

SOME
Social Meeting
Scientific Journal
Scientific editor: Dr. Evandro Prestes Guerreiro

N°02 / JULHO - 2021



ANO 2 / VOLUME 2

SÃO PAULO - BRASIL

ISBN 978-65-991619-0-2 . ORCID ID: 0000-0001-5061-8755.



FICHA TÉCNICA

Conselho Científico



Argentina

Mag. Felix Luciano Bustos (Universidad Nacional del Comahue - Ciências Sociais)

Dr^a. Margott Gladys Flores (Universidad Nacional de la Rioja Centro de Investigación y Innovación Tecnológica)

Dr. Miguel E. V. Trotta (Universidad Nacional de Lanús Departamento de Ciencia Política)

Social Meeting Scientific Journal
ISBN 978-65-991619-0-2
ORCID id: 0000-0001-5061-8755
e-mail: contato@socialmeeting.info
Homepage: www.esocialbrasil.periodikos.com.br
www.socialmeeting.info

Rua México, 156 - 121
Guarujá-SP/ Brasil
CEP. 11410-350

Brasil

Dr^a. Cely de Oliveira (Universidade de São Paulo - Ciências da Saúde)
Dr^a. Thalita Lacerda Nobre (Universidade Católica de Santos - Ciências Humanas)
Dr^a. Giselle Silva Soares (Centro Universitário São Judas Tadeu - Ciências Humanas)
Dr^a. Olivia Cristina Perez (Universidade Federal do Piauí - Ciências Humanas)
Dr^a. Eva Cristina de Carvalho Souza Mendes (Universidade Paulista - Ciências Humanas)
Dr. José Alberto Yemal (Instituto Paulista de Excelência da Gestão - Ciências Sociais Aplicadas)
Dr. José de França Bueno (Universidade Paulista - Ciências Exatas)
Dr. Jorge Monteiro Junior (Faculdade de Tecnologia Rubens Lara - Ciências Econômicas).
Dr. Júlio Cesar Raymundo (Faculdade de Tecnologia Rubens Lara - Ciências Sociais Aplicadas).
Dr. Luiz Guilherme da Costa Wagner Junior (Universidade Presbiteriana Mackenzie - Ciências Sociais Aplicadas)
Dr. Marcos de Oliveira Moraes (Estácio São Paulo - Ciências Sociais Aplicadas).

Ano 2 - Volume 2
Nº 2 - Julho 2021
Edição Especial

Editor-Chefe

Dr. Evandro Prestes Guerreiro (Brasil)

Editor-Adjunto

Mag. Félix Luciano Bustos (Argentina)

Revisão Editorial

Thaynna V. dos Santos de Oliveira (Brasil)
Mauro Agustín Rodríguez (Argentina)

Publicada por

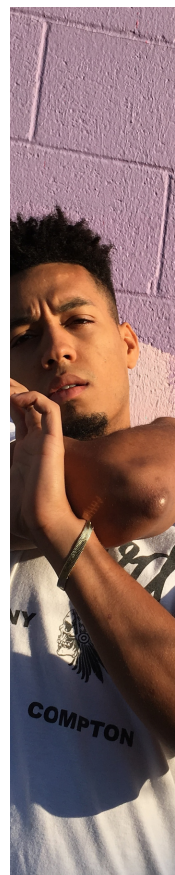


eSocial Brasil - todos os direitos reservados. Capa, imagens e designe produzidos com recursos digitais do canva.

SOMESJ - Social Meeting Scientific Journal - Revista multidisciplinar internacional publicada pela eSocial Brasil, em formato digital (ISBN 978-65-991619-0-2).

SUMÁRIO

- 04** Editorial: Educação Remota
Por: Evandro Prestes Guerreiro
- 07** El caso del "CEAER"- Argentina
Por: María Inés Garcia e Félix Luciano Bustos
- 25** Una propuesta de investigación Acción Participante
Por: Ariel Roberto Bernatene
- 38** Estrategias Didácticas en el Nivel Superior en
Pandemia y Pos Pandemia
Por: Margott Flores Fuentes
- 44** Aspectos psicológicos acerca do suicídio e o
fenômeno da pandemia.
Por: Thalita Lacerda Nobre.



- 54** Mineração de Texto e Redes Sociais: A
COVID-19 no Twitter.
Por: José de França Bueno.

Análise de riscos no terminal de cruzeiros
do Porto de Santos em período de
pandemia da COVID19.

- 68** Por: Felipe Scarpelli Andrade
Júlio Cesar Raymundo
Daniel Alves dos Santos
Matheus de Oliveira Paixão

Ensaio político

- 83** A morte do Messiah - ato I: Nietzsche -
Deus (não) é brasileiro.
Por: Evandro Prestes Guerreiro.

S(CO)
SOCIAL MEETING SCIENTIFIC JOURNAL

VOLUME 2 . ANO 2 . Nº 02 . SÃO PAULO, BRASIL . JULHO 2021

MINERAÇÃO DE TEXTO E REDES SOCIAIS: A COVID-19 NO TWITTER.

POR DR. JOSÉ DE FRANÇA BUENO.
UNIVERSIDADE PAULISTA - BRASIL.

MINERAÇÃO DE TEXTO E REDES SOCIAIS: A COVID-19 NO TWITTER.

POR: DR. JOSÉ DE FRANÇA BUENO.

