Research, Society and Development, v. 11, n. 13, e364111335490, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i13.35490

Mineração de textos e análise de sentimentos aplicados a postagens do Twitter acerca das vacinas contra a Covid-19

Text mining and sentiment analysis applied to Twitter posts about Covid-19 vaccines

Minería de texto y análisis de sentimiento aplicado a publicaciones de Twitter sobre vacunas Covid-19

Recebido: 19/09/2022 | Revisado: 29/09/2022 | Aceitado: 03/10/2022 | Publicado: 09/10/2022

Franciele Leal Farias
ORCID: https://orcid.org/0000-0002-4380-4987
Universidade Federal dos Vales do Jequitinhonha e Mucuri, Brasil
E-mail: franleal09@hotmail.com
Lorena Sophia Campos de Oliveira
ORCID: https://orcid.org/0000-0003-0704-1828
Universidade Federal dos Vales do Jequitinhonha e Mucuri, Brasil
E-mail: lorena.sco@gmail.com

Resumo
A pandemia da Covid-19 já é considerada por muitos estudiosos o maior problema sanitário do século XXI e ceifando a vida de milhares de pessoas. A rapidez com que a doença se espalhou e modificou a vida da população mundial gerou uma grande quantidade de emoções e sentimentos nas pessoas. Desde a descoberta do novo coronavírus, iniciou-se uma corrida pelo desenvolvimento de uma vacina que fosse eficaz para o combate da doença, crescendo o anseio da população pela sua chegada. O trabalho realiza a análise dos sentimentos que a população brasileira desenvolveu em relação às vacinas criadas para o combate da Covid-19, por meio da utilização das técnicas de análise de sentimento e mineração de dados. A construção do banco de dados ocorreu através da captação de postagens públicas disponibilizadas pela API do Twitter. O algoritmo desenvolvido durante a pesquisa é baseado na linguagem de programação Python e implementado na plataforma Jupyter Notebook. O processo de análise de sentimentos foi realizado através da análise semântica, com uso do dicionário de léxicos para a língua portuguesa SentiLex-PT.
Palavras-chave: Pandemia; Vacina; Covid-19; Mineração de textos; Análise de sentimentos.

Abstract
The Covid-19 pandemic is already considered by many scholars to be the biggest health problem of the 21st century and claiming the lives of thousands of people. The speed with the disease spread and changed the lives of the world's population generated a lot of emotions and feelings in people. Since the discovery of the new coronavirus, a race began to develop a vaccine that would be effective to combat the disease, growing the population's desire for its arrival. The work analyzes the feelings that the Brazilian population has developed in relation to vaccines created to combat Covid-19, through the use of sentiment analysis and data mining techniques. The construction of the database took place through the capture of public posts made available by the Twitter API. The algorithm developed during the research is based on the Python programming language and implemented on the Jupyter Notebook platform. The sentiment analysis process was carried out through semantic analysis, using the lexicon dictionary for the Portuguese language SentiLex-PT.
Keywords: Pandemic; Vaccine; Covid-19; Text mining; Sentiment analysis.

Resumen
Muchos académicos ya consideran que la pandemia de Covid-19 es el mayor problema de salud del siglo XXI y se cobra la vida de miles de personas. La velocidad con la que la enfermedad se propagó y cambió la vida de la población mundial generó muchas emociones y sentimientos en las personas. Desde el descubrimiento del nuevo coronavirus se inició una carrera por desarrollar una vacuna que fuera eficaz para combatir la enfermedad, y ha crecido el deseo de la población por su llegada. El trabajo analiza los sentimientos que la población brasileña ha desarrollado en relación a las vacunas creadas para combatir el Covid-19, mediante el uso de técnicas de análisis de sentimiento y minería de datos. La construcción de la base de datos se realizó a través de la captura de publicaciones públicas disponibles a través de la API de Twitter. El algoritmo desarrollado durante la investigación está basado en el lenguaje de programación Python e implementado en la plataforma Jupyter Notebook. El proceso de análisis de sentimiento se llevó a cabo a través del análisis semántico, utilizando el diccionario de léxico para la lengua portuguesa SentiLex-PT.
Palabras clave: Pandemia; Vacuna; Covid-19; Extracción de textos; Análisis de los sentimientos.

1. Introdução

Um vírus novo foi responsável por desencadear uma doença respiratória agressiva, com grande taxa de transmissibilidade que rapidamente se espalhou por todo o planeta, não permitindo que medidas de contenção efetivas fossem descobertas antes que este se alastrasse por todos os países do mundo, fazendo milhares de vítimas fatais. O processo de transmissão do vírus ocorre de pessoa para pessoa, por meio do contato com gotículas de saliva ou superfícies contaminadas. O período de incubação do vírus pode variar de 2 a 14 dias (Cardoso, et al., 2021). De tal modo, a pandemia acarretou problemas nas esferas social, política, econômica e educacional, gerando um aglomerado de sentimentos e emoções na população. Dentro desse contexto mundial, o Brasil se encontra atualmente incluído não apenas em um ambiente hostil gerado pela doença como também tem sofrido com uma instabilidade gerada por crises políticas e governamentais (Souza, et al., 2020).

