

UM ESTUDO DA ACEITAÇÃO DAS VACINAS CONTRA A COVID-19 NA REDE SOCIAL TWITTER UTILIZANDO FERRAMENTAS DE ANÁLISE SENTIMENTAL

Beatriz Angélica Torres¹, Larissa Costa De Moraes¹,
Daniele Aparecida Cicillini Pimenta¹

¹Faculdade de Tecnologia de Ribeirão Preto (Fatec)

Ribeirão Preto, SP – Brasil

beatriz.torres2@fatec.sp.gov.br,

larissa.moraes4@fatec.sp.gov.br,

daniele.pimenta@fatec.sp.gov.br

Resumo. *As redes sociais têm se tornado objeto de estudo de várias áreas, principalmente voltadas a produtos e serviços, devido à alta quantidade de opiniões geradas por usuários dessas plataformas. Sendo assim, por meio da análise de sentimento é feita a extração do sentimento expresso nesses dados. Portanto este trabalho tem o intuito de comparar métodos de aprendizagem de máquina e métodos baseados no dicionário léxico, e utilizá-los para classificar como os usuários do Twitter se sentem a respeito das vacinas que estão sendo desenvolvidas contra o Covid-19. A partir dos resultados das comparações, foi visto que o método Vader mostrou melhor taxa de acurácia e precisão; seus resultados mostraram que 33,28% dos tweets eram positivos, e 66,73% eram negativos.*

Abstract. *Social networks have become the object of study of several areas, mainly focused on products and services, due to the high amount of opinion generated by users of these platforms. Thus, through sentiment analysis, the sentiment expressed in these data is extracted. Therefore, this work aims to compare machine learning methods and methods based on the lexicon dictionary and use them to classify how Twitter users feel about the vaccines being developed against Covid-19. From the results of the comparisons, it was seen that the Vader method showed better accuracy and precision, its results showed that 33.28% of tweets were positive, and 66.73% were negative.*

1. Introdução

Em um mundo cada vez mais digital, as redes sociais vêm se tornando uma plataforma de comunicação cada vez mais fundamental na rotina das pessoas, intensificada pelo surto da Covid-19; nesta era foi introduzido um novo uso da tecnologia, e os dados assumiram papéis mais significativos em seus cotidianos (DATA..., 2020).

A rede social Twitter recebe em média mais de 350.000 *tweets* enviados por minuto, 500 milhões de *tweets* por dia e cerca de 200 bilhões de *tweets* por ano (TWITTER..., 2013). No segundo trimestre de 2020, durante a pandemia da Covid-19, houve um crescimento de 34 % de usuários na plataforma em comparação com um ano antes (ONLINE, 2020).

O número de informações geradas pelos usuários dessas plataformas é extremamente grande; entretanto, essas inúmeras fontes de opiniões raramente são acessíveis, principalmente por estarem escondidas em *posts* de *blogs* ou fóruns, ou até em grandes quantidades de comentários em redes sociais, tornando, assim, inviável a sua verificação manual. Diante desta realidade, faz-se necessária uma busca automática e específica de opinião e a classificação destas quanto a sua polaridade, a partir da análise de sentimento, que é a tarefa de compreender opiniões, sentimentos e emoções que foram expressas em linguagem textual. O objetivo é realizar a extração de elementos do texto que apresentam as opiniões das pessoas a respeito de um determinado assunto, e determinar a classificação dela em positivo ou negativo.

Devido ao aumento na quantidade de mensagens e à geração de discussões no Twitter, este trabalho teve como objetivo a análise e a classificação das opiniões dos usuários extraídas da plataforma a respeito das vacinas que estão em desenvolvimento contra a Covid-19, por meio de métodos de aprendizado de máquina e dicionário léxico.

2. Trabalhos Relacionados

A análise de sentimentos e mineração de opiniões tem tido um grande destaque, devido à grande quantidade de textos com opiniões pessoais disponíveis em redes sociais, *blogs* e fóruns, entre outras fontes da *Web*. Com o desenvolvimento dessa área, hoje é possível aplicar diferentes técnicas para a extração e análise dessas opiniões. A seguir, serão apresentadas sínteses de diferentes técnicas que podem ser utilizadas para a análise de sentimentos em textos.