Resumo

Neste trabalho utilizamos de técnicas de Mineração de Texto (também conhecidas como Text Mining) para buscar padrões em mensagens publicadas na Rede Social Virtual Twitter tendo como referência a palavra-chave COVID-19. O objetivo principal do trabalho é identificar padrões de reação com respeito à palavra-chave COVID-19. Para tal finalidade, como metodologia são utilizados algoritmos para análise das opiniões nos textos e reações publicados por um conjunto selecionados de usuários do Twitter. Este conjunto de algoritmos é conhecido na área de Text Mining como Análise de Sentimentos. Com este conjunto de ferramentas é possível elaborar um quadro que permite identificar o padrão da amostra de mensagens publicadas por aquele usuário específico no período considerado. Neste trabalho foram considerados como usuários para análise: @dcm_online, @g1, @uol, @tvsenado, @NexoJornal, @BandJornalismo, @jornalggn, @revistaforum. As publicações referem-se ao mês de setembro de 2020, efetuadas em língua portuguesa, cuja abrangência geográfica foi território brasileiro. Como conclusões deste trabalho destacamos: 1. A nota média obtida para o usuário @g1 é inferior às notas médias para outros usuários, tradicionalmente associados como críticos do atual governo; 2. O destaque identificado pelas trigramas para a questão da Covid-19 e as terras dos Yanomami; 3. A relevância da discussão sobre bidirecionalidade entre agentes emissores de notícias nas Redes Sociais.

Palavras Chaves: Text Mining, Análise de Sentimentos, Twitter, software R, COVID-19.

TEXT MINING AND SOCIAL MEDIA: COVID-19 ON TWITTER.

POR: DR. JOSÉ DE FRANÇA BUENO.

Abstract

In this work we use Text Mining techniques (also known as Text Mining) to search for patterns in messages published on the Virtual Social Network Twitter using the keyword COVID-19 as a reference. The main objective of the work is to identify reaction patterns with respect to the keyword COVID-19. For this purpose, as methodology, algorithms are used to analyze opinions in texts and reactions published by a selected set of Twitter users. This set of algorithms is known in the area of Text Mining as Sentiment Analysis. With this set of tools it is possible to elaborate a table that allows to identify the pattern of the sample of messages published by that specific user in the considered period. In this work, the following users were considered for analysis: @dcm_online, @ g1, @uol, @tvsenado, @NexoJornal, @BandJornalismo, @jornalggn, @revistaforum. The publications refer to the month of September 2020, made in Portuguese, whose geographical coverage was Brazilian territory. As conclusions of this work, we highlight: 1. The average score obtained for the user @ g1 is lower than the average scores for other users, traditionally associated as critics of the current government; 2. The highlight identified by the trigrams for the Covid-19 issue and the Yanomami lands; 3. The relevance of the discussion about bidirectionality among news issuing agents on Social Networks.

Keywords: Text Mining, Sentiment Analysis, Twitter, Software R, COVID-19.

1. Introdução

Com o advento da Internet e, na sequência, das Redes Sociais virtuais, a produção de dados adquiriu escala e velocidade nunca antes vista. Concomitantemente com o avanço da Internet foram e continuam a ser desenvolvidas técnicas e algoritmos para efetuar a análise e interpretação desta vasta produção de dados. Além da quantidade de dados produzidos ser enorme muitas vezes estes dados não estão estruturados como em um Banco de Dados tradicional (conhecido como Banco de Dados Relacional).

São textos de tuítes, reações no Facebook, imagens no Instagram, comentários no LinkedIn bem como outras formas de representação da informação que não estão estruturadas de uma forma convencional. A este padrão de registro da informação é dado o nome de informação não-estruturada ou NoSQL (em oposição aos Bancos de Dados Relacionais, conhecidos por representar a informação de forma estruturada (em tabelas com linhas e colunas, com as variáveis indicadas nas colunas e os registros individuais indicados nas linhas))[1].

A ferramenta desenvolvida para analisar a informação em forma de texto nas Redes Sociais é conhecida como Text Mining (Mineração de Texto). Inserida neste contexto existe um conjunto de ferramentas algorítmicas que têm o propósito de avaliação das opiniões emitidas nos textos ou reações. Este conjunto de ferramentas é conhecido como Análise de Sentimentos.

Até um passado recente, para se conhecer a opinião pública, sentimentos, emoções e reações, seja com respeito ao consumo de bens ou o posicionamento sócio-político eram realizadas pesquisas quantitativas ou qualitativas (os grupos focais são uma técnicas frequentemente adotada neste caso).

Na atualidade é possível investigar a opinião, sentimentos e reações das pessoas a determinados eventos, produtos, marcas ou pessoas praticamente em tempo real.

Além disso, deve ser reconhecido que existem diferenças fundamentais entre os meios tradicionais de comunicação (Rádio, televisão, jornais e revistas impressos) e os meios proporcionados pela Internet. Se o Rádio levou cerca de quarenta anos para alcançar um público de cinquenta milhões de pessoas no mundo e a televisão levou treze anos, a Internet levou apenas quatro anos. Além da velocidade de propagação outras diferenças fundamentais entre os meios tradicionais de comunicação e a Internet são [2]: i) a bidirecionalidade (a comunicação não é apenas do tipo Broadcasting como em uma televisão ou um jornal tradicional) mas bidirecional; ii) a comunicação não é mais um-para-muitos mas temos a possibilidade de interações todos-para-todos; iii) na Redes Sociais virtuais surgem circuitos de reafirmação ou oposição a determinados assuntos que se retroalimentam, em algumas delas favorecendo a criação de “bolhas” sociais virtuais; iv) as Redes Sociais Virtuais possuem uma característica de auto-organização que não é encontrada nos meios tradicionais de comunicação.

A importância deste conjunto de características está no fato de quanto influenciados e somos influenciados pelos nossos “vizinhos” virtuais (aquelas pessoas que estão mais próximas de nós na nossa Rede Virtual). Nossos valores, crenças, escolhas e percepções da realidade, com seus vieses de confirmação associados, são condicionados por este aspecto de auto-organização das Redes Sociais Virtuais. Apenas como exemplo recorde-se dos sites de Reclamação de Produtos/Serviços.

Tais comportamentos podem resultar em movimentos sociais com propósitos políticos sociais ou ainda reformatar negócios bem como empresas podem multiplicar o retorno do capital investido a partir de viralizações de determinadas postagens.