Desde o aparecimento do novo coronavírus, as redes mundiais de comunicação iniciaram uma cobertura massiva sobre as informações obtidas acerca da doença divulgando cada nova descoberta e como esta poderia auxiliar no combate a expansão da pandemia. O Brasil apresentou um comportamento semelhante ao restante das redes mundiais, contudo a divulgação de publicações, notícias e reportagens foram recebidas inicialmente pela população com certo tom de incredibilidade, não acreditando na gravidade da doença. Com o decorrer do tempo e o avanço desordenado da doença, a percepção da realidade por parte da população foi alterada, favorecendo a adoção de medidas de segurança e o surgimento de novas emoções e sentimentos em relação a pandemia (Pimentel & Silva, 2020).

Durante este mesmo período de tempo, onde o caos entre as pessoas era disseminado, cientistas e estudiosos desenvolviam diversos estudos científicos em busca de uma vacina eficaz no combate ao novo coronavírus. O fato de quatro dessas pesquisas terem sido realizadas no Brasil aumentava nos brasileiros a proporção das emoções que eram produzidas e compartilhadas diariamente através das redes sociais (Castro, 2021).

Nos últimos anos as redes socias tem aumentado a sua adesão, tornando-se algumas das principais plataformas de comunicação utilizadas para compartilhar informações e para que as pessoas possam se expressar (Silva & Malheiros, 2019). Dentro do contexto da pandemia e do período de isolamento social durante o *lockdown* realizado em quase todos os países, o volume de dados produzidos através das redes sociais aumentou significantemente e as tornaram preciosas fontes de dados a serem analisados.

Ao se considerar a grande quantidade de emoções produzidas na população, tem se a análise de sentimentos, uma área da ciência da computação, que se tornou de grande popularidade nos trabalhos em que se desejava encontrar opiniões e sentimentos que poderiam ser associados a produtos, personagens ou lugares, e que fossem expostos em mídias sociais. Porém, atualmente, esta área pode ser aplicada a trabalhos de âmbito social e que beneficiem a população (Ramos & Freitas, 2019).

Para realizar tal trabalho, a análise de sentimentos está sendo associada a mineração de textos, um conjunto de métodos multidisciplinares utilizados para extrair regularidades, encontrar tendências ou padrões em dados textuais, identificando assim informações úteis que dificilmente seriam descobertas utilizando métodos tradicionais de consulta (Morais & Ambrósio, 2007).

Dentro de tais contextos o presente trabalho realiza a análise dos sentimentos que a população brasileira desenvolveu em relação às vacinas criadas para o combate da Covid-19, por meio da utilização das técnicas de análise de sentimento e mineração de dados. A construção do banco de dados ocorreu através da captação de postagens públicas disponibilizadas pela API do Twitter durante o período que abrange 03 de novembro de 2020 a 20 de fevereiro de 2021.

É importante destacar que foi construído um banco de dados que represente uma amostra aleatória das postagens realizadas neste período de tempo, não compreendendo a totalidade das postagens realizadas. Dentro deste período se encontram a data de aprovação e aplicação da primeira vacina contra Covid-19 no mundo e no Brasil.

2. Metodologia

O algoritmo desenvolvido durante a pesquisa foi construído na linguagem de programação Python. A linguagem Python é simples, clara e objetiva, contudo é também muito poderosa e forte, possibilitando o seu uso em diversas aplicações. A utilização desta vem crescendo em áreas da computação como inteligência artificial, banco de dados, jogos, aplicativos para celulares, entre outros (Menezes, 2010). As linguagens de programação necessitam de um Ambiente de Desenvolvimento Integrado, ou *Integrated Development Environment* (IDE), sendo utilizada nesta pesquisa o *Jupyter Notebook*, uma ferramenta que possui códigos iterativos baseados na *Web*, os denominados notebooks, que oferecem suporte para dezenas de linguagens de programação, incluindo o Python (Mckinney, 2019).

Para a construção do banco de dados a ser analisado, o Twitter foi escolhido como a rede social a ser utilizada devido ao fato de suas publicações serem de acesso público e possuir diversas formas de se obter tais publicações, como por exemplo, através da sua API. O Twitter é uma plataforma popular nos dias atuais e se destaca pelas interações praticamente imediatas. Neste, cada usuário possui uma espécie de microblog onde pode escrever mensagens de até 140 caracteres, os *'tweets'* (Franco & Adaniya, 2018). Atualmente a plataforma do Twitter já disponibiliza opções de contas onde o usuário possui a capacidade de escrever mensagens com uma quantidade maior de caracteres. Serão apresentados nos tópicos a seguir uma breve contextualização da pandemia juntamente com as técnicas de mineração de textos e análise de sentimentos que foram utilizadas na pesquisa.

2.1 A Pandemia da Covid-19

Em dezembro de 2019 uma nova espécie do coronavírus, denominada SARS-CoV-2, foi identificada pela primeira vez em Wuhan, uma cidade localizada na província de Hubei, na China. Tal espécie de vírus era responsável por desencadear uma doença respiratória agressiva, denominada Covid-19, que rapidamente se espalhou por todo o planeta tornando-se o maior problema sanitário do século XXI (Cardoso, et al., 2021). Segundo a Organização Mundial de Saúde (OMS) a doença pode se manifestar de diversas formas e níveis, sendo os principais: pessoas assintomáticas, quadros gripais leves e síndrome respiratória aguda. Deve-se ainda salientar que existe maior suscetibilidade de contagio e quadros graves da doença em pessoas idosas, acima de 60 anos de idade, e pessoas com comorbidades (OMS, 2020).

