scientists rise up against statistical significance	76	144	14
Measles, Mumps, Rubella Vaccination and Autism	27	28	0
World Scientists' Warning of a Climate Emergency	23	22	2
A Style-Based Generator Architecture for Generative Adversarial Networks	1	0	0
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	7	3	1
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	11	3	1
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	21	15	4
The global tree restoration potential	4	8	0
Civic honesty around the globe	5	1	0
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	21	35	19

Quantum supremacy using a programmable superconducting processor	1	9	2
Worldwide decline of the entomofauna: A review of its drivers	13	24	3
Ecological memory modifies the cumulative impact of recurrent climate extremes	2	4	1
Restoration of brain circulation and cellular functions hours post-mortem	6	12	0
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	4	5	0
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	7	4	21
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	10	4	0
Cast of the Right Bronchial Tree	40	5	1
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	12	7	3

Fonte: Elaborado pelo autor.

Obtidas as pontuações de atenção, gerou-se um ranque alternativo qualitativo, organizado de modo decrescente, a partir da pontuação de atenção (*altmetric score*), e de acordo com o sentimento positivo de cada título.

As Tabelas 23, 24, 25 e 26 apresentam o ranque de acordo com os métodos de sentimentos Vader, SenticNet, SentiStreght, SentiWordnet e Textblob, respectivamente.

Tabela 23-Ranque de sentimento segundo o método Vader.

RANQUE	TÍTULO	VADER
1°	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8670
2°	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	8086
3°	Scientists rise up against statistical significance	5328
4°	The global tree restoration potential	5057
5°	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	4254
6°	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	3896
7°	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	3795

8°	World Scientists' Warning of a Climate Emergency	3743
9°	Cast of the Right Bronchial Tree	3652
10°	Quantum supremacy using a programmable superconducting processor	3420
11°	Ecological memory modifies the cumulative impact of recurrent climate extremes	3027
12°	A Style-Based Generator Architecture for Generative Adversarial Networks	2856
13°	Civic honesty around the globe	2646
14°	Measles, Mumps, Rubella Vaccination and Autism	2121
15°	Worldwide decline of the entomofauna: A review of its drivers	2074
16°	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	1821
17°	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	1444
18°	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	855
19°	Restoration of brain circulation and cellular functions hours post-mortem	678
20°	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	371

Fonte: Elaborado pelo autor.

Tabela 24 - Ranque de sentimento segundo o método Senticnet.

RANQUE	TÍTULO	SENTICNET
1	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	8574
2	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	7919
3	World Scientists' Warning of a Climate Emergency	7481
4	The global tree restoration potential	7010
5	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	6952
6	A Style-Based Generator Architecture for Generative Adversarial Networks	6083
7	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	5252
8	Ecological memory modifies the cumulative impact of recurrent climate extremes	5231
9	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	5148

10	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	4411
11	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	4320
12	Measles, Mumps, Rubella Vaccination and Autism	4318
13	Quantum supremacy using a programmable superconducting processor	4245
14	Worldwide decline of the entomofauna: A review of its drivers	4170
15	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	4027
16	Civic honesty around the globe	4023
17	Scientists rise up against statistical significance	4016
18	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	3705
19	Restoration of brain circulation and cellular functions hours post-mortem	3465
20	Cast of the Right Bronchial Tree	1660

Fonte: Elaborado pelo autor.

Tabela 25 - Ranque de sentimento segundo o método Sentistreght.

RANQUE	TÍTULO	SENTISTRENGTH
1°	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8398
2°	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	3246
3°	Scientists rise up against statistical significance	2247
4°	The global tree restoration potential	2887
5°	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	1469
6°	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	2153
7°	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	1723
8°	World Scientists’ Warning of a Climate Emergency	1998
9°	Cast of the Right Bronchial Tree	705
10°	Quantum supremacy using a programmable superconducting processor	2983
11°	Ecological memory modifies the cumulative impact of recurrent climate extremes	3869
12°		2594

A Style-Based Generator Architecture for Generative Adversarial Networks		
		1801
13°	Civic honesty around the globe	1098
14°	Measles, Mumps, Rubella Vaccination and Autism	522
15°	Worldwide decline of the entomofauna: A review of its drivers	756
16°	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	549
17°	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	820
18°	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	194
19°	Restoration of brain circulation and cellular functions hours post-mortem	202
20°	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	

Fonte: Elaborado pelo autor.

Tabela 26 - Ranque de sentimento segundo o método Sentiwordnet.

RANQUE	TÍTULO	SENTIWORDNET
1	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	9174
2	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	8835
3	The global tree restoration potential	8010
4	World Scientists' Warning of a Climate Emergency	7320
5	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	6993
6	Scientists rise up against statistical significance	6788
7	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	6470
8	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	6258
9	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	6212
10	Ecological memory modifies the cumulative impact of recurrent climate extremes	5292
11	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	5095
12	Cast of the Right Bronchial Tree	4536
13	Measles, Mumps, Rubella Vaccination and Autism	4380

14	Quantum supremacy using a programmable superconducting processor	4143
15	Civic honesty around the globe	3826
16	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	3671
17	Worldwide decline of the entomofauna: A review of its drivers	3171
18	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	3076
19	A Style-Based Generator Architecture for Generative Adversarial Networks	2022
20	Restoration of brain circulation and cellular functions hours post-mortem	1847

Fonte: Elaborado pelo autor.

Tabela 27-Ranque de sentimento segundo o método Textblob.

RANQUE	TÍTULO	TEXTBLOB
1	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	6293
2	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	6007
3	World Scientists' Warning of a Climate Emergency	5755
4	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	5093
5	The global tree restoration potential	4444
6	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	4443
7	A Style-Based Generator Architecture for Generative Adversarial Networks	4051
8	Scientists rise up against statistical significance	4012
9	Cast of the Right Bronchial Tree	3896
10	Measles, Mumps, Rubella Vaccination and Autism	3865
11	Ecological memory modifies the cumulative impact of recurrent climate extremes	3734
12	Civic honesty around the globe	2891
13	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	2784
14	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	2761
15	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	2670
16	Worldwide decline of the entomofauna: A review of its drivers	2597
17	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	2579

18	Quantum supremacy using a programmable superconducting processor	2352
19	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	2222
20	Restoration of brain circulation and cellular functions hours post-mortem	834

Fonte: Elaborado pelo autor.

A partir deste ranque alternativo qualitativo, foi possível caracterizar os trabalhos qualitativamente, de acordo com os sentimentos existentes nas menções dos usuários, distinguindo-se dos critérios quantitativos alométricos.

A etapa final do modelo DSR, é a avaliação da implementação, onde foi realizada a prova de conceito comparando o ranque quantitativo com os ranques qualitativos gerados.

4.3 CONSIDERAÇÕES FINAIS

Nesta seção, foi apresentado o modelo qualitativo desenvolvido no presente trabalho de doutorado. Para isso, utilizou-se o método Design Science Research.

O modelo baseia-se em técnicas de IA para a criação de uma perspectiva qualitativa, onde a partir desta classificação alternativa, foi possível caracterizar os trabalhos de acordo com os sentimentos existentes nas menções de usuários, distinguindo-os dos critérios tradicionais de seleção de trabalhos, como aqueles publicados em revistas com maior fator de impacto, ou o autor com maior índice-h.

Para avaliar o modelo proposto, houve uma preocupação de manter os detalhes em um nível geral, deixando-se pontos como detalhes de experimentação para a próxima seção. Neste, serão apresentados os resultados obtidos em cada um dos estudos e, por fim, apresenta-se um comparativo dos resultados por meio de uma prova de conceito.

AVALIAÇÃO DA IMPLEMENTAÇÃO

Seguindo o preconizado pela metodologia DSR, a avaliação da implementação teve por objetivo validar se o problema da pesquisa foi resolvido de modo satisfatório através do uso do artefato gerado, e se este corroborou às conjecturas iniciais. Assim, após a execução das etapas do ciclo de desenvolvimento descritas anteriormente, foram analisados os ranques qualitativos de sentimento propostos neste trabalho, com o já conhecido ranque quantitativo Altmétrico (*altmetric score*).

Neste cenário, foram realizados três estudos distintos. O primeiro, consistiu na análise e comparação dos trabalhos no ranque qualitativo de sentimento que mais subiram e desceram posições em relação ao top 20 trabalhos com maior *altmetric score* (quantitativo). No segundo estudo, investigou-se os trabalhos que mais ganharam e perderam posições em um cenário específico, ou seja, de acordo com a área. Por fim, o último estudo demonstrou a construção de uma técnica intitulada IA Altmetric, que teve como base a combinação de técnicas de Inteligência Artificial para aferir a qualidade das menções.

Vale ressaltar, que os estudos e analogias apresentadas nesta seção, tem o intuito apenas de demonstrar novas possibilidades e cenários no contexto da Altmtria, não desmerecendo o modelo tradicional quantitativo de aferir a importância de um título.

Nesta seção, foram verificados e discutidos os resultados preliminares obtidos, a fim de provar o conceito do modelo proposto, e detectar possíveis ocorrências que possam auxiliar na construção de indicadores alternativos qualitativos.

5.1 ESTUDO 1: ANÁLISE DA QUALIDADE DOS TOP 20 TRABALHOS COM MAIOR “ALTMETRIC SCORE” CONSIDERANDO O SENTIMENTO POSITIVO

Neste primeiro estudo, analisou-se os 20 trabalhos com maiores pontuação de atenção (*altmetric score*) da plataforma Altmetric em relação à classificação alternativa segundo a técnica de análise de sentimento. Para este estudo, comparou-se os três trabalhos que mais ganharam e perderam posições no ranque de sentimento positivo.

Primeiramente, para que seja possível a comparação e discussão dos resultados, foi realizado o processo de cálculo para os top 20 trabalhos com maior *altmetric score* mencionados

na seção 4.2.4. Para isso, considerou-se os pesos das menções coletadas das fontes: blogs, notícias, twitter e facebook.

Por exemplo, o trabalho intitulado “*Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*” possui 14 menções em Blogs, 86 em portais de notícias, 9458 tweets e 8 menções no facebook. Sabendo que o peso para blogs é 5; para portais de notícias, 8; twitter, 1; e 0,25 para facebook, chegamos à pontuação de 10118. Tal processo fez-se necessário pelo fato desta pesquisa investigar apenas as principais fontes que compõem a pontuação de atenção, desconsiderando fontes como por exemplo: patentes, weibo, wikipedia, entre outros que de algum modo interferem no calcula da pontuação de atenção.

Em vista disso, foi gerado um novo ranque com base nas fontes investigadas. A Tabela 28 ilustra os trabalhos com maior *altmetric score*, ordenados do maior para o menor. Deste modo, foi possível elaborar e comparar cinco experimentos distintos, nos quais foram gerados ranques de sentimentos produzidos pelas técnicas investigadas.

Tabela 28-Classificação de acordo com os títulos com maiores *altmetric score*.

RANQUE	TÍTULO	ALTMETRIC SCORE
1	World Scientists’ Warning of a Climate Emergency	13235
2	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	12418
3	Ecological memory modifies the cumulative impact of recurrent climate extremes	10432
4	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	10340
5	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	10183
6	The global tree restoration potential	10139
7	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	10118
8	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	8974
9	Scientists rise up against statistical significance	8615
10	Measles, Mumps, Rubella Vaccination and Autism	8454
11	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	8361
12	A Style-Based Generator Architecture for Generative Adversarial Networks	7741
13	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	7733

14	Worldwide decline of the entomofauna: A review of its drivers	6946
15	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	6746
16	Civic honesty around the globe	6137
17	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	5844
18	Quantum supremacy using a programmable superconducting processor	5679
19	Cast of the Right Bronchial Tree	5644
20	Restoration of brain circulation and cellular functions hours post-mortem	4217

Fonte: Elaborado pelo autor.

Para o primeiro experimento utilizou-se a técnica Vader, mencionada na seção 2.3. Deste modo, foi possível elaborar a Tabela 29, que analisa os ranques dos trabalhos com maior *altmetric score* com aqueles investigados pela técnica Vader.

Comparando os ranques, obteve-se a Tabela 30, onde é possível visualizar que o décimo nono título com maior *altmetric score*, intitulado “*Cast of the Right Bronchial Tree*”, foi o nono título que mais recebeu sentimentos positivos nas menções dos leitores, subindo 10 posições. Já os títulos “*Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems*” e “*Quantum supremacy using a programmable superconducting processor*” subiram exatamente 8 posições no ranque de análise de sentimento Vader.

Por sua vez, a Tabela 31 mostra os 3 trabalhos com maior *altmetric score* que perderam posições no ranque quando classificados pelo critério de sentimento Vader. Nesta tabela, é possível visualizar que o segundo título mais lido pelos usuários da plataforma Altmetric, intitulado “*New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding*”, foi somente o décimo oitavo título que mais recebeu sentimento positivo nas menções dos leitores. Já o quarto trabalho da lista, intitulado “*Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior*”, foi apenas o décimo sexto trabalho que mais recebeu sentimentos positivos Vader nas menções dos leitores.