No trabalho de Araújo, Gonçalves e Benevenuto (2013), é feita a comparação de oito métodos utilizados em análise de sentimentos, SentiWordNet, SASA, PANAS-t, Emoticons, SentiStrength, LIWC, SenticNet e Happiness Index, com o objetivo de avaliar a eficácia de cada método. As métricas usadas para avaliar os métodos foram a *acurácia* que representa a taxa de sentimentos classificados corretamente pelo método, *precisão* que calcula a taxa de proximidade entre os métodos, e *F- measure*, taxa usada para comparar os métodos. A partir dos resultados obtidos foi determinado que os métodos propostos tiveram diferentes graus de acurácia; sendo assim, foi considerado que nem sempre um método vai ter os melhores resultados. E, por meio dos resultados, tentaram encontrar um método com os quais os resultados fossem satisfatórios, o que os levou à construção de um novo método, o Método Combinado, que consiste na combinação dos oito métodos apresentados.

Em *Análise de sentimentos no Facebook utilizando o SenticNet* (LIMA, 2013), foi efetuada a coleta de opiniões contidas na rede social *Facebook*; os dados então foram classificados utilizando-se o *software* SenticNet; em seguida, o autor criou um formulário *online* para voluntários classificarem os dados para a avaliação da eficácia, tendo obtido um resultado positivo com relação ao *software*.

O diferencial neste trabalho, em relação aos demais citados, foi a criação de um conjunto de dados, feito a partir da extração de *tweets*, com informações específicas relacionadas à Covid-19 e às vacinas que estão sendo desenvolvidas contra a Covid-19, para classificá-los tanto por alguns métodos citados nos trabalhos anteriores como em outros.

3. Ferramentas

Esta seção apresenta uma breve introdução dos conceitos de análise de sentimento, assim como todas as ferramentas utilizadas para a elaboração do trabalho.

3.1 Twitter

O Twitter é um serviço que permite a comunicação de várias pessoas por meio de mensagens, ou *tweets*, que são as mensagens publicadas nos perfis dos usuários e enviadas para seus seguidores (outros usuários do serviço); muitas dessas mensagens podem conter fotos, vídeos, *links* e textos. Para publicações de mensagens no Twitter existe uma limitação de 280 caracteres por mensagem, e podem ser incluídos até 4 fotos, um GIF ou um vídeo (PERGUNTAS..., 2020).

Essa plataforma disponibiliza uma série de APIs (Interface de programas e aplicativos) para a comunicação da aplicação com o Twitter, facilitando a troca de informações para análises, aprendizado e interação com *tweet* (SOBRE..., 2020).

3.2 Google Colaboratory (Colab)

O Google Colaboratory (Colab), criação baseada no Jupyter Notebook, permite ao usuário escrever e executar na linguagem Python através do navegador. Para utilizá-lo não é necessária nenhuma configuração, além de ser gratuito, ter acesso a GPUs e a possibilidade de salvar e compartilhar as análises de maneira fácil. Foi criado para o incentivo à pesquisa de Inteligência Artificial (WELCOME..., 2020).

3.3 Linguagem Python

Python é uma linguagem de programação de alto nível, interpretada, que utiliza abordagens de programação orientada a objetos; o desenvolvimento de aplicações se torna mais fácil devido a sua sintaxe simples e tipagem dinâmica (THE PYTHON..., 2020). Foi criada por Guido van Rossum em 1991; além de ser uma linguagem multiplataforma, seu código é *open source*, ou seja, seus usuários podem estudar, modificar e fazer a distribuição do código, até mesmo para fins comerciais (ANDRADE, 2020). As bibliotecas de Python são componentes opcionais que podem ser incluídos nas distribuições; esses componentes contêm funções e exceções prontas, que podem ser usadas por meio do *import*; existem bibliotecas padrões que não têm a necessidade de ser incluídas (A BIBLIOTECA..., 2020).

3.4 DeepL

DeepL é um serviço *online* de tradução de texto, publicado em 2017; fazendo o uso de inteligência artificial e redes neurais para expandir as possibilidades das linguagens, promete superar as ofertas de concorrentes como a Google (DEEPL..., 2020).

3.5 Google Docs

Ferramenta de edição e estilo, gratuita da Google, para facilitar a formatação de textos. Essa ferramenta possibilita acesso, criação e edição dos documentos em várias plataformas, mesmo sem o acesso a uma rede, através de aplicativos de extensão instaladas diretamente do Google (GOOGLE..., 2020).