A Mineração de Texto e a Análise de Sentimentos são aplicadas em áreas tão diversas quanto jurídica (consulte a Associação Brasileira de Jurimetria[4]), estudos literários ou de linguística[5], estudos de Ciência Política (análise de discursos de políticos ou das postagens nas Redes Sociais Virtuais)[6] e em Publicidade/Marketing[7].

Deve ser lembrado que a Mineração de Texto e a Análise de Sentimentos são técnicas quantitativas, implementadas pela contagem de palavras que aparecem no texto-alvo. Contudo, não há uma oposição entre tal abordagem quantitativa e abordagens qualitativas. Podemos usar ambas as abordagens de forma complementar e não excludente.

Nesse contexto, o objetivo deste trabalho é, utilizando-se de técnicas de Mineração de Texto e Análise de Sentimentos, comparar

as publicações no Twitter de usuários que representam sites de notícias, sejam grandes grupos corporativos ou blogs noticiosos não associados a grandes empresas. Restringimo-nos às publicações que contivessem a palavra-chave COVID-19.

O uso do Twitter para acompanhar tendências, grandes eventos catastróficos e problemas de saúde pública não é exatamente novo [8]. Assim, justifica-se efetuar estudos utilizando-se esta Rede Social Virtual dada sua efetiva demonstração de indicador rápido de variações no “humor” das populações. Dentre os fatos que motivaram a execução deste trabalho está a análise da reação de diferentes usuários do Twitter aos efeitos da COVID-19 no Brasil.

2. Método

Neste artigo estudamos publicações no Twitter, efetuadas nos meses de agosto e setembro de 2020, em língua portuguesa, no território brasileiro, com a palavra-chave #COVID-19. Uma das vantagens do Twitter é o fato de possuir um comprimento limitado de texto a 140 caracteres. Estamos interessados não apenas nos dados (o texto em si) como nos metadados (outras informações sobre aquele tuíte que o Twitter disponibiliza para o analista. Os usuários do Twitter dos quais coletamos dados foram: @dcm_online, @g1, @uol, @tvsenado, @NexoJornal, @BandJornalismo, @jornalggn, @revistaforum

Para que a Mineração de Texto e Análise de Sentimentos possa ser efetuada, é necessário que seja aplicado ao texto o que se denomina de “limpeza do texto”.

Os passos para a limpeza do texto são: 1. Converter o texto em uma estrutura de armazenamento conhecida como Corpus (estrutura que preserva os metadados); 2. Tornar todas as letras minúsculas (para evitar dupla contagem das palavras); 3. Remover stop words (palavras que não contribuem com significado, tais como “e”, “para”, “pelo”, etc). Há certo grau de subjetividade na seleção das stop words. A lista de stop words varia conforma a língua de estudo e o tópico de interesse; 4. Dependendo do estudo pode ser necessário excluir números, sinais de pontuação e caracteres especiais; 5. Exclusão de URL’s; 6. Exclusão de espaços em branco repetidos.

Adotou-se aqui uma conta gratuita do Twitter. Observamos que para contas gratuitas do Twitter há limitação do período de tempo nas pesquisas de até os sete dias anteriores. Além disso há um limite de até 18.000 tweets extraídos por solicitação. Outro ponto é que há um limite, nas contas gratuitas do Twitter, de se efetuar até quinze buscas via searchTwitter a cada quinze minutos.

Para que os dados do Twitter possam ser baixados é necessária uma conta de desenvolvedor no Twitter para se utilizar uma API de acesso aberto ao Twitter. Está fora do escopo deste trabalho apresentar os detalhes técnicos deste procedimento. Uma referência para tais informações é [9] <https://dev.twitter.com/apps>].

Neste trabalho utilizamos para a implementação dos algoritmos de Mineração de Texto e Análise de Sentimentos a linguagem R. Alternativamente poderia ter sido utilizada também a Linguagem Python.

As técnicas específicas de mineração de Texto e Análise de Sentimentos aqui utilizadas foram: histogramas de palavras mais frequentes, nuvens de palavras, uso de léxicos em língua portuguesa para análise de sentimentos, identificação dos comentários mais positivos e mais negativos, dendogramas, bigramas, trigramas, Redes de Palavras e contagem de seguidores. Para maiores informações técnicas sobre Mineração de Texto e Análise de Sentimentos consulte [1,2, 3, 10, 11, 12, 13, 14, 15, 16]

4. Resultados

Nesta seção apresentamos os resultados obtidos em nosso levantamento. Efetuando-se uma busca das palavras mais frequentes dos tuítes em língua portuguesa, publicados no Brasil no dia 15/08/2020 que contivessem a hashtag #covid19 temos a tabela 1 a seguir com as palavras mais frequentes neste conjunto de tuítes e as respectivas frequências.

Tabela 1 - Palavras mais frequentes, 15/08/20, com a hashtag #covid19

	Palavra	Frequência
1	covid	10827
2	coronavírus	1975
3	brasil	1573
4	pandemia	1516
5	casos	1385
6	mortes	1312
7	httpstco	1269
8	coronavírus	997
9	saúde	796
10	hoje	700

Tabela 1 – Palavras mais frequentes em tuítes publicados em língua portuguesa no Brasil, 15/08/2020. (fonte: autor, 2020).

Observando-se a lista de palavras mais frequentes notamos que não constam as palavras cloroquina, vacina, nem hashtags como “#fiqueemcasa”, “#usemascara” nas menções pesquisadas no Twitter no dia 15/08/2020. Também não aparecem menções positivas ou negativas a políticos (com mandato ou sem no momento da pesquisa), nem órgãos como Ministério da Saúde, Secretaria da Saúde ou FIOCRUZ.

O que sugere que tais palavras poderiam ter saído do discurso majoritário no país nesse momento. É claro que para confirmar esta hipótese é necessário um acompanhamento por um período mais longo de tempo.