Tabela 29 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Vader.

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO
		VADER
1	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8670
2	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	8086
3	Scientists rise up against statistical significance	5328
4	The global tree restoration potential	5057
5	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	4254
6	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	3896
7	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	3795
8	World Scientists’ Warning of a Climate Emergency	3743
9	Cast of the Right Bronchial Tree	3652
10	Quantum supremacy using a programmable superconducting processor	3420
11	Ecological memory modifies the cumulative impact of recurrent climate extremes	3027
12	A Style-Based Generator Architecture for Generative Adversarial Networks	2856
13	Civic honesty around the globe	2646
14	Measles, Mumps, Rubella Vaccination and Autism	2121
15	Worldwide decline of the entomofauna: A review of its drivers	2074
16	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	1821
17	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	1444
18	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	855

19	Restoration of brain circulation and cellular functions hours post-mortem	678
20	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	371

Fonte: Elaborado pelo autor.

Tabela 30 - Relação dos trabalhos que mais subiram posições no ranque de análise de sentimento Vader quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO VADER	QTD. POSIÇÕES GANHAS
Cast of the Right Bronchial Tree	19	9	10
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	15	7	8
Quantum supremacy using a programmable superconducting processor	18	10	8

Fonte: Elaborado pelo autor.

Tabela 31 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Vader quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO VADER	QTD. POSIÇÕES PERDIDAS
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	2	18	16
Large-scale GWAS reveals insights into the	4	16	12

genetic architecture of same-sex sexual behavior			
Ecological memory			
modifies the cumulative impact of recurrent climate extremes	3	11	8

Fonte: Elaborado pelo autor.

No segundo experimento realizado utilizou-se uma técnica que se baseia em IA e Web Semântica, intitulada SenticNet. Tal técnica foi descrita na seção 2.3. Na Tabela 32 são apresentados e comparados o ranque dos trabalhos com maiores altmetric score e valores de sentimento positivo SenticNet.

Após análise da Tabela 33, é possível notar que os trabalhos que mais ganharam posições foram “*A Style-Based Generator Architecture for Generative Adversarial Networks*” e “*US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must*”. Ambos subiram 6 posições no ranque de sentimento, passando do décimo segundo, para o sexto lugar, e do décimo sétimo, para décimo primeiro no ranque de análise de sentimento Senticnet, respectivamente. Os trabalhos “*Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*” e “*Quantum supremacy using a programmable superconducting processor*” aparecem empatados na terceira posição com cinco posições recebidas.

Em contrapartida, a Tabela 34 apresenta os 3 trabalhos com maior altmetric score que desceram no ranque quando classificados pela técnica de sentimento Senticnet. Em relação a esta tabela, analisa-se que o nono trabalho (“*Scientists rise up against statistical significance*”) com maior altmetric score foi apenas o décimo sétimo no ranque de sentimento Senticnet. Em seguida os trabalhos intitulados “*Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial*” e “*Ecological memory modifies the cumulative impact of recurrent climate extremes*” completam o ranque dos trabalhos que mais perderam posições, sendo que o primeiro desceu sete, e o segundo, cinco posições, de acordo com o ranque de sentimento Senticnet.

Tabela 32- Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Senticnet

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO SENTICNET
1	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	8574
2	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	7919
3	World Scientists' Warning of a Climate Emergency	7481
4	The global tree restoration potential	7010
5	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	6952
6	A Style-Based Generator Architecture for Generative Adversarial Networks	6083
7	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	5252
8	Ecological memory modifies the cumulative impact of recurrent climate extremes	5231
9	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	5148
10	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	4411
11	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	4320
12	Measles, Mumps, Rubella Vaccination and Autism	4318
13	Quantum supremacy using a programmable superconducting processor	4245
14	Worldwide decline of the entomofauna: A review of its drivers	4170
15	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	4027
16	Civic honesty around the globe	4023
17	Scientists rise up against statistical significance	4016

18	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	3705
19	Restoration of brain circulation and cellular functions hours post-mortem	3465
20	Cast of the Right Bronchial Tree	1660

Fonte: Elaborado pelo autor

Tabela 33 - Relação dos trabalhos que mais ganharam posições no ranque de análise de sentimento Senticnet quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO SENTICNET	ALTERAÇÃO DE POSIÇÃO
A Style-Based Generator Architecture for Generative Adversarial Networks	12	6	+6
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	17	11	6
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	7	2	5
Quantum supremacy using a programmable superconducting processor	18	13	5

Fonte: Elaborado pelo autor.

Tabela 34 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Sentic quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO SENTICNET	QTD. POSIÇÕES PERDIDAS
Scientists rise up against statistical significance	9	17	8
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	11	18	7
Ecological memory modifies the cumulative impact of recurrent climate extremes	3	8	5

Fonte: Elaborado pelo autor.

Para o terceiro experimento, foi utilizado o método Sentistrength apresentado na seção 2.3. Esse método tem como característica atribuir um grau de valor de sentimento entre -4 e 4, onde frases com emoções positivas recebem valores entre 1 e 4. Por outro lado, frases de cunho negativo recebem valores entre -4 e -1. Neste trabalho, as frases foram consideradas como positivas, neutras ou negativas, independente do grau de sentimento. Consoante a isso, apresenta-se a Tabela 35, que compara o ranque de maior altmetric score com aquele gerado pela técnica Sentistrength.

Analisando-se a Tabela 36, observa-se que, o décimo oitavo trabalho com maior altmetric score (“*Quantum supremacy using a programmable superconducting processor*”), foi o quarto trabalho no ranque de sentimento Sentistrength, subindo precisamente 14 posições. Em seguida, os trabalhos “*Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*”, “*A Style-Based Generator Architecture for Generative Adversarial Networks*” e “*Civic honesty around the globe*” subiram exatamente 6 posições.

Por outro lado, a Tabela 37 apresenta os trabalhos com maior altmetric score que desceram posições no ranque quando classificados pela técnica Sentistrength, são eles: “*New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding*”, “*Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior*”

e “*World Scientists’ Warning of a Climate Emergency*”, os quais perderam 12, 11 e 8 posições respectivamente.

Tabela 35 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Sentistrength.

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO
		SENTISTRENGTH
1	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	8398
2	Ecological memory modifies the cumulative impact of recurrent climate extremes	3869
3	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	3246
4	Quantum supremacy using a programmable superconducting processor	2983
5	The global tree restoration potential	2887
6	A Style-Based Generator Architecture for Generative Adversarial Networks	2594
7	Scientists rise up against statistical significance	2247
8	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	2153
9	World Scientists’ Warning of a Climate Emergency	1998
10	Civic honesty around the globe	1801
11	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	1723
12	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	1469
13	Measles, Mumps, Rubella Vaccination and Autism	1098
14	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	820
15	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	756
16	Cast of the Right Bronchial Tree	705

17	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	549
18	Worldwide decline of the entomofauna: A review of its drivers	522
19	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	202
20	Restoration of brain circulation and cellular functions hours post-mortem	194

Fonte: Elaborado pelo autor.

Tabela 36 - Relação dos trabalhos que mais ganharam posições no ranque de análise de sentimento Sentistrength quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO SENTISTRENGTH	QTD. POSIÇÕES GANHAS
Quantum supremacy using a programmable superconducting processor	18	4	14
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	7	1	6
A Style-Based Generator Architecture for Generative Adversarial Networks	12	6	6
Civic honesty around the globe	16	10	6

Fonte: Elaborado pelo autor.

Tabela 37 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Sentistrength quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO SENTISTRENGTH	QTD. POSIÇÕES PERDIDAS
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	2	14	12
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	4	15	11
World Scientists' Warning of a Climate Emergency	1	9	8

Fonte: Elaborado pelo autor.

No quarto experimento utilizou-se da técnica intitulada SentiWordNet, que tem como base os synsets do dicionário lexical WordNet; a seção 2.3 descreve a técnica empregada neste experimento. Assim, foi possível construir a Tabela 38, que analisa os ranques dos trabalhos com maiores altmetric score e dos trabalhos com maior pontuação de sentimento SentiWordNet em suas menções. Nota-se que, para a maior parte dos trabalhos não houve alteração significativa de posição no ranque quando classificados pelo critério de análise de sentimento Sentiwordnet. Neste experimento foi constatado um grupo de 4 trabalhos que ganharam posições, além de 3 trabalhos que perderam posições significativas na comparação do ranque de maior altmetric score.

A Tabela 39 mostra a relação dos trabalhos que mais subiram posições no ranque quando classificados pela técnica Sentiwordnet. Nesta tabela, observa-se que o décimo nono trabalho (penúltimo do Top20) de maior altmetric score, intitulado “*Cast of the Right Bronchial Tree*”, foi o décimo segundo que mais recebeu sentimentos positivos nas menções dos usuários. O sétimo trabalho do ranque altmetric score, intitulado “*Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*”, foi o primeiro que mais recebeu sentimentos positivos segundo a técnica investigada. Além disso, os trabalhos “*Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial*”, “*Food in the*

Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems” e *“Quantum supremacy using a programmable superconducting processor”* subiram exatamente quatro posições no ranque qualitativo.

Enfim, a Tabela 40 mostra os três trabalhos com maior altmetric score da plataforma Altmetric que mais desceram posições no ranque, quando classificados pela técnica de análise de sentimento Sentiwordnet. Visualiza-se, por meio desta tabela, que os títulos *“A Style-Based Generator Architecture for Generative Adversarial Networks”* e *“Ecological memory modifies the cumulative impact of recurrent climate extremes”* foram somente o décimo nono e décimo trabalho que mais receberam sentimentos positivos nas menções da web social, ambos representando uma perda de sete posições. O título *“Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017”* aparece em seguida, descendo cinco posições no ranque qualitativo de sentimento.

Tabela 38 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Sentiwordnet.

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO SENTIWORDNET
1	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	9174
2	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	8835
3	The global tree restoration potential	8010
4	World Scientists’ Warning of a Climate Emergency	7320
5	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	6993
6	Scientists rise up against statistical significance	6788
7	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	6470
8	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	6258
9	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	6212
10	Ecological memory modifies the cumulative impact of recurrent climate extremes	5292

11	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	5095
12	Cast of the Right Bronchial Tree	4536
13	Measles, Mumps, Rubella Vaccination and Autism	4380
14	Quantum supremacy using a programmable superconducting processor	4143
15	Civic honesty around the globe	3826
16	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	3671
17	Worldwide decline of the entomofauna: A review of its drivers	3171
18	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	3076
19	A Style-Based Generator Architecture for Generative Adversarial Networks	2022
20	Restoration of brain circulation and cellular functions hours post-mortem	1847

Fonte: Elaborado pelo autor.

Tabela 39 - Relação dos trabalhos que mais ganharam posições no ranque de análise de sentimento Sentiwordnet quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO SENTIWORDNET	QTD. POSIÇÕES GANHAS
Cast of the Right Bronchial Tree	19	12	7
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	7	1	6

Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	11	7	4
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	15	11	4
Quantum supremacy using a programmable superconducting processor	18	14	4

Fonte: Elaborado pelo autor.

Tabela 40 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Sentiwordnet quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO SENTIWORDNET	QTD. POSIÇÕES PERDIDAS
A Style-Based Generator Architecture for Generative Adversarial Networks	12	19	7
Ecological memory modifies the cumulative impact of recurrent climate extremes	3	10	7
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	13	18	5

Fonte: Elaborado pelo autor.

Por fim, o último experimento realizado neste estudo consistiu na utilização da biblioteca Textblob apresentada na seção 2.3. Desta maneira, elaborou-se a Tabela 41, que compara o ranque dos trabalhos com maiores altmetric score e dos trabalhos com maior pontuação de sentimento, de acordo com a biblioteca TextBlob.

Da mesma forma, como realizado nos demais experimentos, foram agrupados os trabalhos que mais subiram e desceram posições no ranque quando classificados pelo critério de análise de sentimento Textblob.

A Tabela 42 apresenta a relação dos trabalhos que mais subiram posições no ranque, sendo possível observar que, o trabalho intitulado “*Cast of the Right Bronchial Tree*” foi o que mais ganhou posições, seguido dos títulos “*Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial*” e “*A Style-Based Generator Architecture for Generative Adversarial Networks*”.

Finalmente, na Tabela 43 são apresentados os três trabalhos que mais desceram posições no ranque de sentimento se comparado ao critério de maior Altmetric Score. Ao analisar os trabalhos relacionados nesta tabela, é possível observar que o segundo trabalho com maior altmetric score, intitulado “*New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding*”, foi somente o décimo quarto que mais recebeu sentimentos positivos nas menções dos leitores. Já o sétimo título da lista, intitulado “*Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*”, foi o penúltimo, ou seja, décimo nono trabalho que mais recebeu sentimentos positivos nas menções de leitores.

Tabela 41 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Textblob.