3.6 SenticNet 4

SenticNet4 é um método que consiste em um conjunto de ferramentas que utilizam técnicas de inteligência artificial e semântica, para análise de sentimento e mineração de opinião, com base em dicionários léxicos. A SenticNet4 é capaz de classificar um sentimento em cinco rótulos na análise de sentimento, *strong positive*, *positive*, *neutral*, *negative* e *strong negative*, com a utilização somente das polaridades, positiva $p > 0.1$ e negativa $p \leq 0$. A Figura 1 mostra um exemplo de como essa classificação é atribuída, para que a Sentic possa interpretar a frase “This car is very old but it is rather not expensive”. Primeiramente, a SenticNet pega a sentença; em seguida, procura pelos conceitos da frase (CAMBRIA et. al., 2016).

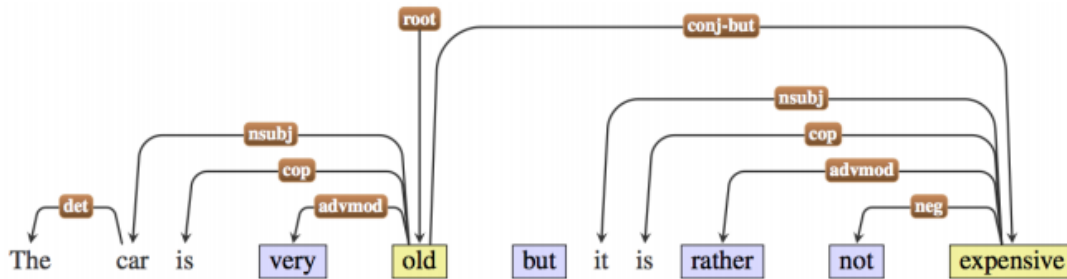


Figura 1. Análise da frase exemplo
Fonte: (CAMBRIA et. al., 2016)

No exemplo da frase, as únicas duas palavras que têm polaridade essencial são mostradas na cor amarela, as palavras que modificam o significado de outras palavras de maneira contextual são mostradas em azul. Após a remoção das palavras não utilizadas para cálculo de polaridade (em branco), um circuito com elementos semelhantes a complementos lógicos e amplificadores eletrônicos é obtido, adquirindo-se, assim, o sentimento da frase analisada, como mostrado na Figura 2 (CAMBRIA et. al., 2016).

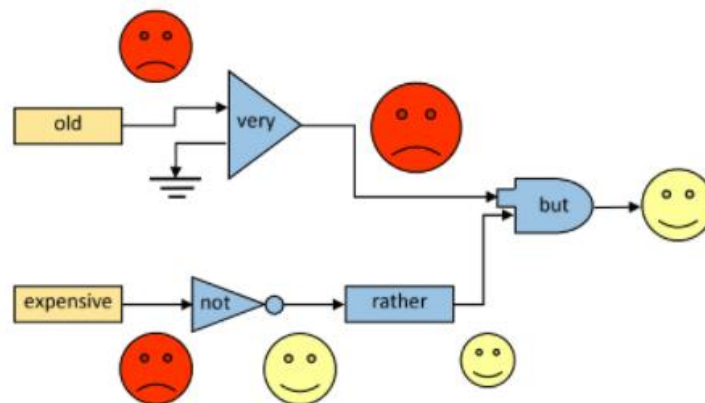


Figura 2. Circuito lógico da análise sentimental
Fonte: (CAMBRIA et. al., 2016)

3.7 SentiStreght

SentiStrength é um software que compara textos com um classificador de sentimentos baseado em dicionário léxico, também mais um método baseado em aprendizagem de máquina. Ele é capaz de determinar o sentimento classificando-o entre -5 a +5. Números positivos indicam contextos favoráveis, enquanto números negativos indicam contextos negativos. O programa oferece uma pontuação para cada palavra dentro de uma frase,

com isso é de extrema importância que os dados sejam limpos. Termos precedidos por palavras de reforço como “*very*” têm seu potencial sentimental positivo ou negativo aumentado, enquanto diminui bastante o potencial de sentimento da palavra seguinte. Essa aplicação também classifica a polaridade de *emoicons*, sendo capaz de processar grande número de dados em segundos; isso torna SentiStreght uma ferramenta muito útil para a análise de sentimentos em redes sociais (SENTISTREGHT..., 2010).

