



Guilherme Lapa de Araújo Cadengue

Análise de Sentimentos dos Tweets Relacionados ao Superior Tribunal Federal no Ano de 2019

Recife

2022

Guilherme Lapa de Araújo Cadengue

Análise de Sentimentos dos Tweets Relacionados ao Superior Tribunal Federal no Ano de 2019

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE
Departamento de Estatística e Informática
Curso de Bacharelado em Sistemas de Informação

Orientador: Ermeson Andrade
Coorientador: Silvana Bocanegra

Recife
2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

C122a Cadengue, Guilherme Lapa de Araújo
Análise de sentimentos dos tweets relacionados ao Superior Tribunal Federal no ano de 2019 /
Guilherme Lapa de Araújo Cadengue. - 2022.
43 f. : il.

Orientador: Ermeson Carneiro de Andrade.
Coorientadora: Silvana Bocanegra.
Inclui referências.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Sistemas da Informação, Recife, 2022.

1. Análise de sentimentos. 2. Twitter. 3. STF. 4. Superior Tribunal Federal. 5. Redes Sociais. I. Andrade,
Ermeson Carneiro de, orient. II. Bocanegra, Silvana, coorient. III. Título

CDD 004

Guilherme Lapa de Araújo Cadengue

Análise de Sentimentos dos Tweets Relacionados ao Superior Tribunal Federal no Ano de 2019

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em: 10 de Novembro de 2022.

BANCA EXAMINADORA

Ermeson Andrade
Departamento de Computação-DC
Universidade Federal Rural de Pernambuco

Silvana Bocanegra
Departamento de Estatística e Informática-DEINFO
Universidade Federal Rural de Pernambuco

Dedico este trabalho primeiramente a Deus, por ele ser sempre meu apoio nos momentos difíceis; e à minha família e aos meus mestres por toda a ajuda e incentivo.

“Tendo tudo contra si, que ele preserve a si mesmo e que isso lhe baste.”
(A.D. Sertillanges)

Resumo

As redes sociais desde suas origens afetaram todos os usuários da Internet. Redes sociais, como o Twitter fornecem um novo modo de comunicação, interação e principalmente uma forma de expressar opiniões sobre os diversos eventos da vida em sociedade, conseqüentemente possibilitam geração de conteúdo. Conhecer as opiniões dos brasileiros sobre instituições públicas é muito importante para o engajamento das pessoas na sociedade, como forma de agentes participantes nas decisões que afetam todos os indivíduos, ou seja, é uma forma de inclusão social. A aplicação da Análise de sentimentos é realizada em diversas áreas para extrair o teor da opinião pública. Este trabalho visa identificar os sentimentos da população brasileira sobre o Superior Tribunal Federal do Brasil através dos conteúdos de tweets publicados entre Janeiro e Dezembro de 2019. Para isso foram coletados os tweets no período, os quais foram pré-processados, classificados e analisados. Os resultados mostram opiniões bastante polarizadas, mas que predominam, de forma geral, opiniões negativas em relação ao STF (estimativa em 51,7%).

Palavras-chave: Análise de sentimentos, Twitter, STF, Superior Tribunal Federal, Redes Sociais.

Abstract

The Social media since its inception has affected all Internet users. Networks such as Twitter provide a new form of communication, interaction and, above all, a way of expressing opinions about the different events of life in society, consequently enabling the generation of content. Knowing the opinions of Brazilians about public institutions is very important for engaging people in society, as agents participating in decisions that affect all individuals, that is, it is a form of social inclusion. The application of Sentiment Analysis is carried out in several areas in order to extract the content of public opinion. The objective of this work is to identify the feelings of the Brazilian population about the Superior Federal Court of Brazil through the content of published tweets between January and December 2019. For this, the tweets in the period were collected, which were pre-processed, classified and then analyzed. The results show highly polarized opinions, but generally negative opinions regarding the STF are predominant (estimate at 51.7%).

Keywords: Sentiment analysis, Twitter, STF, Superior Tribunal Federal, Social Media.

Lista de ilustrações

Figura 1 – Processo da Etapa de Coleta.	21
Figura 2 – Exemplo da palavra "stf" em fonte personalizada.	25
Figura 3 – Nuvem de palavras inicial.	27
Figura 4 – Evidenciando os Outliers.	29
Figura 5 – Gráfico de Dispersão sem os outliers.	29
Figura 6 – Nuvem de Palavras sem os outliers.	30
Figura 7 – Gráfico de Barras com as 10 Palavras mais Frequentes sem os outliers.	31
Figura 8 – Quantidade de tweets por mês no ano de 2019.	31
Figura 9 – Quantidade de palavras positivas e negativas do total dos tweets.	32
Figura 10 – Quantidade de tweets positivos e negativos por mês em 2019.	33

Lista de tabelas

Tabela 1 – Diferentes formas de ocorrência da palavra STF.	25
Tabela 2 – Detalhamento dos registros removidos.	26
Tabela 3 – Tabela de frequência inicial.	28
Tabela 4 – Exemplos de Tweets Positivos.	32
Tabela 5 – Exemplos de Tweets Negativos.	32
Tabela 6 – Resumo dos dados por polaridade em 2019.	33
Tabela 7 – Evidenciando Tweets de cunho irônico e sua classificação.	34
Tabela 8 – Evidenciando Tweets com Crítica Indireta e sua classificação.	34
Tabela 9 – Análise de palavras-chave de temas polêmicos.	34
Tabela 10 – Análise de palavras relevantes com alta frequência.	35
Tabela 11 – Análise de sentimentos relacionado aos ministros do STF.	35

Sumário

Lista de ilustrações	6
1 INTRODUÇÃO	9
1.1 Contexto	9
1.2 Justificativa	10
1.3 Objetivos	11
1.4 Estrutura Monografia	11
2 TRABALHOS RELACIONADOS	12
3 FUNDAMENTAÇÃO TEÓRICA	14
3.1 Twitter	14
3.2 STF	15
3.3 Mineração de Texto	16
3.4 Análise de Sentimentos	17
3.5 Análise Léxica	17
4 METODOLOGIA	19
4.1 Coleta de Dados	19
4.2 Pré-processamento	21
4.3 Classificação	23
4.4 Análise	24
5 RESULTADOS	25
6 CONCLUSÕES	36
6.1 Trabalhos Futuros	37
REFERÊNCIAS	38

1 Introdução

1.1 Contexto

As redes sociais tiveram seu início na década de 1990, com o objetivo de estabelecer contato e aproximar as pessoas de acordo com seus valores e interesses através da Internet. De forma mais abrangente, as redes sociais representam “*um conjunto de participantes autônomos, unindo ideias e recursos em torno de valores e interesses compartilhados*” (MARTELETO, 2001). Com a disseminação da Internet a partir dos anos 2000, devido à facilidade de acesso, foram surgindo novas redes sociais, como Orkut, Facebook, Youtube, MySpace e LinkedIn, as quais diferem entre si, mas definitivamente modificaram a forma de comunicação e interação das pessoas.

A interação entre os usuários das redes sociais varia conforme o propósito e o tipo da rede social. Porém, de forma geral, se dá através de postagens, curtidas, compartilhamentos e comentários sobre postagens ativas. Consequentemente, as redes sociais possibilitam também aos seus usuários a geração de conteúdo e a formação de opinião de um determinado público. Direcionada principalmente para a emissão de opinião, em 2006 surge a rede social **Twitter**, onde cada conta de usuário seria como um microblog com um serviço de mensagens que se limitariam à 140 caracteres, ou seja, a interação se dava com a geração de conteúdo e opinião apenas em textos curtos (PATEL, 2019).

No Brasil, no ano de 2013 os eventos que ficaram conhecidos como *Jornadas de Junho* foram marcantes e decisivos tanto no que se refere à política quanto na influência das redes sociais como instrumento de interação social. As *Jornadas de Junho* foram a consequência do acúmulo da insatisfação de diversos setores e camadas sociais com relação, principalmente, ao aumento das tarifas de transporte, ao aumento da violência urbana e os altos custos da Copa do Mundo de 2014 (PACELLI, 2013). Entretanto, o estopim para o início do movimento foi o aumento das tarifas do transporte público em grandes capitais do país, como São Paulo, Rio de Janeiro, Maceió, Salvador, Belém, Curitiba e Recife que reuniram milhares de manifestantes com as ações programadas através das redes sociais (G1, 2013). O Twitter, por sua vez, contabilizou cerca de 11 milhões de tweets com a palavra “Brasil” e 2 milhões mencionando “protesto” entre os dias 6 e 26 de junho (COSTA, 2013).

O movimento *Occupy Wall Street* iniciado em 16 de setembro de 2011, também teve sua origem nas redes sociais de forma descentralizada e sem liderança. Esse movimento conseguiu mobilizar milhares de manifestantes não só dos EUA, mas tam-

bém do mundo através de usuários, principalmente, do Facebook e do Twitter (BERKOWITZ, 2011). Eventos como estes geram um grande volume de dados através dos comentários na rede. A análise de sentimentos pode ser usada como ferramenta de extração de informação, para reconhecer emoções e avaliar o que realmente está sendo expresso, se possui uma conotação positiva ou negativa.

1.2 Justificativa

A história política recente do Brasil mostra-se cada vez mais surpreendente, de uma "guinada à esquerda" no ano de 2016, passando novamente por uma destituição de um presidente, até a prisão de um ex-presidente. Associado a esses eventos tem-se a não menos importante atuação **Supremo Tribunal Federal(STF)**, que é uma instituição do poder judiciário que representa a mais alta corte do Brasil.

Nesse cenário político conturbado, o Supremo Tribunal Federal tem um grande papel a desempenhar como Guardião da Constituição Federal (BRASIL, 2019). Entretanto, decisões contraditórias bem como decisões referentes a assuntos polêmicos (ex.: legalização do aborto) vem pondo em dúvida a isenção política e o desempenho do STF e tem levado a população brasileira a questionar suas decisões.

Pode-se citar como exemplo, a destituição da presidente Dilma Rousseff, sob acusação de crime de responsabilidade fiscal. Durante o julgamento no senado federal, em 31 de Agosto de 2016, o então presidente do Superior Tribunal Federal, o ministro Ricardo Lewandowski, deferiu o fracionamento da votação, estabelecendo uma votação para a perda do cargo da presidência da república e outra para julgar a inabilitação para assumir cargos públicos pelo período de 8 anos (G1, 2016). Tal decisão por fracionamento da votação contradiz a Constituição Federal, em seu artigo 52 (BRASIL, 2019), parágrafo único. Este exemplo é apenas um recorte de inúmeras decisões contraditórias do STF e têm sido causa de muita insatisfação por parte da população e de inúmeras discussões, tanto nos meios de comunicação tradicional quanto em mídias alternativas, como Youtube, Facebook e Twitter. Com o crescimento do engajamento político da população que vêm se posicionando e demonstrando suas opiniões através das redes sociais, é importante saber os sentimentos da população com relação ao STF.

