



Análise de sentimentos sobre as vacinas contra covid-19: um estudo com algoritmo de *machine learning* em postagens no *twitter*

SENTIMENT ANALYSIS ON COVID-19 VACCINES: A STUDY WITH MACHINE LEARNING ALGORITHM ON TWITTER POSTS

Cecilio Merlotti Rodas¹, Suellen Elise Timm Barros², Rodrigo Ananias da Silva Souza³, Silvana Aparecida Borsetti Gregorio Vidotti⁴

¹ Doutor em Ciência da Informação Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp). Docente do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) e do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp).

ORCID: <https://orcid.org/0000-0002-4856-066X>

Email: cecilio.rodas@ifsp.edu.br

² Mestranda em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp)

ORCID: <https://orcid.org/0000-0001-9124-0242>

Email: suellen.timm@unesp.br

³ Mestrando em Ciência da Informação pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp).

ORCID: <https://orcid.org/0000-0002-5778-5158>

Email: rodrigo.ananias@unesp.br

⁴ Doutora em Educação pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp). Docente dos cursos de graduação em Arquivologia e Biblioteconomia e do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp).

ORCID: <https://orcid.org/0000-0002-4216-0374>

Email: silvana.vidotti@unesp.br

Correspondência: Faculdade de Filosofia e Ciências - Av. Hygínio Muzzi Filho, 737, Campus Universitário, 17525-900, Marília, SP.

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Conflito de interesses: os autores declaram que não há conflito de interesses.

Como citar este artigo

Rodas, CM; Barros, SET; Souza RA da Silva; Vidotti SABG. Análise de sentimentos sobre as vacinas contra covid-19: um estudo com algoritmo de *machine learning* em postagens no twitter. Revista de Saúde Digital e Tecnologias Educacionais. [online], volume 7, número especial III. Editor responsável: Luiz Roberto de Oliveira. Fortaleza, fevereiro de 2022, p. 24-44. Disponível em: <http://periodicos.ufc.br/resdite/index>. Acesso em "dia/mês/ano".

Data de recebimento do artigo: 17/08/2021

Data de aprovação do artigo: 28/01/2022

Data de publicação: 14/02/2022

Resumo

Introdução e objetivo: As vacinas contra a Covid-19 foram um dos assuntos mais comentados durante a pandemia na rede social *Twitter*. Assim, este trabalho realizou uma Análise de Sentimentos a partir das postagens dos usuários do *Twitter* nesse contexto. **Método:** Foi utilizada a linguagem de programação *Python 3* aplicando bibliotecas específicas para a coleta, manipulação e visualização dos *tweets* e dos dados. Realizou-se um mapeamento dos sentimentos dos usuários brasileiros no período de 17 de janeiro a 6 de março de 2021, classificando-os em positivos, neutros e

negativos. **Resultados:** Foi percebida uma significativa variação na polaridade dos sentimentos, com a redução dos sentimentos positivos e consequente crescimento dos negativos e dos neutros durante o decorrer do período. Esse tipo de análise adquire relevância quando se considera a grande quantidade de informações e a rapidez com que estas são compartilhadas nas redes. Os resultados confirmaram que os *tweets* referentes à Covid-19 influenciaram o sentimento dos usuários. **Conclusão:** O sentimento que a informação produz nas pessoas pode ter impacto direto ou indireto em sua vida e na sociedade em geral apresentando significativa relevância, pois o sentimento pode ser um fator importante na experiência dos usuários.

Palavras-chave: Análise de sentimento; Covid-19. Vacina. Twitter. Machine Learning.

Abstract

Introduction and objective: Covid-19 vaccines were among the most talked-about topics during the pandemic on the social network Twitter. In this paper a Sentiment Analysis was performed

on Twitter users' posts. **Methods:** The Python 3 programming language was used, applying specific libraries to collect, manipulate and visualize the tweets and the data. A mapping of the sentiments of Brazilian users was performed for the period January 17 to March 6, 2021, classifying them into positive, neutral, and negative. **Results:** A significant variation in sentiment polarity was perceived, with a reduction in positive sentiments and consequent growth in negative and neutral ones during the period. This type of analysis acquires research relevance when one considers a large amount of information and the speed with which it is shared on the networks. The results confirm that the tweets referring to Covid-19 influenced the users' sentiment. **Conclusion:** The feeling that information produces in people can directly or indirectly impact their lives and society in general, presenting significant relevance, since sentiment can be an important factor in the user experience.

Keywords: Sentiment Analysis. Covid-19. Vaccine. Twitter. Machine Learning.

1. Introdução

A Internet acarretou várias mudanças na maneira como consumimos e compartilhamos informações. O grande volume de informações produzidas diariamente por profissionais e usuários influenciou a complexidade de recuperação e categorização das informações nesses ambientes. Desse modo:

Conseguir conhecimento útil nesse ambiente é uma tarefa cada vez mais desafiadora, pois é impossível analisar manualmente esse volume de dados disponíveis na rede, mas o desenvolvimento de novas tecnologias tem possibilitado a análise e visualização de informações a partir de uma grande quantidade de dados¹.

O avanço das Tecnologias de Comunicação e Informação (TIC) e o desenvolvimento dos processos computacionais de Inteligência Artificial (IA), como algoritmos e modelos de Aprendizagem de Máquina, possibilitaram o surgimento de novos mecanismos e técnicas para obtenção e análise de dados digitais. Segundo Liu², é possível realizar a categorização de informações de textos como fatos ou opiniões, e recorrendo à área de mineração de

dados textuais ou Análise de Sentimentos é possível categorizar sentimentos em textos publicados em plataformas digitais como as redes sociais, blogs e fóruns.

A importância de pesquisas que analisem os conteúdos produzidos nessas plataformas digitais se deve ao fato de que a popularização do acesso à Internet e a criação de ferramentas de publicação melhoraram a usabilidade dos meios que possibilitaram os usuários a também produzirem conteúdos informacionais nesses ambientes. Assim, de passivo, o usuário assumiu um papel mais crítico e de produtor de conteúdo. Entretanto, ressalta-se que, com o excesso de informações à disposição, o usuário muitas vezes opta pelo recebimento e compartilhamento de informação de fácil leitura e assimilação. Ou seja, um tipo conteúdo que frequentemente é compartilhado nas redes sociais. Considerando que os usuários podem se tornar influenciadores digitais, percebe-se o perigoso alcance que suas postagens e opiniões ali expressas podem chegar a ter.

Nesse contexto, a rede social *Twitter*¹ passou a ser considerada um poderoso canal de rápida difusão de opinião e informação no qual usuários conseguem ter contato direto com empresas e influenciadores sem precisar de mediadores. Recuero e Zago³ descrevem o *Twitter* como uma plataforma de difusão de informações e conhecimento que proporciona a publicação de mensagens (*tweets*) de até 280 caracteres para uma rede de seguidores, bem como a troca de mensagens entre usuários de forma pública (*replies*) e privada (*messages*).

Além de ser uma fonte de informação, o *Twitter* passou a ser uma relevante ferramenta de pautas para jornalistas e de contato para empresas. A rede social se tornou um importante espaço para o empoderamento do consumidor, já que os usuários não hesitam em compartilhar opiniões sobre fatos e produtos. Murthy⁴ aponta que é por esse motivo que empresas se empenham em estudar os sentimentos dos consumidores no *Twitter* a partir de, por exemplo, *hashtags*² e palavras-chave, já que esses podem fornecer *insights* relevantes.