3.8 PySentiment 0.2

Uma biblioteca Python usada na área da análise de sentimento, utilizando-se a estrutura de dicionário. São fornecidos na biblioteca dois dicionários, Harvard IV-4 e McDonald Financial Sentiment Dictionaries, que são dicionários de sentimento para análise de sentimento geral e financeiro (PYSENTIMENT..., 2020).

Os dados são classificados em positivo e negativo, tendo também a polaridade e a subjetividade, inspirados no sistema Lydia. São usadas pelo método fórmulas para calcular a polaridade e para calcular a subjetividade:

$$\text{Polarity} = (\text{Pos}-\text{Neg}) / (\text{Pos}+\text{Neg}) \quad (1)$$

$$\text{Subjectivity} = (\text{Pos}+\text{Neg}) / \text{count} (*) \quad (2)$$

3.9 TextBlob

O TextBlob é uma biblioteca da linguagem Python que visa fornecer acesso às operações mais comuns de processamento de texto por meio de uma interface familiar. Os objetos do TextBlob são tratados como se fossem *strings* de Python com o diferencial de que agora possuem as técnicas de processamento de linguagem natural é uma biblioteca simples que suporta análises e operações complexas em dados textuais (TEXTBLOB..., 2020).

A análise de sentimento Textblob retorna polaridade e subjetividade. A Figura 3 mostra como é realizado o processo de classificação da subjetividade, que quantifica a quantidade de opiniões pessoais e informações factuais contidas no texto. A maior subjetividade significa que o texto contém opinião pessoal em vez de informações factuais. Tanto a polaridade quanto a subjetividade estão entre [-1,1], -1 define um sentimento negativo e 1 define um sentimento positivo.

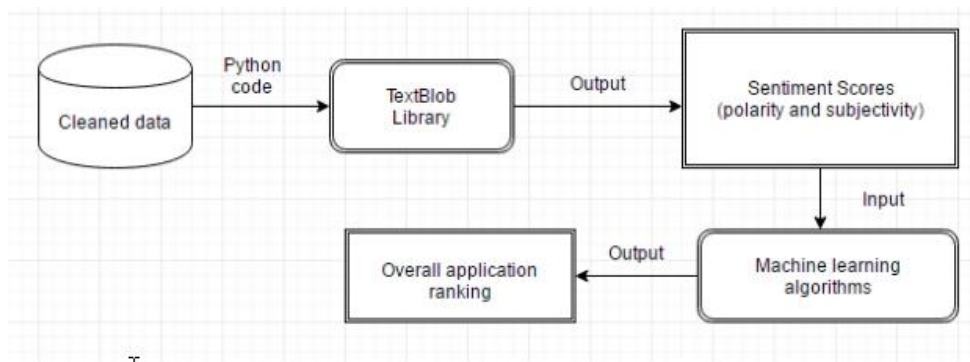


Figura 3. Processo de classificação de polaridade e subjetividade do TextBlob.
Fonte:(KUZMINYKH, 2020)

Por exemplo, para a frase “I do not like this example at all, it is too boring”, são

retornadas a polaridade -1 e a subjetividade 1. Porém, no exemplo “This was a helpful example but I would prefer another one”, Textblob irá retornar 0,0 para subjetividade e polaridade, o que com certeza não é a melhor resposta, visto que a frase contém uma opinião pessoal. Contudo, uma das melhores coisas sobre o TextBlob é que ele permite que o usuário escolha um algoritmo para a implementação das tarefas de NLP de alto nível, como o *PatternAnalyzer*, um classificador padrão construído na biblioteca de padrões, e o *NaiveBayesAnalyzer*, um modelo NLTK treinado em um *corpus* de resenhas de filmes (KUZMINYKH, 2020).