A Internet alcançou 49% da população mundial no ano de 2018 e em 2019 esse número saltou para 54% (THE WORLD BANK GROUP, 2021) (INTERNATIONAL TELECOMMUNICATION UNION (ITU), 2021). No Brasil cerca de 73,91% da população já utilizava a Internet, o que corresponde a aproximadamente 156,53 milhões de pessoas (OUR WORLD IN DATA, 2020a). O Twitter contava com 330 milhões de usuários em 2019 (OUR WORLD IN DATA, 2020b), segundo (DOMO, 2020) em seu infográfico, os

usuários do Twitter enviaram 511.200 tweets por minuto, sendo 8,28 milhões usuários brasileiros (TI INSIDE, 2019).

O uso da análise de sentimentos considerando o STF pode ser uma ferramenta de importância para avaliar o momento político e analisar a opinião da população com relação as decisões que foram tomadas por essa instituição, devido as grandes repercussões geradas.

Este estudo visa aplicar mineração de texto e análise de sentimentos em uma base de dados de tweets sobre o STF, e então descobrir a opinião predominante dos usuários sobre essa instituição e seus ministros.

1.3 Objetivos

O objetivo deste trabalho é identificar os sentimentos da população brasileira sobre o Superior Tribunal Federal do Brasil através dos conteúdos de tweets entre Janeiro e Dezembro de 2019. Esse período foi escolhido por se tratar de uma histórica mudança ideológica no poder executivo federal, bem como por apresentar propostas de governo polêmicas em um contexto político/social bastante polarizado.

Este trabalho tem os seguintes objetivos específicos:

- Gerar uma base de dados com os tweets relacionados ao STF entre Janeiro e Dezembro de 2019.
- Realizar tratamento dos dados coletados.
- Sumarizar os sentimentos do brasileiro relativos aos STF.
- Investigar se a opinião dos brasileiros no Twitter tem relação com os acontecimentos sociopolíticos do período analisado.

1.4 Estrutura Monografia

No Capítulo 2, são apresentados os trabalhos relacionados.

No Capítulo 3, são introduzidos os conceitos fundamentais a serem utilizados na monografia, tais como: Twitter, STF e Mineração de Texto.

No Capítulo 4, é descrito a metodologia adotada para o desenvolvimento do trabalho.

No Capítulo 5, são apresentados os resultados obtidos.

Finalmente, no Capítulo 6, são apresentadas as conclusões e possibilidades de trabalhos futuros relacionados.

2 Trabalhos Relacionados

Nesta seção serão descritos alguns trabalhos que utilizaram dados da rede social Twitter para tentar compreender a opinião a respeito de alguém ou de uma instituição, analisar o engajamento dos usuários através de suas opiniões durante e após determinado evento e até tentar prever o resultado de algum evento tomando como base apenas o texto escrito pelos usuários dessa rede social.

O trabalho realizado por (SILVA et al., 2021) buscou identificar o sentimento da população brasileira a respeito do sistema SUS (Sistema Único de Saúde) no contexto da pandemia de COVID-19. Utilizando a ferramenta Twint realizou-se a extração de dados através da Twitter API no período de dezembro de 2019 até Outubro de 2020, obtendo uma base de dados de 25 mil Tweets. Para o processo de Análise de sentimentos foi utilizada a biblioteca Syuzhet (JOCKERS, 2020). A classificação dos Tweets foi limitada aos sentimentos Raiva, Nojo, Medo, Alegria, Tristeza e Confiança os quais enquadram-se melhor ao contexto do trabalho. Para a realização da análise foram estabelecidos períodos, para cada qual foi gerado uma nuvem de palavras utilizando-se a biblioteca TM (FEINERER, 2020b) em conjunto com a Wordcloud (FELLOWS, 2018). Logrou identificar que no período anterior a pandemia, de Dezembro de 2019 até Fevereiro de 2020, predominavam os sentimentos positivos. A partir de Março de 2020, quando efetivamente iniciou a pandemia no Brasil, os sentimentos negativos ultrapassaram os positivos destacando-se o sentimento de Medo. Entre Maio e o fim do mês de Agosto, os sentimentos positivos voltaram a sobrepor os negativos e no mês de Outubro houve o pico máximo de sentimentos positivos. Conclui que o fator principal que levou à ocorrência de sentimentos negativos estavam associados ao medo de uma possível sobrecarga no sistema público de saúde e não a uma insatisfação com o serviço prestado à população.

Com o objetivo de analisar as opiniões dos eleitores com relação aos candidatos nas eleições brasileiras em 2018 (MATOS; MAGALHÃES; SOUZA, 2020) foi utilizada a ferramenta Orange Canvas (BIOINFORMATICS LABORATORY, 2020). Foram coletados 1000 tweets sobre cada candidato em 6 dias diferentes no contexto do 2º turno das eleições. O processamento dos dados coletados foi feito usando as técnicas de geração de corpus, stemming, remoção de stopwords e a separação dos dados por relevância utilizando tabelas de frequência (para diminuição de ruído), e a análise de sentimentos através de um léxico. Por meio da geração de nuvem de palavras e gráficos de distribuição de frequência constatou-se que a intensidade dos sentimentos para os dois candidatos (Jair Bolsonaro e Fernando Haddad) foram semelhantes, não demonstrando uma posição positiva ou negativa, apesar de os sentimentos de

surpresa, medo e alegria destacarem-se. Muitos usuários demonstraram neutralidade nas opiniões sobre ambos candidatos. Dessa forma, os estudos não tiveram êxito em determinar qual candidato tinha a preferência dos usuários, não sendo possível realizar a previsão dos resultados.

Em (CONOVER et al., 2013), os autores apresentam um estudo sobre a evolução do movimento Occupy Wall Street, onde um grande conjunto de dados do Twitter foi usado para analisar o engajamento, os interesses e a conectividade social dos participantes ao longo de um período de quinze meses. Através da API de Streaming do Twitter, os autores utilizaram termos como #ows e #occupy para estabelecer uma medida do tráfego total na plataforma Twitter, de Setembro de 2011 a Setembro de 2012. Foram utilizados aproximadamente 1,85 milhão de tweets produzidos por 447.241 postagens distintas. Como resultado, foi observado uma queda relativa ao pico na atividade do movimento ao final de 2011, mesmo havendo nessa época tentativas de reviver o movimento. Verificou-se que o tráfego relativo ao movimento em geral foi baixo, caracterizado por picos no início dos protestos com uma significativa diminuição nos 11 meses seguintes, onde no período de 17 de setembro de 2011 a 31 de agosto de 2012 houve uma diminuição de 80,8%.

Nos trabalhos citados acima, é possível perceber dificuldades comuns enfrentadas durante os seus desenvolvimentos, como obter uma massa de dados com volume que traga relevância para a análise, o processamento nos dados brutos visando reduzir o ruído, a identificação de outliers¹ e até mesmo situações em que os dados em formato de texto trazem interpretação dúbia, o que pode inverter completamente o sentido da opinião expressada durante a classificação.

Nenhum dos trabalhos aborda o tema a respeito do STF, cujas decisões ao longo dos anos vêm afetando toda a população, nem procuraram realizar a análise de sentimentos a partir da montagem de um léxico próprio em português reaproveitando as funcionalidades de ferramentas já existentes. Com a exceção de (CONOVER et al., 2013) nenhum outro trabalho apresenta uma base de dados com volume consistente com o tamanho da que se propõe analisar, ou seja, utilizam um volume pequeno comparado ao que foi adquirido nesse trabalho, mais de 250 mil tweets.

¹ "Um outlier é qualquer valor que seja muito distante de outros valores em um conjunto de dados."(WICKHAM; GROLEMUND, 2019)

3 Fundamentação Teórica

Neste capítulo serão apresentados os conceitos básicos para o desenvolvimento do trabalho.

3.1 Twitter

Segundo (TWITTER, 2020a), o Twitter é "o lugar certo para saber mais sobre o que está acontecendo no mundo e sobre o que as pessoas estão falando agora.". A rede social foi criada em 15 de julho de 2006 por Jack Dorsey, Evan Williams, Biz Stone e Noah Glass, sendo Williams e Stone ex-funcionários do Google. Inicialmente o Twitter foi concebido para ser um microblog e os usuários incentivados a interação entre eles através da troca de mensagens de texto curtas (chamados de tweets). No início, os tweets limitavam-se à apenas postagens de 140 caracteres, o que forçava os usuários a serem bastante concisos. Com o passar do tempo, a rede social foi amadurecendo e em 2017 o limite das mensagens foi aumentado para 280 e atualmente é possível publicar fotos e vídeos nas postagens.

Ao longo dos anos, o Twitter veio se estabelecendo na Internet como um ambiente ideal para emissão de opinião. O seu potencial de influência política começou a ser revelado em 2011, nos eventos conhecidos como Primavera Árabe e no mesmo ano com Occupy Wall Street, nos Estados Unidos (GRUPO NZN, 2018). Foi no Twitter que surgiu uma funcionalidade que revolucionaria a forma de criar e divulgar conteúdo em redes sociais. Em 23 de Agosto de 2007, os usuários passaram a poder utilizar o conceito de hashtag, representada pelo símbolo cerquilha #. Através desse novo conceito, foi possível criar um hiperlink com as palavras após o símbolo, a qual direciona para todo o conteúdo que foi marcado com aquela mesma hashtag, ou seja, criando uma nova forma de organizar, contextualizar e até divulgar as publicações na rede social. Com o sucesso das hashtags, outras redes sociais, como Instagram e Facebook também adotaram tal funcionalidade. Outra funcionalidade de bastante sucesso, são os Trending Topics listando as 10 hashtags mais utilizadas no dia (PATEL, 2018). Essa funcionalidade foi criada em 2010 pelo Twitter.

O Twitter também disponibiliza como forma de segurança e de credibilidade, a verificação de contas. Tal funcionalidade é usada para contas que eram consideradas de interesse público, por exemplo, contas de usuários nas áreas de música, teatro/cinema/TV, moda, governo, política, religião, jornalismo, mídia, esportes e negócios (TWITTER, 2020d). Note que ela é usada como uma forma pública de autenticar a identidade daquela conta de interesse público, mostrando para os outros usuários

que a conta em questão realmente representa aquela pessoa. O Twitter identifica essa funcionalidade através de um selo de verificação azul que passa a ser exibido ao lado do nome no perfil. A rede social estipula uma série de critérios que os usuários devem seguir. O descumprimento resulta na perda do selo, por exemplo:

- Promover o ódio ou a violência, atacar diretamente ou ameaçar outras pessoas com base em raça, etnia, nacionalidade, religião, orientação sexual, sexo, identidade de gênero, idade ou deficiência. Apoiar organizações ou indivíduos que promovam o acima mencionado.
- Incitar ou se envolver em assédio a outras pessoas.
- Enganar pessoas intencionalmente no Twitter ao alterar o próprio nome de exibição ou bio (TWITTER, 2020b).

No primeiro trimestre de 2019, a empresa Twitter registrou 330 milhões de usuários ativos mensalmente (PATEL, 2019).