Um estudo realizado por Wu *et al.*⁵ destaca que os usuários com o maior número de seguidores não são instituições ou empresas de comunicação, mas pessoas que se tornaram celebridades e que se comunicam diretamente com os seus seguidores. Para os pesquisadores, o *Twitter* representa todo o "[...] espectro das comunicações, desde as

¹ www.twitter.com

² *Hashtags* são compostas pelo símbolo de cerquilha (#acompanhado de tópicos ou palavras-chave associados a uma informação ou discussão. O uso facilita que publicações com a mesma *hashtag* sejam identificadas com mais facilidade nas redes sociais como, por exemplo, o *Twitter*.

peçoais e privadas até as 'masspeçoais', passando pelos meios de comunicação de massa tradicionais" (WU *et al.*, 2011, p. 706, tradução nossa)⁵.

A pandemia provocada pelo coronavírus SARS-CoV-2, o causador da Covid-19 e que segundo a Organização Mundial da Saúde infectou 121.209.510 pessoas em todo o mundo (dados de 19 de março de 2021 - 11h30 CET), tornou-se pauta recorrente dos mais prestigiados meios de comunicação e um dos assuntos mais comentados nas redes sociais, como o *Twitter*. Isso porque, segundo Bracons⁶, “[...] a situação de pandemia já tem um impacto social, econômico e político, e mais concretamente ao nível da saúde pública”. O imunologista do Instituto Oswaldo Cruz da Fiocruz, Wilson Savino, concorda ao destacar que é importante entender a fisiopatologia da Covid-19, bem como o comprometimento da saúde mental dos pacientes e cidadãos em distanciamento social⁷. Para ele, se a questão emocional não for priorizada, poderemos ter uma futura 'pandemia' de saúde mental.

Assim, esta pesquisa, por meio de técnicas e ferramentas de recuperação da informação, busca realizar uma Análise de Sentimentos sobre o tema referente à vacina da Covid-19 no *Twitter*, tendo em vista a popularidade da plataforma. Nela, os usuários podem opinar sobre diversos assuntos e expressar sentimentos utilizando textos, *emoji's*, vídeos e imagens.

Para alcançar os objetivos previstos nesta pesquisa foi necessário traçar os seguintes objetivos específicos: (1) Apresentar o papel das redes sociais como fontes de informação e os desafios da rede social *Twitter* na pandemia da Covid-19, especificamente relacionados à temática de vacinação; (2) Apresentar o avanço de estudos que envolvem Análise de Sentimentos e *Machine Learning* em redes sociais, especificamente na rede social *Twitter*; (3) Realizar a coleta de *tweets* e fazer a aplicação do algoritmo para a Análise de Sentimentos dos *tweets*; (4) Gerar gráficos para ilustrar a análise realizada por meio do algoritmo aplicado, com a finalidade de facilitar a interpretação dos resultados.

Foi feita a coleta dos dados para categorização e uso na pesquisa durante o período de 17 de janeiro a 6 de março de 2021, momento que corresponde ao início da vacinação contra a Covid-19 no Brasil. O algoritmo para coleta dos *tweets* foi adaptado de forma com que os *tweets* fossem concentrados em publicações realizadas especificamente no território brasileiro, para que o estudo fosse realizado apenas nos sentimentos de usuários residentes no Brasil.

D'Ancona⁸ afirma que os algoritmos em suas formas atuais “são indiferentes à verdade”. Então, considerando a enorme quantidade de conteúdo informacional que é gerada a todo instante nas mais diversas plataformas digitais, como o *Twitter*, pode-se

inferir o impacto que as redes sociais permitem alcançar sobre as pessoas e a sociedade em geral. Afinal essas informações podem ser consideradas insumo precioso para gerar ainda mais informação para outras máquinas, algoritmos, pessoas e plataformas, num ciclo sem fim.

E por que avaliar o sentimento que essas plataformas geram nas pessoas pode ser um importante tema de pesquisa? Justamente por considerar esse ciclo infindável e a facilidade com que as informações são transmitidas e retransmitidas por meio das tecnologias digitais. Essas informações compartilhadas de forma aleatória ou muito bem-organizadas podem criar teorias conspiratórias de forma virulenta, capazes inclusive de serem incontestadas, afinal: “Sua força popular depende não da evidência, mas do sentimento; a essência da cultura da pós-verdade”⁸.

1.1 O excesso de informação e as redes sociais

A grande quantidade de informação diariamente produzida, compartilhada e divulgada no ambiente digital aponta a importância de estudos que se dediquem a compreender o comportamento informacional dos usuários em tais interfaces. Não é recente a preocupação com o excesso de conteúdo disponível e a influência deste sobre o comportamento de busca informacional. Pesquisadores como Jacoby⁹, Schneider¹⁰ e Malhotra¹¹, já discutiam o que chamavam de *information overload*, em português sobrecarga informacional, e as consequências para o usuário.

Se no passado o desafio dos comunicadores era divulgar o maior número de informações no menor tempo possível, há alguns anos a preocupação passou a ser o comportamento de busca informacional, visto que o usuário precisa filtrar o que julga relevante no grande volume de conteúdo. Como lembra o sociólogo Bauman¹², o excesso de informação propiciou a pouca reflexão nos deixando “inundados por informação e famintos por sabedoria”.

Ao mesmo tempo que a abundância de informações pode ser útil para a construção de saberes, o excesso pode representar um problema. Diante da grande quantidade de informação disponível, a sensação é que nunca será possível atingir o conhecimento essencial e necessário. Wurman¹³, criador do termo “arquitetura da informação”, aponta a sua preocupação com a constante ansiedade dos indivíduos por informação, já que o usuário acredita que seu poder está relacionado à quantidade informacional a que ele tem acesso. E de fato, pesquisas relacionam inclusive distúrbios psicológicos decorrentes da ânsia informacional como, por exemplo, ansiedade, depressão, solidão e o medo de ficar de

fora (em inglês conhecido pela sigla FOMO - *Fear Of Missing Out*), que caracteriza uma constante necessidade de acompanhar a vida de outras pessoas¹⁴.

Além das questões psicológicas relacionadas à sobrecarga informacional, o usuário lida com outros desafios como o de filtrar o que julga confiável, relevante e verídico. Isso porque “[...] o exagero na quantidade de informação começa a nublar as diferenças marcantes entre dados e informação, entre fatos e conhecimentos fazendo com que nossos canais de percepção entrem em curto-circuito”¹³. Redes sociais, agregadores de conteúdo e sistemas de busca, então, se propõe a dar suporte na busca informacional para que o usuário encontre o conteúdo que julga relevante, no menor tempo possível.

1.2 Redes sociais como fontes de informação

Hassan¹⁵ afirma que cada faceta da vida é mediada ou influenciada por tecnologia. As redes sociais são um exemplo disso, já que se tornaram interfaces relevantes não apenas para relacionamentos, mas para a divulgação e compartilhamento de informações. Porém, a grande quantidade informacional trouxe consequências como a dificuldade de separar fatos de opiniões. Isso porque essas redes produziram uma fragilidade que leva os usuários a constantemente competirem por informação, mesmo que a quantidade de conteúdo disponível diariamente seja mil vezes maior do que a capacidade cerebral humana assimilar¹².