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO TEXTBLOB
1	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	6293
2	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	6007
3	World Scientists’ Warning of a Climate Emergency	5755
4	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	5093

5	The global tree restoration potential	4444
6	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	4443
7	A Style-Based Generator Architecture for Generative Adversarial Networks	4051
8	Scientists rise up against statistical significance	4012
9	Cast of the Right Bronchial Tree	3896
10	Measles, Mumps, Rubella Vaccination and Autism	3865
11	Ecological memory modifies the cumulative impact of recurrent climate extremes	3734
12	Civic honesty around the globe	2891
13	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	2784
14	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	2761
15	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	2670
16	Worldwide decline of the entomofauna: A review of its drivers	2597
17	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	2579
18	Quantum supremacy using a programmable superconducting processor	2352
19	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	2222
20	Restoration of brain circulation and cellular functions hours post-mortem	834

Fonte: Fonte elaborado pelo autor.

Tabela 42 - Relação dos trabalhos que mais ganharam posições no ranque de análise de sentimento Textblob quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO TEXTBLOB	QTD. POSIÇÕES GANHAS
Cast of the Right Bronchial Tree	19	9	10
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	11	2	9
A Style-Based Generator Architecture for Generative Adversarial Networks	12	7	5

Fonte: Elaborado pelo autor.

Tabela 43 - Relação dos trabalhos que mais desceram posições no ranque de análise de sentimento Textblob quando comparados com o critério de trabalhos com maior altmetric score.

TÍTULO	POSIÇÃO ALTMETRIC SCORE	POSIÇÃO ANÁLISE DE SENTIMENTO TEXTBLOB	QTD. POSIÇÕES PERDIDAS
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	2	14	12
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	7	19	12

Ecological memory modifies the cumulative impact of recurrent climate extremes	3	11	8
---	---	----	---

Fonte: Elaborado pelo autor.

Por fim, a Tabela 44 apresenta a quantidade de posições ganhas/perdidas dos 20 trabalhos investigados neste estudo, organizados de acordo com as técnicas de IA. Em Vermelho, está o destaque para os 3 trabalhos que mais perderam posições no ranque qualitativo. Enquanto a cor azul, indica os trabalhos que mais ganharam posições no ranque.

Analisando a tabela, nota-se que, em relação às cinco técnicas de sentimento, 6 trabalhos não tiveram alteração significativa de posição no ranque quando classificados pelo critério qualitativo de maior percentual de sentimentos positivos, são eles: “*Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target*”, “*The global tree restoration potential*”, “*A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy*”, “*Worldwide decline of the entomofauna: A review of its drivers*” e “*Restoration of brain circulation and cellular functions hours post-mortem*”.

Observou-se, que apesar dos títulos “*Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*” e “*A Style-Based Generator Architecture for Generative Adversarial Networks*” serem apresentados por 3 técnicas como que mais ganharam posições, destacam-se pelo fato de uma técnica de IA classificá-los entre os que mais perderam posições.

Como mencionado nas análises realizadas anteriormente e constatado na tabela apresentada, dentre as cinco técnicas de IA investigadas, o trabalho “*Quantum supremacy using a programmable superconducting processor*” aparece em quatro técnicas de IA como sendo o que mais ganhou posição no ranque qualitativo. Por outro lado, o título “*Ecological memory modifies the cumulative impact of recurrent climate extremes*” surge como o que mais perdeu posições, sendo representado por 4 técnicas.

A partir dos dados compilados dos ranques de sentimentos gerados pelas técnicas de Inteligência artificial, é possível notar que cada técnica produziu um cenário / ranque distinto em relação às posições dos títulos com maior altmetric score.

Apesar dos cenários apresentados serem subjetivos, tendo em vista que a classificação pode mudar de acordo com a técnica aplicada, é possível analisar como as menções nas mídias sociais podem interferir no cálculo quantitativo da pontuação de atenção de um trabalho

Tabela 44 - Posições ganhas ou perdidas dos trabalhos de acordo com cada técnica de IA.

RANQUE ALTMETRIC SCORE	TÍTULO	POSIÇÕES GANHAS / PERDIDAS				
		VADER	SENTICNET	SENTISTRENGTH	SENTIWORDNET	TEXTBLOB
1	World Scientists' Warning of a Climate Emergency	-7	-2	-8	-3	-2
2	New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	-16	-3	-12	-3	-12
3	Ecological memory modifies the cumulative impact of recurrent climate extremes	-8	-5	1	-7	-8
4	Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	-12	3	-11	-4	3
5	Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	3	-4	2	3	-1
6	The global tree restoration potential	2	2	1	3	1
7	Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	6	5	6	6	-12
8	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	3	1	-4	-1	4
9	Scientists rise up against statistical significance	6	-8	2	3	1
10	Measles, Mumps, Rubella Vaccination and Autism	-4	-2	-3	-3	0

11	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	5	-7	3	4	9
12	A Style-Based Generator Architecture for Generative Adversarial Networks	0	6	6	-7	5
13	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	-4	3	-4	-5	-4
14	Worldwide decline of the entomofauna: A review of its drivers	-1	0	-4	-3	-2
15	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	8	0	4	4	2
16	Civic honesty around the globe	3	0	6	1	4
17	US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	-3	6	-2	1	2
18	Quantum supremacy using a programmable superconducting processor	8	5	14	4	0
19	Cast of the Right Bronchial Tree	10	-1	3	7	10
20	Restoration of brain circulation and cellular functions hours post-mortem	1	1	0	0	0

Fonte: Elaborado pelo autor.

5.2 ESTUDO 2: ANÁLISE QUALITATIVA DAS MENÇÕES DE ACORDO COM A ÁREA DE ESTUDO

Para o segundo estudo, utilizou-se também o modelo apresentado no capítulo 3. Entretanto, nesse estudo, os 20 títulos com maiores pontuações de atenção foram agrupados por área e classificados de acordo com as técnicas de Inteligência Artificial, sendo que foi investigado o trabalho que mais perdeu e o que mais ganhou posição.

Inicialmente, neste experimento os 20 trabalhos com maiores pontuações (*altmetric score*) foram agrupados por área, conforme informações disponibilizadas no site *Altmetric.com*. O Quadro 11 ilustra as respectivas áreas de publicação de cada trabalho.

Quadro 11 - Área de publicação dos top 20 trabalhos com maiores altmetric score.

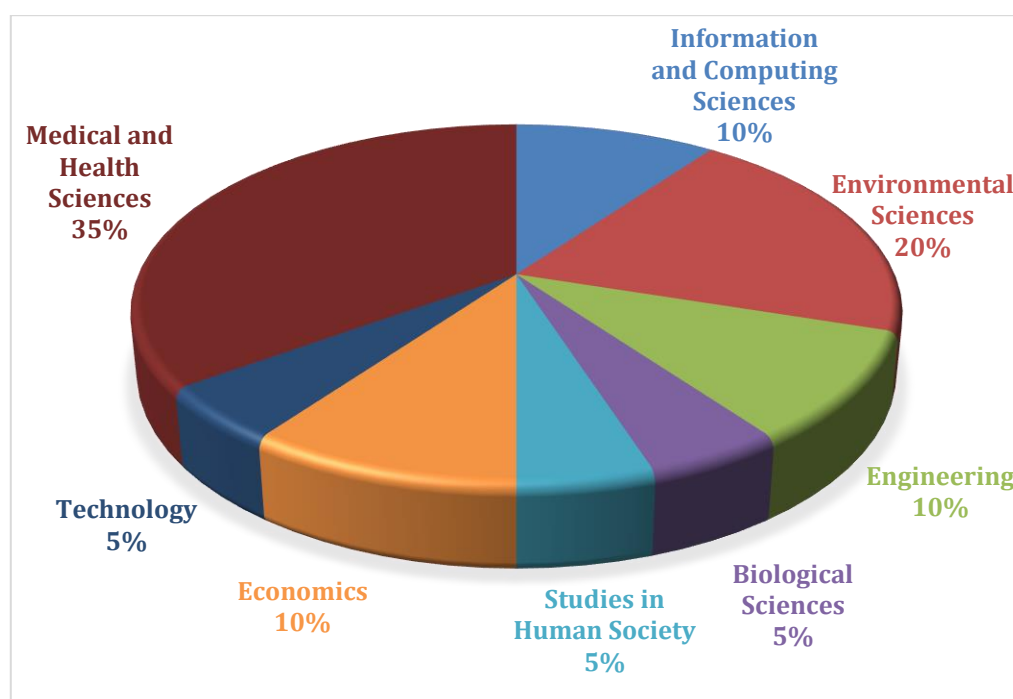
TRABALHO	ÁREA
Few-Shot Adversarial Learning of Realistic Neural Talking Head Models	<i>Information and Computing Sciences</i>
Scientists rise up against statistical significance	<i>Mathematical Sciences</i>
Measles, Mumps, Rubella Vaccination and Autism	<i>Medical and Health Sciences</i>
World Scientists' Warning of a Climate Emergency	<i>Environmental Sciences</i>
A Style-Based Generator Architecture for Generative Adversarial Networks	<i>Information and Computing Sciences</i>
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior	<i>Biological Sciences</i>
New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding	<i>Engineering</i>
Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	<i>Medical and Health Sciences</i>
The global tree restoration potential	<i>Environmental Sciences</i>
Civic honesty around the globe	<i>Economics</i>
Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	<i>Medical and Health Sciences</i>
Quantum supremacy using a programmable superconducting processor	<i>Technology</i>
Worldwide decline of the entomofauna: A review of its drivers	<i>Environmental Sciences</i>

Ecological memory modifies the cumulative impact of recurrent climate extremes	<i>Environmental Sciences</i>
Restoration of brain circulation and cellular functions hours post-mortem	<i>Medical and Health Sciences</i>
Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target	<i>Engineering</i>
US military is a bigger polluter than as many as 140 countries – shrinking this war machine is a must	<i>Studies in Human Society</i>
Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	<i>Medical and Health Sciences</i>
Cast of the Right Bronchial Tree	<i>Medical and Health Sciences</i>
A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	<i>Medical and Health Sciences</i>

Fonte: Elaborado pelo autor.

Observando-se o quadro é possível verificar a presença de 8 áreas distintas, dentre elas: Engineering, Environmental Sciences, Information and Computing Sciences, Biological Sciences, Economics, Technology, Studies in Human Society e Medical and Health Sciences. O Gráfico 2 ilustra a abrangência das áreas, de acordo com os títulos investigados.

Gráfico 2 - Percentual dos 20 trabalhos investigados de acordo com a área.



Fonte: Elaborado pelo autor.

A área *Medical and Health Sciences* destaca-se por apresentar o maior número de trabalhos entre os top 20, contendo cerca de 7 trabalhos (35 por cento do total). Por conta disso, neste estudo considerou-se apenas trabalhos da área *Medical and Health Sciences*, devido a sua abrangência no número de trabalhos em relação as demais.

Selecionado a área, para o próximo passo foram ordenados os 7 trabalhos de acordo com a maior pontuação de atenção (*altmetric score*) e comparados com o ranque gerado pelas técnicas de IA.

As Tabelas 45, 46, 47, 48 e 49 apresentam o comparativo entre o ranque dos títulos com maior altmetric score e maior pontuação de sentimento positivo para as técnicas Vader, Senticnet, Sentistrength, Sentiwordnet e Textblob, respectivamente.

Tabela 45 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Vader de acordo com a área Medical and Health Sciences.

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO VADER
1	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	4254
2	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	3896
3	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	3795
4	Cast of the Right Bronchial Tree	3652
5	Measles, Mumps, Rubella Vaccination and Autism	2121
6	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	1444
7	Restoration of brain circulation and cellular functions hours post-mortem	678

Fonte: Elaborado pelo autor.

Tabela 46 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Senticnet de acordo com a área Medical and Health Sciences

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO SENTICNET
1	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	5252
2	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	4411
3	Measles, Mumps, Rubella Vaccination and Autism	4318
4	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	4027
5	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	3705
6	Restoration of brain circulation and cellular functions hours post-mortem	3465
7	Cast of the Right Bronchial Tree	1660

Fonte: Elaborado pelo autor.

Tabela 47 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Sentistrength de acordo com a área Medical and Health Sciences

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO SENTISTRENGTH
1	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	2153
2	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	1723
3	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	1469
4	Measles, Mumps, Rubella Vaccination and Autism	1098

5	Cast of the Right Bronchial Tree	705
6	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	549
7	Restoration of brain circulation and cellular functions hours post-mortem	194

Fonte: Elaborado pelo autor.

Tabela 48 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Sentiwordnet de acordo com a área Medical and Health Sciences

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO SENTIWORDNET
1	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	6470
2	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	6212
3	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	5095
4	Cast of the Right Bronchial Tree	4536
5	Measles, Mumps, Rubella Vaccination and Autism	4380
6	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	3076
7	Restoration of brain circulation and cellular functions hours post-mortem	1847

Fonte: Elaborado pelo autor.

Tabela 49 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo Textblob de acordo com a área Medical and Health Sciences

RANQUE	TÍTULO	ANÁLISE DE SENTIMENTO TEXTBLOB
1	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	6007
2	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	5093
3	Cast of the Right Bronchial Tree	3896
4	Measles, Mumps, Rubella Vaccination and Autism	3865
5	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	2784
6	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	2579
7	Restoration of brain circulation and cellular functions hours post-mortem	834

Fonte: Elaborado pelo autor.