Desde o aparecimento do novo coronavírus, as redes mundiais de comunicação iniciaram uma cobertura massiva sobre as informações obtidas acerca da doença divulgando cada nova descoberta e como esta poderia auxiliar no combate a expansão da pandemia. O Brasil apresentou um comportamento semelhante ao restante das redes mundiais, inclusive no que abrange a divulgação de publicações, notícias e reportagens que foram recebidas inicialmente pela população com certo tom de incredibilidade, não acreditando na gravidade da doença. Com o decorrer do tempo e o avanço desordenado da doença, a percepção da realidade por parte da população foi alterada, favorecendo a adoção de medidas de segurança e o surgimento de novas emoções e sentimentos em relação a pandemia (Pimentel & Silva, 2020).

De acordo com o avanço da doença e sua expansão no cenário global, as populações foram submetidas a um período de intenso temor, acarretando pensamentos irracionais e atitudes antes consideradas inimagináveis. A população brasileira além do temor pelo desconhecido e de angustia por medo da morte, enfrentava ainda um agravante causado pelas inúmeras fake News espalhadas pelos meios de comunicação e pelas redes sociais. Assim, sentimentos como ansiedade, estresse, pânico, medo, raiva, entre outros, se tornaram extremamente comuns no cotidiano da população (Pimentel & Silva, 2020)

Os problemas gerados pela falta de informações precisas no Brasil se tornaram tão amplos que diversos meios de comunicação desenvolveram seus próprios painéis para acompanhar o avanço da pandemia, visto que o Ministério da Saúde não divulgava as informações dentro dos prazos que eram necessários. O Conselho Nacional de Secretários de Saúde

desenvolveu o Painel CONASS de acompanhamento ao avanço da Covid 19 no Brasil, sendo que em 30 de março de 2022 já eram contabilizados quase 30 milhões de brasileiros infectados pelo vírus e 659.504 vidas perdidas para a doença (CONASS, 2022).

Desde a descoberta do novo coronavírus, iniciou-se uma corrida pelo desenvolvimento de uma vacina que fosse eficaz para o combate da doença. Foram registrados junto a OMS cerca de 200 projetos de vacinas para a Covid-19. Dentre estes, quatro foram efetuados no Brasil, fazendo com que a população estivesse familiarizada com as etapas do processo de desenvolvimento (Castro, 2020).

A existência dessa grande quantidade só foi possível devido ao gigantesco investimento realizado pelos governos dos países desenvolvidos, de algumas organizações não governamentais e das empresas farmacêuticas, que por muitas vezes trabalharam entre si para obter uma vacina no menor período de tempo possível. Uma medida inédita foi adotada em relação a vacina da Covid 19: uma ação liderada pela OMS denominada *Covax Facility*, que possuía o objetivo de acelerar o desenvolvimento e a fabricação de vacinas contra a Covid-19 para garantir que todos os países tivessem acesso a elas toda a população mundial fosse vacinada (Domingues, 2021).

Ainda no ano de 2020, as primeiras vacinas obtiveram autorização para uso emergencial em alguns países europeus e nos Estados Unidos. Contudo, no Brasil o uso emergencial das vacinas só foi autorizado pela Agência Nacional de Vigilância Sanitária, a ANVISA, em 17 de janeiro de 2021. Após alguns minutos a enfermeira Mônica Calazans foi a primeira brasileira a ser vacinada em território nacional (Castro, 2020).

2.2 Mineração de Textos

A Mineração de Textos (MT), ou *text mining*, é considerada pelos estudiosos como um Processo de Descoberta de Conhecimento, onde são utilizadas técnicas de análise e extração de dados a partir de documentos textuais como textos, frases ou apenas palavras. A mineração de textos é um método multidisciplinar que utiliza técnicas de Estatística, Informática, Programação e linguística para processar textos identificando informações úteis que estavam implícitas nos textos e que dificilmente seriam descobertas utilizando métodos tradicionais de pesquisa (Morais & Ambrósio, 2007).

O principal objetivo da mineração de texto é encontrar termos relevantes em documentos textuais com grande volume de dados e assim estabelecer padrões e relacionamentos entre eles. Tais dados se encontram arquivados de forma estruturada, não estruturada ou semi estruturada. É importante destacar que a mineração de textos não é um mecanismo de busca, como os encontrados na internet, pois a mineração auxilia o usuário a descobrir informações que antes eram desconhecidas. Já nos mecanismos de busca, o usuário já sabe o que eles desejam procurar (Pezzini, 2016).

Existem duas abordagens principais para o processo de mineração em dados textuais que são: a Análise Estatística, que está diretamente ligada a frequência de aparição de um determinado termo, não se atentando ao contexto em que este se encontra inserido no texto; e a Análise Semântica, que se preocupa com a funcionalidade dos termos nos textos por meio dos significados morfológico, sintático, semântico, entre outros que podem ser empregados no contexto da pesquisa realizada. Tais abordagens podem ser aplicadas sozinhas ou em associação (Brito, 2017).

A aplicação inicialmente é realizada em um grupo teste para que o algoritmo seja treinado, e se o mesmo não alcançar um resultado satisfatório, as etapas devem ser repetidas até que o objetivo seja alcançado. Após o algoritmo se encontrar apto, este pode ser aplicado aos dados que se deseja analisar (Brito, 2017).

A Figura 1 apresenta as etapas do processo de mineração de texto segundo o trabalho de Brito (2017), em formato de diagrama para facilitar a visualização das mesmas.

Figura 1. Diagrama do Processo de Mineração de Texto

Fonte: Adaptado de Brito (2017).