Nesses ambientes, os usuários são incentivados a constantemente produzirem e divulgarem conteúdos para garantirem a sua visibilidade. Com isso, as redes sociais criaram uma explosão de conteúdo e dados gerados por usuários, em um ambiente virtual que já era saturado de conteúdo¹⁵. Assim, o cidadão não apenas consome, mas constantemente produz e dissemina conteúdos julgados por ele como relevantes.

Com o conteúdo distribuído e divulgado por usuários, é difícil precisar informações de validade e veracidade. Outro ponto vastamente estudado por Pariser¹⁶ é que agregadores de conteúdo, como redes sociais e sistemas de busca, filtram e personalizam o conteúdo, limitando o acesso dos usuários às informações. O autor ressalta que esse filtro disponibiliza conteúdo sob medida baseado nos interesses de cada usuário criando realidades individualizadas. Nesse contexto, é enevoada a separação entre o real e o falso, a opinião e o fato, a informação relevante para construir a cidadania e a que serve apenas para entretenimento.

1.3 Twitter e os desafios na pandemia da covid-19

Junto com o rápido alastramento do vírus Covid-19 (SARS-CoV-2) iniciado no final de 2019, percebeu-se o aumento do número de buscas relacionadas ao tema, inclusive nas redes sociais. Segundo o *Insider Report* do *Google*³, um dos impactos provocados pela pandemia é que o conteúdo informativo se tornou mais importante do que distrações como consumismo e entretenimento. E de fato outra pesquisa divulgada pela *Comscore*⁴ aponta a intensificação em 42,6% (725 milhões para 1,03 bilhão de acessos) do consumo de conteúdo noticioso com a busca de informações sobre a Covid-19.

Diante da quantidade de informações compartilhadas sobre a Covid-19 também nas redes sociais, o diretor-geral da Organização Mundial da Saúde (OMS), Tedros Ghebreyesus⁵, destaca que “Não estamos lutando só contra uma epidemia. Combatemos também uma infodemia”. Kuzmanovic (*apud* PIERRO, 2020)¹⁸, gerente de mídias sociais da OMS, concorda e completa que “[...] informações imprecisas sobre a Covid-19, a doença respiratória causada pelo coronavírus, estão se espalhando mais rapidamente do que o próprio vírus” e que, por isso, a organização busca trabalhar diretamente com empresas como *Facebook*⁶, *Google*⁷, *Pinterest*⁸ e *Twitter*.

A preocupação da OMS com as redes sociais é justificada em estudos que destacam a relevância das comunidades virtuais para a obtenção de informação e formação da opinião dos cidadãos. Desde seu lançamento, a rede social *Twitter* procura ser uma grande plataforma que garante a disseminação de fatos e informações, porém opiniões e desinformações também podem ser compartilhadas. Para se ter uma ideia do impacto que esses ambientes podem ter sobre a opinião de seus leitores, Vosoughi, Roy e Aral¹⁷ alertam que as *fake news* têm 70% mais chances de serem retuitadas do que as notícias reais.

Pesquisadores da área médica Kouzy *et al.*¹⁹ avaliaram 673 *tweets* com informações relacionadas à Covid-19. Os estudiosos constataram que contas individuais apresentaram 33,8% de informações não verídicas e que desinformações médicas sobre a Covid-19 têm sido propagadas em quantidade alarmante nas redes sociais. Alguns dos casos de desinformação sobre a pandemia viralizaram, como a tese publicada no *Twitter* de um

³ Disponível em: https://www.mmaglobal.com/files/328._brazil_1_5010688669204349081.pdf. Acesso em: 01 mar. 2021.

⁴ Disponível em: <https://www.comscore.com/por/Insights/Blog/Consumo-de-midia-durante-a-pandemia-de-coronavirus-no-Brasil>. Acesso em: 05 mar. 2021.

⁵ Disponível em: <https://saude.estadao.com.br/noticias/geral,fake-news-de-coronavirus-incluem-arma-biologica-e-uisque-de-prevencao-veja-as-mais-comuns,70003225279>. Acesso em: 10 julho 2020.

⁶ www.facebook.com

⁷ www.google.com

⁸ www.pinterest.com

pseudocientista atribuindo relação entre as redes móveis de 5G e a propagação do coronavírus. Essa tese foi compartilhada por celebridades e, a partir do boato, antenas nas cidades de Birmingham, Liverpool, Melling (Meyerside) e Belfast foram voluntariamente incendiadas. A rede social *Twitter* rapidamente reagiu com a medida de apagar qualquer postagem que relacionasse 5G à Covid-19⁹.

Diante das desinformações, redes sociais como o *YouTube*, *Facebook* e o próprio *Twitter* passaram a adotar medidas para evitar a proliferação de conteúdo duvidoso. O *Twitter* passou a avaliar e adotar medidas para informações enganosas, conteúdo com veracidade duvidosa ou não confirmada. Entre as atitudes adotadas estão a inclusão de links para informações confiáveis, aviso de conteúdo que pode conter informações questionadas por autoridades, ou até mesmo a exclusão do *post*.

1.3 O Twitter e a Vacinação

Enquanto profissionais da área de saúde destacam a importância da vacinação para conter a pandemia da Covid-19, as desinformações sobre o assunto continuaram a ser espalhadas em redes sociais. Uma pesquisa realizada pelo Instituto Brasileiro de Opinião Pública e Estatística (IBOPE) apontou que um quarto da população brasileira pode decidir não ser vacinada contra a Covid-19¹⁰. No estudo, cerca de 34% dos motivos apontados para a decisão estavam relacionados à desinformação.

Porém, essa não é a primeira vez que autoridades de saúde precisam lutar contra a desinformação. Doenças como o sarampo, poliomielite e rubéola em 2019 preocuparam especialistas devido ao significativo risco de voltarem a contaminar a população. O motivo era uma forte campanha anti-vacina baseada em desinformações espalhadas nas redes sociais. A pesquisa “As Fake News estão nos deixando doentes?”¹¹, realizada pela Sociedade Brasileira de Imunizações (SBIm) e pela ONG Avaaz, constatou que aproximadamente 67% dos brasileiros acreditam em pelo menos uma informação inverídica sobre a vacinação. Para Sacramento (2018)²⁰, isso pode ser explicado pelo fato de, na contemporaneidade, o regime de confiança nas instituições ser substituído por um baseado em dogmas e experiência pessoal:

A experiência tem legitimado o conhecimento sobre a verdade. É intensamente valorizado um outro tipo de autoridade: a autoridade

⁹ Disponível em: <https://www.bbc.com/news/uk-england-52164358>. Acesso em: 08 maio 2020.

¹⁰ Disponível em: <https://saude.estadao.com.br/noticias/geral,um-em-cada-quatro-brasileiros-resiste-a-ideia-de-tomar-vacina-contr-a-covid-19,70003427273>. Acesso em: 20 fev. 2021.

¹¹ Disponível em: <https://sbim.org.br/acoes/as-fake-news-estao-nos-deixando-doentes>. Acesso em: 20 fev. 2021.

experiencial. Ela enfatiza o caráter testemunhal: eu vivi, eu sei. Produz na primeira pessoa (naquele que viu, viveu, sentiu) da experiência e da narrativa de um determinado acontecimento a origem da verdade ou um documento de que o narrado realmente existiu²⁰.