3.2 STF

O Supremo Tribunal Federal, STF, tem suas origens no antigo Superior Tribunal de Justiça conforme determinado na antiga constituição de 1824 em seu artigo 163: "composto de Juízes letrados, tirados das Relações por suas antiguidades" (BRASIL, 2012). Com a promulgação de uma nova constituição em 1891, passou a ser denominado de Supremo Tribunal Federal, sendo instalado em 28 de fevereiro do mesmo ano e desde então, estabeleceu-se como responsável pelo controle da constitucionalidade das leis (BRASIL, 1891).

A nova e vigente constituição de 1988, dedicou seus artigos 101 até 103 ao Supremo Tribunal Federal. Nela fica estabelecido que o STF deve ser composto por onze ministros, os quais devem ter entre trinta e cinco e sessenta e cinco anos de idade. Além disso, cada ministro assume seu cargo no STF após ser indicado pelo presidente da república mediante aprovação por maioria absoluta no Senado Federal (BRASIL, 2019). Dentre as várias competências do STF, pode-se citar:

- Processar e julgar originariamente, a ação direta de inconstitucionalidade de lei ou ato normativo federal ou estadual e a ação declaratória de constitucionalidade de lei ou ato normativo federal; nas infrações penais comuns, o Presidente da República, o Vice-Presidente, os membros do Congresso Nacional, seus próprios Ministros e o Procurador-Geral da República; o pedido de medida cautelar das ações diretas de inconstitucionalidade (BRASIL, 2019).

- Julgar em recurso ordinário o crime político (BRASIL, 2019).

Em 14 de Agosto de 2002, o STF inova sua forma de comunicação passando a transmitir ao vivo as sessões plenárias através da TV Justiça, sendo o primeiro tribunal do mundo a transmitir seus julgamentos ao vivo e as sessões plenárias na íntegra por meio de canais próprios (STF, 2011). Ou seja, através de um canal de televisão público, de caráter não-lucrativo e coordenado pelo próprio Supremo Tribunal Federal (STF, 2020). Nessa mesma data, o STF passou também a utilizar o Youtube para suas transmissões e, em agosto de 2009, ingressou no Twitter de forma a se aproximar dos cidadãos com a utilização de novas mídias (TWITTER, 2020).

Em 2019, o STF era composto pelos ministros Luiz Fux, Rosa Weber, Marco Aurélio, Gilmar Mendes, Ricardo Lewandowski, Cármen Lúcia, Dias Toffoli, Roberto Barroso, Edson Fachin e Alexandre de Moraes.

3.3 Mineração de Texto

Mineração de Dados é a extração de informação a partir de grandes quantidades de dados. Pode-se considerar a mineração de dados como uma fase do processo de Descoberta de Conhecimento em Bases de Dados (Knowledge Discovery in Databases). O objetivo da mineração de dados é obter informações relevantes e novas "algo anteriormente não conhecido e que tenha valor para o domínio em que é aplicada" (CARVALHO et al., 2017).

No mundo real, as informações são encontradas de forma dispersa, ou seja, não estruturadas, onde na maioria das vezes elas se encontram no formato de texto, principalmente quando trabalha-se com a Internet e sua infinidade de blogs e também redes sociais. Se os dados a serem trabalhados consistem em dados especificamente no formato de texto, o processo é denominado de Mineração de Textos. A Mineração de Textos consiste na utilização de algoritmos capazes de analisar conjuntos de documentos de texto e classificá-los, de modo a extrair conhecimento deles (GONÇALVES, 2012). Após realizada a mineração dos dados, ou seja, a organização e a classificação deles, defronta-se com um grande conjunto de dados que a princípio não revela nada. Isto é, a mineração de texto ainda não atingiu seu objetivo de revelar informações desconhecidas. A partir desse momento é realizada a Extração de Conhecimento.

A Mineração de Textos está intrinsecamente associada à Extração de Conhecimento devido a forma como se identifica o novo conhecimento, o qual pode ser caracterizado de acordo com a técnica empregada, que pode ser através de consulta (baseada em documento ou em consulta) ou através de análise (baseada em estatística ou semântica). A recuperação da informação baseada em consulta, consiste em

inicialmente estabelecer os critérios sobre que tipo de informação que deseja-se obter para então realizar uma busca em tais critérios. Por outro lado, a técnica baseada em documento, busca por dados semelhantes na massa recuperada, algo que correlacione o conjunto dos textos, para que seja possível classificá-los segundo algum critério. A técnica de análise por semântica classifica cada palavra do texto segundo sua relevância, após realizar uma análise sintática e semântica. A técnica por estatística normalmente extrai palavras isoladas ou palavras compostas do texto analisado. Essas palavras são identificadas pela frequência de coocorrência ou pela presença em dicionários (CARVALHO et al., 2017).

3.4 Análise de Sentimentos

A análise de sentimentos também é conhecida como mineração de sentimentos e extração de opinião, é conceituada por (LIU, 2012) como o campo que estuda as opiniões e emoções das pessoas em relação a produtos, serviços, organizações, indivíduos e eventos. Pode ser aplicada para encontrar e monitorar opiniões de consumidores em grandes volumes de dados de texto, produzidos nas mídias sociais ou em dados internos de corporações como feedbacks coletados de clientes através de e-mails ou resultados de pesquisas realizadas.

Ao considerar os textos gerados nas redes sociais e blogs, fica mais claro como essa etapa da Extração de Conhecimento pode ser considerada e aplicada. Nesse contexto o objetivo da análise de sentimentos é classificar o texto, geralmente em multirrótulo, como por exemplo, sentimentos positivo, negativo, neutro ou até medo e raiva, dependendo da forma como foi feito o treinamento para a classificação (GONÇALVES, 2012). De forma resumida, pode-se definir a análise de sentimentos, como a análise e classificação de opiniões e sentimentos expressos em textos (GOMES, 2013).

Existem duas técnicas para realizar a análise de sentimentos: aprendizado de máquina e análise léxica. Neste trabalho será utilizada a análise léxica.

3.5 Análise Léxica

A comparação de padrões é uma atividade comum na computação, a análise léxica em sua essência consiste em identificar padrões. Segundo (SEBESTA, 2011) *"um analisador léxico é essencialmente um casador de padrões, que tenta encontrar uma sub-cadeia de uma cadeia de caracteres que case com um padrão de caracteres."*

O processo de análise léxica consiste em, a partir de uma entrada de um conjunto de textos, realizar uma coleta dos caracteres, agrupá-los logicamente e atribuir-lhes códigos internos de acordo com sua categoria (token), esses caracteres agru-

padros em categorias possuem então um significado coletivo. Essa segmentação e o agrupamento em categorias é o processo chamado de tokenização, os tokens são os padrões que ocorrem no texto, ou seja, as categorias que dão significado ao conjunto de caracteres.

A Análise Léxica é amplamente utilizada na construção de compiladores, porém, no contexto da Análise de Sentimentos ela atua na etapa da Classificação, de modo a categorizar os textos e então extrair o sentimento ali expressado. A extração dessa informação (o sentimento), é realizada considerando indicadores nos textos, ou seja, palavras que indicam sentimentos, como por exemplo, bom, maravilhoso, e incrível indicam sentimento positivo, enquanto que ruim, pobre e terrível indicam sentimento negativo. A tabela contendo as palavras que indicam sentimentos é chamada de léxico de sentimentos (LIU, 2012). Assim, na execução da Classificação através da análise léxica, o léxico figura como um recurso de referência para o processamento da informação (FREITAS, 2013).

Neste trabalho, para realizar a classificação dos tweets através da análise léxica, foi utilizado um dicionário de polaridades (léxico) como referência.

4 Metodologia

A execução deste trabalho foi dividida em quatro etapas, a coleta de dados, o pré-processamento, a classificação e a análise.

4.1 Coleta de Dados

A fase inicial do trabalho consiste na coleta dos dados, cujo o objetivo é justamente minerar os dados da rede social Twitter, ou seja, coletar os tweets dos usuários. Para a execução de tal tarefa foi utilizada uma API do Twitter ([TWITTER, 2021](#)).

As APIs do Twitter são formas de proporcionar a interação com outros programas e aplicativos fornecidas pela rede social. Através delas o Twitter permite a desenvolvedores, usuários ou empresas acessarem dados públicos da rede social, ([TWITTER, 2020c](#)). O principal ponto de acesso da API utilizada na coleta é o *Search Tweets*, esse serviço fornece os dados dos tweets de acordo com os critérios especificados na consulta. Tais critérios são definidos através dos seguintes parâmetros: *q*, *until*, *since*, *language* e *min_position*. Segue abaixo uma breve descrição de cada um deles ([RDOCUMENTATION, 2020](#)):

- **q** - corresponde à *querysearch*, ou seja, os termos de busca a serem considerados.
- **until** - corresponde aos tweets criados antes da data informada, que deve ser formatada como YYYY-MM-DD.
- **since** - corresponde aos tweets criados a partir da data informada, que deve ser formatada como YYYY-MM-DD.
- **language** - limita a busca para tweets na língua informada.
- **min_position** - limita a busca para tweets com posição superior ao informado, utilizado como parâmetro de controle para iterar para as próximas páginas.

A API oficial do Twitter possui certas limitações, as quais impactam diretamente na realização do trabalho, como o limite de tempo de apenas sete dias anteriores para a busca de tweets. Para contornar tal problema, foi desenvolvido um algoritmo próprio, baseado em *GetOldTweets-R* ([MEHMET, 2019](#)), para realizar a coleta dos tweets. O código está disponível em ([LAPA, 2020](#)).

O código para realizar a coleta dos dados baseia-se em uma função que a partir dos parâmetros da busca **q**, **until**, **since**, **language** e **min_position**, executa a busca e a coleta dos tweets.

Para realizar a consulta os argumentos utilizados foram: a string "stf" como termo da busca (correspondendo ao parâmetro **q**), assim fica garantido que todos os tweets estejam relacionados ao STF; pt para a linguagem (correspondendo a português para o parâmetro **language**), conseqüentemente, considerou-se apenas tweets escritos em português visto que tais tweets possuem maior relevância para este trabalho; e o período de tempo de 01/01/2019 (correspondendo ao parâmetro **since**) até 31/12/2019 (correspondendo ao parâmetro **until**), na primeira request à API não é passado nenhum argumento para o parâmetro **min_position**. A primeira parte do código nessa etapa do trabalho consiste em, a partir dessas informações, montar a URL para executar a request ao endpoint do Search Tweets da API.

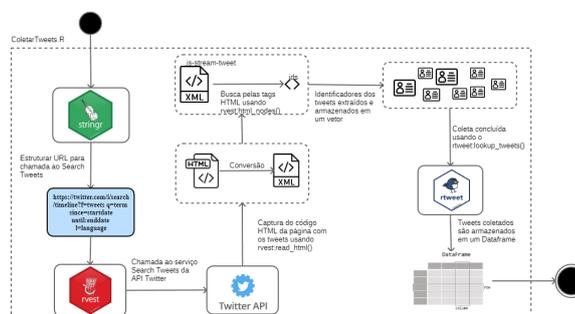
Como auxiliar para essa atividade o pacote **stringr** (WICKHAM, 2020b) do Tidyverse foi amplamente utilizado. Com funções como a *str_c()* para preparar vetores unindo-os elemento a elemento em um único vetor, a função *str_sub()* para extrair substrings para um determinado intervalo passado, a *str_count()* para contar o número de vezes que determinado padrão aparece nos textos e a *str_detect()* para extrair termos que contém um padrão procurado. O pacote **stringr** é construído com base no pacote padrão da linguagem R para processamento de texto, que fornece um conjunto de funções para uma fácil e rápida manipulação de strings.