Cada uma das etapas apresentadas na Figura 1 tem grande importância dentro do processo e possui as suas peculiaridades. A etapas podem ser definidas como:

Seleção – Se trata da extração e coleta dos dados que serão utilizados na pesquisa, criando uma base de dados textuais que represente toda a população que será analisada, ou seja, a criação do corpus da pesquisa.

Pré-processamento – Consiste na filtragem e limpeza dos dados, eliminando informações desnecessárias ao algoritmo.

Transformação – Se trata da classificação dos dados, separando-os de acordo com as características que se pretende observar.

Mineração – Esta etapa está diretamente ligada as técnicas de aprendizagem de máquina para obtenção de novos conhecimentos.

Avaliação – É a etapa onde será validada a eficiência do processo como um todo, analisando os dados que foram obtidos após a aplicação dos algoritmos.

2.3 Análise de Sentimentos

A análise de sentimentos, também denominada pelos estudiosos como Mineração de Opinião, é um método utilizado para identificar emoções e opiniões em textos, a partir do uso das técnicas de Mineração de Textos, Informática, Linguística, entre outros (Brito, 2017). A análise de sentimentos é considerada pelos estudiosos como um subcampo do Processamento de Linguagem Natural, uma área definida como conjunto de técnicas computacionais criadas para analisar e representar textos naturais, de forma a obter um processamento de linguagem o mais próximo possível do humano (Olenscki, et al., 2020).

Sentimento é definido por Liu (2012) como uma opinião ou uma avaliação sobre algum aspecto ou objeto. Já a análise do sentimento, envolve o PLN, para que extrair as emoções expressas principalmente em textos online, onde tal área tem sido bastante utilizada para verificar a popularidade de pessoas, objetos, lugares e situações (Brito, 2017).

A classificação dos sentimentos identifica a polaridade do texto analisado, mesmo que este seja um documento, uma frase ou apenas uma palavra, classificando-o em positivo, negativo ou neutro. Para realização de uma classificação utilizando sentimentos mais amplos, é necessário realizar alterações nos algoritmos para que estes sejam capazes de determinar tais emoções (Liu, 2012).

Segundo Silva (2016) a análise de sentimentos aplicada a redes sociais pode ser caracterizada em quatros tipos:

- Supervisionada – Esta técnica é realizada por meio do uso de algum algoritmo de aprendizagem de máquina e exige a existência de uma base de dados para treinamento que seja previamente rotulada por um especialista.

- Guiada pelo uso de léxico – Neste modelo é utilizada uma lista de termos classificados, por alguém especializado, em positivos e negativos que direcionará o processo de identificação da polaridade da base de dados.
- Supervisionada Híbrida - Nesta abordagem são aplicados simultaneamente um algoritmo de aprendizagem de máquina e uma lista de léxicos.
- Baseada em Grafos – Esta técnica utiliza as características especificas das redes socias, como os relacionamentos e interações entre os usuários para realizar a análise.

A análise de sentimentos é um método que utiliza a mineração de textos em sua execução, além de apresentar também suas técnicas especificas de acordo com a aplicação desejada (Brito, 2017).

3. Resultados e Discussão

Inicialmente foi realizada a construção do primeiro algoritmo, sendo este responsável por realizar a coleta de dados, através da aplicação da biblioteca para linguagem Python, *Tweepy*. Para realizar o processo de coleta é necessário possuir uma conta no Twitter para desenvolvedores. Após a aprovação da conta, o usuário pode criar um aplicativo para ser vinculado a IDE escolhida pelo mesmo. O Twitter criará automaticamente credenciais para liberar o acesso do usuário a uma quantidade pré-determinada de tweets, de acordo com o modelo de conta solicitado. As credenciais fornecidas pela plataforma são individuais e não podem ser compartilhadas com outras pessoas.

A construção da base de dados para testes foi realizada no período de primeiro de abril de 2021 a 07 de abril de 2021. A coleta foi realizada através da busca por *endpoints*, por meio do *Tweepy*. Nesse momento é importante destacar que a API do *Twitter* passou por atualizações recentes e atualmente o retorno para cada pesquisa realizada é de no máximo 100 *tweets*, selecionados de acordo com perfis de interação próprios da plataforma.

Os *endpoints* utilizados para esta busca foram: ‘pandemia’ e ‘Covid19’. Não foram consideradas as *hashtags* nesse momento. Como resultado desta busca inicial, obteve-se 700 *tweets*, nos quais foram realizados os tratamentos iniciais de limpeza, onde foram excluídas as publicações que não possuíam classificação de sentimentos, como os informativos publicados por diversos perfis de canais de comunicação.

Após a exclusão restaram 248 *tweets* que foram manualmente classificados, dos quais 150 são negativos, 52 positivos e 46 nulos. A presença de um número de publicações negativas muito superior as demais pode ser explicada pelo fato dos sentimentos que a pandemia produzia na população eram em sua essência de medo, tristeza, angústia, entre outros já apresentados neste trabalho. A Tabela 1 traz exemplos dos tweets classificados na base de dados para testes.