Segundo Sacramento²⁰, as redes sociais são um espaço fechado de confiança e segurança onde usuários confiam nos amigos e familiares que compartilharam as informações. Assim, é provável que as pessoas prefiram confiar na experiência e testemunho de quem conhecem, do que nas instituições e na própria ciência.

Para conter o compartilhamento de informações enganosas e apresentar fontes confiáveis, em 16 de dezembro de 2020 o *Twitter* divulgou a abordagem que seria adotada em relação ao compartilhamento de informações sobre a vacina¹². A própria rede social justifica que, enquanto o mundo iniciava o processo de vacinação, os usuários procuravam o *Twitter* para conversar e procurar informações.

Já em 1 de março de 2021, o *Twitter* lançou uma atualização que, então, incluiria avisos e alertas em *tweets* que apresentassem informações possivelmente enganosas. As medidas são apresentadas no idioma de exibição da conta do usuário e são baseadas em três categorias, conforme a Figura 1 a seguir.

Figura 1 - Medidas adotadas pelo Twitter com base em três amplas categorias

Informações enganosas	Aviso	Remoção
Afirmções questionáveis	Aviso	Alerta
Afirmções não confirmadas	Sem medida	Sem medida*
	Moderada	Grave
	Propensão a dano	

Fonte: elaborado pelos autores com base no *Twitter*.

Avisos são usados para direcionar os usuários para uma página de curadoria da rede social ou uma fonte de credibilidade, já os alertas informam a presença no *tweet* de

¹² Disponível em: https://blog.twitter.com/pt_br/topics/company/2020/covid-19-nossa-abordagem-para-informacoes-enganosas-sobre-vacinas.html. Acesso em: 20 fev. 2021.

informações divergentes com a de especialistas. Além disso, continua a prática de remoção de conteúdo enganoso iniciada em dezembro de 2020, já que desde o lançamento da abordagem cerca de 2.400 contas foram suspensas e 11,5 milhões de contas foram desativadas¹³. O *Twitter* em março também passou a adotar um sistema de progressão de penalidades que inclui tempo de bloqueio da conta para casos de incidência de desinformação sobre a vacina de Covid-19. A rede social usa revisões automatizadas por máquina e humanas para avaliar o conteúdo.

1.4 *Machine learning* e análise de sentimentos

Com o grande volume de dados disponíveis on-line, pesquisas com *Machine Learning*, ou Aprendizado de Máquina em português, passaram a ser essenciais para identificar padrões e conectar informações, de forma que o algoritmo consiga aprender e se aperfeiçoar através das informações coletadas. De acordo com Mitchell (1997)²¹, *Machine Learning* trata da questão de desenvolvimento de algoritmos que automaticamente melhorem sua performance com base na sua própria experiência, ou seja, os algoritmos acabam por ter a possibilidade de melhorar sua assertividade de forma automática.

Os algoritmos de *Machine Learning* (ML) podem aperfeiçoar os processos de recuperação da informação permitindo avanços significativos na automatização desse tipo de tarefa. "O Aprendizado de Máquina é a área de maior destaque em análise de *Big Data*, pois permite a criação de modelos computacionais para processamento automático ou semiautomático"¹. Nesse contexto, os algoritmos de ML têm sido utilizados para colaborar no processamento e identificação de sentimentos em textos digitais postados por usuários em redes sociais, por exemplo.

Segundo Magalhães, Carvalho e Moraes¹, a Análise de Sentimento é uma área da Mineração de Textos tendo como um dos seus objetivos a classificação de textos com base na emoção contida em determinado documento. Um dos principais estágios no processo de Análise de Sentimento é a classificação²². Dessa forma, é possível utilizar técnicas de *Machine Learning* para realizar a classificação dos sentimentos de usuários através de algoritmos e bibliotecas¹⁴ específicas.

A Análise de Sentimentos é um campo dentro do Processamento de Linguagem Natural (PLN) que é constituída de sistemas que distinguem e extraem opiniões dentro de

¹³ Disponível em: https://blog.twitter.com/pt_br/topics/company/2020/atualizacao-no-trabalho-desinformacao-sobre-vacina-covid-19.html. Acesso em: 10 mar. 2021.

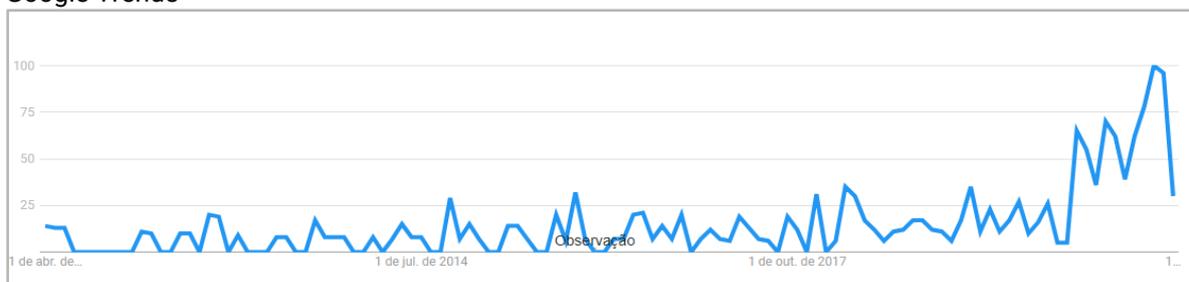
¹⁴ Na ciência da computação, biblioteca é uma coleção de subprogramas utilizados no desenvolvimento de *software*.

textos. Entre uma de suas aplicações, está o algoritmo que permite criar recomendações baseadas nos dados gerados pelos usuários de *stream* de vídeos e músicas a partir do histórico de visualizações registrado em plataformas como o *YouTube*¹⁵. Segundo Tyagi e Tripathi²³, na atualidade as plataformas de mídia social, como *Twitter*, *Facebook* e *YouTube*, são uma grande fonte de informação conhecida como dados sociais.

A pandemia causada pela COVID-19 tem sido um dos assuntos mais discutidos e disseminados em todo o mundo. Desse modo, torna-se pertinente realizar estudos que classificam o estado emocional dos usuários em plataformas digitais, por esses ambientes possibilitarem a expressão de opiniões através de texto. De acordo com o estudo realizado por Esuli²⁴, a Análise de Sentimento é uma área da linguística computacional cuja intenção é descobrir a opinião que um documento expressa sobre tal tópico ou produto ao qual se refere. Para Liu², a área de Análise de Sentimentos é de interesse tanto para aplicações destinadas para o uso de organizações, quanto para uso pessoal.

De acordo com Foschiera²⁵, “[...] o tratamento computacional da opinião vem representando um desafio na atualidade, quando temos à disposição um grande número de informações relevantes que nos permite entender o ponto de vista dos outros”. Por isso, Magalhães, Carvalho e Moraes¹ alegam que o campo da Análise de Sentimentos possui grande potencial em pesquisas científicas, de mercado e para o campo empresarial devido à variedade de informações que esta pode revelar. Saber a opinião das pessoas sobre um determinado assunto influencia positivamente na tomada de decisões, além de permitir a detecção de tendências ou comportamentos sociais em geral. Em uma busca na ferramenta *Google Trends* utilizando como palavra-chave o termo “Análise de Sentimentos”, é possível observar o aumento do interesse por esse assunto nos últimos 10 anos, conforme ilustrado na Figura 2.