Dentre os 7 trabalhos analisados, o trabalho intitulado “*Cast of the Right Bronchial Tree*” e “*Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems*” ganharam 2 posições no ranque de sentimento Vader quando comparados aos de maior altmetric score por área. Por outro lado, o trabalho que mais caiu posições no ranque foi “*Measles, Mumps, Rubella Vaccination and Autism*”, perdendo cerca de 3 posições.

Posteriormente, avaliou-se o ranque de sentimento gerado pela técnica Senticnet. Em um primeiro momento, é possível notar que para os trabalhos investigados, o trabalho “*Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017*” foi o que subiu mais posições no ranque por área (2 posições). Em contrapartida, o título “*Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial*” foi o que mais desceu posições (2 posições). Com exceção do trabalho classificado como primeiro do ranque (“*A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy*”), os demais sofreram ao menos uma mudança de posição, seja subindo, ou descendo posição no ranque de sentimento Senticnet.

Para o ranque de sentimento da técnica Sentistrength observa-se que o título “*Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems*” conquistou o maior número de posições (3 posições). Enquanto 3 trabalhos perderam exatamente o mesmo número de posições (2 posições), são eles: “*A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy*”, “*Measles, Mumps, Rubella Vaccination and Autism*” e “*Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017*”. Nota-se que, apesar do trabalho “*A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy*” possuir maior número de menções coletadas dentre os trabalhos investigados, obteve apenas a terceira colocação no ranque de sentimento Sentistrength.

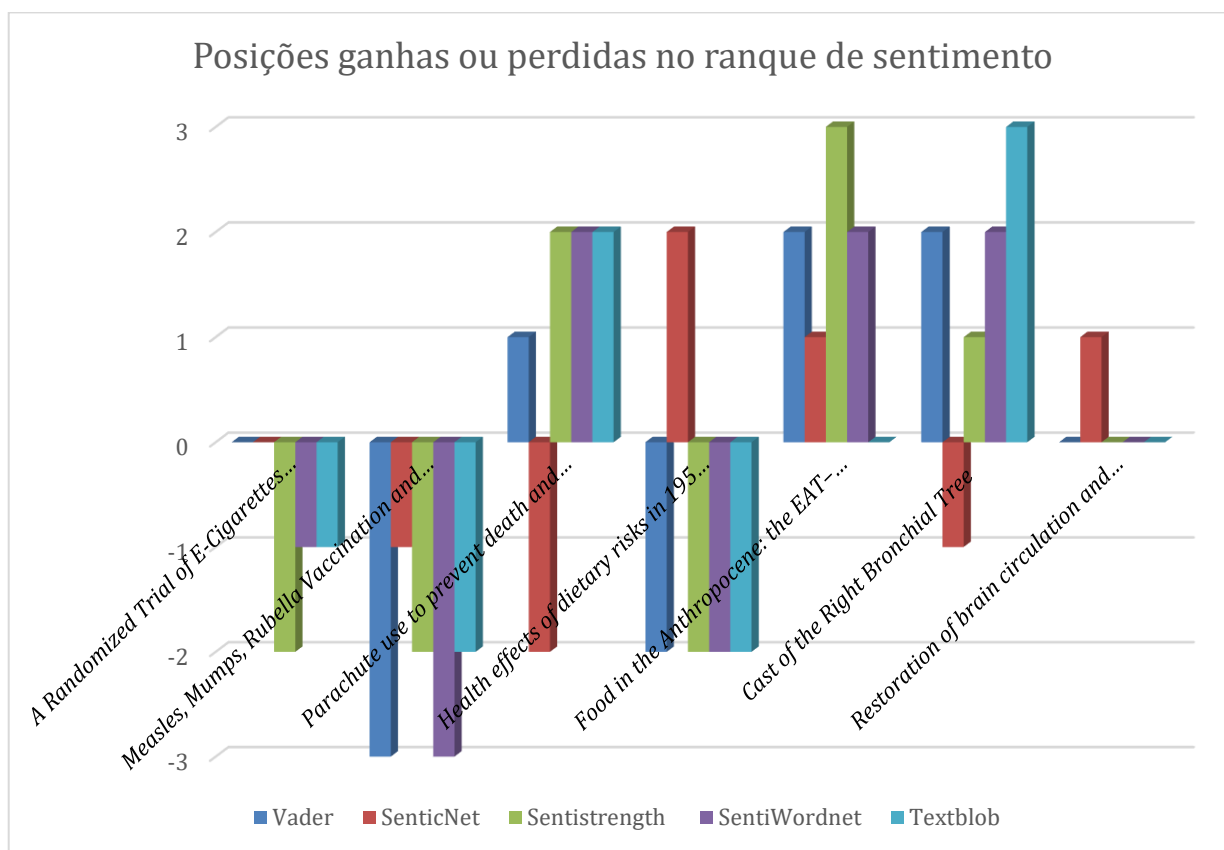
Os dados apresentados pela técnica Sentiwordnet demonstram que 3 títulos que estavam em posições inferiores no ranque de maior altmetric score subiram 2 posições, sendo estes: “*Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial*”, “*Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems*” e “*Cast of the Right Bronchial Tree*”. Ao contrário dos trabalhos apresentados, o trabalho intitulado “*Measles, Mumps, Rubella Vaccination and Autism*” perdeu exatamente 3 posições, sendo o que mais perdeu posição no ranque de sentimento Sentiwordnet.

Ao analisar o ranque de sentimento Textblob, observa-se que o trabalho intitulado “*Cast of the Right Bronchial Tree*” ganhou o maior número de posição, passando da sexta colocação

para a terceira no ranque de sentimento. Já analisando os trabalhos que mais perderam posições, verifica-se que os títulos “*Measles, Mumps, Rubella Vaccination and Autism*” e “*Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017*” foram os que mais caíram (2 posições).

A fim de avaliar se a maioria das técnicas investigadas geraram resultados semelhantes, elaborou-se o Gráfico 3, com a relação entre a quantidade de posições ganhadas ou perdidas dos trabalhos para cada técnica investigada. Ressalta-se que no gráfico estão todos os trabalhos, independentemente se eles não ganharam ou perderam posições.

Gráfico 3 - Posições ganhas ou perdidas de acordo com a técnica de IA.



Fonte: Elaborado pelo autor.

Os dados do gráfico revelam que a maioria das técnicas tiveram resultados semelhantes, tanto para os trabalhos que ganharam, quanto para os que perderam posições, sendo que o trabalho “*Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems*” foi o que mais ganhou posições para os diferentes ranques de sentimentos gerados.

Com exceção da técnica Senticnet, os títulos “*Cast of the Right Bronchial Tree* ” e “*Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial*” destacaram-se também pelo número de posições ganhas.

Como já mencionado, dentre os trabalhos investigados, o que mais perdeu posição entre as técnicas investigadas é o intitulado “*Measles, Mumps, Rubella Vaccination and Autism*”.

Em seguida, o segundo trabalho que mais perdeu posição, de acordo com o ranque de sentimento construído por área, foi o trabalho “*Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017*”.

Apesar do trabalho “*A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy*” possuir o maior altmetric score (8974) dentre os trabalhos investigados, observa-se que para as técnicas Sentistrength, Sentiwordnet e Textblob, tal trabalho aparece apenas na segunda ou terceira posição do ranque de sentimento.

No que diz respeito ao título com menor altmetric score (*Restoration of brain circulation and cellular functions hours post-mortem*), nota-se que no ranque da técnica Senticnet tal trabalho conquistou uma posição, diferentemente das outras abordagens testadas, onde esse trabalho manteve a última colocação no ranque.

Tendo em vista os diferentes ranques por área gerados pelas técnicas de IA, um terceiro estudo que consiste na combinação das técnicas de IA foi apresentado.

5.3 ESTUDO 3: ANÁLISE DA COMBINAÇÃO DE TÉCNICAS DE IA NA ALTMETRIA: IA ALTMETRIC

O último estudo deste documento de tese consistiu na combinação das técnicas de Inteligência Artificial investigadas. Para isso, utilizou-se a polaridade resultante das menções classificadas por cada técnica, gerando um novo ranque e uma nova perspectiva de cunho qualitativo para mensurar as menções da web social, assim, tal métrica foi intitulada como IA Altmetric.

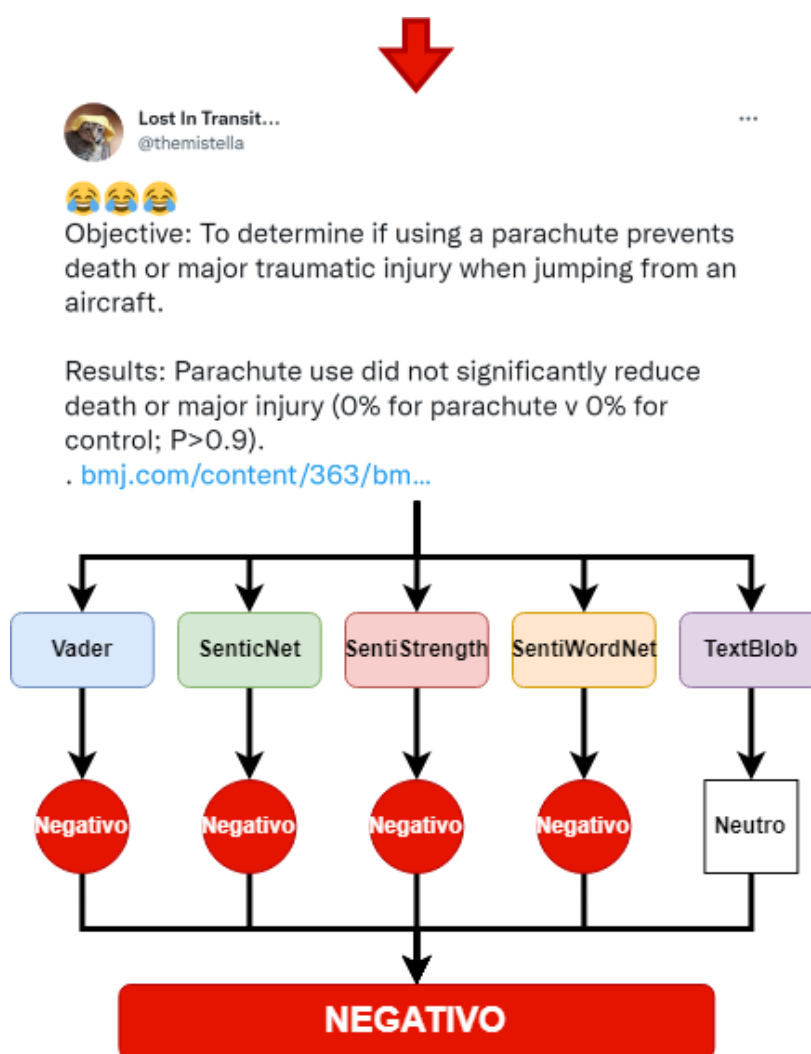
O IA Altmetric utiliza o número de ocorrências com que diferentes técnicas concordam na polaridade de um conteúdo. Por exemplo, quando três técnicas detectam sentimentos em um texto, há a possibilidade de verificar se esses sentimentos foram idênticos, o que seria capaz de fortalecer a confiança da classificação. Para isso, são utilizados os resultados das seguintes técnicas: Vader, SenticNet, SentiStrength, SentiWordNet e TextBlob. Assim, analisa a

classificação de cada técnica, assumindo a polaridade que obteve maior número de ocorrências dentre o grau positivo, neutro e negativo.

Faz-se importante mencionar que nas situações em que não houve uma predominância da polaridade (positivo, negativo ou neutro), classificou-se a menção como sendo neutra.

A título de exemplo, a Figura 14 demonstra o processo de classificação da técnica IA Altmetric para uma menção extraída do twitter, referente ao trabalho intitulado “*Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial*”.

Figura 14 - Processo de classificação da técnica combinada IA Altmetric.



Fonte: Elaborado pelo autor.

Como pode ser visto na Figura 14, ao analisar a menção pelos classificadores de sentimento, gerou-se um conjunto de resultados onde as técnicas Vader, SenticNet, SentiStrength e SentiWordNet classificaram a respectiva menção como sendo negativa. Por

outro lado, a técnica TextBlob classificou a menção como neutra. Deste modo, para obter a classificação IA Altmetric, combinou-se os cinco resultados gerados, de modo que o resultado foi a polaridade negativa (4 negativos e 1 neutro).

Em complemento, uma possível maneira de representar a combinação dos resultados das técnicas de IA é por meio de um pseudocódigo, que consiste em uma linguagem simples, sendo possível escrever, a partir de regras predefinidas, os passos a serem seguidos, sem que haja a necessidade de conhecimento de linguagens de programação.

A Figura 15 descreve o pseudocódigo da técnica IA Altmetric.

Figura 15 - Pseudocódigo da técnica IA Altmetric.