Com a URL devidamente estruturada, é necessário que o algoritmo acesse o endereço para realizar a captura das páginas dos tweets. No desenvolvimento dessa parte do código foi utilizado o pacote **rvest** (RSTUDIO, 2020a), que é um pacote do Tidyverse que auxilia na extração de dados da Internet, inspirado na biblioteca BeautifulSoup (RICHARDSON, 2022). O **rvest** fornece métodos para acessar e ler uma página HTML (HyperText Markup Language) (*read_html()*), encontrar elementos HTML através de seletores CSS (Cascading Style Sheets) ou expressões XPATH (XML Path Language) (*html_nodes()*), obter dados dos atributos HTML (*html_attr()*).

Assim, usando a função *read_html()*, a página é acessada e é feita a captura do seu código HTML, esse código é então convertido para uma aproximação correspondente de um XML válido. Em seguida, através da função *html_nodes()*, é feita uma busca nesse XML baseada no CSS, pelas tags que correspondem aos identificadores dos tweets, daí então essas tags com os identificadores são armazenados em um vetor. Para cada uma das tags contidas no vetor é recuperado o conteúdo, ou seja, os identificadores dos tweets em si com a função *html_attr()*. Finalmente, utilizando o pacote **rtweet** (ROPENSCI, 2020), um cliente implementado em R para acessar a API REST do Twitter, passando os identificadores dos tweets como argumento para o

método `lookup_tweets()`, os dados dos tweets são coletados. A Figura 1 representa os passos da etapa de coleta.

Figura 1 – Processo da Etapa de Coleta.



O processo foi repetido para cada dia de 2019, começando em 01/01/2019 e sendo incrementado pelo algoritmo em 1 dia até 31/12/2019. Assim, os parâmetros **since** e **until** da API, foram inicializados em 01/01/2019 e 02/01/2019 e incrementados até chegar em 30/12/2019 e 31/12/2019, respectivamente. Os dados foram organizados em uma estrutura dataframe. Um dataframe é uma estrutura de dados semelhante à uma matriz, uma coleção retangular onde as variáveis correspondem às colunas e as observações correspondem às linhas (WICKHAM; GROLEMUND, 2019). Dessa maneira, é obtido uma estrutura muito semelhante à uma view de um banco de dados relacional comum, a partir dessa estrutura foi gerada uma base de dados em um arquivo em formato .csv, contendo todos os dados da coleta.

4.2 Pré-processamento

Com a conclusão da fase de coleta, foram obtidos dados brutos, os quais necessitam de uma "limpeza". A fase de pré-processamento consiste exatamente em tratar esses dados obtidos do Twitter para que seja descartado tudo que seja irrelevante no processo de análise de sentimentos. As tarefas a seguir fazem parte da etapa de pré-processamento:

- Transformação de todos os caracteres em minúsculos, de modo a criar um padrão, pois alguns usuários de redes sociais costumam escrever texto em letras maiúsculas para enfatizar sua opinião.
- Remoção de pontuação e números, bem como espaços extras em branco. Isto é, foram removidos elementos que não sejam letras em português e espaço, pois não agregam valor à análise do texto.
- Remoção de URLs, pois não possuem valor semântico para serem analisadas.

- Remoção de palavras ruído (também chamadas de stopwords) são palavras que não possuem significado semântico para o texto, como por exemplo, artigos e preposições. Exemplos de stopwords: "pra", "porque", "pro" ou "com" (MULLER; GRANATYR; LESSING, 2015).
- Realização do Stemming, que é o processo no qual reduzimos as palavras variantes, mas com mesmo significado ao seu radical, consistindo na maioria das vezes na remoção de afixos (STANFORD NLP GROUP, 2002)(GONÇALVES, 2012)(SILVA, 2016).

Para a execução dessas tarefas foram utilizados basicamente os pacotes **dplyr** (WICKHAM et al., 2020) e **tm** (FEINERER, 2020b). O pacote **dplyr** serve para manipulação de dados. Especificamente esse pacote foi necessário para selecionar variáveis a partir de seus nomes (*select()*), filtrar linhas para análise a partir de condições específicas (*filter()*) e ordenar colunas (*arrange()*). O pacote **tm**(Text Mining) fornece um conjunto de funcionalidades para gerenciar a manipulação de documentos de textos heterogêneos suportando a leitura de diversos formatos de documentos como por exemplo, texto simples, PDFs ou arquivos XML. Esse pacote foi utilizado para criar uma estrutura que representa uma coleção de documentos textos, a qual é manipulada no pré-processamento (*corpus()*), para criar um vetor de coleções de textos (*VectorSource()*), aplicar funções que alteram o conteúdo da estrutura corpus (*tm_map()*), criar uma matriz termo-documento, a qual é uma estrutura que matriz que descreve a frequência de termos que ocorrem em uma coleção de documentos (*TermDocumentMatrix()*), remover termos esparsos na matriz termo-documento (*removeSparseTerms()*) e remover stopwords na matriz termo-documento (*stopwords()*).

Inicialmente foi feita uma avaliação do dataframe obtido, e dentre as 90 variáveis obtidas, apenas as seguintes foram utilizadas: *user_id*, *status_id*, *created_at*, *screen_name*, *text*, *source*, *display_text_width*, *reply_to_status_id*, *reply_to_user_id*, *reply_to_screen_name*, *is_quote*, *is_retweet*, *favorite_count*, *retweet_count* e *location*. Para isso, foi utilizada a função *select()* do pacote **dplyr** para realizar essa separação e obter um novo dataframe com apenas as variáveis de interesse.

Após selecionadas as variáveis manualmente, o dataframe foi transformado em um **Corpus**, que é correspondente às coleções de documentos de textos em linguagem natural (FEINERER, 2020a). A partir dessa nova estrutura de dados foram executadas as tarefas acima descritas utilizando as funções do pacote **tm**. É importante observar que o próprio pacote **tm** fornece um conjunto de stopwords pré-definidas para cada língua, por isso foi feita a remoção de stopwords considerando apenas esse conjunto.

4.3 Classificação

Essa é a etapa onde foi utilizado o algoritmo para classificar os textos dos tweets, ou seja, atribuir um valor ao sentimento expresso, Positivo ou Negativo. Para realizar essa tarefa foi montada uma estrutura dataframe para representar um Léxico. Segundo (FREITAS, 2013), Léxicos "no âmbito do processamento computacional da língua, se referem ao componente de um sistema que contém informação (semântica e/ou gramatical) sobre palavras ou expressões". Para montar o Léxico foram utilizados os dados do Léxico do ReLi (PUC-RIO, 2020) e o léxico para português criado por (CHEN; SKIENA, 2014) disponibilizado na rede Kaggle (KAGGLE INC., 2018).

O ReLi foi elaborado pelo Departamento de Informática da PUC-Rio para o projeto Anotadores Semânticos baseados em Aprendizado Ativo, o qual também fornece como resultado um léxico de substantivos, adjetivos, verbos e expressões já classificados em positivos ou negativos em arquivos de texto, cujo processo de elaboração é detalhado por (FREITAS, 2013). Em (CHEN; SKIENA, 2014) são apresentados léxicos em 136 idiomas, tendo conseguido uma aceitação de 95,7% com relação à outros léxicos já publicados, disponibilizando conjuntos de 1399 palavras positivas e 2554 palavras negativas no léxico em português.

De posse dos arquivos do ReLi, foi necessário fazer modificações manualmente no conteúdo para retirar caracteres '[', ']', espaços em brancos e textos descritivos sobre as palavras. Assim, foram importadas todas as palavras a partir dos arquivos de texto e então organizadas em um dataframe, o qual contém as colunas: palavra, polaridade, tipo e sentimento. Para a polaridade convencionou-se atribuir (-1) para negativo e (1) para positivo, a coluna tipo identifica as palavras em substantivo, adjetivo, verbo ou expressão.

Com o léxico de polaridades estruturado foram removidos os termos duplicados, e então gerado um dicionário de palavras positivas e negativas utilizando o pacote **Quanteda**, que é um pacote para usuários da linguagem R desenvolvido para aplicar processamento de linguagem natural a textos (BENOIT, 2020). Tal dicionário, serviu como parâmetro para diferenciar palavras positivas e negativas e de posse do Corpus foi aplicada a função *dfm()* do **Quanteda**, a qual faz a "tokenização" dos documentos do Corpus e tabula os recursos extraídos em uma matriz de documentos, gerando assim uma matriz contendo a frequência total das polaridades, ou seja, o total de palavras Positivas e Negativas do Corpus. O que fornece uma visão geral de polaridade dos tweets relacionados ao termo "stf".

Dessa forma, a classificação foi realizada considerando apenas o Corpus, então foi executado o algoritmo desenvolvido para realizar a classificação de cada tweet individualmente, utilizando mesmo método feito para o Corpus. Como resultado, foi

obtido um novo dataframe com uma nova coluna chamada **classificacao** com cada registro de tweet rotulado como Positivo ou Negativo.

4.4 Análise

Realizada a classificação foram obtidos os sentimentos dos tweets como Positivos e Negativos. Em seguida, iniciou-se então a etapa de análise a qual consiste em, a partir desses sentimentos, buscar sua correlação com o objeto de estudo. Para estabelecer essas conexões entre as entidades, foram utilizados gráficos, tabelas de frequência (absoluta) e nuvem de palavras. Esses recursos de visualização são detalhados abaixo.

- A frequência absoluta corresponde ao número de ocorrências de cada termo. Já as tabelas de frequência consistem na montagem de uma tabela simples onde são listados os termos (palavras) e suas respectivas quantidades de ocorrência.
- A nuvem de palavras, é uma representação gráfica que considera a frequência com que cada palavra ocorre no texto de modo que, quanto maior for a ocorrência da palavra maior destaque ela terá na nuvem, ou seja, maior será sua fonte e sua evidência na imagem da nuvem.
- Gráficos para auxiliar a visualização da distribuição dos dados foram usados como, por exemplo, gráficos de barras, gráficos de boxplot e gráficos de dispersão.

As nuvens de palavras foram criadas utilizando-se o pacote ([FELLOWS, 2018](#)), já os gráficos foram gerados com o auxílio do **ggplot2** ([WICKHAM, 2020a](#)), um pacote do Tidyverse para criação de gráficos com mais de 10 anos e amplamente utilizado.