A construção da base de dados para a pesquisa foi realizada no período de 03 de novembro de 2020 a 20 de fevereiro de 2021. Esta segunda coleta também foi realizada através da busca por *endpoints*, por meio da biblioteca *Tweepy*. A base de dados para pesquisa foi dividida em 3 grandes grupos, onde cada grupo compreende o período de 5 semanas epidemiológicas que abrangem em três momentos específicos, sendo eles:

- Primeiro período (03 de novembro de 2020 – 07 de dezembro de 2020) – este período compreende o momento entre a descoberta oficial da primeira vacina no mundo e um dia antes da primeira pessoa ser vacinada contra a Covid-19.
- Segundo período (08 de dezembro de 2020 – 16 de janeiro de 2021) – neste ponto é importante verificar a expectativa dos brasileiros, pois já existia uma vacina sendo aplicada pelo mundo, porém não havia a liberação da ANVISA para que o processo de vacinação fosse liberado no Brasil.
- Terceiro período (17 de janeiro de 2021 – 20 de fevereiro de 2021) – a vacinação é liberada pela ANVISA e é a

primeira pessoa é vacinada no Brasil contra a Covid-19. Os momentos após essa liberação são de suma importância para serem visualizados devido as ramificações dos problemas que foram ocorrendo dia a dia neste processo.

Tabela 1. Exemplo de *tweets* da Base de Testes.

Tweets	**Polaridade**
só quero acordar desse pesadelo chamado pandemia	Negativo
milhões de pessoas em situação de imigração se veem desumanizadas e desprotegidas, cenário ainda mais grave no meio de uma pandemia que hierarquiza vidas	Negativo
Nessa pandemia to aprendendo valorizar cada vez mais o abraço, as risadas, as alegrias as tristezas a oportunidade de congregar e de conviver. Quando tudo isso passar não quero perder meu tempo com coisas fúteis, quero investir meu tempo em pessoas!	Positivo
E em meio a essa pandemia, só tenho a agradecer e ficar feliz por poder estar com minha família e ter uma ceia feliz	Positivo
Fome atingiu 19 milhões de brasileiros durante a pandemia, diz pesquisa.	Neutro
Gilmar Mendes contraria Kassio Nunes e respalda fechamento de igrejas e templos durante pandemia	Neutro

Fonte: Autoria Própria (2022).

Ao se analisar a Tabela 1 é importante salientar que durante a pesquisa, bem como na composição do artigo, os erros de ortografia e gramática que as publicações possuíam foram mantidos.

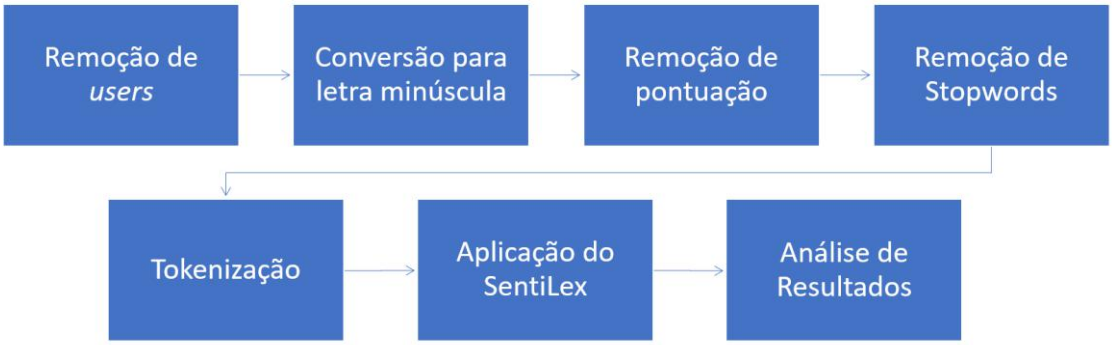
Os *endpoints* utilizados para a construção do banco de dados da pesquisa foram: ‘vacina’, ‘coronavírus’ e ‘Covid19’. Nesta busca também não foram consideradas as hashtags. Como resultado desta busca inicial, obteve-se 11000 *tweets*, nos quais foram realizados os tratamentos abaixo especificados. Após o tratamento inicial, a base de dados para a pesquisa era constituída de 5041 *tweets.*

3.1 Algoritmo de Analise de Sentimentos

A Figura 2 apresenta a estrutura de execução do algoritmo construído nesta pesquisa. O algoritmo foi aplicado no conjunto de dados obtidos na etapa de coleta de dados, classificado como Banco de Dados não Relacional, pois não existe uma relação entre pré-estabelecida entre os dados.

O algoritmo é iniciado com a limpeza dos tweets mantendo apenas aqueles que possam verdadeiramente fornecer uma informação quanto a opinião dos usuários, e para isso foram excluídas as publicações que possuíam conteúdos informativos, como as realizadas pelas páginas de meios de comunicação.

Figura 2. Estrutura do Algoritmo



Fonte: Autoria Própria (2022)

A Figura 2 apresenta sobre o modelo de um fluxograma as etapas a serem executadas no algoritmo. É importante ressaltar que tal fluxograma mostra uma versão do fluxograma apresentado na Figura 1, porém com as técnicas aplicadas no algoritmo criado nesta pesquisa.

Figura 3. Exemplo de *tweet* a ser excluído.

Fonte: Twitter (2022).

A Figura 3 representa um exemplo de tweet a ser excluído, pois apesar de ele possuir uma informação relevante, este não apresenta a opinião ou sentimento de nenhum usuário, não sendo assim considerado apto para constituir o banco de dados da pesquisa.

Para possibilitar esta etapa de exclusão foram identificados os *users* das páginas de comunicação e os *tweets*

retornados por tais usuários eram excluídos, até que restassem apenas as publicações que continham opiniões. A Figura 4 apresenta alguns dos *users* excluídos.

Figura 4. *Users* excluídos.

Fonte: Autoria Própria (2022).