Existe um recurso na Mineração de Texto para representar visualmente as palavras mais frequentes conhecido como Nuvem de Palavras. No caso do levantamento anterior, com a hashtag #covid19, a nuvem de palavras obtida está apresentada na figura 1 a seguir.



Figura 1 - Nuvem de palavras para o levantamento do dia 15/08/2020 com a hashtag #covid19. (fonte: autor, 2020).

Observamos que nessa nuvem de palavras não aparecem palavras relacionadas com Economia: “desemprego”, desoneração fiscal”, “trabalhadores”, “empresários”, “empresas”, dentre outras sugerindo a separação entre o bloco conceitual covid19 e temas afeitos à Economia do Brasil.

4.1 Análise de Sentimentos

A Análise de Sentimentos consiste em estabelecer um conjunto de palavras na língua-alvo com pontuações para sentimentos positivos o, negativos. Após a prévia limpeza do texto é feita uma contagem dessas palavras no texto estudado. Neste estudo utilizaram-se dois léxicos construídos em língua portuguesa para esta finalidade: o oplexicon_v3.0 e o sentiLex_lem_PT02. Mais referências em [10]. Com tais léxicos obtivemos os tuítes com pontuação mais elevada de palavras associadas nestes léxicos a sentimentos positivos e a sentimentos negativos. Assim, podemos identificar os tuítes mais positivos e mais negativos. São eles:

Sentimento mais positivo: “O UNICEF vai usar sua força no fornecimento de #vacinas para garantir que todos os países tenham acesso seguro, rápido e equitativo às doses iniciais da vacina contra a #covid19, quando disponíveis”.

Sentimento mais negativo: “Se meu filho perder um ano na escola, isso não vai fazer diferença para ele lá na frente. Mas se eu perder um filho, sim. Prefiro perder um ano a perder um filho”.

Observe que, obtendo-se os tuítes pontuados com palavras associadas a sentimentos mais positivos ou mais negativos permite-se que se acompanhe, praticamente em tempo real, os sentimentos, emoções e reações de grandes grupos. Tal ferramenta possui interesse, como já mencionado para as mais diversas áreas: Ciência Política, Publicidade/Marketing, Jornalismo, Economia e Saúde Pública.

A partir da pontuação dos sentimentos com os léxicos acima mencionados, podemos construir histogramas para cada um deles. Tais histogramas estão nas figuras 2 e 3 a seguir.

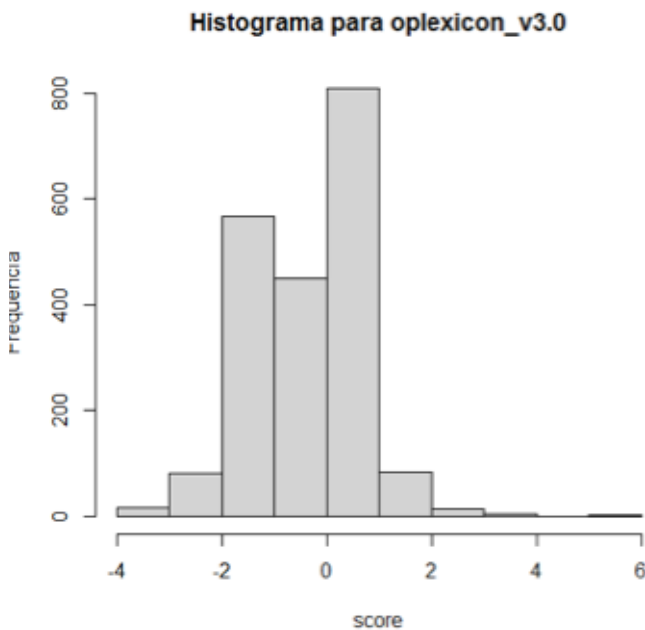


Figura 2 - Histograma obtido com o léxico olexicon_v3.0, palavras em língua portuguesa, hashtag #covid19, levantamento realizado dia 15/08/2020. (fonte: autor, 2020).

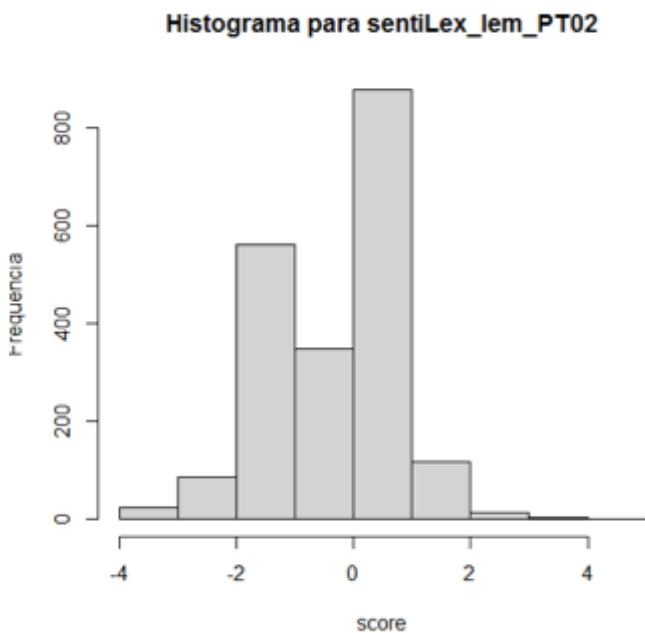


Figura 2 - Histograma obtido com o léxico olexicon_v3.0, palavras em língua portuguesa, hashtag #covid19, levantamento realizado dia 15/08/2020. (fonte: autor, 2020).

Observe que o perfil do histograma para os dois léxicos é bastante similar, sugerindo uma distribuição normal em torno da média zero. No histograma obtido com o léxico olexicon_v3.0 temos uma palavra que obteve pontuação 6. No histograma construído a partir da pontuação dada pelo sentiLex_lem_PT02 não existe palavra com essa pontuação.