Figura 2 - Progressão de pesquisas por Análise de Sentimentos nos últimos 10 anos - Extraído do Google Trends



Fonte: elaborado pelos autores (dados acessados em 18 mar. 2021).

¹⁵ Disponível em: www.youtube.com. Acesso em: 21 jul. 2020.

Para a representação da linha de construção metodológica foram abordados a priori estudos que trazem um embasamento justificável para a escolha da rede social *Twitter* como Serviço de Redes Sociais *On-line* (SRSO) e fornecedor de dados. Em sequência são apresentados estudos que abordam a Análise de Sentimentos e que fazem uso ou não de técnicas de *Machine Learning* para justificar a escolha do método *Naive Bayes Multinomial* (NBM).

De acordo com estudo realizado por Recuero e Zago³, mais de 50% dos *tweets* são de conteúdo informativo e 25,3% são de opinião. Tal estudo também aponta que o tipo de capital social que é buscado pelo usuário é do tipo conectivo, onde a prioridade é a conexão com diversos usuários e não o aprofundamento de comunicações pessoais. Tendo em vista o cenário atual onde as informações são compartilhadas em sua maior parte nas redes sociais onde a disseminação de desinformação é recorrente, fazer uso do *Twitter* como SRSO para analisar o impacto da temática de vacinação durante a pandemia da Covid-19 é ter um ambiente de investigação para a coleta de dados de pesquisa.

Considerando que o *Twitter* se enquadra como um SRSO e com o objetivo de realizar Análise de Sentimentos, Moraes e Magalhães²⁶ elaboraram uma metodologia para realizar a extração de opiniões a partir de *hashtags* e para possibilitar que um *tweet* seja buscado e recuperado. O trabalho de Miranda Filho, Carvalho e Pappa²⁷ aponta que o estudo utilizando o modelo classificador de dados *Naive Bayes Multinomial* (NBM) para aprendizagem obteve resultados satisfatórios e apresentou melhor eficiência, atingindo acurácias próximas de 90% e 80%. Assim, considerando os bons resultados apresentados, esse modelo foi selecionado para a realização dos experimentos e avaliações propostas neste trabalho.

1.5 Algoritmo Naive Bayes Multinomial

De acordo com Sombra²⁸, “[...] o algoritmo *Naive Bayes* é considerado um dos algoritmos mais simples, porém completos, para classificação de dados. Seu conceito baseia-se na teoria das probabilidades que, segundo Rocha, Cortez e Neves²⁹, é considerada uma das principais fontes técnicas para classificação de bases de dados”. Segundo Baeza-Yates e Ribeiro-Neto³⁰, podemos abordar os algoritmos Bayesianos de diferentes formas, por exemplo modelos binários e multinomial. O classificador *Naive Bayes* pressupõe que os termos de uma instância são condicionalmente independentes entre si, sem exercer influência um sobre o outro, apenas têm influência sobre as classes a que se remetem³¹.

Ainda, segundo Schmitt³¹, no modelo multinomial cada documento é caracterizado por um vetor de números do tipo inteiro que indica a quantidade de vezes que cada termo ocorre no documento. Enquanto no modelo binário cada documento é caracterizado por um vetor binário e a existência ou ausência de um termo são representadas pelos valores 1 ou 0, respectivamente. Nessa técnica, a proximidade entre uma palavra e outra é ignorada, pois considera que as palavras não possuem relacionamento entre si.

Com base no trabalho desenvolvido por Magalhães, Carvalho e Moraes¹, utilizamos o modelo multinomial para o desenvolvimento da pesquisa, portanto consideramos a ocorrência ou não da palavra em um texto e a quantidade de ocorrências de cada palavra na base de treinamento de aprendizagem. "Ao rotular uma sentença, todas as palavras recebem uma ocorrência da classe correspondente ao rótulo. Ao final da leitura da base de treino, o algoritmo define a classe de cada palavra e 'aprende' a classificar novos períodos não rotulados"¹.

A base de dados disponível para treino deve conter uma quantidade significativa de dados e atributos das classes 'positivas', 'negativas', e 'neutro' para que o algoritmo consiga aprender e fazer a predição de uma coleção não rotulada. Como exemplo de *tweets* com esses atributos e que foram usados na presente pesquisa durante o período de investigação e coleta dos *tweets* a respeito da vacinação em combate a Covid-19 no início de 2021, temos: positivo - "Instituto Butantan comprova eficácia da *CoronaVac* contra as novas variantes da Covid-19 <https://t.co/seGZolZMBW> via... <https://t.co/Y0XmQmgAmU> "; negativo - "Covid-19: Pfizer diz que governo rejeitou ofertas de 70 milhões de doses de vacina" #coronavirus #covid19 #vacina <https://t.co/g958j00Lsq>"; e neutro - "ENTENDA - Como está a situação das vacinas da covid-19 no Brasil? LEIA E SAIBA MAIS <https://t.co/ZiudluKj0n...> <https://t.co/XzPbTFgb9F>".

2. Metodologia

Para a construção do algoritmo nesta pesquisa foram utilizadas: a linguagem de programação *Python 3*¹⁶, a qual oferece bibliotecas pré-estabelecidas que auxiliam na compreensão e no desenvolvimento do código; e as documentações oficiais do *Twitter* sobre coleta e análise de dados. Em especial, foram empregadas as bibliotecas: *Tweepy* para a comunicação com a *Application Programming Interface (API)*¹⁷ do *Twitter* e a coleta

¹⁶ Disponível em: <https://www.python.org/download/releases/3.0/>. Acesso em: 18 mar. 2021.

¹⁷ Uma API permite a integração entre dois sistemas ou mais, em que um deles fornece informações e serviços que podem ser utilizados por outro sistema.

de *tweets*, *Pandas* e *Machine Learning Scikit-learn* para manipulação, leitura e visualização de dados; e a *Matplotlib* para a elaboração de gráficos destinados à interpretação dos resultados obtidos.

O algoritmo desenvolvido foi baseado no código desenvolvido por Magalhães, Carvalho e Moraes¹ e ajustado de forma a atender as especificidades desta pesquisa. Para a coleta de dados, o SRSO escolhido foi a rede social *Twitter*. Foi utilizada a API disponibilizada pela plataforma *Twitter* para realizar a coleta de publicações feitas na plataforma, os *tweets* que apresentavam palavras-chaves inseridas pelo usuário foram coletados e armazenados em arquivo com a extensão 'TXT'.

Os dados coletados necessitam estar em formato de lista para que os mesmos possam ser transferidos para o modelo de predição usando a linguagem de programação *Python 3*. Para transformar o arquivo em lista, foram utilizadas técnicas desenvolvidas por Magalhães, Moraes e Carvalho¹, segundo as quais:

[...] inicialmente o arquivo foi armazenado em uma *string* para, em seguida, a sequência de caracteres ser transformada em uma lista de *tweets*. Primeiramente, os dados da *string* passaram por um processo de planificação, onde foram retiradas todas as quebras de linha, pois elas são danosas para o modelo de predição. Com a *string* planificada, a lista foi criada e o caractere “\n” foi usado para finalizar cada *tweet*¹.