Para realizar o Processamento de Linguagem Natural foi utilizada a biblioteca NLTK. Esta biblioteca está disponível para a linguagem de programação Python e contém por volta de 35 módulos. Cada modulo possui seus diversos submódulos que realizam múltiplas tarefas de PLN, como *tokenização*, remoção de *stopwords*, e possui inclusive um módulo para análise de sentimentos, entre outros módulos (Araújo, 2017).

Os módulos aplicados neste algoritmo, na etapa de Pré-Processamento são: a remoção de caracteres especiais e pontuação, a remoção de *stopwords*, a *tokenização* e a conversão para letras minúsculas.

Após a etapa de pré-processamento é realizada a aplicação do dicionário de léxicos SentiLex-PT aos *tweets*. A aplicação deste consiste na busca de cada palavra que compõe o *tweet* dentro da estrutura do dicionário. Neste modelo de avaliação cada tweet é entendido como um vetor, onde as palavras são suas estruturas. Cada palavra recebe um valor e o resultado obtido é o somatório do valor das palavras. A classificação e o valor atribuído as palavras estão expressos na Tabela 2. Entretanto, é necessário destacar que o dicionário já apresenta as palavras com seus devidos valores definidos.

Tabela 2. Valor atribuído as palavras de acordo com o dicionário.

Tipo de palavra	**Valor**
Palavras que expressam sentimentos negativos	-1
Palavras que expressam sentimentos positivos	+1
Palavras que não expressam sentimentos	0

Fonte: Adaptado de Silva (2016).

A Tabela 2 informa os valores que são concedidos a cada palavra de acordo com o dicionário SentiLex-PT. Desta maneira um *tweet* vai ter sua classificação expressa pelo somatório dos valores individuais de cada uma das suas palavras

Na sequência das etapas da MT, é realizada a avaliação dos resultados obtidos. Tal avaliação será realizada através da utilização da métrica acurácia. A acurácia encontrada no algoritmo aplicado é 68,5 %. O valor pode ser considerado pequeno quando comparado aos algoritmos que utilizam técnicas de aprendizagem de máquina, entretanto, é considerado adequado ao se tratar da aplicação de dicionários de léxicos.

A Tabela 3 e a Figura 5 apresentam os resultados obtidos para a análise de sentimento aplicada no banco de dados da

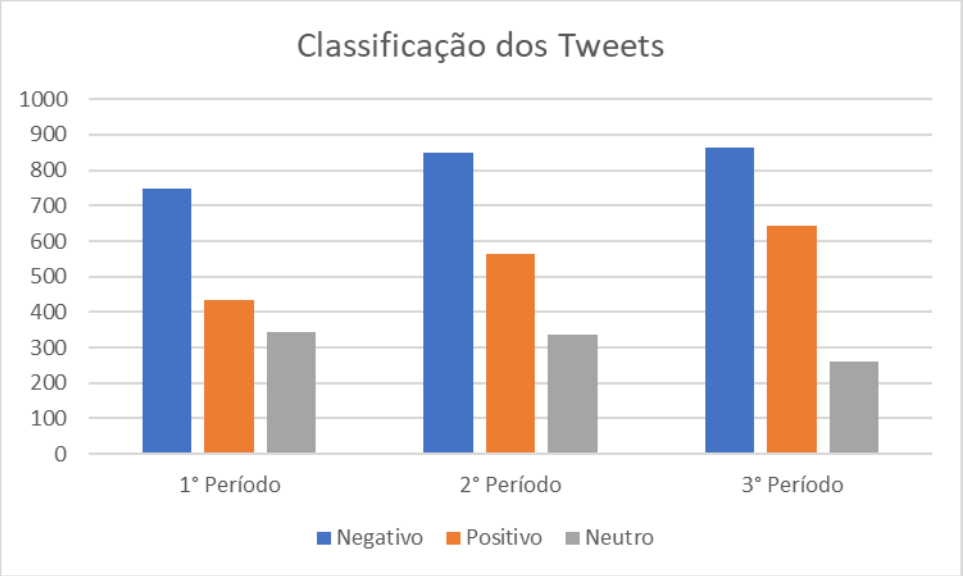
pesquisa, conforme a divisão de períodos supracitada. Para justificar certos comportamentos encontrados, como a presença majoritária de tweets negativos em ambos os períodos, é importante salientar que essa negatividade pode não estar diretamente associada a vacina, ou seja, a pessoa pode não estar defendendo a não aplicação da vacina e sim estar discorrendo sobre algum acontecimento acerca da vacinação que causa sentimentos negativos na pessoa que realizou a publicação.

Tabela 3. Valores obtidos pelo algoritmo.

Período	Positivo	Negativo	Neutro	Total
Primeiro	435	747	343	1525
Segundo	562	851	336	1749
Terceiro	643	865	259	1767

Fonte: Autoria Própria (2022)

Figura 5. Resultado da pesquisa.



Fonte: Autoria Própria (2022).

Verificando os resultados apresentados na Tabela 3 e na Figura 5, pode-se perceber a tendência predominante dos *tweets* em serem negativos, porém ao se analisar os *tweets* positivos, é visível um aumento considerável na quantidade de ocorrência dos mesmos. Tal aumento pode representar que a população se encontrava otimista quanto ao fato de vacinas terem tido o seu uso autorizado, bem como o início da vacinação no Brasil.