Tabela 2 - Medidas de tendência central com os dois léxicos

Medidas com o léxico olexicon_v3.0					
Mínimo	1o. Quartil	Mediana	Média	3o. Quartil	Máximo
- 4,0000	- 1,0000	0,0000	0,1808	1,0000	5,0000
Medidas com o léxico SentiLex_lem_PT02					
Mínimo	1o. Quartil	Mediana	Média	3o. Quartil	Máximo
- 4,0000	- 1,0000	0,0000	0,1314	1,0000	6,0000

Tabela 2 - Medidas de tendência central com os dois léxicos, tuítes publicados em língua portuguesa no Brasil, 15/08/2020. (Fonte: autor, 2020).

Na tabela 2 é possível observar as medidas de tendência central obtidas para os dois léxicos, demonstrando a similaridade já vista nos histogramas. Notamos que os valores para a média são positivos e muito próximos entre si para os dois léxicos, exibindo a consistência das medidas obtidos pelos léxicos. Destacamos ainda que neste momento (agosto/2020) o Brasil estava apresentando cerca de 1000 mortes oficiais devido à Covid-19 há cerca de dois meses.

Ou seja, há um sentimento de "positividade" nas publicações no Twitter associadas com a hashtag #covid19 apesar das mortes associadas ao vírus, sugerindo que a Pandemia não era um problema com sentimentos negativos. Neste momento ainda não existiam informações seguras de quando vacinas contra a Covid-19 estariam prontas nem com respeito à sua eficiência.

4.3 Dendogramas

Outra ferramenta de visualização da estrutura do texto oferecida pela Técnica de Mineração de texto é o dendograma. Um dendograma agrupa as palavras na forma de um diagrama de árvore formando grupos de palavras em função de uma medida de distância entre as mesmas. O grau de similaridade é medido no eixo vertical.

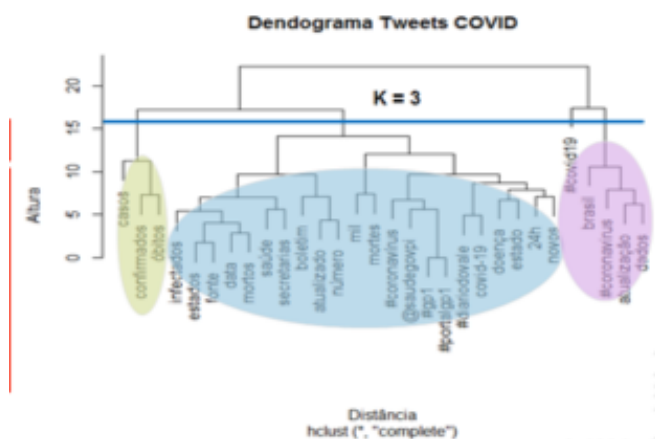


Figura 4 - Dendograma para o conjunto de dados obtido no Twitter, dia 15/08/2020, em língua portuguesa, território brasileiro, hashtag #covid19, levantamento realizado dia 15/08/2020. (fonte: autor, 2020).

Se considerarmos três grupos de palavras, veremos à esquerda as palavras “casos”, “confirmados”, “óbitos” (destacado em verde claro). No agrupamento da direita (destacado em lilás) as palavras “#covid19”, “brasil”, “coronavirus”, “atualização”, “dados”. E um grande grupo central (destacado em azul) com as palavras restantes. O dendograma permite inferir o padrão de proximidade entre as palavras no texto. Assim, as palavras “casos”, “confirmados” e “óbitos” estão mais próximas entre si que com outras palavras no texto.

4.4 Bigramas e trigramas

Os algoritmos de Mineração de texto também possibilitam que encontremos os pares de palavras mais frequentes no texto.

No caso do conjunto de dados já considerado nas tabelas e figuras anteriores os bigramas e respectivas frequências estão apresentados na tabela 3 a seguir.

Palavra 1	Palavra 2	Frequência
coronavírus	covid19	758
covid19	coronavírus	420
Mortes	Hoje	264
Coronavírus	Covid19	261
Casos	confirmados	230
Pandemia	Covid19	220

Tabela 3 - bigramas mais frequentes em tuítes publicados em língua portuguesa no Brasil, 15/08/2020. (fonte: autor, 2020)

De forma análoga aos bigramas também podemos considerar os trigramas. Neste caso o conjunto de trigramas, com respectivas frequências está na tabela 4.

Palavra 1	Palavra 2	Palavra 3	Frequência
terra	indígena	yanomami	76
covid_19	covid19brasil	coronavirusnobrasil	57
anos	rt	pcr	52
rt	pcr	detectável	52
mortes	hoje	sábado	46
yanomami	precisam	sobreviver	46

Tabela 4 - trigramas mais frequentes em tuítes publicados em língua portuguesa no Brasil, 15/08/2020.

Vemos uma informação diferente na tabela 4: “terra”, “indígena” e “yanomami”. Essa terna de palavras ainda não havia aparecido.

4.5 Redes de Palavras

Outra ferramenta frequentemente utilizada na Mineração de Texto é a Rede de Palavras. A Rede de palavra exige os conjuntos de palavras mais fortemente

associadas do texto. É outra ferramenta para identificação de padrões em um texto.