ALGORITMO *ia_altmetric*

```

1      var  vader, senticnet, sentistrenght, sentiwordnet, textblob: string
2      var  positivo, negativo, neutro, ia: string
3  INÍCIO
4      positivo, negativo, neutro = 0
5      SE vader = “positivo” FAÇA
6          positivo <- positivo + 1
7      SE senticnet = “positivo” FAÇA
8          positivo <- positivo + 1
9      SE sentistrenght = “positivo” FAÇA
10         positivo <- positivo + 1
11     SE sentiwordnet = “positivo” FAÇA
12         positivo <- positivo + 1
13     SE textblob = “positivo” FAÇA
14         positivo <- positivo + 1
15     SE vader = “negativo” FAÇA
16         negativo <- negativo + 1
17     SE senticnet = “negativo” FAÇA
18         negativo <- negativo + 1
19     SE sentistrength = “negativo” FAÇA
20         negativo <- negativo + 1
21     SE sentiwordnet = “negativo” FAÇA
22         negativo <- negativo + 1

```

```

23      SE textblob = “negativo” FAÇA
24          negativo <- negativo + 1
25      SE positivo >=3 FAÇA
26          ia <- positivo
27      SENÃO
28      SE negativo >=3 FAÇA
29          ia = negativo
30      SENÃO
31          ia = neutro
32  FIM

```

Fonte: Elaborado pelo autor.

Com o objetivo de identificar vantagens, desvantagens e possíveis limitações da IA Altmétric, aplicou-se tal técnica para os trabalhos da área *Medical and Health Sciences* investigados no estudo 2, comparando-os com os trabalhos de maior altmetric score. Neste estudo, analisou-se os trabalhos que mais subiram e perderam posições no ranque. A seguir, a Tabela 50 apresenta os resultados encontrados e compara-os aos do ranque de maior altmetric score.

Observa-se que para o ranque da técnica IA Altmétric o trabalho “*Cast of the Right Bronchial Tree*” conquistou o maior número de posições, subindo exatamente 3 posições. Enquanto o trabalho intitulado “*Measles, Mumps, Rubella Vaccination and Autism*” foi o que mais perdeu posição no ranque, descendo 3 posições. Nota-se, também, que os trabalhos “*A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy*” e “*Restoration of brain circulation and cellular functions hours post-mortem*” não tiveram perda ou ganho de posição no ranque.

Ao comparar o trabalho que mais desceu posições da técnica IA Altmétric, com os que mais perderam nos ranques das demais técnicas de IA, é possível notar uma ligeira semelhança, visto que quatro das cinco técnicas de IA investigadas apontaram que o trabalho intitulado “*Measles, Mumps, Rubella Vaccination and Autism*” foi o que teve a pior repercussão na web social.

Apesar do trabalho “*Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems*” ser o único trabalho que ganhou posições em todos ranques

de sentimento investigados, o trabalho “*Cast of the Right Bronchial Tree*” foi o que mais ganhou posição no ranque IA Altmetric.

Tabela 50 - Comparativo entre os ranques de trabalhos com maior altmetric score e de maior pontuação de sentimento positivo IA Altmetric de acordo com a área Medical and Health Sciences

RANQUE	TÍTULO	IA Altmetric
1	A Randomized Trial of E-Cigarettes versus Nicotine-Replacement Therapy	4664
2	Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial	4216
3	Cast of the Right Bronchial Tree	3403
4	Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems	3013
5	Measles, Mumps, Rubella Vaccination and Autism	2758
6	Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017	1584
7	Restoration of brain circulation and cellular functions hours post-mortem	631

Fonte: Elaborado pelo autor.

5.4 CONSIDERAÇÕES FINAIS

Com o objetivo de qualificar a etapa de avaliação da implementação proposto pela metodologia DSR, foi apresentado o comportamento do modelo proposto organizado em três estudos distintos, na quais foram possível provar o conceito, tornando a proposta factível. Assim, de acordo com o modelo e os resultados apresentados nesta seção, verificou-se que o objetivo de viabilizar uma perspectiva de análise qualitativa no contexto da Altmatria, que utilize técnicas de Inteligência Artificial foi atendido. Tal afirmação é realizada, tendo como base o desenvolvimento do modelo teórico e os resultados dos estudos que demonstraram diferentes cenários em relação ao nível de atenção de um trabalho.

Os estudos demonstraram que a proposta criada é passível de ser executada, viabilizando a aplicação de técnicas de Inteligência Artificial dentro do contexto da Altmatria. Outro ponto importante está no conteúdo das menções realizadas por leitores, pelo fato de tais menções possuírem textos de cunho opinativo possibilitando proposta qualitativa.

Vale destacar que os estudos realizados possibilitaram a prova de conceito, sendo possível estender a aplicação da análise qualitativa em outros trabalhos nas quais não foram mencionados nesta tese. A partir dessa premissa, é possível a construção de um protótipo que faça tal análise de modo automatizado e instantâneo.

PROTÓTIPO

As atividades realizadas ao longo desta tese foram divididas em 3 partes. A primeira estava direcionada a fazer toda a revisão bibliográfica a respeito da qualidade das menções na Altmtria, Tal fase teve como objetivo a extração de ideias relacionadas com o tema e obtenção de conhecimento necessários para posteriormente desenvolver o sistema proposto.

A segunda fase teve como intuito provar o conceito por meio de estudos, nessa etapa procurou-se avaliar a melhor forma de processar as menções da web social. Também foi analisado nessa etapa os diferentes cenários gerados pelas diferentes técnicas de inteligência artificial.

Por fim, a terceira parte do desenvolvimento da tese foi a implementação do protótipo qualitativo altmétrico denominado IA Altmtric. O IA Altmtric foi baseado em um modelo no qual o pesquisador e usuário podem registrar e buscar informações referentes a um trabalho existente na plataforma altmetric por meio do identificador DOI. Assim, foi desenvolvido uma aplicação web que conta com algumas tecnologias que ligadas entre si formam todo o sistema. Essas tecnologias envolvem, interfaces de usuários, banco de dados e servidor web.

Todas as etapas da tese foram importantes tanto para a formação do conhecimento a respeito das menções na web social, como para a extração de informações relevantes para o desenvolvimento desta tese.

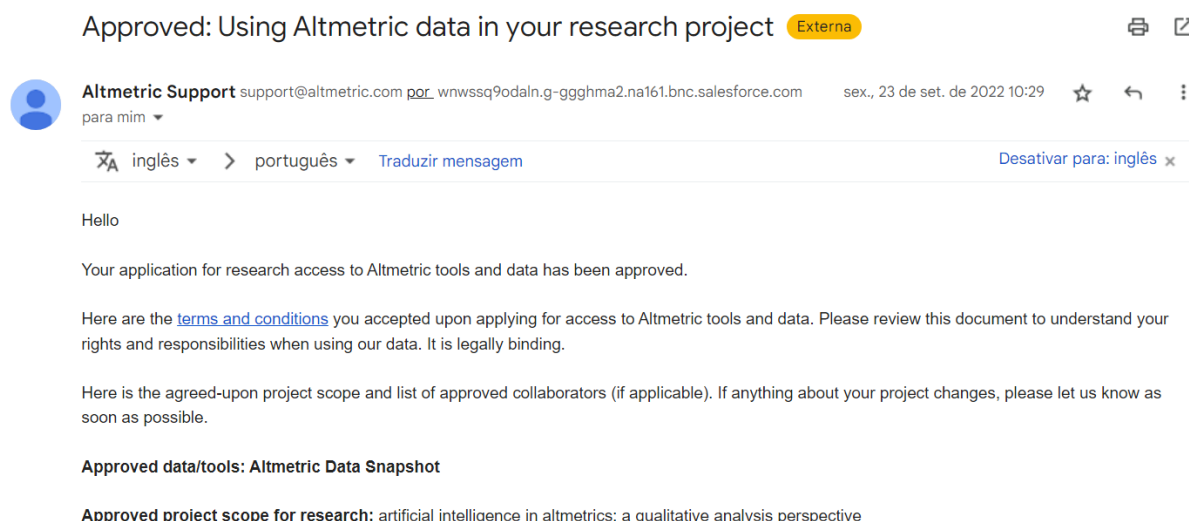
A seção a seguir são apresentadas as ferramentas auxiliares utilizadas para o desenvolvimento do protótipo com o intuito de mostrar uma visão ampla de todo o sistema.

6.1 FERRAMENTAS AUXILIARES

Para a construção do protótipo IA Altmtric, foi utilizado a base de dados completa da plataforma Altmtric.com intitulada “Altmtric Data Snapshot” (*atualização de 23 de abril de 2022*). Para ter acesso a tal base foi necessário contactar a empresa detalhando a proposta de pesquisa e o intuito da utilização dos dados.

A Figura 16 apresenta o email recebido pela Altmtric, autorizando o uso dos dados na pesquisa.

Figura 16 - Autorização do uso dos dados da Altmetric.



Fonte: Elaborada pelo autor.

Em conjunto com a aprovação da utilização dos dados para a pesquisa, foi recebido um arquivo possuindo cerca de 160 Gygabytes de dados pertencentes a plataforma Altmetric. Nesta base, há dados como: pontuação de atenção, menções de blogs, notícias, facebooks, identificador do tweet, título do trabalho, DOI, revista publicada, entre outras informações.

Para recuperar os textos dos tweets na base foi necessário utilizar a API do Twitter, de modo que ao inserir o identificador do tweet fosse recuperada a menção, sendo utilizada a biblioteca tweepy desenvolvida em Python.

Em relação a parte visual do protótipo foi utilizado o Bootstrap, que é um framework para desenvolvimento de sistemas web utilizando HTML (Hyper Text Markup Language), CSS (Cascading Style Sheets) e Javascript, contendo elementos de interface com usuário como: botões, títulos, tabelas, textos e etc. Além disso, para a apresentação dos dados foi utilizada a linguagem PHP (*Hypertext Preprocessor versão 8*). Tais tecnologias foram utilizadas devido ao conhecimento prévio das ferramentas e também levando em consideração o alto potencial que essas tecnologias fornecem para a construção de sistemas web dinâmicos. O SGBD utilizado na aplicação foi o MySQL, que é um gerenciador de banco de dados muito utilizado na web

Como já mencionado, o protótipo seguiu os passos do modelo qualitativo apresentado, portanto foram utilizadas as bibliotecas de análise de sentimento: vadersentiment, Sentic, Sentistrength, Textblob e Sentiwordnet. Além disso, o protótipo limitou-se as menções do idioma inglês, sendo selecionadas por meio da biblioteca langdetect. Tais bibliotecas são implementadas na linguagem Python.

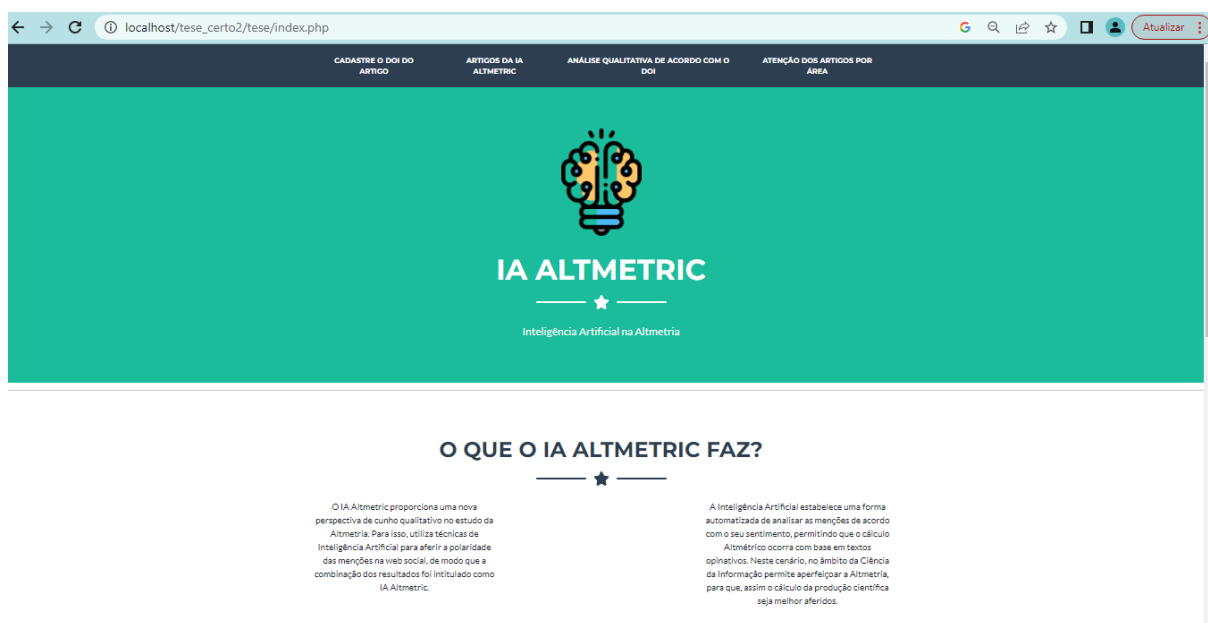
6.2 RESULTADOS

Um dos artefatos gerado nesse trabalho foi uma aplicação web que analisa qualitativamente as menções da plataforma altmetric de acordo com um determinado DOI selecionado, de modo que é apresentado ao usuário a quantidade de menções pertencentes ao sentimento positivo, negativo e neutro.