3.10 VaderSentiment

Vader (Valence Aware Dictionary Sentiment Reasoner) é mais um recurso utilizado para a mineração de opinião baseado em regras léxicas especialmente sintonizadas com os sentimentos expressos nas redes sociais. Além das redes sociais, é muito utilizado na análise de editoriais do NY Times, em resenhas de filmes e análises de produtos. Vader é capaz de classificar uma frase com quatro tipos de métrica: positiva, negativa, neutra e *compound*. Como mostra a Figura 4, a pontuação do *compound* é calculada a partir da soma das pontuações de valência de cada palavra no léxico; elas são ajustadas de acordo com as regras, em seguida normalizadas para ficarem entre -1 (negativo mais extremo) e +1 (positivo mais extremo). Esse rótulo se mostra útil se os dados desejados forem uma única medida unidimensional de sentimento para uma determinada frase; também se mostra adequada para pesquisadores que desejam definir limites padronizados para classificar sentenças como positivas, neutras ou negativas (HUTTO, 2020).

1. sentimento positivo : `compound pontuação >= 0,05`
2. sentimento neutro : (`compound pontuação > -0,05`) e (`compound pontuação < 0,05`)
3. sentimento negativo : `compound pontuação <= -0,05`

Figura 4. Exemplo de valores limite típicos, para a pontuação do *compound*.
Fonte:(HUTTO, 2020)

O método Vader se destaca em análise de redes sociais devido a sua adaptação de análise nos textos. Pontuação como (!), palavras em maiúsculo, intensificadores, conjunções, emojis e gírias afetam no resultado final da análise de sentimento. Por exemplo, na frase “VADER is smart, handsome, and funny”, a saída da análise é de 74% positivo, 25% neutro, 0% negativo e 83% para *compound*; porém, na mesma frase, adicionando-se letras maiúsculas, gírias, pontuação e emojis, “VADER is VERY SMART, uber handsome, and FRIGGIN FUNNY!!! 😊”, as métricas retornadas são de 70% positivo, 23% neutro, 0% negativo e 94% para *compound* (HUTTO, 2020).

4. Metodologia

4.1 Base de dados

Neste trabalho foram utilizados um conjunto de dados relacionados a Covid-19, criado pelos autores, e uma base de dados com os dados rotulados retirada da *Web*.

4.2 Criação do conjunto de dados

Para a criação do conjunto de dados foi desenvolvida uma aplicação utilizando a linguagem Python, por meio da biblioteca TwitterSearch, que é um conjunto de

ferramentas para coleta de dados que não implementa toda a API do Twitter, fazendo o uso somente da API de pesquisa que busca mensagens na rede social através de palavras-chaves e da API de linha de tempo do usuário. Para realizar a coleta dos dados referente a Covid-19 e a vacina, as palavras-chave colocadas na busca foram “covid” e “vacina”, sendo, inicialmente, o maior número de dados sobre as vacinas da China, de Oxford e da Rússia e, posteriormente, as vacinas da Pfizer e da BioNTech. Continuando com o método de busca, também foi especificado o idioma que os dados deveriam conter; nesse caso, “português”. À localidade dos dados, foi atribuído todo o território brasileiro. A aplicação foi executada semanalmente, através do Colaboratory, com a busca de dados e armazenamento deles em um arquivo de texto, organizado com os dados do usuário, a mensagem e a data de publicação do *tweet*.

Antes e após o armazenamento, foram criados alguns critérios de avaliação do texto. Primeiramente, antes de armazenar o dado, a aplicação checava se, no início das mensagens, havia um ‘RT’, que significa *retweet*, palavra usada pelo Twitter para mostrar que a publicação já havia sido compartilhada por outro usuário; em seguida, ela checava o nome do usuário, e então a comparava com um *array* de nomes de usuários do Twitter, que foram observados pelos autores como usuários que transmitiam notícias, como estadão, jornaldarecord, GlobalPandemics; então essas mensagens eram descartadas. Após os dados terem sido armazenados em arquivos de texto, eles eram agrupados em um único arquivo com a relação ao mês de sua coleta. A tabela abaixo mostra um exemplo dos dados recolhidos.