Observando a figura 5

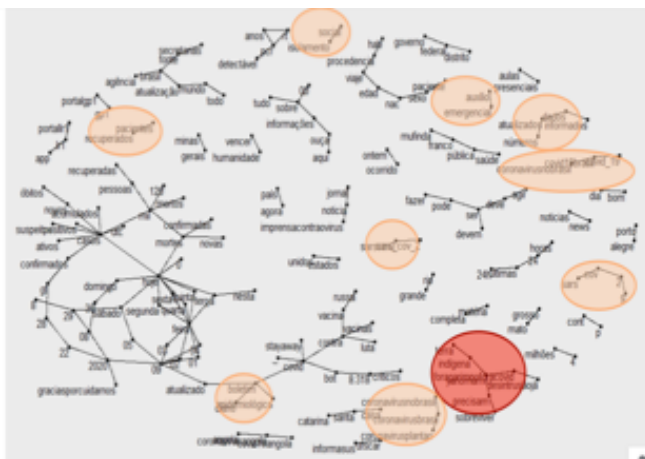


Figura 5 – Rede de Palavras para o conjunto de dados obtido no Twitter, em língua portuguesa, território brasileiro, hashtag #covid19, levantamento realizado dia 15/08/2020. (fonte: autor, 2020).

É possível observar alguns elementos da estrutura do texto. Nesta representação é possível identificar as palavras com as quais cada palavra se relaciona bem como aquelas com as quais ela não se relaciona.

Destacamos:

“jornal-notícia-imprensacontraovirus”,
“minas-gerais”,
“vacina-russa-luta-contra”,
**“4-milhões”, “terra-indigena-foragarimpo-
 yanomani-precisam-sobreviver”,**

dentre outras.

4.6 Contagem de seguidores

Um dos elementos mais importantes para anunciantes, agências de publicidades e analistas políticos é o número de seguidores e de amigos de um determinado usuário do Twitter. O número de seguidores indica a popularidade daquele usuário e é um indicador de sua influência na mídia social.

Já uma baixa proporção de seguidores/amigos sugere que o usuário estaria mais aberto a uma comunicação bidirecional. Na tabela 5 vemos os números para alguns usuários noticiosos selecionados.

Tabela 5 - Contagem de amigos (seguindo) e seguidores, 15/08/2020

screen-name	amigos	seguidores	seguidores/amigos
g1	106	11.609.280	109.522
FoxNews	266	19.648.342	73.866
UOL	96	1.707.056	17.782
BBC	203	1.713.550	8.441
BandJornalismo	156	722.946	4.634
NExoJornal	133	549.437	4.131
RFI	690	2.556.996	3.706
CBS	818	1.120.920	1.370
atribuna	1	1.105	1.105
tvsenado	82	43.090	525
TV5MONDE	2.309	979.114	424
revistaforum	805	315.472	392
franceculture	2.085	700.032	336
JornalGGN	235	67.364	287
DCM_online	1.078	223.590	207

Tabela 5 - Seguidores e razão seguidores/amigos para usuários selecionados. 15/08/2020. (Fonte: autor, 2020).

De forma análoga aos bigramas também podemos considerar os trigramas. Neste caso o conjunto de trigramas, com respectivas frequências está na tabela 4.

4.7 Análise por usuários selecionados

Na sequência apresentamos as nuvens de palavras, histogramas e tuítes mais positivos e negativos para usuários selecionados: @g1, @tvsenado, @NexoJornal. Em primeiro lugar vemos na Figura 6 a nuvem de palavras do @g1 com a hashtag #covid19.

Observe que não há menção a desemprego, descumprimento do isolamento social, crise econômica.

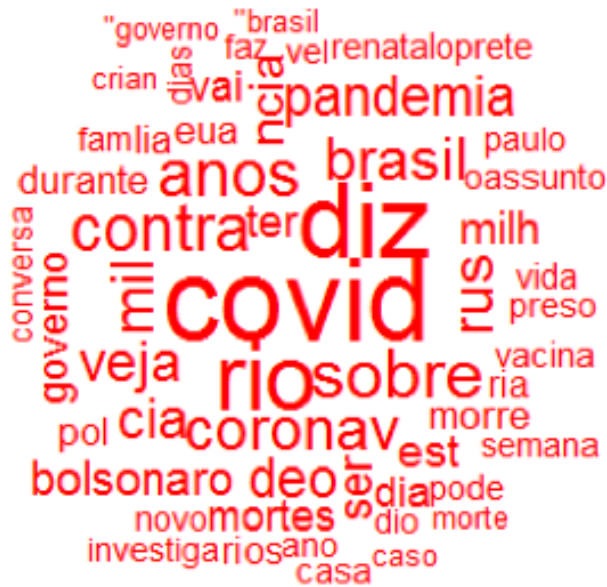


Figura 6 – Nuvem de palavras para o usuário @g1 com a hashtag #covid19, conjunto de dados obtido no Twitter, dia 15/08/2020, em língua portuguesa, território brasileiro. (Fonte: autor, 2020).

Na figura 7 vemos o histograma para os dados do usuário @g1 com o léxico sentiLex_lem_PT02:

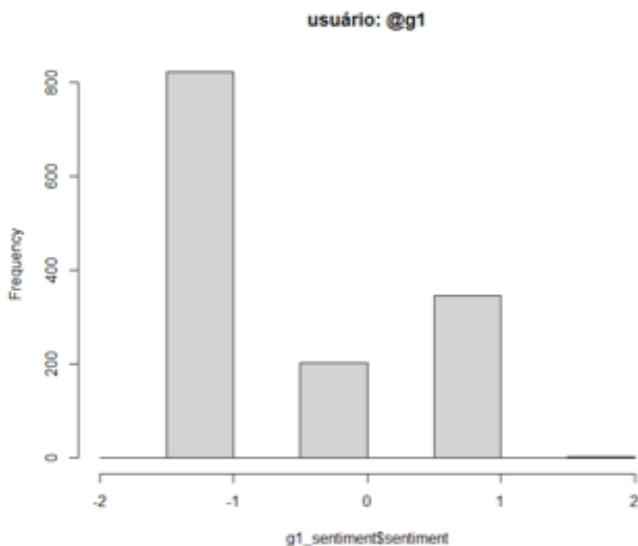


Figura 7 – Histograma de análise de sentimentos para o usuário @g1 com a hashtag #covid19, conjunto de dados obtido no Twitter, dia 15/08/2020, em língua portuguesa, território brasileiro. (Fonte: autor, 2020).