As bases de testes desenvolvidas para esta pesquisa foram formadas por 70 mil *tweets*, que foram divididos semanalmente em grupos de 10 mil *tweets*, recuperados em 17 de janeiro de 2021 no período inicial da vacinação contra a Covid-19. Para tanto, foram usadas as seguintes palavras-chave (ou *hashtags*) para a busca e coleta: '#Vacinação, #Coronavac, #Vacina'. Após esse procedimento, foram efetuadas as análises de polaridade dos *tweets* recuperados com a finalidade da realização da análise dos resultados obtidos.¹⁸

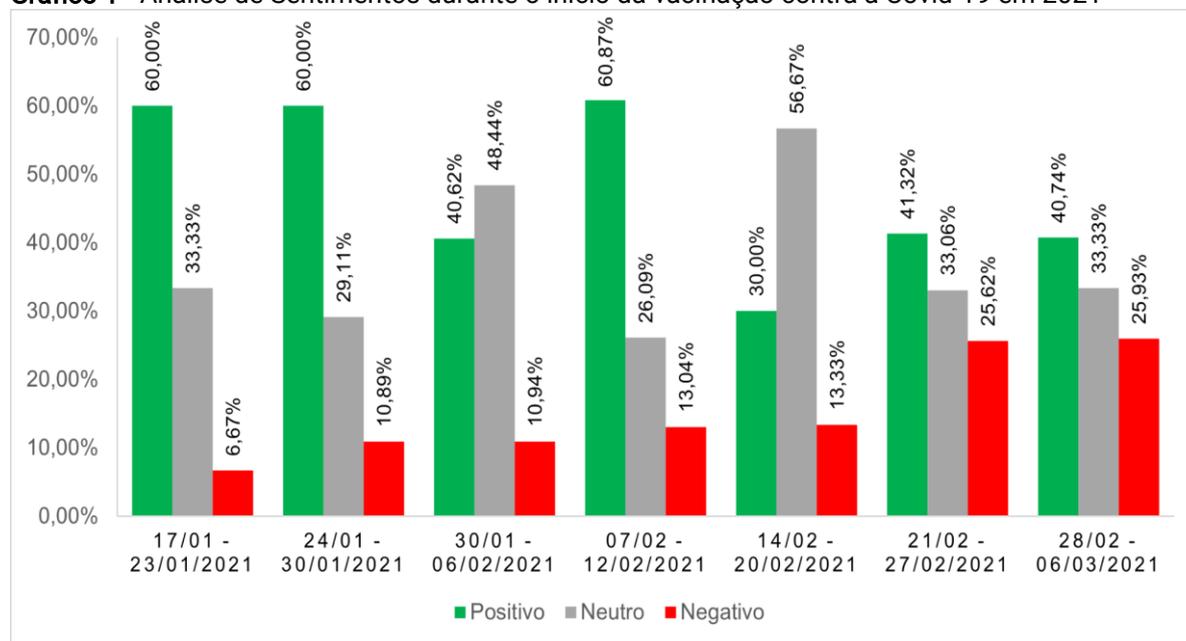
3. Análise dos resultados

Em 17 de janeiro de 2021 iniciou a campanha de vacinação contra a Covid-19 no Brasil utilizando a vacina denominada de *CoronaVac*, desenvolvida em uma parceria da empresa *Sinovac Biotech* e o Instituto Butantan. O Gráfico 1, a seguir, destaca o período das sete primeiras semanas do início da campanha de vacinação contra a Covid-19 no País (17 de janeiro a 6 de março de 2021). Foi notada uma significativa variação no sentimento dos

¹⁸ Os dados da pesquisa serão publicizados no Repositório Institucional da Universidade de origem dos autores.

usuários, especialmente em quatro momentos específicos que serão especificados a seguir.

Gráfico 1 - Análise de Sentimentos durante o início da vacinação contra a Covid-19 em 2021



Fonte: elaborado pelos autores.

Na primeira e segunda semanas os *tweets* considerados positivos se mantiveram em 60%, porém os neutros e negativos apresentaram uma pequena variação comparando a primeira semana com a segunda - sendo um declínio de 4,22% nos neutros (de 33,33 para 29,11%) e um aumento de 4,22% nos negativos (de 6,67 para 10,89%). Analisando o conteúdo dos *tweets* recuperados, na primeira semana foi percebido um alto número de publicações referentes à efetivação da vacinação. Já na segunda semana verificou-se o crescimento de *tweets* neutros e negativos com comentários antivacina, além de outros que mencionam questões gerais da pandemia como aumento do número de pessoas infectadas e da quantidade de óbitos registrados em decorrência da Covid-19.

Foram percebidas nas publicações recuperadas, não apenas nas duas primeiras semanas, mas durante toda a pesquisa, que alguns *tweets* não fazem referência direta à campanha de vacinação. Porém, por utilizarem as *hashtags* de vacinação foram recuperados pelo algoritmo que os classificou como publicações neutras ou negativas. É importante lembrar que a base de treino utilizada para a análise dos dados baseou-se em *tweets* recuperados que fazem associação com a campanha de vacinação. Assim, é possível concluir que alguns acontecimentos como o aumento de interações em

decorrência da Covid-19, a manifestação de uma nova variante do vírus e o aumento de casos durante o período de recuperação dos dados podem influenciar na classificação dos *tweets* obtidos nas semanas analisadas nesta pesquisa.

Na terceira semana (30 de janeiro a 6 de fevereiro de 2021) foi a primeira vez que os *tweets* neutros superaram o número dos positivos. Situação semelhante foi percebida na quinta semana de análise (14 a 20 de fevereiro de 2021). Ambas as semanas foram as únicas do período deste estudo em que os sentimentos neutros representaram a maioria, portanto apresentaram um aumento considerável - com 48,44% dos *tweets* na terceira semana e 55,67% na quinta semana.

Na quarta semana de vacinação os *tweets* classificados com a polaridade positiva tiveram um considerável aumento se comparado com a semana anterior, chegando a um total de 60.87%. Porém, na semana anterior (30 de janeiro a 6 de fevereiro de 2021) e na semana seguinte (14 a 20 de fevereiro de 2021), como mencionado anteriormente, os sentimentos neutros representaram a maioria dos *tweets* analisados. Portanto, possivelmente a quantidade de cidadãos vacinados contra a Covid-19 até a quarta semana atingindo um total de 4.9 milhões, um valor que representa 2.32% da população, pode ter afetado a positividade no sentimento dos usuários.

Nas duas últimas semanas analisadas nesta pesquisa (21 a 27 de fevereiro e 28 de fevereiro a 6 de março de 2021) foi percebida uma proximidade entre os valores das publicações classificadas com as polaridades positivas, neutras e negativas - as positivas de 41.32% e 40.74%; as neutras de 33.06% e 33.33%; e as negativas de 25.62% e 25.93%. Importante destacar que os sentimentos negativos praticamente duplicaram em comparação com as cinco semanas anteriores analisadas nesta pesquisa. Inclusive, nessas duas últimas semanas apresentadas no Gráfico 1 o número de contaminados e a quantidade de óbitos aumentaram significativamente.