Tabela 1. Exemplos de tweets coletados

<i>Tweet</i>
sonhei que a vacina da covid ja tinha saído... https://t.co/P3OwVJDFpv
migos compartilhem cmg aí a vacina da covid na moral quero ir p rolês com voces poxa
Não vou tomar, e quero ver quem vai me obrigar. Se a vacina do Dória fosse boa e funcionasse ele não teria pego
Tantos meses procurando a vacina pro covid e a cura era as eleições
Quem tomar essa vacina vái morrer mais depressa do que com covid 🤔🤔
CONTRA A OBRIGATORIEDADE DA VACINA CHINESA

Fonte:(Autores, 2020)

Após a coleta viu-se necessária a tradução dos dados para o inglês, visto que, dos cinco métodos de mineração apresentados, somente dois possuíam compatibilidade para a língua portuguesa. Para a tradução dos dados foi usada a biblioteca Googletrans, por meio da API do Google desenvolvida para Python que usa a API Ajax do Google Translate para a chamada de métodos como a detecção e a tradução de dados; além dele foi feito o uso do tradutor de documento DeepL, para analisar os textos. Após a conversão dos arquivos de português para inglês, uma comparação foi realizada de ambos os arquivos traduzidos, por meio do método de comparação de arquivos do Google Docs; foram mantidos apenas os dados com a melhor tradução. A Tabela 2 mostra a diferença das frases traduzidas pelo DeepL e pelo GoogleTrans, e a Figura 6 mostra como era feita a correção dessas frases traduzidas por meio do Google Docs.

4.3 Base Rotulada

A base de dados rotulados contém um conjunto de dados em inglês para a detecção de sentimentos, extraídos por meio da API do Twitter, e rotulados, pela Sentiment140, em positivo, negativo e neutro; ao todo são 1.600,00 *tweets*. O conjunto é organizado pelo “*target*”, que é a polaridade do *tweet*, sendo 0 para negativo, 2 para neutro e 4 para positivo, “*Id*”, o id do *tweet*, “*date*”, data de publicação do *tweet*, “*flag*”, que sinaliza se o *tweet* já foi consultado (caso já tenha sido consultado mostrará “*lyx*”; se não tenha sido mostrará “*NO_QUERY*”), “*user*”, nome do usuário que compartilhou o *tweet*, e “*text*” o texto do *tweet* (GO; BHAYANI; HUANG, 2009).

Para que fosse possível avaliar a *acurácia* e *precisão* dos métodos, foi utilizada a base de dados rotulada. Como o banco continha grande quantidade de dados rotulados em negativo, positivo e neutro, e apenas alguns métodos classificavam em neutro, foram usados apenas 10.000 *tweets* rotulados como positivo e 10.000 *tweets* classificados como negativo, reorganizados pela “*classificação*” como NEG para negativo ou POS para positivo.

5. Resultados

5.1 Análise da taxa de acurácia dos métodos

A avaliação de desempenho dos métodos foi feita a partir dos critérios de avaliação usados no trabalho de Araújo, Gonçalves e Benevenuto (2013).

$$P = a / (a + b) \quad (3)$$

$$A = (a+d) / (a+b+c+d) \quad (4)$$

Onde, A: acurácia de cada método e P: a precisão entre os métodos; a: quantidade de mensagens positivas classificadas como positivas; b: quantidade de mensagens negativas classificadas como positivas; c: quantidade de mensagens positivas classificadas como negativas; d: quantidade de mensagens negativas classificadas como negativas.

Tabela 3. Resultado dos cálculos da acurácia e da precisão de cada método

<i>Métodos</i>	<i>Acurácia</i>	<i>Precisão</i>
VaderSentiment	64.75%	59.65%
TextBlob	61.07%	53.81%
SentiStreght	60.35%	40.86%
SenticNet	55.74%	56.15%
PySentiment2	53.05%	13.03%

Fonte: (Autores, 2020)

A Tabela 3 mostra os resultados obtidos utilizando-se as fórmulas de acurácia e precisão. Para avaliar a acurácia foi usado o banco de dados rotulados, em que foram desenvolvidas duas aplicações, uma para verificar a quantidade de *tweets* positivos que cada método classificou como “positivo” na comparação com os dados classificados do banco e outra para verificar a quantidade de *tweets* negativos.

5.2 Análise resultados da vacina

As Figuras 7,8,9 mostram os resultados em gráficos sobre a classificação do conjunto de

dados da Covid-19, em positivo e negativo, utilizando-se os diferentes métodos de análise de sentimentos apresentados anteriormente.

