



ANÁLISE DE SENTIMENTOS UTILIZANDO DADOS DO TWITTER:

um estudo sobre as reações ao corte de recursos do CNPq

Bianca Savegnago de Mira¹

Resumo: Trata-se de uma pesquisa exploratória cujo objetivo consiste em analisar as reações ao corte de recursos do Conselho Nacional de Desenvolvimento Científico e Tecnológico no Twitter por meio da Análise de Sentimento utilizando o *software Orange Canvas*. A análise classificou os *tweets* em positivos, negativos e neutros e apontou para a predominância de neutralidade. Apesar disso, *tweets* classificados como positivos pela análise, em sua maioria, sequer abordavam o tema o que mostra que as reações exibidas na plataforma flutuaram entre neutras e negativas.

Palavras-Chave: Análise de Sentimentos. CNPQ. Twitter.

1 INTRODUÇÃO

As redes sociais são recursos da *web 2.0* que promovem a participação, colaboração e interatividade entre seus membros. Nesse espaço as trocas info-comunicacionais ocorrem entre diversos atores de forma complexa. Os atores participantes, sejam eles pessoas ou instituições, estão envolvidos nos processos de produção, distribuição e recepção das mensagens (ARAÚJO, 2014). O *Twitter*, rede social criada em 2006 no formato de *microblog*, acompanha o imediatismo informacional proporcionado pelo compartilhamento nas redes e concentra um grande contingente de informações que carregam as reações instantâneas de seus usuários sobre diversos temas.

A plataforma supre a necessidade de uma comunicação ágil e atualizada e seus membros a alimentam diariamente com mensagens que contém desde o simples compartilhamento de links a opiniões sobre as mais diversas temáticas. Monitorar e explorar o conteúdo produzido e compartilhado no *Twitter* se tornou uma tarefa que provoca o interesse de indivíduos, pesquisadores e, especialmente, organizações (ALVES, 2011). Dada a profusão de opiniões expostas no *Twitter* e a possibilidade de recuperá-las e analisá-las automaticamente, a presente pesquisa buscou, de maneira exploratória, analisar os sentimentos e emoções

¹ Universidade Estadual Paulista (UNESP)

expressos pelos *tweets* que continham reações às notícias de corte nos recursos do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

2 METODOLOGIA

Entender o comportamento emocional coletivo expresso em posts, *tweets*, updates, entre outros, se tornou um desafio para o campo de análise computacional (MALINI *et al.*, 2017). Um dos métodos mais utilizados para tal é a análise de sentimento, também chamada de mineração de opinião. Ela provém de um desdobramento da mineração de textos em que o conteúdo textual é classificado pelo sentimento ou opinião emitidos. A classificação costuma dividir-se entre sentimentos positivos e negativos (SANTOS *et al.*, 2010).

Na produção brasileira em Ciência da Informação (CI) a Análise de Sentimento tem sido aplicada aos mais diversos contextos, alguns exemplos são os trabalhos de Souza e Café (2018), Silva e Silva (2019), França *et al.* (2020), Machado *et al.* (2020) e Dutra e Gosling (2021). Em todos estes estudos o uso de ferramentas automáticas foi aplicado total ou parcialmente durante a análise de sentimento. A presente pesquisa utiliza o método aplicado aos *tweets*.

Para a coleta dos *tweets* fez-se uso do *software Orange Canvas* (versão 3.30.2), *software* livre que conta com ferramentas específicas para a extração e processamento de *tweets* e opera com a linguagem *Python*. Previamente foi necessário criar uma conta de desenvolvedor no *Developer Portal*² do Twitter e solicitar a API (*Application Program Interface*). Ao solicitar a API recebe-se uma chave que permite o acesso programático ao Twitter. A coleta de dados foi realizada no dia 18 de outubro de 2021 com a ferramenta *Twitter* no *software Orange Canvas*. No próprio *software* há a informação de que a recuperação de *tweets* pelo API consegue abranger até 15 dias anteriores à data da coleta, no entanto em testes realizados em diferentes dispositivos observou-se que a captura foi de no máximo 10 dias anteriores.

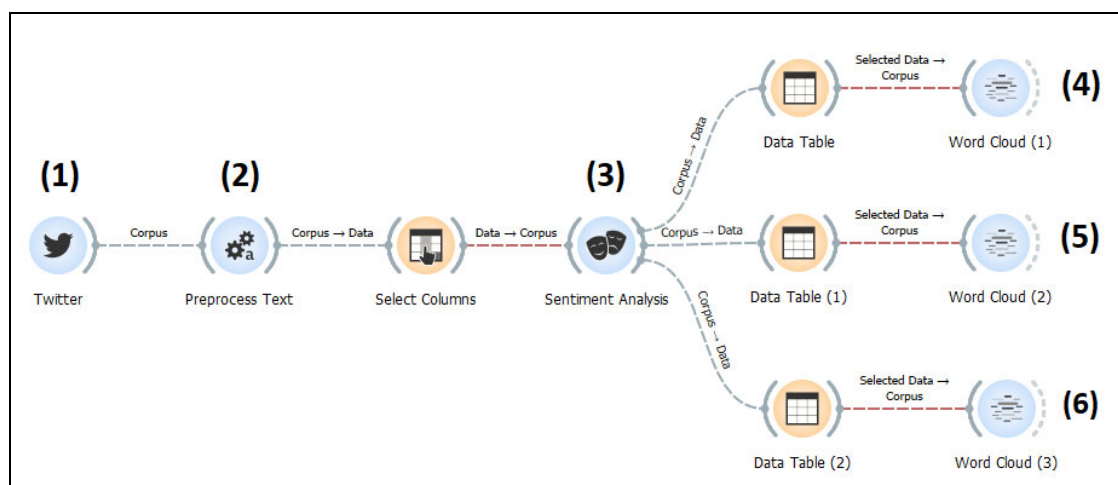
Os *tweets* foram coletados no dia 18 de outubro com abrangência de dez dias. Salienta-se que outras ferramentas gratuitas para a extração de *tweets* também possuem limitação no período que pode ser recuperado, que no geral varia de 7 a 10 dias. O objetivo da coleta foi captar *tweets* sobre o CNPq, a fim de captar reações às notícias publicadas entre os dias 8 e 17 de outubro que comunicavam sobre a aprovação do corte de 92% dos recursos do CNPq pelo Congresso Nacional a pedido do Ministério da Economia e posterior sanção pelo presidente