Destaca-se ainda que a plataforma *Twitter* teve uma atualização nas normas e diretrizes a respeito da desinformação durante a vacinação entre as semanas epidemiológicas citadas anteriormente. Essa atualização procurou marcar *tweets* e buscou publicações que continham informações falsas a respeito da vacinação. Nas publicações identificadas são inseridos *links* que redirecionam os usuários para órgãos competentes, como, por exemplo, a Organização Pan-Americana de Saúde para fornecer informações de credibilidade aos usuários a respeito da imunização contra a Covid-19. Na Figura 3 é possível observar um exemplo de marcação.

Figura 3 - Exemplo de marcação feita pela plataforma *Twitter* vinculada a busca de *tweets* sobre a vacinação contra a Covid-19.



Fonte: elaborado pelos autores (baseado no recorte do *Twitter*).

Foi realizada uma busca e uma análise nos *tweets* recuperados por intermédio do algoritmo desenvolvido para conferir a existência de marcações atribuídas às publicações. Importante ressaltar que a abordagem do *Twitter*, que passou a ser aplicada em 21 de dezembro de 2020, consiste na remoção de *tweets* com conteúdo enganoso sobre a vacina, portanto os *tweets* deletados da base de dados da plataforma impedem a sua recuperação. Os avisos e alertas foram incluídos pelo *Twitter* apenas a partir de março de 2021, o que coincide apenas com a última semana desta pesquisa. Entretanto, não foi identificado no conjunto de publicações recuperadas nenhum *tweet* que continha a marcação. O *Twitter* usa tanto monitoramento por máquina e por humanos, além de destacarem que estão constantemente aperfeiçoando o sistema. Inclusive, mesmo que em nenhuma publicação recuperada para esta pesquisa foram identificados os avisos e alertas, uma pesquisa realizada nove dias após a coleta para este estudo identificou os avisos em algumas publicações e na busca por *tweets* relacionados à vacina.

4. Considerações finais

O trabalho mostra que o assunto referente à pandemia do SARS-CoV-2 produziu uma grande quantidade de informações em diversas plataformas disponíveis no ambiente *Web*. Verificou-se que essas informações foram capazes de impactar no comportamento informacional e no sentimento de usuários, muitos dos quais estavam constantemente em busca de compreender e absorver mais informações relevantes a respeito da campanha de vacinação, por exemplo.

Este trabalho apresentou resultados sobre uma Análise de Sentimentos com dados que foram obtidos durante o início da vacinação no Brasil, ou seja, entre 17 de janeiro a 6 de

março de 2021 a partir de postagens dos usuários na plataforma *Twitter*. No início desse período, grande parte dos *tweets* recuperados eram classificados com a polaridade positiva. Porém com o avanço da vacinação e algumas atualizações nas diretrizes da plataforma *Twitter*, as polaridades neutras e negativas apresentaram uma variação considerável de forma que foi possível notar o impacto das atualizações do conteúdo sobre o sentimento dos usuários.

Ao se levar em consideração as *hashtags* utilizadas para a coleta de dados, foi possível identificar que algumas publicações fizeram uso de palavras-chave que tiveram uma ampliação nas redes sociais durante o período de vacinação para garantir mais destaque ao *tweet* na busca na plataforma e de *softwares* de terceiros. Notou-se que alguns dos *tweets* analisados não estavam diretamente associados à campanha de vacinação contra a Covid-19, porém se relacionavam com outras questões referentes ao tema da pandemia. Por exemplo, algumas dessas publicações registravam o aumento dos óbitos em decorrência da Covid-19 e a manifestação de uma nova variante de origem brasileira que poderia tornar a infecção ainda mais letal.

Durante o período de análise dos *tweets* foi possível concluir que existe uma limitação na atribuição das marcações utilizadas em publicações relacionadas à campanha de vacinação. Ou seja, grande parte dos *tweets* recuperados não tinham a marcação que redireciona os usuários para órgãos oficiais a respeito da vacinação. Porém, é necessário considerar a grande quantidade de *tweets* publicados diariamente sobre o tema, a limitação humana dos revisores do *Twitter* e de processamento da máquina para realizar tal tarefa, além de que os avisos e alertas foram incluídos pelo *Twitter* apenas a partir de março de 2021, o que coincide apenas com a última semana desta pesquisa.

A mineração de texto e a classificação do mesmo para a realização de Análise de Sentimentos em redes sociais traz alguns desafios adicionais se comparado a outros formatos de dados. No caso de informações recuperadas da rede social *Twitter*, existem outros desafios, como o limite de 280 caracteres por postagem, fator que pode dificultar a análise do texto efetuada pelo algoritmo. Outra dificuldade encontrada na Análise de Sentimentos em postagens do *Twitter* é o uso recorrente de *emojis*, os quais podem identificar a percepção e sentimento do usuário sobre determinado assunto.

Finalmente, considerando o sentimento como um fator importante para a experiência do usuário na sociedade atual⁸, capaz inclusive de comprometer a saúde mental das pessoas, ressaltamos a relevância do aperfeiçoamento de técnicas e tecnologias que permitam a análise de conteúdo nas redes sociais considerando a Análise de Sentimentos.

Referências

1. Magalhães, L. H. de, Carvalho, R. R. A., Moraes, I. C., Santos, E. B. D. Análise de sentimentos utilizando o algoritmo naive bayes multinomial. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 20., 2019, Florianópolis. Anais [...]. Florianópolis: Ancib, Universidade Federal de Santa Catarina, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/1390/649>. Acesso em: 10 jan. 2021.
2. Liu, B. Sentiment analysis and subjectivity. *In: INDURKHAYA, N.; DAMERAU, F. Handbook of Natural Language Processing*. 2. ed. [Londres]: Chapman and Hall/CRC, 2010. p. 1-38. Disponível em: <https://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>. Acesso em: 30 maio 2020.
3. Recuero, R.; Zago, G. A Economia do retweet: redes, difusão de informações e capital social no Twitter. *Contracampo*, Niterói, v. 24, n. 1, p. 19-43, jul. 2011.
4. MURTHY, D. Twitter: "social communication in the twitter age". *In: MURTHY, D. Twitter: social communication in the twitter age (digital media and society)*. 2. ed. Medford: Polity Press, Apple Books, 2018. 220 p.
5. Wu, S. *et al.* Who says what to whom on Twitter. *In: WWW '11 - INTERNATIONAL CONFERENCE ON WORLD WIDE WEB*, 20, 2011, New York. Proceedings [...]. New York: Association for Computing Machinery, 2011. p. 705-714. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/1963405.1963504>. Acesso em: 15 fev. 2021.
6. Bracons, H. Percepção dos estudantes finalistas de Serviço Social face ao contexto de emergência Covid-19. *Revista Internacional de Educación para la Justicia Social*, [s. l.], v. 9, n. 3, p. 1-6, maio 2020. Disponível em: <https://revistas.uam.es/riejs/article/view/12124/>. Acesso em: 01 fev. 2021.
7. Menezes, W. Artigo analisa os impactos da Covid-19 na saúde mental. Portal Fiocruz, Manguinhos, [online], jun. 2020. Disponível em: <https://portal.fiocruz.br/noticia/artigo-analisa-os-impactos-da-covid-19-na-saude-mental>. Acesso em: 21 jan. 2021.
8. D'ancona, M. Pós-verdade: a nova guerra contra os fatos em tempos de fake news. Barueri: Faro Editorial, 2018.
9. Jacoby, J. Perspectives on information overload. *Journal of Consumer Research*, Oxford University Press, v. 10, n. 4, p. 432-435, mar. 1984. Disponível em: <https://academic.oup.com/jcr/article-abstract/10/4/432/1822431?redirectedFrom=fulltext>. Acesso em: 05 fev. 2021.
10. Schneider, S. C. Information overload: causes and consequences. *Human Systems Management*, Amsterdam, v. 7, n. 2, p. 143-153, 1987. Disponível em: <https://content.iospress.com/articles/human-systems-management/hsm7-2-07>. Acesso em: 10 fev. 2021.
11. Malhotra, N. K. Reflections on the Information Overload Paradigm in Consumer Decision Making. *Journal of Consumer Research*, [s. l.], v. 10, n. 4, p. 436-440, 1984. Disponível em: https://www.jstor.org/stable/2488913?casa_token=3TyO4qaPqoQAAAAA%3AKDxypxueEhinoOMIbI9rTHkh9gB-IScRSkBOUWTFJEgZi4LR5ik-BKMU01EWtcoPJD90IuvUGmnYA7A70x5Yqcw_cVFgY8MYP8AdkYMT9PEO6QRORdI&seq=1#metadata_info_tab_contents. Acesso em: 15 fev. 2021.
12. Bauman, Z. Avalanche informativa. [Entrevista cedida a] Alberto Dines. *Observatório da Imprensa*, [s. l.], edição 872, 2015. Disponível em:

- <http://observatoriodaimprensa.com.br/programa-do-oi-na-televisao/zygmunt-bauman-fala-sobre-o-google-e-a-avalanche-informativa/>. Acesso em: 30 dez. 2020.
13. Wurman, R. S. Ansiedade de informação: como transformar informação em compreensão. 2. ed. São Paulo: Cultura, 2001. 384 p.
 14. Hunt, M. G. *et al.* No More FOMO: limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, [s. /], v. 37, n. 10, 2018. Disponível em: <https://guilfordjournals.com/doi/abs/10.1521/jscp.2018.37.10.751>. Acesso em: 22 fev. 2021.
 15. Hassan, K. Bursting the filter bubble: information literacy and questions of valuation, navigation, and control in a digital landscape. 2018. 115 p. BA Thesis (Major Program in Writing and Rhetoric) - University of Central Florida, Orlando, Florida, 2018. Disponível em: <https://stars.library.ucf.edu/honorsthesis/326/>. Acesso em: 12 jan. 2021.
 16. Pariser, E. *The Filter Bubble: what the internet is hiding from you*. New York: Penguin Press, 2011. 294 p.
 17. Vosoughi, S.; Roy, D.; Aral, S. The Spread of True and False News Online. *Science*, [s. /], v. 359, n. 6380, p. 1146-1151, 2018.
 18. Pierro, B. de. Epidemia de *fake news*. *Revista Pesquisa Fapesp*, São Paulo, [online], abr. 2020. Disponível em: <https://revistapesquisa.fapesp.br/epidemia-de-fake-news/>. Acesso em: 10 fev. 2021.
 19. Kouzy, R. *et al.* Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. *Cureus Journal of Medical Science*, [s. /], v. 12, n. 3, p. 1-9, 2020. Disponível em: <https://www.cureus.com/articles/28976-coronavirus-goes-viral-quantifying-the-covid-19-misinformation-epidemic-on-twitter>. Acesso em: 05 jan. 2021.
 20. Sacramento, I. A saúde numa sociedade de verdades. *Revista Eletrônica da Comunicação, Informação & Inovação em Saúde, Manguinhos*, v. 12, n. 1. p. 1-5, jan./mar. 2018. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1514/2201>. Acesso em: 06 jan. 2021.
 21. Mitchell, T. M. *Machine learning*. [S. /]: McGraw-Hill Science, 1997. 421 p. Disponível em: <https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20Tom%20Mitchell.pdf>. Acesso em: 15 jan. 2021.
 22. Becker, K.; Tumitan, D. Introdução à mineração de opiniões: conceitos, aplicações e desafios. *In: FERREIRA, J. E. (org.). Lectures of the 28th Brazilian Symposium on Databases*. Recife: Universidade Federal de Pernambuco, 2013. p. 27-52.
 23. Tyagi P., Tripathi R. J. A. a. S. A Review towards the sentiment analysis techniques for the analysis of Twitter data. *In: INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING AND SOFTWARE ENGINEERING (ICACSE), 2., 2019, Sultanpur. Proceedings [...]. Sultanpur: SSRN, 2019.*
 24. Esuli, A. *Opinion Mining*. Pisa: Istituto di Scienza e Tecnologie dell'Informazione, 2006. Disponível em: https://www.researchgate.net/publication/228394667_Opinion_Mining. Acesso em: 22 jan. 2021.
 25. Foschiera, S. M. P. A semântica da emoção: um estudo contrastivo a partir da FrameNet e da roda das emoções. 2012. 290 p. Tese (Doutorado em Linguística Aplicada) - Universidade do Vale do Rio dos Sinos, São Leopoldo, 2012. Disponível em: <http://www.repositorio.jesuita.org.br/handle/UNISINOS/4220>. Acesso em: 29 jan. 2021.
 26. Moraes, I. C., Magalhães, L. H. de. Análise de sentimentos de usuários a partir de hashtags no Twitter: um estudo de caso do Oscar de 2017. *In: WORKSHOP DE INICIAÇÃO CIENTÍFICA EM SISTEMAS DE INFORMAÇÃO, 4., 2017, Lavras. Anais [...].*

- Lavras: Universidade Federal de Lavras, 2017. Disponível em: <http://sbsi2017.dcc.ufla.br/download/wicsi.pdf>. Acesso em: 06 dez. 2020.
27. Miranda filho, R., Carvalho, A. I., Pappa, G. L. Inferência de sexo e idade de usuários no Twitter. *In*: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO - CSBC, 34., 2014, Brasília. Anais [...]. Brasília, DF: Universidade de Brasília, 2014. p. 200-211. Disponível em: <http://each.uspnet.usp.br/digiampietri/BraSNAM/2014/p17.pdf>. Acesso em: 30 jan. 2021.
 28. Sombra, T. Reconhecimento de padrões em rede social científica: Aplicação do algoritmo Naive Bayes para classificação de papers no Mendeley. 2018. 198 p. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal do Rio de Janeiro e Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2018. Disponível em: <http://ridi.ibict.br/handle/123456789/968>. Acesso em: 30 jan. 2021.
 29. Rocha, M.; Cortez, P.; Neves, J. M. Análise inteligente de dados: algoritmos e implementação em Java. Lisboa: FCA, 2008. 204 p.
 30. Baeza-Yates, R; Ribeiro-Neto, B. Recuperação de Informação: conceitos e tecnologia das máquinas de busca. 2. ed. Porto Alegre: Bookman, 2013.
 31. Schmitt, V. F. Uma análise comparativa de técnicas de aprendizagem de máquina para prever a popularidade de postagens no Facebook. 2013. 56 p. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <https://lume.ufrgs.br/handle/10183/86407>. Acesso em: 20 jan. 2021.