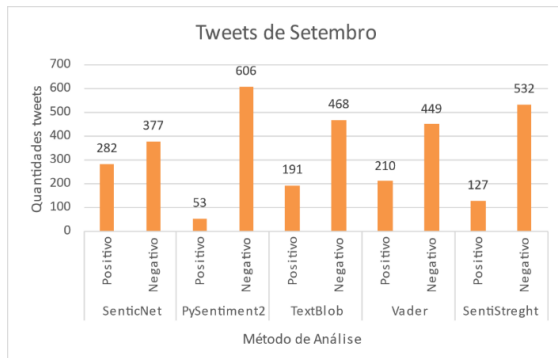


Figura 7. Resultados mês de setembro
Fonte:(Autores, 2020)

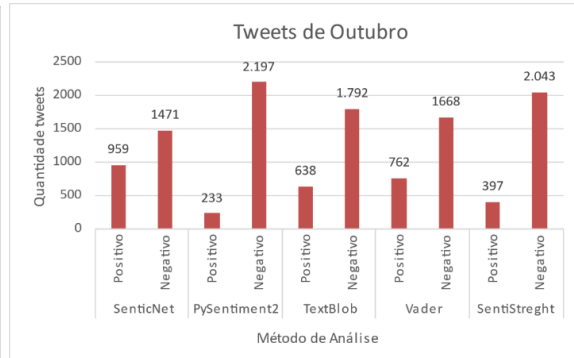


Figura 8. Resultados mês de outubro
Fonte:(Autores, 2020)

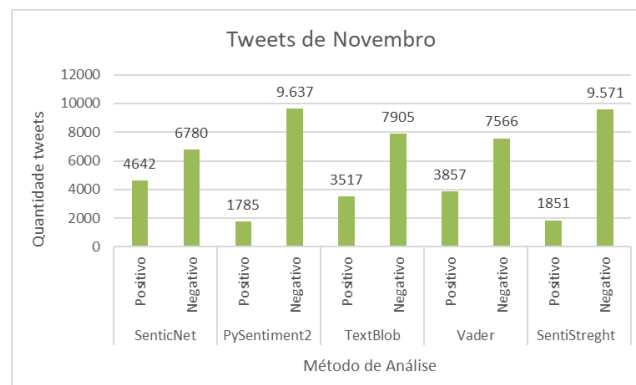


Figura 9. Resultados mês de novembro
Fonte:(Autores, 2020)

Após ser realizada uma breve pesquisa das palavras mais comentadas de cada mês, foi observado que a maior parte das palavras nos *tweets* tinha relação com as notícias informadas no determinado mês em que foram feitas a coleta e a classificação. Em setembro, após a classificação dos *tweets*, as palavras mais comentadas foram “revolta”, “anti-vacina”, “população”, “morte”, tendo relação com a discussão que ocorreu no Twitter mencionando uma segunda “Revolta da vacina em 2020”. Em seguida, as palavras mais recorrentes de outubro foram “contra”, “obrigatoriedade”, “Bolsonaro”, “Doria”, “China”, “vacina chinesa”, notando-se que os maiores comentários são em relação às notícias do movimento antivacina promovido pelo presidente Jair Bolsonaro e da compra que o governador de São Paulo João Doria fez da vacina desenvolvida pela China. Por fim, as palavras mais comentadas em novembro foram, de forma disparada, “Pfizer”, “BioNTech”, mostrando a relação dos dados com a notícia da eficácia da vacina anunciada pelos laboratórios.

6. Conclusão

Após uma comparação dos *tweets*, foi notado que, apesar de os resultados serem diferentes, todos os métodos mostraram que a polaridade em relação à vacina da Covid é, em sua maior parte, negativa. É notório que as ferramentas Vader e TextBlob tiveram a maior acurácia com a base de dados rotulada e na análise dos *tweets* sobre a vacina da Covid; seus valores se mostraram bem próximos, devido a ambas calcularem não somente

a polaridade como também a subjetividade, tendo maior facilidade na detecção de dados com opinião verdadeira.

Foram apresentados cinco métodos de classificação de, e a partir de, resultados obtidos através da análise da taxa de acurácia e precisão e da vacina. Concluímos que o melhor método utilizado neste trabalho para a classificação do conjunto de dados da Covid-19 foi o método Vader, com 64,75% de taxa de acurácia em relação à polaridade dos *tweets* da base rotulada. Em relação aos dados de teste, o método classificou que 33,28% dos *tweets* eram positivos a respeito das vacinas em desenvolvimento, e 66,73% dos *tweets* eram negativos.