Note o destaque para valores negativos.

Usando esse mesmo léxico o tuíte mais negativo neste conjunto de dados para o usuário @g1 foi: “Padrao é preso suspeito de matar garoto que foi afogado em lama em Goiás.”. Os tuítes considerados os mais positivos neste conjunto de dados do usuário @g1 foram: “Em pronunciamento, Bolsonaro diz ter compromisso com Constituição, soberania, democracia e liberdade” e “Antonio Banderas diz estar curado da Covid-19 após testar positivo”.

Apresentemos agora elementos do usuário @jornalggn.

Na figura 8 vemos a nuvem de palavras. Observe a diversidade de temas exibidos: Banco Pactual BTG e Paulo Guedes, tendo como elemento central as palavras Covid e Brasil.



Figura 8 – Nuvem de palavras para o usuário @jornalggn com a hashtag #covid19, conjunto de dados obtido no Twitter, dia 15/08/2020, em língua portuguesa, território brasileiro. (fonte: autor, 2020).

Na figura 9 vemos o histograma para o usuário @jornalggn.

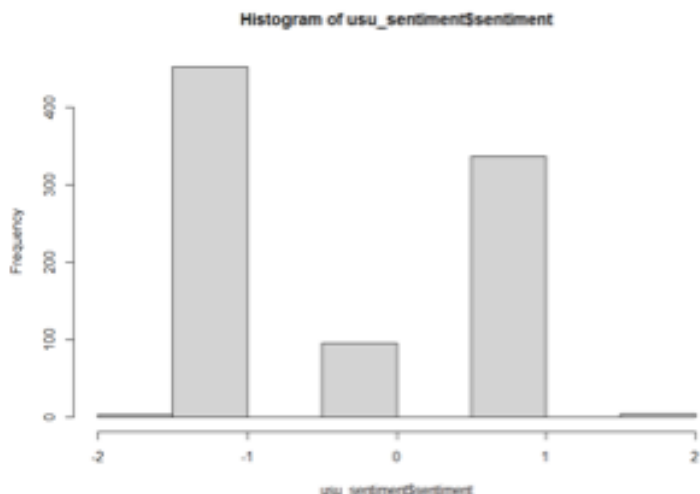


Figura 9 – Histograma para o usuário @jronalggn com a hashtag #covid19, conjunto de dados obtido no Twitter, dia 15/08/2020, em língua portuguesa, território brasileiro. (fonte: autor, 2020).

Neste histograma podemos observar a frequência de notas negativas um pouco acima das notas positivas.

O tuíte mais negativo para o usuário @jornalggn, considerando o léxico sentiLex_lem_PT02 foram: “Como a flexibilização precoce do isolamento está promovendo um desastre nos EUA”. Já o tuíte mais positivo foi “Tal decisão ofende devido processo legal, os tratados internacionais, a Constituição que garantem a liberdade de expressão como direito fundamental.

É um ataque à Democracia e ao Estado de Direito, pois faz da Censura um método para calar quem denuncia.” Vejamos agora a nuvem de palavras para o usuário @tvsenado na figura 10.

Como esperado para este usuário, a palavra com mais destaque é a palavra senado, envolta por palavras como comissão, projeto, senador, presidente, coronav, pandemia e outras.



Figura 10 – Nuvem de palavras para o usuário @tvsenado com a hashtag #covid19, conjunto de dados obtido no Twitter, dia 15/08/2020, em língua portuguesa, território brasileiro (Fonte: autor, 2020).

Na figura 11 observamos o histograma para @tvsenado. Vemos uma maior frequência significativa para as mensagens com palavras associadas a sentimentos positivos.

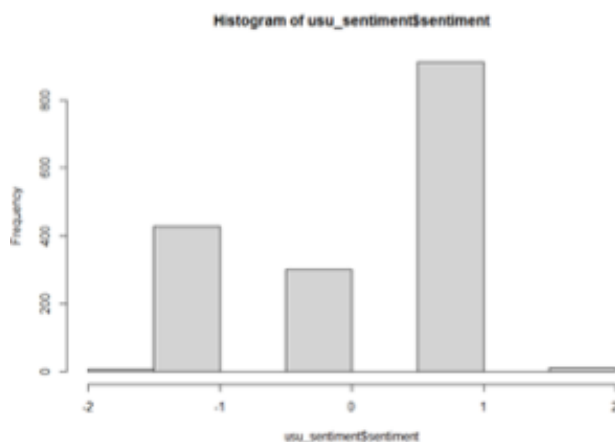


Figura 11 – Histograma para o usuário @tvsenado com a hashtag #covid19, conjunto de dados obtido no Twitter, dia 15/08/2020, em língua portuguesa, território brasileiro. (Fonte: autor, 2020).

Por fim, o tuíte mais negativo neste conjunto de dados, segundo o léxico sentiLex_lem_PT02 foi “@barbosaaps obrigado por sua audiência”. Já o tuíte mais positivo foi: “A sessão foi suspensa por falha técnica, mas já foi resolvido”.

Compilando os dados da análise de sentimentos para cada um dos usuários a seguir: @g1, @dcm_online, @uol, @NexoJornal, @jornalggn, @BandJornalismo, @tvsenado, @revistaforum obtemos a tabela.

tabela 6 - Medidas de tendência central para usuários selecionados

usuário	Mínimo	Mediana	Média	Máximo
@g1	-2	-1	-0,3462	2
@dcm_online	-2	-1	-0,2681	2
@uol	-3	-1	-0,2845	2
@NexoJornal	-4	0	-0,01453	3
@jornalggn	-2	-1	-0,1274	2
@BandJornalismo	-2	0	-0,103	2
@tvsenado	-2	-1	0,2968	2
@revistaforum	-2	0	-0,1239	2

Tabela 6 - Medidas de tendência central para usuários selecionados. Tuítes em língua portuguesa, território brasileiro, 15/08/2020.