Para cada etapa foram criados algoritmos, cujo código está disponibilizado em ([LAPA, 2020](#)). O código foi implementado na linguagem R em conjunto com a IDE RStudio ([RSTUDIO, 2020b](#)). Neste trabalho foi utilizado principalmente o Tidyverse ([RSTUDIO, 2020c](#)), que é um conjunto de pacotes para trabalhar especificamente com análise de dados.

5 Resultados

A quantidade de tweets obtidos foi de 255824 registros, onde cada registro possuía 90 variáveis correspondentes aos seus metadados, sendo apenas uma dessas variáveis correspondendo ao conteúdo do texto do tweet. Utilizando-se da base de dados mencionada na Seção 4.1, foi feita uma validação manual no campo `text` de cada tweet, com a intenção de verificar se de fato continham a palavra "stf". Consequentemente, foram realizadas consultas utilizando as funções `str_count()` e `str_detect()` e foi constatado que a palavra apresenta-se várias formas diferentes conforme a Tabela 1.

Tabela 1 – Diferentes formas de ocorrência da palavra STF.

Forma	Frequência
sTF	31
StF	53
STf	109
Stf	1112
sTf	7
stF	11
STF	223049
stf	9906
Total:	234278

Verifica-se, então, que há uma diferença de 21546, ou seja, a possibilidade de 21546 tweets fora do escopo da consulta, sem menção ao STF. Fazendo novas consultas e desconsiderando as formas de escrita da Tabela 1, foi constatado que tweets com textos escritos em fontes personalizadas, diferentes do padrão, mesmo que possuam a palavra "STF", não estavam sendo contabilizados pela função `str_count()`, como no exemplo da Figura 2.

Figura 2 – Exemplo da palavra "stf" em fonte personalizada.

"D e r r u b a r o s t f"

Além desse caso da fonte também foi verificado que alguns tweets apesar de não possuírem a palavra "stf" no campo `text`, possuíam no campo `screen_name` e por isso foram coletadas. O campo `screen_name` corresponde ao nome de usuário que criou o tweet, onde, aparentemente, correspondiam apenas ao usuário STF_oficial. Para solucionar o problema e identificar quantos registros realmente não mencionam STF, foram criados scripts para identificar quantos tweets não possuíam a palavra "stf"

no campo *text* e quantos possuíam no campo *screen_name*. Após a execução dos scripts foi concluído que:

- Havia 145 tweets que não possuíam a palavra "stf" no campo *text*.
- Dos 145 tweets que não possuíam a palavra "stf" no campo *text*, 136 deles tinham a *screen_name* igual a "STF_oficial" e por isso foram coletados.
- Dos 145 tweets que não possuíam a palavra "stf" no campo *text*, 9 deles não possuíam *screen_name* igual a "STF_oficial", mas mesmo assim foram coletados.
- Havia 205 tweets do usuário *STF_oficial*.

Os 136 tweets com *screen_name* igual a "STF_oficial" foram desconsiderados para este trabalho. Logo, foram removidos todos os registros com campo *screen_name* igual "STF_oficial", pois não refletem opiniões de usuários sobre o STF. Tais tweets representavam apenas postagens do próprio STF em seu canal no Twitter. No total, foram removidos 205 tweets do usuário *STF_oficial*.

Dos 9 tweets que não possuem o campo *screen_name* igual a "STF_oficial", apenas o registro de índice 50384 possuía o conteúdo de texto sobre o STF e não tinha o *screen_name* igual "STF_oficial". Os outros 8 registros só continham no campo *text* a seguinte frase "*account is temporarily unavailable because it violates the Twitter Media Policy.*", indicando que os usuários excluíram seu tweet ou o próprio Twitter, bloqueou devido as suas políticas. Sendo assim, esses 8 registros foram excluídos para este trabalho.

Como resultado dessas alterações foram removidos 213 registros, correspondendo a 0.08326% do total coletado e que, desse total, 255.611 de fato correspondem a opiniões de usuários do Twitter com relação ao STF. A Tabela 2, exibe os valores identificados e as porcentagens com relação ao total que foi coletado:

Tabela 2 – Detalhamento dos registros removidos.

	Total	Porcentagem
Tweet's do Usuário Stf_Oficial	205	0.08013%
Tweet's Bloqueados	8	0.00313%
Tweet's Removidos Manualmente	213	0.08326%
Tweets Restantes	255611	99.91674%

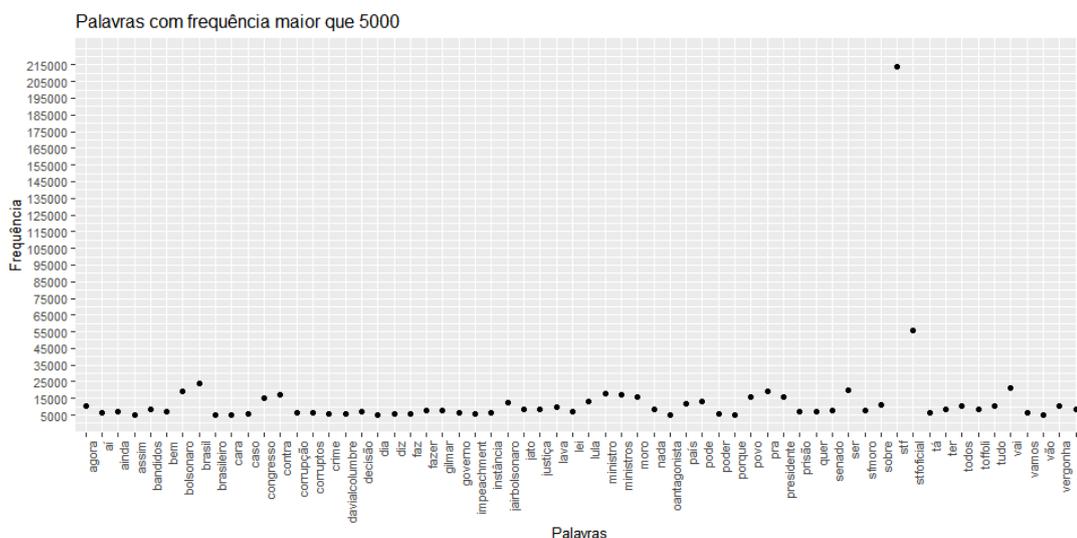
A tabela de frequência obtida é composta por 146125 palavras. Na Tabela 3 estão apresentadas as primeiras palavras mais frequentes. Como pode ser observado, as palavras com maior frequência são "stf" e "stfoficial", com 213761 e 55539 ocorrências, respectivamente. Considerando a ocorrência da terceira palavra mais frequente

Tabela 3 – Tabela de frequência inicial.

Palavra	Frequência
stf	213792
stfocial	55572
bolsonaro	19256
ministro	18289
contra	17194
ministros	17122
presidente	15990
moro	15667
povo	15666
congresso	15021
lula	13255
pode	13038
jairbolsonaro	12485
país	11943
sobre	11044
vergonha	10651
tudo	10475
lava	9824
bandidos	8580
toffoli	8417
justiça	8389
nada	8140
jato	8113
sfmoro	7746
gilmar	7696
senado	7631
decisão	7402
bem	6907
prisão	6907
lei	6894
instância	6608
corrupção	6494

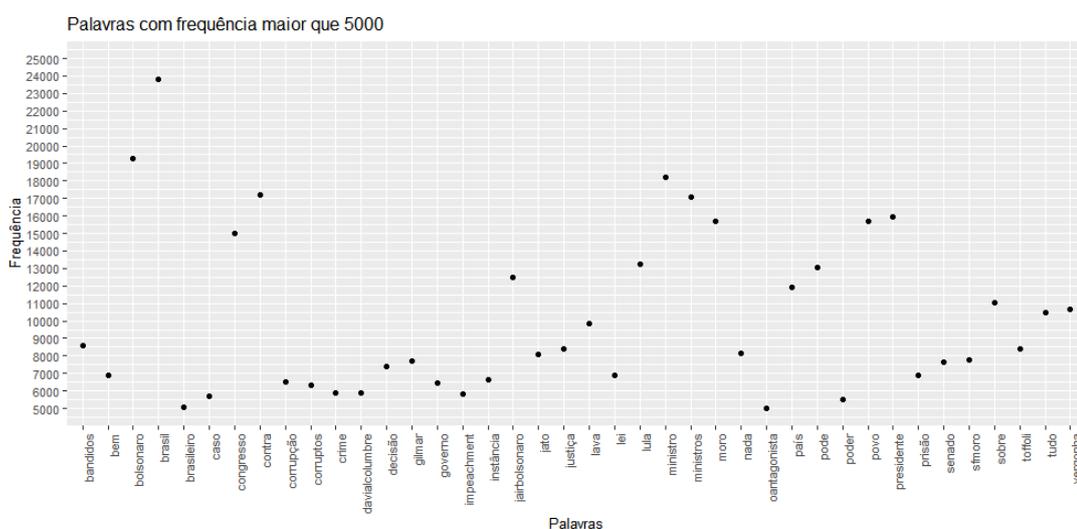
A ocorrência da palavra "stf" tem baixa relevância por ser termo de busca. Para que se tenha uma visualização melhor desses dois outliers, basta observar o gráfico de dispersão apresentado na Figura 4, onde foi considerado apenas ocorrências acima de 5000.

Figura 4 – Evidenciando os Outliers.



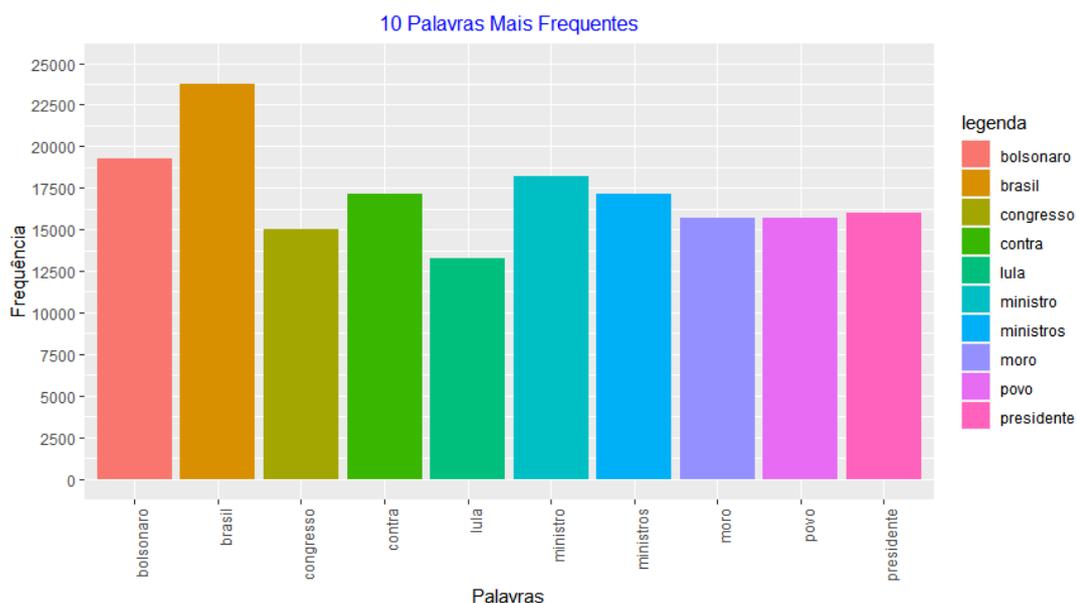
Também foi possível constatar que haviam 82485 palavras com ocorrência igual a 1, o que correspondia a 56,45% das palavras. Tais palavras foram removidas, visto que provocaria uma disfunção para a análise dos dados e por sua baixa ocorrência teriam baixa relevância. Assim, foram desconsiderados os outliers e as palavras com apenas uma ocorrência e tornando a distribuição das palavras mais uniformes, conforme descrito na Figura 5. Ao final da etapa de pré-processamento, já havendo desconsiderado os outliers, as ocorrências de baixa relevância e as palavras-ruído que ainda persistiam, foi realizada a análise dos dados.