O protótipo IA Altmetric foi desenvolvido e testado localmente (localhost), na máquina do desenvolvedor, tal fato justifica-se por motivos financeiros, pois seria necessário carregar o sistema em um servidor web com alta capacidade de armazenamento e processamento, inviabilizando a implementação em ambiente público.

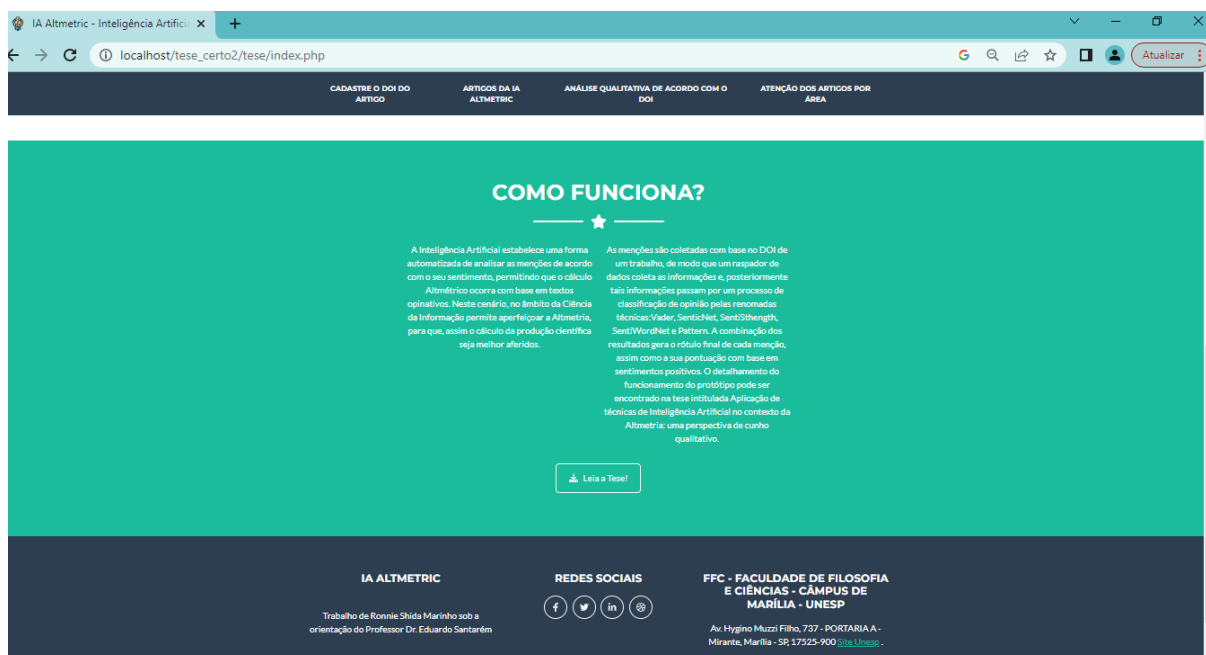
As Figura 17 e 18 apresentam a tela inicial do protótipo. Nessa tela são apresentadas informações de como utilizar a ferramenta.

Figura 17 - Tela inicial do protótipo.



Fonte: Elaborada pelo autor.

Figura 18 - Tela inicial do protótipo 2.



Fonte: Elaborada pelo autor.

O sistema foi baseado em um modelo no qual o pesquisador pode cadastrar informações de um determinado DOI solicitando o processamento das menções que compõem a nota de atenção na plataforma Altmetric.

A Figura 19 ilustra a tela de cadastro do trabalho, onde o usuário, informará alguns dados solicitados, tais como: DOI, título do trabalho, link onde o trabalho encontra-se disponível e a área de estudo, sendo que destes dados o único campo obrigatório é o identificador DOI, pois como mencionado anteriormente é o atributo identificador de um título. Já a Figura 20 é demonstrado um pequeno exemplo de como preencher os dados do formulário para o título *Instagram photos reveal predictive markers of depression*. Tal título apresenta o DOI *10.1140/epjds/s13688-017-0110-z*.

Figura 19 - Cadastro das informações do título a ser processado.

The screenshot shows a web browser at the URL `localhost/tese_certo2/tese/form_trabalho.php`. The page has a dark blue header with navigation links: "CADASTRE O DOI DO ARTIGO", "ARTIGOS DA IA ALTMETRIC", "ANÁLISE QUALITATIVA DE ACORDO COM O DOI", and "ATENÇÃO DOS ARTIGOS POR ÁREA". Below the header is a green banner with the "IA ALTMETRIC" logo and the tagline "Inteligência Artificial na Altmtria". The main content area is titled "CADASTRO DO DOI DIGITAL OBJECT IDENTIFIER" and contains four input fields: "Digite o DOI", "Digite o Título", "Insira a URL do trabalho", and a dropdown menu labeled "--Selecione a área do trabalho--". A green "Enviar informações" button is at the bottom.

Fonte: Elaborado pelo autor.

Figura 20 - Exemplo de preenchimento de cadastro do título.

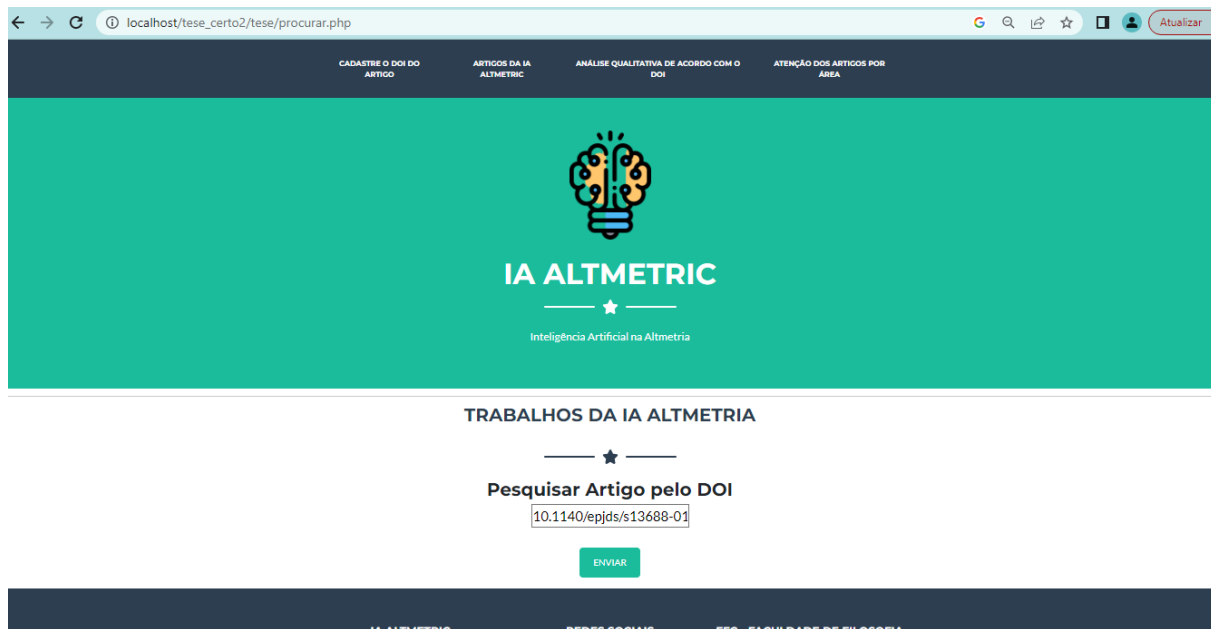
This screenshot shows the same registration form as Figure 19, but with example data entered. The "Digite o DOI" field contains "10.1140/epjds/13688-017-0110-z", the "Digite o Título" field contains "Instagram photos reveal predictive markers of depression", and the "Insira a URL do trabalho" field contains "asscience.springeropen.com/articles/10.1140/epjds/13688-017-0110-z". The dropdown menu remains at "--Selecione a área do trabalho--". The "Enviar informações" button is still present at the bottom.

Fonte: Elaborada pelo autor.

Após cadastrar os dados do título alvo no protótipo, é necessário aguardar o processamento das menções e atualização do sistema. Para verificar se a análise qualitativa do trabalho se encontra disponível, é necessário acessar a aba “*Análise qualitativa de acordo com o doi*” e inserir o DOI do trabalho alvo.

Após clicar no botão “enviar” é exibida uma tela Além disso, essa funcionalidade permite o usuário buscar por um determinado DOI que possivelmente já tenha sido cadastrado no sistema. Tal etapa está exemplificada e ilustrada na Figura 21.

Figura 21 - Pesquisa de artigo por meio do DOI.

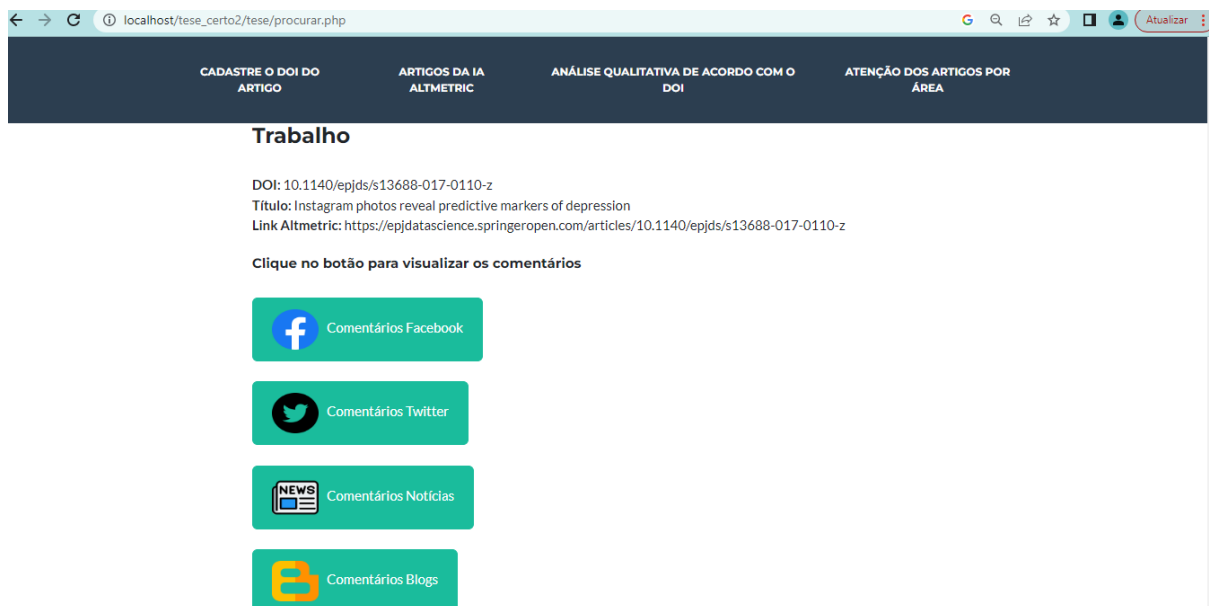


Fonte: 2 Elaborada pelo autor.

Após processar e verificar a existência do título na base de dados são apresentadas informações referentes as menções, tais como a quantidade de menções classificadas como positivas, neutras e negativas para as técnicas de Inteligência artificial investigadas. É possível visualizar as menções coletadas de cada fonte como twitter, blogs, notícias e facebook com sua respectiva classificação.

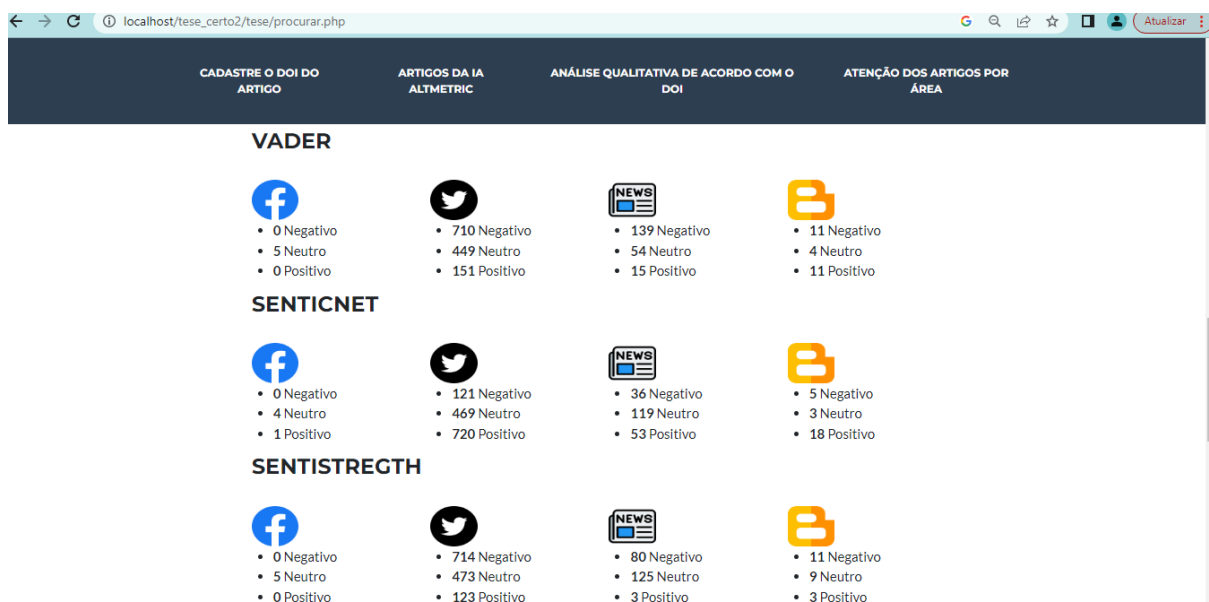
As Figura 22, 23 e 24 ilustram as informações disponibilizadas para um determinado DOI (*Instagram photos reveal predictive markers of depression*).