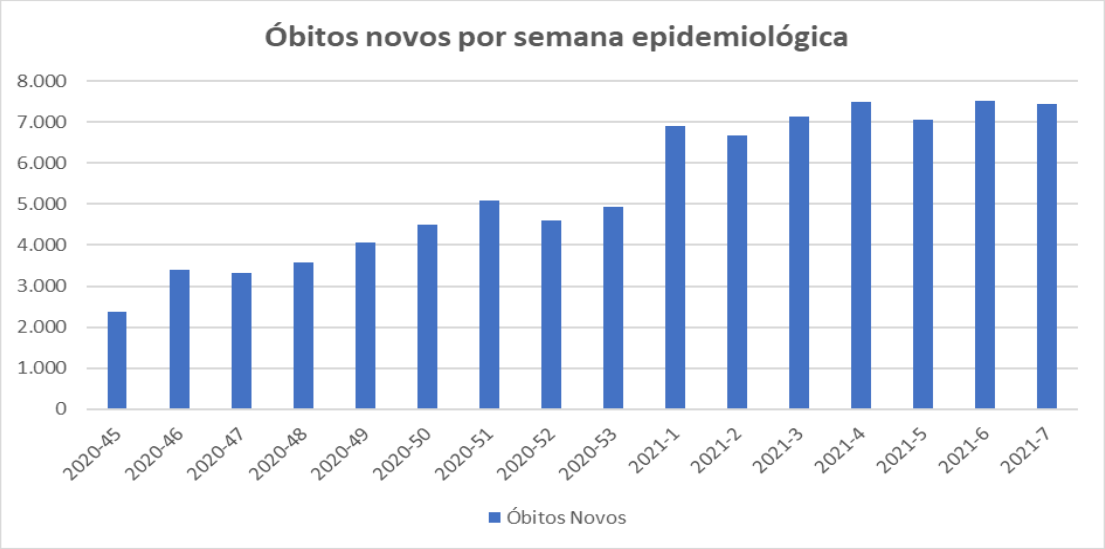
A seguir serão apresentados alguns temas e notícias presentes no período de coleta de dados e que possivelmente seriam responsáveis por algumas das variações que geraram essa negatividade predominante:

- A ocorrência de eleições municipais no Brasil.
- A interrupção que a ANVISA realizou nos estudos das vacinas realizados no território brasileiro.
- Caos instaurado em Manaus devido à falta de oxigênio e de leitos para atendimento dos pacientes.
- Declarações diversas realizadas pelo presidente Jair Bolsonaro, que causaram grande alvoroço nas redes socias, como a aplicação da vacina que supostamente transformaria as pessoas em ‘jacaré’.
- Com o início da vacinação, surgem novos problemas como as pessoas que burlavam as filas de pessoas

prioritárias para receberem as vacinas antes da sua convocação oficial.

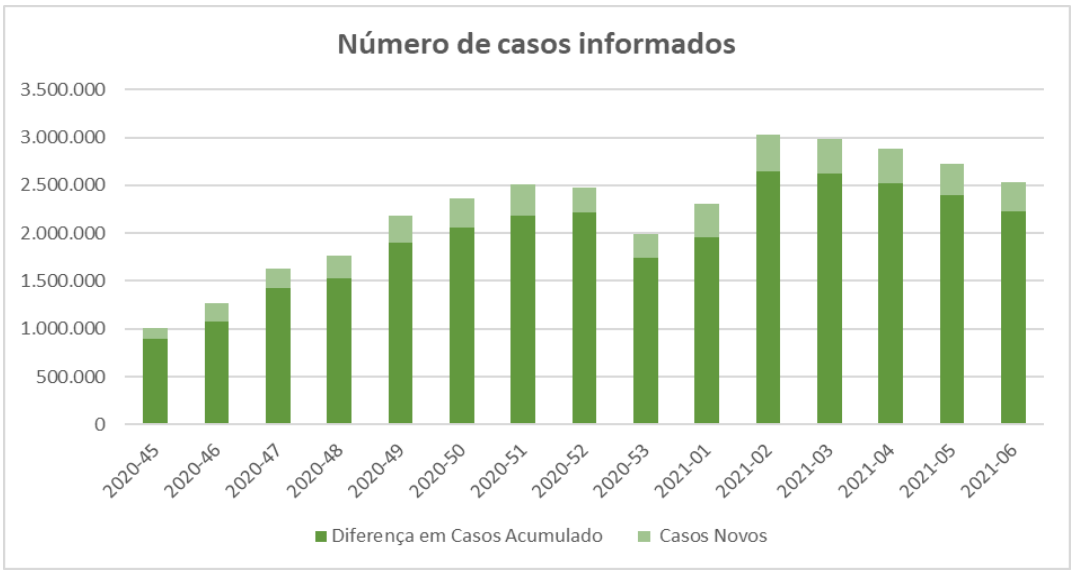
- Profissionais da saúde que foram gravados no momento da vacinação e não injetaram nenhum medicamento nos pacientes.

Figura 6. Quantidade de óbitos por semana epidemiológica.



Fonte: Adaptado CONASS (2022).

Figura 7. Quantidade de casos novos por semana epidemiológica.



Fonte: Adaptado CONASS (2022).

Para proporcionar uma visão mais clara do cenário em que se encontrava o Brasil no momento em que tais dados foram coletados, a Figura 6 apresenta a quantidade de óbitos contabilizados nas semanas epidemiológicas em que se realizou a pesquisa e a Figura 7 expõe a quantidade crescente de casos de pessoas infectadas. A vacina teve seu uso emergencial liberado pela ANVISA no momento que o Brasil iniciava a curva ascendente de casos da segunda onda de contaminação. É importante ainda destacar que o ápice de óbitos no país ocorreu após a liberação da vacina, entretanto não havia imunizante o suficiente para que toda a população fosse vacinada em um curto período de tempo.