Figura 5 – Gráfico de Dispersão sem os outliers.



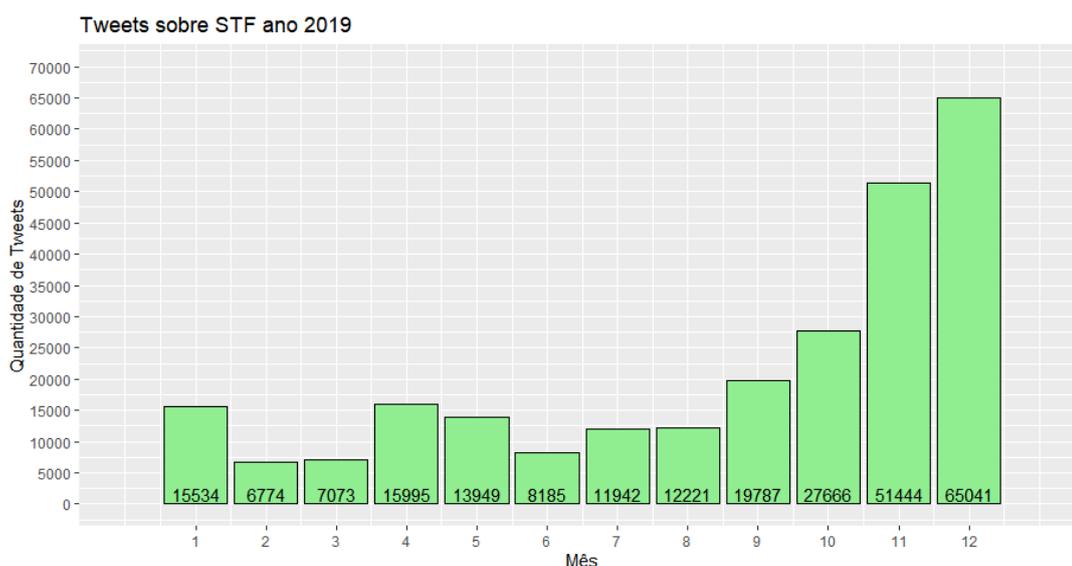
Na nova nuvem de palavras, ilustrada na Figura 6, pode-se observar que a palavra mais destacada é "brasil" ao centro, entretanto é importante observar algumas outras palavras como "ministro" e "bolsonaro", o então presidente da república, "moro",

Figura 7 – Gráfico de Barras com as 10 Palavras mais Frequentes sem os outliers.



Inicialmente, foi verificado a evolução da quantidade de tweets durante o ano de 2019 (ver Figura 8), constatando que durante o período de Janeiro até Agosto, o número de tweets permanece abaixo do 20000 e seguiram uma média de 11459 tweets por mês. A partir de Setembro, inicia-se um aumento nas postagens chegando a 19787, uma diferença de 7566 em relação ao mês anterior. Já entre Setembro e Dezembro há uma média 40984 tweets por mês, sendo que o pico de postagens sobre o STF ocorreu no mês de Dezembro com o total de 65041 tweets.

Figura 8 – Quantidade de tweets por mês no ano de 2019.



Após a realização da classificação de dados constatou-se um total de 187085 palavras positivas e 285174 palavras negativas no corpus (Figura 9). Tais dados reve-

lam um indício sobre a predominância da polaridade do sentimento a respeito do que foi postado nos tweets sobre o STF, no ano de 2019. As Tabelas 4 e 5 apresentam exemplos de tweets positivos e negativos.

Figura 9 – Quantidade de palavras positivas e negativas do total dos tweets.

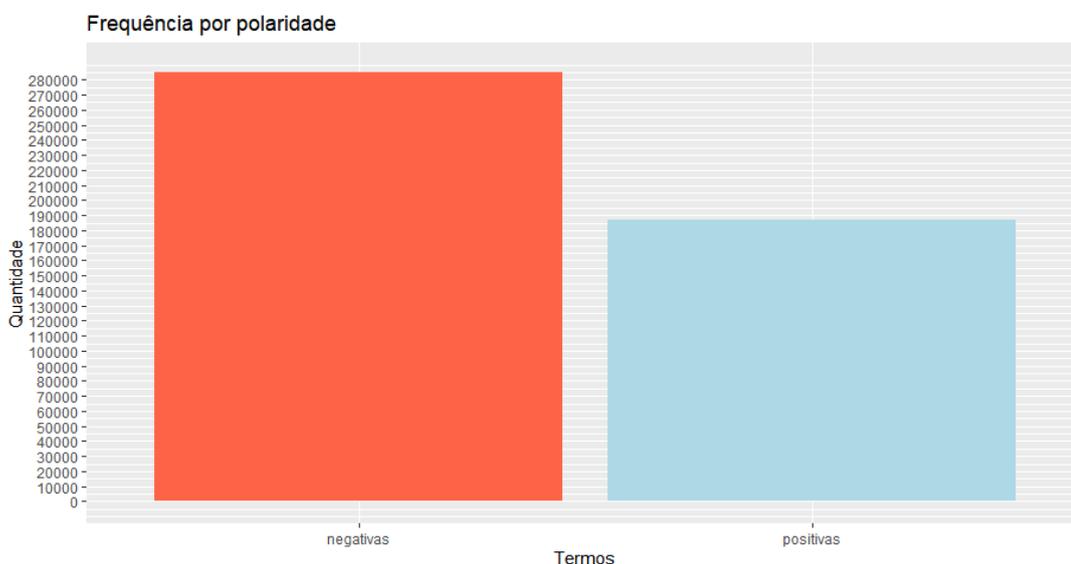


Tabela 4 – Exemplos de Tweets Positivos.

Tweets Positivos
"Raquel Dodge pede ao STF que Gleisi Hoffmann perca mandato de senadora, que seja condenada por corrupção e devolva R\$ 4 milhões aos cofres públicos"
"@STF_Oficial @CIS_br Muito bom! Nada melhor que organizar a casa para 2019, 2020, 2021, 2022...Parabéns aos envolvidos."

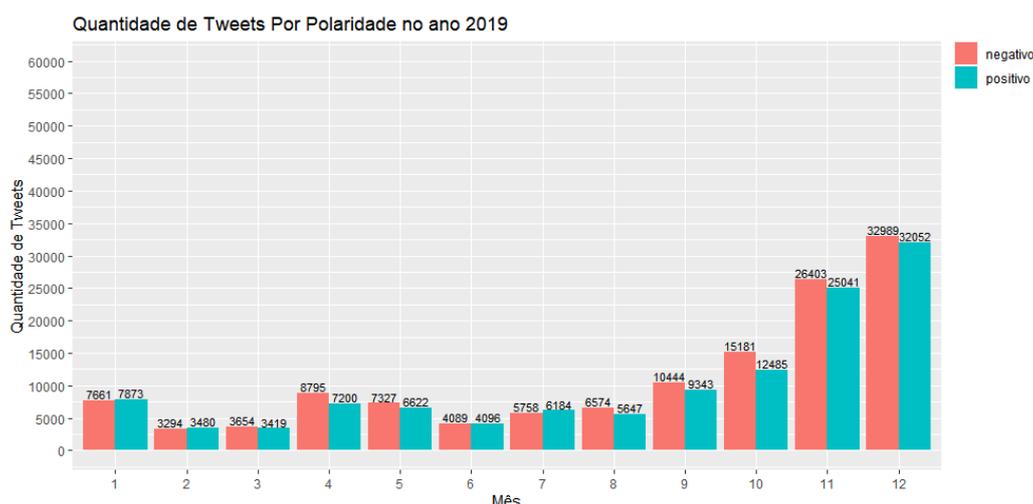
Tabela 5 – Exemplos de Tweets Negativos.

Tweets Negativos
"Eu assisti ela discursando! Golaço! Angela Gandra Martins será secretária da Família no governo Bolsonaro STF cometerá aborto jurídico se legalizar o aborto, diz a jurista Ângela Gandra Martins https://t.co/2TSYgLoHyw "
"E por falar cores,cade o STF.?Aqurle cartel de senhores inúteis e bem assalariados,nababescos,com aquela fantasia horrenda.Calado,acovardado,aquartelado, só vira notícia quando quer aumento.Idiotas mimados,de cu virado prá lua.Vamos trabalhar,VAGABUNDOS."

A Figura 10 apresenta a evolução dos sentimentos positivos e negativos ao longo dos meses. A partir da análise dos dados e do gráfico obtido, pode-se observar

que no total houveram 132169 tweets negativos e 123442 positivos, ou seja, 51,7% são negativos. Nesse gráfico fica evidenciado que de uma forma geral os tweets negativos se sobressaíram durante o ano, apesar de se apresentarem em quantidades relativamente próximas considerando o total do mês, em média a diferença entre a quantidade de tweets positivos e negativos por mês foi de 816.

Figura 10 – Quantidade de tweets positivos e negativos por mês em 2019.



A Tabela 6 resume os dados por polaridade de todo o ano e nos períodos de Janeiro à Agosto e de Setembro à Dezembro.

Tabela 6 – Resumo dos dados por polaridade em 2019.

Período	Positivos	Negativos
Todo o ano	132169	123442
Janeiro até Agosto	47152	44521
Setembro até Dezembro	85017	78921

Foi constatado nessa fase, de forma manual, dois casos específicos. O primeiro caso, onde alguns textos de tweets com classificação tanto positiva quanto negativa continham conteúdo de cunho irônico, como mostra a Tabela 7. E o segundo caso, que alguns tweets com classificação positiva mencionavam o STF, mas não necessariamente faziam uma crítica direta à instituição, como mostra a Tabela 8.

Considerando que em geral as decisões polêmicas do STF giravam em torno de temas como o aborto, o aumento salarial do STF no início de 2019 e a controversa decisão a respeito de prisão em segunda instância, foi investigada alguma correlação da polaridade dos tweets com esses temas. Foram considerados como palavras-chave as palavras "aborto", "segunda" e "instância", "aumento" e "salário". Em seguida, foi analisada a polaridade dessas palavras em cada tweet. Foi observado que a palavra "aborto" ocorre em 164 tweets dos quais 34 são positivos e 130 são negativos. As

Tabela 7 – Evidenciando Tweets de cunho irônico e sua classificação.