Figura 22- Apresentação das informações qualitativas para o DOI investigado (10.1140/epjds/s13688-017-0110-z).



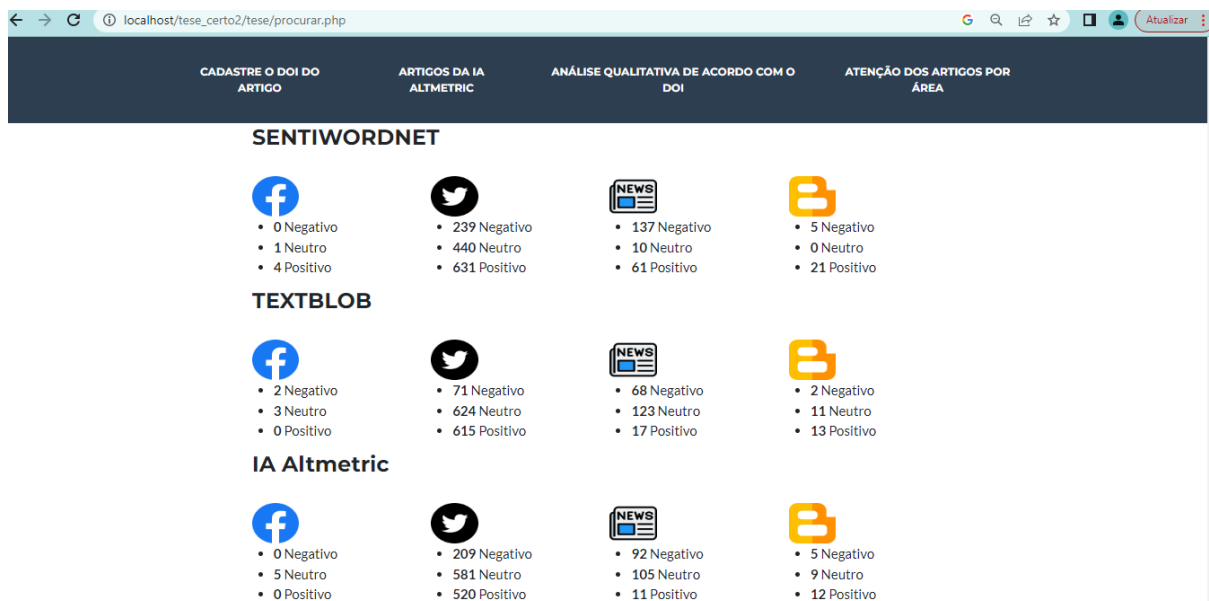
Fonte: Elaborada pelo autor.

Figura 23 - Apresentação da quantidade de menções classificadas como positivas, neutras e negativas de acordo com a técnica.



Fonte: Elaborada pelo autor.

Figura 24 - Apresentação da quantidade de menções classificadas como positivas, neutras e negativas de acordo com a técnica. (parte 2)



Fonte: Elaborada pelo autor.

6.3 CONSIDERAÇÕES FINAIS

Com o intuito de gerar uma ferramenta altmétrica qualitativa, foi apresentado o protótipo do modelo proposto, tornando a proposta possível. Desta forma, de acordo com o protótipo e os resultados apresentados nesta seção, analisou-se que o objetivo de desenvolver uma perspectiva de análise qualitativa no cenário da Altmetria, por meio de técnicas de Inteligência Artificial foi respondido. Tal afirmação é realizada, tendo como base o desenvolvimento do modelo teórico e os resultados do protótipo que demonstraram diferentes cenários em relação ao nível de atenção de um trabalho.

A criação do protótipo demonstrou que a proposta desta tese foi possível de ser executada devido o desenvolvimento do modelo qualitativo, assim como a análise e prova de conceito realizada no decorrer dos estudos desta pesquisa.

CONCLUSÃO

A aferição qualitativa da importância de um trabalho utilizando menções de redes sociais como fonte de dados é uma forma de elevar a compreensão dos pesquisadores, visando não somente valores quantitativos. Assim, apesar da Altmtria ser amplamente utilizada como um indicador de atenção, uma pontuação elevada pode não refletir as opiniões dos usuários que mencionam um determinado trabalho.

A partir desse cenário, apresentou-se o modelo qualitativo de sentimento, como uma forma de analisar a qualidade do conteúdo opinativo nas menções que compõem a pontuação de atenção na Altmtria. Para isso, a aplicação de técnicas de Inteligência Artificial possibilitou diferentes resultados no processo qualitativo dos dados, proporcionando cenários interessantes para estudo. Deste modo, frente aos resultados ora apresentados nos estudos 1,2 e 3, acredita-se no potencial da Inteligência Artificial, em especial, nos métodos provenientes da análise de sentimento.

Ao analisar os diferentes cenários gerados pelo modelo proposto, percebemos que as cinco técnicas de IA investigadas possuem variadas limitações, vantagens e desvantagens, dificultando a análise de qual técnica apresentou melhores resultados. Tais resultados levaram à construção de uma nova técnica, intitulada IA Altmtric, que consiste na combinação dos métodos Vader, SenticNet, SentiStrenght, SentiWordNet e Textblob, na tentativa de alcançar resultados satisfatórios.

Deste modo, o modelo proposto por esse trabalho teve como objetivo geral o desenvolvimento de uma perspectiva de análise qualitativa no cenário da Altmtria, a partir da aplicação de técnicas de Inteligência Artificial utilizando as opiniões de leitores em menções da web social. Para alcançar tal objetivo, foi utilizada a metodologia DSR, como forma de organizar as etapas e validar o artefato gerado. Ao comparar o comportamento do modelo qualitativo de sentimento com os resultados quantitativos altmétricos, foi possível provar o conceito defendido nessa tese.

Em relação a metodologia, o DSR foi apresentado, relatando uma pesquisa de desenvolvimento de um modelo aplicado à Altmtria. Cada etapa do processo de desenvolvimento foi abordada, desde a investigação do problema até a etapa final de avaliação da implementação.

Os artefatos gerados em conjunto com as justificativas teóricas utilizadas para embasá-los em um ciclo sistemático de desenvolvimento, torna interessante a utilização da metodologia para pesquisas na área de Ciência da Informação.

Analogamente, os objetivos específicos deste trabalho foram alcançados no decorrer das etapas do ciclo DSR. O primeiro visava em analisar o processo de construção dos indicadores na Altmtria, na qual foi atingido na etapa de investigação do problema, quando, detectou opiniões nas menções da web social.

Já o segundo devia propor um modelo qualitativo no contexto da Altmtria, sendo atendido na etapa projeto de soluções, que demonstrou as fases necessárias para solucionar o problema em questão.

O terceiro estava em relacionar como a Inteligência Artificial pode apoiar o estudo da Altmtria, sendo alcançado na etapa validação de projeto, que apresentou a aplicação de técnicas de IA em conjunto com as menções extraídas da Altmtria para validar o projeto. Ainda nesta etapa, alcançou o quarto objetivo específico de viabilizar um novo parâmetro de avaliação que ajude pesquisadores na tomada de decisão sobre a leitura de um trabalho científico.

Posteriormente, o quinto e sexto objetivos específicos consistiam sobre apresentar diferentes cenários qualitativos gerados por meio da aplicação de técnicas de IA, tal objetivo foi constatado na etapa de implementação da solução, onde foram detalhadas as configurações referentes ao desenvolvimento do modelo.

O sétimo objetivo específico consistia em validar o modelo proposto por meio de uma prova de conceito, tal prova foi atingida na etapa / seção avaliação da implementação em que demonstrou o comportamento e resultados do modelo qualitativo. O oitavo objetivo específico visava em demonstrar de forma prática a técnica combinada intitulada IA Altmtric, na qual também foi alcançada na etapa de avaliação da implementação, em que descreveu os passos realizados para o funcionamento de tal abordagem.

Por fim, o nono e último objetivo específico visava a implementação de um protótipo utilizando o modelo Altmétrico qualitativo, sendo atendido na etapa de avaliação da implementação.

Analisando os objetivos específicos alcançados, o trabalho mostrou que o tema desta pesquisa ainda é pouco explorado na literatura. Entretanto, refere-se a uma lacuna importante de pesquisa, que pode beneficiar a Altmtria ao propor um modelo que utiliza técnicas de Inteligência Artificial para aferir de modo qualitativo as menções na Web social, favorecendo a medição da produção científica no contexto de mídias sociais, bem como a criação de uma nova perspectiva da área da Altmtria.

Adicionalmente, aponta-se que a Inteligência Artificial, quando compreendida de forma conceitual e enquanto um campo de estudos associado à Ciência da Computação, é possível de colaborar com a Ciência da Informação a uma evolução, em que as tecnologias computacionais

possibilitem um aperfeiçoamento e a criação de novos campos de estudos. Assim, a Inteligência Artificial pode ser redescoberta como um campo de estudos associado a Ciência da Informação.

Outro elemento importante que é parte central do trabalho desenvolvido diz respeito a Altmtria, por meio de seu conceito, ferramentas e aplicações. A Altmtria foi utilizada como insumo para compreensão das menções na web social. A partir das técnicas de Inteligência Artificial, a Altmtria pode ser explorada com mais profundidade, podendo efetivamente sugerir uma nova análise qualitativa.

Ressalta-se que este trabalho demonstra a viabilidade de uma nova perspectiva no contexto da altmtria, ao indicar possibilidades de aferir o sentimento contido em uma menção da web social. Neste contexto, há uma análise da quantidade de menções de cunho opinativos possibilitando que o modo como o usuário encontre a informação possa ser com maior significado e mais expressivo.

Portanto, por meio desta tese, foi possível validar e definir uma nova perspectiva no contexto da Altmtria, que parte dos conceitos e dos elementos da Ciência da Informação, apoiado por técnicas de Inteligência Artificial, permitindo assim um avanço e uma evolução nesses campos de estudos.

Destaca-se que a partir dessa nova visão no estudo da Altmtria, realiza-se uma mudança no modo como a Ciência da Informação se relaciona com os conceitos de Inteligência Artificial nesse processo.

Vale ressaltar que a relação entre Ciência da Informação e Ciência da Computação guiou o desenvolvimento desta pesquisa, se complementando e utilizando elementos que permitiram a construção do modelo proposto. O ponto de vista técnico da Ciência da Computação ao demonstrar técnicas e teorias da Inteligência Artificial, junto as discussões e aplicações da Ciência da Informação, foram expandidos, possibilitando a interdisciplinaridade desta tese.

Em relação as limitações do presente trabalho, aponta-se que a depender do número de menções de um determinado título cadastrado no protótipo, pode levar um certo tempo para os dados serem processados e disponibilizados na plataforma.

Enquanto trabalhos futuros, pretende-se realizar um estudo detalhado com as menções classificadas como negativas e neutras. Além disso, busca-se disponibilizar o protótipo do modelo em um servidor web, de modo que tal protótipo contribua com a Altmtria, sendo uma complementação das métricas tradicionais altmétricas já conhecidas.

REFERÊNCIAS

ADIE, Euan; ROE, William. Altmetric: enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, v. 26, n. 1, p. 11-17, 2013.

ALPERIN, Juan Pablo. Ask not what Altmetrics can do for you, but what Altmetrics can do for developing countries. *Bulletin of the American Society for Information Science and Technology*, v. 39, n. 4, p. 18-21, 2013.

ARAUJO, Ronaldo Ferreira; MURAKAMI, Tiago Rodrigo Marçal; PRADO, Jorge Moisés Kroll do. A repercussão de artigos de periódicos brasileiros da ciência da informação no Facebook: um estudo altmétrico. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, v. 16, n. 2, p. 365-379, 2018.

BARBOSA, Daniel Mendes; BAX, Marcello. A Design Science como metodologia para a criação de um modelo de Gestão da Informação para o contexto da avaliação de cursos de graduação. **Revista Ibero-Americana de Ciência da Informação (RICI); De Ciência Da Informação, F., Ed**, p. 32-48, 2017.