Referências

- A BIBLIOTECA, Padrão do Python: Introdução. **Python**. Disponível em: <<https://docs.python.org/pt-br/3/library/intro.html>>. Acesso em: 23 nov. 2020.
- ANDRADE. A. P. O que é Python?. **Treinaweb**. Disponível em: <<https://www.treinaweb.com.br/blog/o-que-e-python/>>. Acesso em: 23 nov. 2020.
- ARAÚJO, M.; GONÇALVES, P.; BENEVENUTO, F. O Métodos para Análise de Sentimentos no Twitter. In **Proceedings** of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13), 2013.
- CAMBRIA, Erik et al. SenticNet 4: A Semantic Resource for Sentiment Analysis Based on Conceptual Primitives. In **Proceedings** of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 2666–2677, Osaka, Japan, 2016.
- DATA, never sleep 8.0. **Domo**. Disponível em: <<https://www.domo.com/learn/data-never-sleeps-8>>. Acesso em: 14 nov. 2020.
- DEEPL, Pro - Perguntas frequentes: que é o DeepL?. **DeepL**. Disponível em: <<https://www.deepl.com/pro-faq.html/>>. Acesso em: 23 nov. 2020.
- GOOGLE, Documentos: crie documentos impactantes. **Google**. Disponível em: <<https://www.google.com/intl/pt-BR/docs/about/>>. Acesso em: 23 nov. 2020.
- GO, A.; BHAYANI, R.; HUANG, L. Twitter sentiment classification using distant supervision. **CS224N Project Report**, Stanford, p.12, 2009. Disponível em: <<https://www.kaggle.com/kazanova/sentiment140>>. Acesso em: 23 nov. 2020.
- HUTTO, C.J. VADER-Sentiment-Analysis. **GitHub**. Disponível em: <<https://github.com/cjhutto/vaderSentiment>>. Acesso em: 10 nov. 2020.
- LIMA, G. A. P. **Análise de sentimentos no facebook utilizando o Senticnet. 2013**. Tese (Bacharel em Sistemas de Informação) – Curso de Bacharelado em Sistemas de Informação do Centro de Ciências Aplicadas e Educação (CCAEE), Universidade Federal da Paraíba, Paraíba.
- ONLINE, V. Mesmo com aumento de usuários, Twitter tem prejuízo de US\$ 1,23 bilhão no 2º trimestre. **G1**. Disponível em: <<https://g1.globo.com/economia/tecnologia/noticia/2020/07/23/mesmo-com-aumento-de-usuarios-twitter-tem-prejuizo-de-us-123-bilhao-no-2deg-trimestre.ghtml>>. Acesso em: 17 nov. 2020.
- PERGUNTAS, frequentes de novos usuários. **Central de Ajuda Twitter**. Disponível em:

<<https://help.twitter.com/pt/new-user-faq>>. Acesso em: 23 nov. 2020.

PYSENTIMENT, 0.2. **The Python Package Index (PyPI)**. Disponível em: <<https://pypi.org/project/pysentiment/>>. Acesso em: 5 out. 2020.

KUZMINYKH, Natalia. Sentiment Analysis in Python with TextBlob. **Stack Abuse**. Disponível em: <<https://stackabuse.com/sentiment-analysis-in-python-with-textblob/>> Acesso em 17 nov. 2020

SENTISTRENGTH, about. **SentiStrength**. Disponível em: <<http://sentistrength.wlv.ac.uk/>>. Acesso em: 01 out. 2020.

SOBRE, as APIs do Twitter. **Central de Ajuda Twitter**. Disponível em: <<https://help.twitter.com/pt/rules-and-policies/twitter-api>>. Acesso em: 17 nov. 2020.

TEXTBLOB, Simplified Text Processing. **TextBlob**. Disponível em: <<https://textblob.readthedocs.io/en/dev/>>. Acesso em: 17 nov. 2020.

THE PYTHON, Tutorial. **Python**. Disponível em: <https://docs.python.org/pt-br/3/tutorial/>. Acesso em: 17 nov. 2020.

TWITTER, Usage Statistics. **Internet Live Stats**. Disponível em: <<https://www.internetlivestats.com/twitter-statistics>>. Acesso em: 14 nov. 2020

WELCOME, To Colaboratory. **Colab Research**. Disponível em: <<https://colab.research.google.com/notebooks>>. Acesso em: 17 nov. 2020.