Tweets Com Cunho Irônico	Classificação
"R\$ 998 o salário mínimo. Queria o mesmo reajuste do STF, poxa!"	Positivo
"@its_AlyMartins @mmenezesedezza @nelsonbrazil @Maluh_Mends @brendpls @felipeneto Agora ficar soltando todo político condenado não dá né senhores do STF"	Positivo
"@o_antagonista Putzz. Os membros do stf tem uma tara por bandidos hein."	Negativo

Tabela 8 – Evidenciando Tweets com Crítica Indireta e sua classificação.

Tweets com Crítica Indireta	Classificação
"@MichelTemer @alex_borges Vaza! Não esqueceremos o aumento ao STF."	Positivo
"@MichelTemer Obrigada Presidente! Fez um bom governo, teria sido ótimo se tivesse vetado o aumento do STF. Que pague por seus erros."	Positivo

palavras "segunda" e "instância" aparecem juntas em 437 tweets, dos quais 155 são positivos e 282 são negativos. Por fim, as palavras "aumento" e "salário" aparecem juntas em 127 tweets dos quais 35 são positivos e 92 são negativos. A Tabela 9 apresenta um resumo de tal análise. Ela indica que os usuários do Twitter tem uma visão bem negativa de tais decisões.

Tabela 9 – Análise de palavras-chave de temas polêmicos.

Palavras	Frequência nos Tweets	Positivos	Negativos
"aborto"	164	34	130
"segunda" e "instância"juntas	437	155	282
"aumento" e "salário"juntas	127	35	92

A mesma análise foi feita para algumas palavras com alta relevância, as quais foram: "presidente", "congresso", "lula", "moro", "jairbolsonaro". Observando a Tabela 10 pode-se concluir que essas palavras estão mais associadas à polaridade negativa com exceção das palavras "moro" e "jairbolsonaro" as quais têm frequência maior em tweets positivos.

Procurou-se também estabelecer com qual polaridade cada ministro do STF está associado de acordo com as ocorrências de seus nomes nos tweets. Para isso, filtrou-se a tabela de frequência em busca dos nomes e sobrenomes dos ministros do STF e as palavras foram organizadas da maior para menor frequência.

A Tabela 11 apresenta a análise de sentimentos relacionado aos ministros do

Tabela 10 – Análise de palavras relevantes com alta frequência.

Palavras	Frequência nos Tweets	Positivos	Negativos
"presidente"	11098	5490	5608
"congresso"	6009	2909	3100
"lula"	1616	577	1039
"moro"	1349	686	663
"jairbolsonaro"	12157	6407	5750

Tabela 11 – Análise de sentimentos relacionado aos ministros do STF.

Palavras	Frequência nos Tweets	Positivos	Negativos
"toffoli"	656	281	375
"gilmar"	5132	2453	2679
"carmen"	26	16	10
"fachin"	48	27	21
"barroso"	56	36	20
"mello"	139	51	88
"weber"	30	12	18
"fux"	114	72	42
"lewandowski"	32	11	21
"moraes"	118	52	66
"marco aurelio"	56	15	41

STF. Foi verificado que dos onze ministros, 7 aparecem com mais frequência nos tweets negativos e 4 nos positivos. Os nomes dos ministros que têm maior frequência nos tweets são "toffoli" (Dias Toffoli) e "gilmar" (Gilmar Mendes), com 656 e 5132 ocorrências respectivamente. A frequência do nome "gilmar" foi maior do que o nome "toffoli", ou seja, o nome do ministro Gilmar é mais mencionado do que o nome do ministro Toffoli. Todavia, ambos aparecem com maior frequência em tweets negativos. Os nomes dos demais ministros possuem uma frequência muito baixa tornando-se irrelevantes para alguma análise. De maneira geral, quando o nome de algum ministro do STF ocorre em um tweet, em sua maioria este é de polaridade negativa.

6 Conclusões

O ano de 2019 no Brasil, foi um ano conturbado nos âmbitos político e jurídico. O povo emitindo opiniões bastante polarizadas, provocou ânimos exaltados em comentários nas redes sociais. Politicamente, marcou o primeiro ano de um governo com ideologia e propostas opostas aos que o precederam nos últimos 30 anos.

No ano de 2019 foram postados 255611 tweets sobre o STF, com uma média de 21301 tweets por mês. O período de janeiro até agosto teve média 11459 tweets por mês e de setembro até dezembro uma média de 40985 tweets por mês. A partir disso, foi possível observar que em setembro houve um aumento de 7566 tweets e em dezembro foi o mês em que mais houve tweets sobre o STF, totalizando 65041.

Foi analisado também, que as palavras com maior frequência foram respectivamente "stf" e "stf oficial", as quais caracterizaram-se como outliers. Com exceção dessas duas últimas, as palavras mais frequentes foram: "brasil", "bolsonaro", "ministro", "contra", "ministros", "presidente", "moro", "povo", "congresso", "lula".

Predominaram, de forma geral, tweets negativos em relação ao STF totalizando 132169 tweets negativos e 123442 positivos, ou seja, com relação ao total de tweets 51,7% são negativos, esses tweets continham 187085 palavras positivas e 285174 palavras negativas. O mês com menor incidência de tweets negativos foi fevereiro com 3294 e o com maior incidência foi dezembro com 32989.

Buscou-se evidenciar a ocorrência dos nomes dos ministros em cada tweet, para verificar o quanto seus nomes são mencionados, e em que condições são mencionados, ou seja, em tweets com qual polaridade. Foi verificado então, que os nomes dos ministros Gilmar e Toffoli foram os mais mencionados. Porém ambos aparecem com maior frequência em tweets negativos e de forma geral quando o nome de um ministro do STF é mencionado mais frequentemente aparece em tweets negativos.

Quando considera-se palavras-chave associadas com decisões polêmicas do STF como "aborto", "segunda" e "instância" juntas, "aumento" e "salário" juntos, apesar de constatar-se que nos tweets em que essas palavras ocorrem predomina a polaridade negativa, a ocorrência delas com relação ao total de tweets é muito baixa, pois tem-se um conjunto universo de 255611 tweets. Logo não é possível estabelecer uma correlação entre as palavras-chave e os temas, deste modo, não é possível nem conjecturar que a causa de tweets negativos é alguma decisão relacionada ao tema da respectiva palavra-chave. Portanto, em geral, pode-se apenas evidenciar que, ao emitir alguma opinião com o termo "stf" no ano de 2019 sempre predomina a polaridade negativa.

6.1 Trabalhos Futuros

Para a realização deste trabalho foi considerada apenas a palavra "stf" como termo de busca, o que ocasionou na obtenção de qualquer tweet que mencione o STF e não somente tweets direcionados diretamente para a instituição STF. Uma avaliação mais específica poderia ser feita considerando-se apenas aqueles que não só mencionam, mas também são dirigidos diretamente para o STF utilizando outro filtro para a coleta. Uma consulta mais minuciosa pode ser feita considerando além do termo "stf", o nome de cada ministro isoladamente, pois seria possível estabelecer uma classificação individualizada.

Para aprimorar a acurácia da classificação, pode-se ainda considerar tweets com conteúdo irônico, o que poderia em alguns casos até inverter a polaridade estabelecida para a classificação individual dos tweets. Uma possibilidade de estudo interessante, seria avaliar a continuidade do engajamento dos usuários opinando sobre o STF após o ano de 2019, para identificar se são opiniões de ocasião, impulsionadas por uma articulação política ou se há de fato uma ocorrência orgânica.

Referências

- BENOIT, K. *Quantitative Analysis of Textual Data - quanteda*. 2020. Disponível em: <<https://quanteda.io/>>. Acesso em: 13 jan. 2021. Citado na página 23.
- BERKOWITZ, B. *From a single hashtag, a protest circled the world*. 2011. Disponível em: <<https://www.reuters.com/article/us-wallstreet-protests-social/from-a-single-hashtag-a-protest-circled-the-world-idUSTRE79G6E420111018>>. Acesso em: 14 set. 2020. Citado na página 10.
- BIOINFORMATICS LABORATORY. *Biolab Orange3*. 2020. Disponível em: <<https://github.com/biolab/orange3/wiki>>. Acesso em: 10 nov. 2020. Citado na página 12.
- BRASIL. *CONSTITUIÇÃO DA REPÚBLICA DOS ESTADOS UNIDOS DO BRASIL (DE 24 DE FEVEREIRO DE 1891)*. 1891. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao91.htm>. Acesso em: 10 nov. 2020. Citado na página 15.
- BRASIL. *Constituição (1824). Constituição da República Federativa do Brasil*. Brasília, DF: Senado Federal: Centro Gráfico, 2012. Bibliografia: p. 83. ISBN 978-85-7018-424-5. Citado na página 15.
- BRASIL. *Constituição (1988). Constituição da República Federativa do Brasil*. Brasília, DF: Senado Federal: Centro Gráfico, 2019. ISBN 978-85-54223-41-0. Citado 3 vezes nas páginas 10, 15 e 16.
- CARVALHO, A. C. P. L. F. de et al. *Inteligência Artificial Uma Abordagem de Aprendizado de Máquina*. Rio De Janeiro: LTC, 2017. 378 p. Bibliografia: p. 331–332. ISBN 978-85-216-1880-5. Citado 2 vezes nas páginas 16 e 17.
- CHEN, Y.; SKIENA, S. Building sentiment lexicons for all major languages. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Baltimore, Maryland: Association for Computational Linguistics, 2014. p. 383–389. Disponível em: <<https://www.aclweb.org/anthology/P14-2063>>. Citado na página 23.
- CONOVER, M. et al. The digital evolution of occupy wall street. *PloS one*, v. 8, p. e64679, 05 2013. Citado na página 13.
- COSTA, C. *Brasileiros ‘descobrem’ mobilização em redes sociais durante protestos*. 2013. Disponível em: <https://www.bbc.com/portuguese/noticias/2013/07/130628_p_rotestos_redes_personagens_cc>. Acesso em: 14 set. 2020. Citado na página 9.
- DOMO. *Data Never Sleeps 7.0 Infographic | Domo*. 2020. Disponível em: <<https://www.domo.com/learn/infographic/data-never-sleeps-7-0>>. Acesso em: 04 out. 2022. Citado na página 10.
- FEINERER, I. *Text Mining Package*. 2020. Disponível em: <<https://cran.r-project.org/web/packages/corpus/index.html>>. Acesso em: 13 jan. 2021. Citado na página 22.

FEINERER, I. *tm - Text Mining Package*. 2020. Disponível em: <<http://tm.r-forge.r-project.org/>>. Acesso em: 13 jan. 2021. Citado 2 vezes nas páginas 12 e 22.

FELLOWS, I. *wordcloud*. 2018. Disponível em: <<https://www.rdocumentation.org/packages/wordcloud/versions/2.6>>. Acesso em: 13 jan. 2021. Citado 2 vezes nas páginas 12 e 24.