BECKER, K.; TUMITAN, D. Introdução à mineração de opiniões: conceitos, aplicações e desafios. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 28., 2013, Recife. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2013. p. 27-52.

BENEVENUTO, Fabrício; RIBEIRO, Filipe; ARAÚJO, Matheus. Métodos para Análise de Sentimentos em mídias sociais. In: *Brazilian Symposium on Multimedia and the Web (Webmedia)*. Manaus, Brazil. 2015.

BISWAS, S. et al. Examining the effects of pandemics on stock market trends through sentiment analysis. *Xi'an Dianzi Keji Daxue Xuebao/Journal of Xidian University*, v. 14, p. 1–14, 06 2020.

BOLLEN, Johan et al. A principal component analysis of 39 scientific impact measures. *PloS one*, v. 4, n. 6, p. e6022, 2009.

BOLLEN, Johan; MAO, Huina; PEPE, Alberto. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In: Fifth International AAAI Conference on Weblogs and Social Media. 2011.

BORBA, Vildeane; MARINHO, Andréa Carla Melo; CAREGNATO, Sônia Elisa. Análise do termo “Repositório Institucional” no Twitter: um estudo altmétrico. *Em Questão*, v. 23, p. 290-308, 2017.

CAMBRIA, Erik et al. Senticnet: A publicly available semantic resource for opinion mining. In: 2010 AAAI fall symposium series. 2010.

CAMBRIA, E.; LIVINGSTONE, A.; HUSSAIN, A. The Hourglass of Emotions. In: ESPOSITO, A.; ESPOSITO, A. M.; VINCIARELLI, A.; HOFFMANN, R.; MÜLLER, V. C. (org.). *Cognitive Behavioural Systems*. Heidelberg: Springer, 2012. Disponível em: http://dx.doi.org/10.1007/978-3-642-34584-5_11. Acesso em: 30 jun. 2021.

CAVALCANTI, D. C., PRUDÊNCIO, R. B., PRADHAN, S. S., SHAH, J. Y., & PIETROBON, R. S. (2012). Análise de sentimento em citações científicas para definição de fatores de impacto positivo. In *Proceedings of the IV International Workshop on Web and Text Intelligence (WTI)* (pp. 1-10). sn.

CHANG, J.; DESAI, N.; GOSAIN, A. Correlation Between Altmetric Score and Citations in Pediatric Surgery Core Journals. *Journal of Surgical Research*, v. 243, p. 52–58, nov. 2019.

DODDS, Peter Sheridan; DANFORTH, Christopher M. Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of happiness studies*, v. 11, n. 4, p. 441-456, 2010.

DRESCH, Aline; LACERDA, Daniel Pacheco; JÚNIOR, José Antonio Valle Antunes. **Design science research: método de pesquisa para avanço da ciência e tecnologia**. Bookman Editora, 2015.

ESULI, Andrea; SEBASTIANI, Fabrizio. Sentiwordnet: A publicly available lexical resource for opinion mining. In: LREC. 2006. p. 417-422.

FANG, Zhichao; COSTAS, Rodrigo. Studying the accumulation velocity of altmetric data tracked by Altmetric. com. Scientometrics, v. 123, n. 2, p. 1077-1101, 2020.

GISBERT, Javier P.; PANÉS, Julián. Índice H de Hirsch: una nueva herramienta para medir la producción científica. **Cir. Esp.(Ed. impr.)**, v. 86, n. 4, p. 193-195, 2009.

GRÁCIO, Maria Cláudia Cabrini et al. (Ed.). Tópicos da bibliometria para bibliotecas universitárias. Editora Oficina Universitária, 2020.

GONTIJO, Marília Catarina Andrade; DE ARAÚJO, Ronaldo Ferreira. Dados bibliométricos e altmétricos de artigos científicos sobre Inteligência Artificial. Múltiplos Olhares em Ciência da Informação, v. 9, n. 2, 2019.

GOUVEIA, Fábio Castro. Altmetria: métricas de produção científica para além das citações | Altmetrics: scientific production metrics beyond citations. Liinc em revista, v. 9, n. 1, 2013.

GOUVEIA, Fabio Castro et al. A Altmetria e a interface entre a ciência e a sociedade. 2016. Trab. Educ. saúde, Rio de Janeiro, v. 14, n. 3, p.643-645.

HIRSCH, Jorge E. An index to quantify an individual's scientific research output. Proceedings of the National academy of Sciences, v. 102, n. 46, p. 16569-16572, 2005.

HUTTO, C. J.; VADER, Gilbert E. A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. 2015 Jan Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014. Ann Arbor, MI URL: <http://eegilbert.org/papers/icwsm14.vader.hutto.pdf>.

KAUFMAN, Dora. A inteligência artificial irá suplantará a inteligência humana?. ESTAÇÃO DAS LETRAS E CORES EDI, 2019.

LACERDA, Daniel Pacheco et al. Design Science Research: método de pesquisa para a engenharia de produção. *Gestão & produção*, v. 20, p. 741-761, 2013.

LIDDY, E. D. Natural language processing. In **Encyclopedia of Library and Information Science**, 2nd Ed. NY. Marcel Decker, Inc. 2001. Disponível em: <<https://surface.syr.edu/cgi/viewcontent.cgi?referer=http://scholar.google.com.br/&httpsredir=1&article=1019&context=cnlp>>. Acesso em: 05 jul. 2022.

LIU, Bing; ZHANG, Lei. A survey of opinion mining and sentiment analysis. In: *Mining text data*. Springer, Boston, MA, 2012. p. 415-463.

LORIA, Steven et al. textblob Documentation. Release 0.15, v. 2, n. 8, 2018.

MARINHO, Ronnie Shida; PEREIRA, Clayton Martins; SEGUNDO, José Eduardo Santarem. Análise de sentimento como apoio à seleção de livros: um estudo aplicado à plataforma Skoob. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 27, p. 1-20, 2022.

MATTOS, Pedro Lincoln CL. " Bibliometria": a metodologia acadêmica convencional em questão. *RAE-eletrônica*, v. 3, n. 2, p. 1-6, 2004.

MCFEDRIES, Paul. Measuring the impact of Altmetrics [Technically Speaking]. *IEEE Spectrum*, v. 8, n. 49, p. 28, 2012.

MEADOWS, Arthur Jack; DE LEMOS LEMOS, Antonio Agenor Briquet. A comunicação científica. Briquet de Lemos/livros, 1999.

MILLER, George A. WordNet: a lexical database for English. *Communications of the ACM*, v. 38, n. 11, p. 39-41, 1995.

NASCIMENTO, Andréa G. Altmetria para bibliotecários: guia prático de métricas alternativas para avaliação da produção científica. *Revolução eBook*, 2016.

NASCIMENTO, Andrea G.; ODDONE, Nanci. Uso de indicadores altmetrics na avaliação de periódicos científicos brasileiros em Ciência da Informação. 2014.

NASCIMENTO, Paula et al. Análise de sentimento de tweets com foco em notícias. In: Brazilian Workshop on Social Network Analysis and Mining. 2012.

NOAH, A.; PEARCE, I.; MODGIL, V. Altmetrics and citations: the impact of journal publications. *The Bulletin of the Royal College of Surgeons of England*, v. 102, n. 7, p. 324–327, 1 out. 2020.

PÅLSSON, Andreas; SZERSZEN, Daniel. Sentiment Classification in Social Media: An Analysis of Methods and the Impact of Emoticon Removal. 2016.

PANG, Bo; LEE, Lillian; VAITHYANATHAN, Shivakumar. Thumbs up? Sentiment classification using machine learning techniques. **arXiv preprint cs/0205070**, 2002.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, Delft, v. 2, n. 1-2, p. 1-94, 2008.

PAULINO, Rita de Cássia Romeiro; EMPINOTTI, Marina Lisboa. Altmertia e jornalismo: análise de comentários do Twitter para verificar a participação do leitor em notícias de impacto. *Ameaças ao Ciberjornalismo*, p. 303.

PEFFERS, Ken et al. A design science research methodology for information systems research. **Journal of management information systems**, v. 24, n. 3, p. 45-77, 2007.

PIMENTEL, Mariano; FILIPPO, Denise; SANTORO, Flávia Maria. Design Science Research: fazendo pesquisas científicas rigorosas atreladas ao desenvolvimento de artefatos computacionais projetados para a educação. *Metodologia de Pesquisa em Informática na Educação: Concepção da Pesquisa*. Porto Alegre: SBC, 2019.

PIRYANI, Rajesh et al. Book impact assessment: A quantitative and text-based exploratory analysis. **Journal of Intelligent & Fuzzy Systems**, v. 34, n. 5, p. 3101-3110, 2018.

PRIEM, Jason et al. Altmetrics: A manifesto. 2010.

PRIEM, Jason; GROTH, Paul; TARABORELLI, Dario. The Altmetrics collection. PloS one, v. 7, n. 11, p. e48753, 2012.

PRITCHARD, Alan et al. Statistical bibliography or bibliometrics. Journal of documentation, v. 25, n. 4, p. 348-349, 1969.

RICH, Elaine et al. Inteligencia artificial. McGraw-Hill, 1994.

ROCHADEL, Willian. As ciências do artificial em uma abordagem entre o conhecimento e a tecnologia. Seminário de Pesquisa, Pós-Graduação e Inovação, 2016.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence**: a modern approach. Malaysia; Pearson Education Limited, 2016.

SANTOS, Raimundo Nonato Macedo dos. Produção científica: por que medir? O que medir?. Revista digital de Biblioteconomia e Ciência da Informação, v. 1, n. 1, 2003.

SANTOS, Leandro Matioli. Protótipo para mineração de opinião em redes sociais: estudo de casos selecionados usando o Twitter. Monografia. Departamento de Ciência da Computação, Universidade Federal de Lavras, 2010.

SERRA, Liliana Giusti et al. Os princípios da descrição e os formatos MARC 21 e ONIX. Ciência da Informação, v. 46, n. 2, 2017.

SILVA, Edna Lucia; MENEZES, Estera Muszkat. Metodologia da pesquisa e elaboração de dissertação. UFSC, Florianópolis, 4a. edição, v. 123, 2005.

SIMON, Herbert A. Cognitive science: The newest science of the artificial. **Cognitive science**, v. 4, n. 1, p. 33-46, 1980.

SMEDT, Tom; DAELEMANS, Walter. Pattern for python. The Journal of Machine Learning Research, v. 13, n. 1, p. 2063-2067, 2012.

STONE, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis.

TABOADA, Maite et al. Lexicon-based methods for sentiment analysis. *Computational linguistics*, v. 37, n. 2, p. 267-307, 2011.

TAUSCZIK, Yla R.; PENNEBAKER, James W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, v. 29, n. 1, p. 24-54, 2010.

TAYLOR, Mike. The challenges of measuring social impact using Altmetrics. *Research trends*, v. 33, p. 11-15, 2013.

TORRES-SALINAS, Daniel; ROBINSON-GARCÍA, Nicolás; GORRAIZ, Juan. Filling the citation gap: measuring the multidimensional impact of the academic book at institutional level with PlumX. *Scientometrics*, v. 113, n. 3, p. 1371-1384, 2017.

THELWALL, M. Heart and soul: Sentiment strength detection in the social web with sentistrength (summary book chapter). *Cyberemotions: Collective emotions in cyberspace*. Berlin, Germany: Springer, p. 119-134.

THELWALL, Mike et al. Do Altmetrics work? Twitter and ten other social web services. *PloS one*, v. 8, n. 5, p. e64841, 2013.

THELWALL, Mike; NEVILL, Tamara. Could scientists use Altmetric. com scores to predict longer term citation counts?. *Journal of informetrics*, v. 12, n. 1, p. 237-248, 2018.

VANTI, Nadia. A cientometria revisitada à luz da expansão da ciência, da tecnologia e da inovação. 2011.

VANTI, Nadia; CASADO, Elías Sanz. Posição das universidades públicas espanholas por meio do cálculo do fator de impacto web alternativo. In: *Evaluación de la Calidad de la Investigación y de la Educación Superior: libro de resúmenes XI FECIES*. Asociación Española de Psicología Conductual AEPC, 2015. p. 132-132.

WAHEED, Hajra et al. A bibliometric perspective of learning analytics research landscape. *Behaviour & Information Technology*, v. 37, n. 10-11, p. 941-957, 2018.

WEBER, Rosina. Intelligent jurisprudence research: a new concept. In: *Proceedings of the 7th international conference on Artificial intelligence and law*. 1999. p. 164-172.

WILLIAMS, C. Altmetrics for books: a guide for editors. 2017. Disponível em: <<https://www.Altmetric.com/blog/Altmetrics-for-books-a-guide-for-editors/>>. Acesso em: 20 ago. 2020.

WU, David; ZHANG, Qiping. An Altmetric Study of Artificial Intelligence in Medicine. In: *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2021. p. 196-198.

VYSAKH, C.; BABU, H. Rajendra. An altmetric approach to measure the social media attention of COVID-19 articles. *Library Philosophy and Practice*, v. 4441, 2020.