4. Conclusão

A pandemia do Covid-19 já é considerada por vários estudiosos como o maior problema sanitário do XXI. A sua rápida expansão no cenário global gerou um estado de caos na população mundial, que de repente se encontra isolada, privada de sua liberdade e do contato social. Tais situações contribuíram para criar um misto de sentimentos nas pessoas, gerado principalmente pelo temor do desconhecido, visto que não se conhecia muito sobre o vírus, o seu comportamento e tratamentos que fossem eficazes.

Assim sendo, tal pesquisa foi desenvolvida com o objetivo de realizar um estudo sobre a percepção dos brasileiros acerca da vacina contra a Covid-19, baseado nas postagens realizadas pela população no Twitter, no período marcado pela aprovação das primeiras vacinas no Brasil e no mundo. O algoritmo desenvolvido apresentou acurácia de 68,5%, sendo considerado um valor adequado para este modelo de pesquisa.

Para dar continuidade ao trabalho, em uma pesquisa futura seria viável o desenvolvimento de um algoritmo baseado na aprendizagem de máquina para que sejam realizadas comparações entre seus respectivos resultados. Tal desenvolvimento necessita de uma base de dados para treinamento do algoritmo que possua uma quantidade elevada de publicações já classificadas de acordo com a polaridade possibilitando uma análise adequada do modelo.

Referências

Araújo, L. G. de A. (2017). *Sentimentall Versão 2: Desenvolvimento de Análise de Sentimentos em Python*. Trabalho de Conclusão de Curso em Ciência da Computação do Centro Universitário Luterano de Palmas. Palmas.

Brito, E. M. N. (2017). *Mineração de Textos: detecção automática de sentimentos em comentários nas mídias sociais*. Dissertação: Programa de Mestrado em Sistemas de Informação e Gestão do Conhecimento da Universidade Fundação Mineira de Educação e Cultura — FUMEC. Belo Horizonte.

Cardoso, R. F. et al. (2021). COVID-19: An epidemiological challenge. *Research, Society and Development.* 10(7), e32110716313. DOI: 10.33448/rsd-v10i7.16313.

Castro, R. (2021). Vacinas contra a Covid-19: o fim da pandemia? *Physis: Revista de Saúde Coletiva*. 31, e310100.

Conselho Nacional de Secretários de Saúde - CONASS. (2022). *Painel Covid-19.* Acesso em 17 de julho de 2022 em https://www.conass.org.br/painelconasscovid19.

Domingues, C. M. A. S. (2021). Desafios para a realização da campanha de vacinação contra a COVID-19 no Brasil. *Cadernos de Saúde Pública*, *37*.

Franco, R. B., & da Costa Adaniya, M. H. A. (2018). Sistemas de análise de sentimentos usando dados do Twitter. *Revista Terra & Cultura: Cadernos de Ensino e Pesquisa*, *34*(esp.), 111-118.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

Mckinney, W. (2019). *Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython*. Novatec Editora.

Menezes, N. N. C. (2010). *Introdução a programação com Python*. *São Paulo:* Novatec.

Morais, E. A. M., & Ambrósio, A. P. L. (2007). *Mineração de textos*. Relatório Técnico: Instituto de Informática da Universidade Federal de Goiás.
Olenscki, J., Xavier, F., Acosta, A., Saraiva, A., & Sallum, M. (2020). Aplicação de análise de sentimentos no Twitter para avaliação da percepção pública quanto a cloroquina. Em *Anais do XX Simpósio Brasileiro de Computação Aplicada à Saúde* (pp. 500-505). SBC.

Organização Mundial da Saúde - OMS. (2020). *Coronavirus disease 2019 (COVID-19): Situation Report –51*. Acesso em 15 de julho de 2022 em https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid19.pdf?sfvrsn=1ba62e57_10.

Pezzini, A. (2017). Mineração de textos: conceito, processo e aplicações. *Revista Brasileira De Contabilidade E Gestão*, *5*(8), 58-61.

Pimentel, A. do S. G., & Silva, M. de N. R. M. de O. (2020). Saúde psíquica em tempos de Corona vírus. *Research, Society and Development*, *9*(7), e11973602. DOI: 10.33448/rsd-v9i7.3602.

Ramos, B., & Freitas, C. (2019). "Sentimento de quê?" uma lista de sentimentos para a Análise de Sentimentos. *STIL*, 15-18.

Silva, E. P. da & Malheiros, Y. (2019). *Um conjunto de dados extraído do Twitter para análise de sentimentos na língua portuguesa*. Trabalho de Conclusão de Curso em Sistemas de Informação da Universidade Federal da Paraíba.

Silva, N. F. F. (2016). *Análise de sentimentos em textos curtos provenientes de redes sociais*. Tese de Doutorado: Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo – ICMC-USP. São Carlos.

Sousa, A. R. D., Carvalho, E. S. D. S., Santana, T. D. S., Sousa, Á. F. L., Figueiredo, T. F. G., Escobar, O. J. V., Mota, T. N. & Pereira, Á. (2020). Sentimento e emoções de homens no enquadramento da doença Covid-19. *Ciência & Saúde Coletiva*, *25*, 3481-3491.

Twitter. (2022). *Página Inicial*. Acesso em 15 de julho de 2022 em https://twitter.com.