O usuário com a média mais negativa foi @g1. O usuário que apresentou tanto o menor mínimo quanto o maior máximo foi @NexoJornal, ou seja, a maior amplitude de sentimentos segundo o léxico adotado. A maior média foi apresentada pelo usuário @tvsenado.

5. Discussões

É importante lembrarmos que as técnicas de Mineração de Texto e Análise de Sentimentos, são quantitativas, mas nem por isso isentas de opções subjetivas por parte do analista. Mesmo que os histogramas apresentados nas figuras 2 e 3 para a Análise de Sentimentos com os dois

léxicos adotados apresentem o mesmo perfil, há sutis diferenças entre eles, resultantes dos conjuntos de palavras e pontuações distintas em cada léxico.

Os tuítes mais negativos ou mais positivos também são função dos léxicos adotados. Além disso, os léxicos também variam conforme a língua. Há mais de um léxico disponível para análises efetuadas em inglês, assim os resultados podem, em princípio, diferir. Uma busca por maior robustez nas avaliações pode ser obtida pelo acompanhamento contínuo ao longo do tempo e por mais de uma mídia social. Assim, o analista se protege de eventuais vieses amostrais.

A partir dos dados apresentados na tabela 6, os blogs noticiosos classificados como sendo abertamente críticos ao atual governo apresentaram nota média superior a do usuário @g1. Também deve ser observado que o usuário @tvsenado (órgão informativo do Senado Federal) apresentou a maior nota positiva, valor discrepante do conjunto de usuários selecionados.

Em pouco mais de quinze anos de expansão contínua da Internet e das Redes Sociais Virtuais foi possível assistir a grandes impactos na forma como as pessoas e organizações se comunicam. Apenas como exemplo deste impacto, podemos citar a geração de novas profissões tais como Analistas de Conteúdo ou Analistas de Redes Sociais: profissionais que acompanham o desempenho de suas organizações nas Redes Sociais Virtuais e buscam multiplicar as notícias com emoções positivas e mitigar danos de eventuais desacertos da organização.

6. Conclusões

Como principais conclusões deste trabalho citamos:

1. A nota média obtida para o usuário @g1 é inferior às notas médias para outros usuários, tradicionalmente associados como críticos do atual governo;
2. O destaque identificado pelos trigramas para a questão da Covid-19 e as terras dos Yanomami. Identificar tendências como esta representa valor para analistas de comportamentos sociais e políticos;
3. A tabela 6 fornece um instrumento, que, se utilizado rotineiramente, permite acompanhar as opiniões e emoções em usuários do Twitter;
4. A tabela 5 fornece um indicador que permite avaliar a atuação a favor de maior bidirecionalidade na comunicação por parte de agentes noticiosos nas Redes Sociais Virtuais.

8. Referências

Livros:

- [1] Amaral, Fernando. Introdução À Ciência de Dados. Mineração de dados e Big Data. Rio de Janeiro: Alta Books, 2016.
- [2] Danneman, Nathan e Heimann, Richard. Social Media Mining with R. Packt Publishing. Birmingham - Mumbai, 2014.
- [3] Lesmeister, Cory e Chinnamgari, Sunil. Advanced Machine Learning with R. Packt Publishing. Birmingham - Mumbai, 2019.

Periódicos:

- [8] Gomide, Camila Sant'anna; Lima, Afonso Augusto Teixeira de Freitas; Gomide, Janaína Sant'anna; Roque, Daiane Medeiros; Silva, Thaís Santos.

O Twitter como instrumento de detecção de epidemias de Dengue e desenvolvimento de Políticas Públicas.

XXXVIII Encontro da ANPAD, RJ 13 a 17 de setembro de 2014. Disponível em:

- http://www.anpad.org.br/diversos/down_zips/73/2014_EnANPAD_APB1234.pdf (acesso 22/09/2020).

Sites:

- [4] Associação Brasileira de Jurimetria: <https://abj.org.br/> (acesso 03/09/2020)
- [5] <https://www2.iel.unicamp.br/rll307/?p=285> (acesso 18/07/2020).
- [6] <https://terracoeeconomico.com.br/usando-r-para-analisar-os-discursos-de-getulio-vargas/> (acesso 10/08/2020).
- [7] file:///C:/Users/Usuario/Desktop/Livros_R_Estatistica/ACMoreno_MSIAD_FAnalise_Sentimentos_R.pdf
- [9] <https://dev.twitter.com/apps> (acesso em 16/08/2020).
- [10] https://rstudio-pubs-static.s3.amazonaws.com/342346_fcc3e8b926c7498285d49099140bd247.html
- [11] <https://abjur.github.io/r4jurimetrics/funcoes-do-tidyr.html> (acesso 15/08/2020).
- [12] <https://p4husp.github.io/material/tutorial11/> (acesso 12/07/2020)
- [13] <https://www.fulljoin.com.br/posts/2019-09-19-modelagem-de-tpicos-com-o-r-analisando-os-textos-de-acordos-comerciais/> (acesso 03/08/2020)
- [14] <https://www.tidytextmining.com/index.html> (acesso 05/07/2020)
- [15] https://www.analyticsvidhya.com/blog/2020/09/how-dbscan-clustering-works/?utm_source=browser&utm_medium=rss_notification&utm_id=25153.339428793497
- [16] Quanteda Tutorials. <https://tutorials.quanteda.io>. (acesso 20/07/2020)

SOME

Social Meeting

Scientific Journal

Scientific editor: Dr. Evandro Prestes Guerreiro



[2021]

Todos os direitos desta edição reservados à ©eSocial Brasil

Rua México, 156 - 121

Pitangueiras - Guarujá SP - 11410-350

+55 (13) 3329-1548

www.socialmeeting.info

contato@socialmeeting.info