FREITAS, C. Sobre a construção de um léxico da afetividade para o processamento computacional do português. *Revista Brasileira de Linguística Aplicada*, scielo, v. 13, p. 1031 – 1059, 12 2013. ISSN 1984-6398. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-63982013000400004&nrm=iso>. Citado 2 vezes nas páginas 18 e 23.

G1. *Protestos pelo país reúnem mais de 250 mil pessoas*. 2013. Disponível em: <<http://g1.globo.com/brasil/noticia/2013/06/protestos-pelo-pais-reunem-mais-de-250-mil-pessoas.html>>. Acesso em: 14 set. 2020. Citado na página 9.

G1. *Lewandowski acolhe pedido para fazer duas votações do impeachment*. 2016. Disponível em: <<http://g1.globo.com/politica/processo-de-impeachment-de-dilma/noticia/2016/08/lewandowski-decide-votacao-em-separado-no-julgamento-de-dilma.html>>. Acesso em: 14 set. 2020. Citado na página 10.

GOMES, H. J. C. Cisti 2013 - 8th iberian conference on information systems and technologies. 2013. Citado na página 17.

GONÇALVES, E. C. Mineração de texto - conceitos e aplicações práticas. *SQL Magazine*, v. 105, p. 31–44, 2012. Disponível em: <https://www.researchgate.net/publication/317912973_Mineraçao_de_texto_-_Conceitos_e_aplicacoes_praticas>. Acesso em: 17 nov. 2020. Citado 3 vezes nas páginas 16, 17 e 22.

GRUPO NZN. *A história do Twitter, a rede social de 140 caracteres*. 2018. Disponível em: <<https://www.tecmundo.com.br/redes-sociais/120893-historia-twitter-rede-social-140-caracteres-video.htm>>. Acesso em: 10 nov. 2020. Citado na página 14.

INTERNATIONAL TELECOMMUNICATION UNION (ITU). *Individuals using the Internet*. 2021. Disponível em: <<https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>>. Acesso em: 04 out. 2022. Citado na página 10.

JOCKERS, M. *Syuzhet*. 2020. Disponível em: <<https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>>. Acesso em: 13 jan. 2021. Citado na página 12.

KAGGLE INC. *Sentiment Lexicons for 81 Languages | Kaggle*. 2018. Disponível em: <<https://www.kaggle.com/rtatman/sentiment-lexicons-for-81-languages/version/1>>. Acesso em: 13 jan. 2021. Citado na página 23.

LAPA, G. *Analise-sentimentos-tweets-stf-2019*. GitHub, 2020. Disponível em: <<https://github.com/GuilhermeLapa/analise-sentimentos-tweets-stf-2019>>. Citado 2 vezes nas páginas 19 e 24.

LIU, B. *Sentiment Analysis and Opinion Mining*. Chicago: Morgan & Claypool Publishers, 2012. 167 p. ISBN 9781608458851. Citado 2 vezes nas páginas 17 e 18.

MARTELETO, R. M. Análise de redes sociais – aplicação nos estudos de transferência da informação. *Ciência da Informação*, scielo, v. 30, p. 71 – 81, 04 2001. ISSN 0100-1965. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000100009&nrm=iso>. Citado na página 9.

MATOS, F.; MAGALHÃES, L.; SOUZA, R. R. Recuperação e classificação de sentimentos de usuários do twitter em período eleitoral. *Informação & Informação*, v. 25, p. 92, 04 2020. Citado na página 12.

MEHMET, s. *GetOldTweets-R*. GitHub, 2019. Disponível em: <<https://github.com/sarikayamehmet/GetOldTweets-R/blob/master/GetOldTweets.R>>. Citado na página 19.

MULLER, E.; GRANATYR, J.; LESSING, O. R. Comparativo entre o algoritmo de luhn e o algoritmo gistsumm para sumarização de documentos. *Revista de Informática Teórica e Aplicada - RITA*, v. 22, p. 75–94, 2015. Disponível em: <https://www.researchgate.net/publication/324790254_Comparativo_entre_o_algoritmo_de_Luhn_e_o_algoritmo_GistSumm_para_sumarizacao_de_documentos>. Acesso em: 13 jan. 2021. Citado na página 22.

OUR WORLD IN DATA. *Number of Internet users, 2019*. 2020. Disponível em: <<https://ourworldindata.org/grapher/number-of-internet-users-by-country?tab=chart&time=2019&country=USA~IND~CHN~BRA~JPN~RUS~MEX~DEU~IDN~GBR>>. Acesso em: 04 out. 2022. Citado na página 10.

OUR WORLD IN DATA. *Number of people using social media platforms*. 2020. Disponível em: <<https://ourworldindata.org/grapher/users-by-social-media-platform?time=2004..2019&country=~Twitter>>. Acesso em: 04 out. 2022. Citado na página 10.

PACELLI, S. *2013 foi marcado por manifestações populares através do uso estratégico das redes sociais*. 2013. Disponível em: <https://www.em.com.br/app/noticia/tecnologia/2013/12/26/interna_tecnologia,482501/2013-foi-marcado-por-manifestacoes-populares-atraves-do-uso-estrategico-das-redes-sociais.shtml>. Acesso em: 14 set. 2020. Citado na página 9.

PATEL, N. *Hashtag (#): O Que É, Como Usar e As Mais Usadas nas Redes Sociais!* 2018. Disponível em: <<https://neilpatel.com/br/blog/como-usar-hashtag/>>. Acesso em: 10 nov. 2020. Citado na página 14.

PATEL, N. *Redes Sociais: O Que São, Para Que Servem e TUDO Sobre Elas*. 2019. Disponível em: <<https://neilpatel.com/br/blog/tudo-sobre-redes-sociais/>>. Acesso em: 14 set. 2020. Citado 2 vezes nas páginas 9 e 15.

PUC-RIO, D. de I. *Corpus ReLi*. 2020. Disponível em: <<https://www.linguateca.pt/Repositorio/ReLi/#agr>>. Acesso em: 13 jan. 2021. Citado na página 23.

RDOCUMENTATION. *search tweets: Get tweets data on statuses identified via search query*. 2020. Disponível em: <https://www.rdocumentation.org/packages/rtweet/versions/0.7.0/topics/search_tweets>. Acesso em: 13 jan. 2021. Citado na página 19.

RICHARDSON, L. *Beautiful Soup*. 2022. Disponível em: <<https://www.crummy.com/software/BeautifulSoup/>>. Acesso em: 02 out. 2022. Citado na página 20.

ROPENSCI. *rtweet*. GitHub, 2020. Disponível em: <<https://github.com/ropensci/rtweet>>. Citado na página 20.

RSTUDIO. *Easily Harvest (Scrape) Web Pages - rvest*. 2020. Disponível em: <<http://rvest.tidyverse.org/>>. Acesso em: 13 jan. 2021. Citado na página 20.

RSTUDIO. *RStudio*. 2020. Disponível em: <<https://rstudio.com/products/rstudio/>>. Acesso em: 13 jan. 2021. Citado na página 24.

RSTUDIO. *Tidyverse*. 2020. Disponível em: <<https://www.tidyverse.org/>>. Acesso em: 13 jan. 2021. Citado na página 24.

SEBESTA, R. W. *Conceitos de Linguagens de Programação*. Porto Alegre: Bookman, 2011. 792 p. Bibliografia: p. 137, 192–193. ISBN 978-85-7780-791-8. Citado na página 17.

SILVA, H. et al. Sentiment analysis of tweets related to sus before and during covid-19 pandemic. *IEEE Latin America Transactions*, v. 100, n. XXX, Oct. 2021. Disponível em: <<https://latamt.ieee9.org/index.php/transactions/article/view/5434>>. Citado na página 12.

SILVA, N. F. F. d. *CAnálise de sentimentos em textos curtos provenientes de redes sociais*. 138 p. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2016. Citado na página 22.

STANFORD NLP GROUP. *Stemming and lemmatization*. 2002. Disponível em: <<https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html>>. Acesso em: 13 jan. 2021. Citado na página 22.

STF. *120 anos do Supremo Tribunal Federal*. 2011. Disponível em: <http://www2.stf.jus.br/portalStfInternacional/cms/verConteudo.php?sigla=portalStfSobreCorte_pt_br&idConteudo=196222&modo=cms>. Acesso em: 10 nov. 2020. Citado na página 16.

STF. *120 anos do Supremo Tribunal Federal*. 2020. Disponível em: <<http://www.tvjustica.jus.br/>>. Acesso em: 10 nov. 2020. Citado na página 16.

THE WORLD BANK GROUP. *Individuals using the Internet (% of population)*. 2021. Disponível em: <<https://data.worldbank.org/indicator/it.net.user.zs?end=2020&start=1960&view=chart>>. Acesso em: 04 out. 2022. Citado na página 10.

TI INSIDE. *Brasil é o 6º país no ranking mundial de usuários do Twitter*. 2019. Disponível em: <<https://tiinside.com.br/11/08/2019/brasil-e-o-6o-pais-no-ranking-mundial-de-usuarios-do-twitter/>>. Acesso em: 04 out. 2022. Citado na página 11.

TWITTER. *O Twitter é*. 2020. Disponível em: <<https://about.twitter.com/pt.html>>. Acesso em: 10 nov. 2020. Citado na página 14.

TWITTER. *Perguntas frequentes sobre contas verificadas*. 2020. Disponível em: <<https://help.twitter.com/pt/managing-your-account/twitter-verified-accounts>>. Acesso em: 10 nov. 2020. Citado na página 15.

TWITTER. *Sobre as APIs do Twitter*. 2020. Disponível em: <<https://help.twitter.com/pt/rules-and-policies/twitter-api>>. Acesso em: 13 jan. 2021. Citado na página 19.

TWITTER. *Sobre contas verificadas*. 2020. Disponível em: <<https://help.twitter.com/pt/managing-your-account/about-twitter-verified-accounts>>. Acesso em: 10 nov. 2020. Citado na página 14.

TWITTER. *STF*. 2020. Disponível em: <<http://www.tvjustica.jus.br/>>. Acesso em: 10 nov. 2020. Citado na página 16.

TWITTER. *Twitter API*. 2021. Disponível em: <<https://developer.twitter.com/en/docs/twitter-api>>. Acesso em: 13 jan. 2021. Citado na página 19.

WICKHAM, H. *ggplot2*. 2020. Disponível em: <<https://ggplot2.tidyverse.org>>. Acesso em: 13 jan. 2021. Citado na página 24.

WICKHAM, H. *Simple, Consistent Wrappers for Common String Operations - stringr*. 2020. Disponível em: <<https://stringr.tidyverse.org/>>. Acesso em: 13 jan. 2021. Citado na página 20.

WICKHAM, H. et al. *A Grammar of Data Manipulation - dplyr*. 2020. Disponível em: <<https://dplyr.tidyverse.org/index.html>>. Acesso em: 13 jan. 2021. Citado na página 22.

WICKHAM, H.; GROLEMUND, G. *R Para Data Science*. [S.l.]: Editora Alta Books, 2019. Bibliografia: p. 4, 11. ISBN 8550803243. Citado 2 vezes nas páginas 13 e 21.