² <http://developer.twitter.com>

Jair Bolsonaro. O termo de busca escolhido (*query word*) no *software*, foi *cnpq*. A busca foi realizada a partir do conteúdo de *tweets* em qualquer idioma e não incluiu *retweets*, mensagens automaticamente repostadas pelos usuários da plataforma.

Ao todo 1.209 *tweets* foram recuperados automaticamente pelo *software*. Após a recuperação dos *tweets* (1) seguiu-se o pré-processamento para que o conteúdo textual fosse padronizado e permitisse a análise. A Figura 1 apresenta todas as ferramentas utilizadas e o fluxo de trabalho (*workflow*) no *software*.

Figura 1 – *Workflow* no *software Orange Canvas*



Fonte: Elaboração da autora.

O pré-processamento (2) é dividido em três etapas: transformação, tokenização e filtragem. Na transformação os dados de entrada são padronizados. O método de tokenização divide o corpus em um modelo próprio para o *Twitter* que considera *hashtags* e outros símbolos especiais. Já a filtragem é responsável por remover ou manter uma seleção de palavras, neste trabalho foram removidas palavras sem valor semântico para o texto como artigos, preposições e pronomes. Após o pré-processamento o corpus foi submetido às análises que produziram os resultados: a análise de sentimentos (3) e nuvens dela derivadas (4, 5 e 6).

Para a análise de sentimentos foi escolhido o método VADER (*Valence Aware Dictionary for Sentiment Reasoning*) obtido pela biblioteca NLTK (*Natural Language Toolkit*) e disponibilizado nas configurações do *software*. Sua escala varia de -1 a 1, sendo -1 o sentimento mais negativo, 0 neutro e 1 o mais positivo. De acordo com Reis *et al.* (2015) VADER foi criado por Hutto e Gilbert (2014) especialmente para mídias sociais. Seu propósito é avaliar mensagens no contexto do *Twitter* e outras mídias sociais sem a necessidade de treinamento, sua análise considera características próprias desse meio, como, por exemplo, o excesso de pontuação. A escolha de VADER também se pautou no artigo

realizado por um grupo ligado à Universidade de Oxford na Inglaterra. As palavras relacionadas à redução de recursos do CNPq estão presentes, mas em menor frequência. Existem dois *tweets* de um mesmo usuário que abordam a questão e utilizam as palavras *denuncia* e *denunciar* solicitando ajuda para divulgar o acontecimento.

A análise de sentimentos identificou 315 *tweets* cujo conteúdo foi considerado negativo e que serviram de base para a construção da nuvem que os representa. Nela sobressaem as palavras *orçamento*, *corte*, *ciencia*, *cnpq*, *brasil*, *@cnpq_oficial*, *bolsas*, *pesquisas* e *milhões*. As palavras *orçamento* e *corte* emergiram com mais força. Em seu conteúdo, os *tweets* classificados como negativos trouxeram lamentações sobre os corte e preocupações com o futuro da ciência brasileira, principalmente sobre as bolsas e projetos que seriam potencialmente afetados pela medida. Nas extremidades das nuvens é possível visualizar algumas palavras utilizadas pelos usuários cuja conotação é negativa, como *penuria*, *assassinato*, *avassalador*, *crime*, entre outras. A classificação negativa abarcou alguns dos *tweets* com as *hashtags* *#sosciencia* e *#liberafndct* que compuseram o *tuitaço*⁴ organizado pela Sociedade Brasileira para o Progresso da Ciência (SBPC). Essas *hashtags* não possuem nenhuma menção entre os *tweets* positivos.

Na manhã do dia 15 de outubro a SBPC organizou um *tuitaço*, uma espécie de mobilização dos usuários da plataforma para aumentarem a visibilidade da pauta contra a medida adotada. Para isso a entidade incentivou o uso das *hashtags* *#sosciencia* e *#liberafndct* em publicações sobre o tema. A *#liberafndct* trata sobre o Fundo Nacional de Desenvolvimento Científico e Tecnológico (FNDCT) que foi afetado em cerca de 655,4 milhões de reais após a decisão governamental.

As *hashtags* levantadas pelo SBPC também apareceram entre os *tweets* classificados como neutros. A presença delas, inclusive, é mais frequente nesta nuvem do que na nuvem correspondente aos *tweets* negativos. Dos 1209 *tweets*, 815 correspondem aos *tweets* neutros, ou seja, representam 67,4% do total. Parte desses *tweets* apenas reproduziu a notícia veiculada a respeito dos cortes, sem necessariamente emitir uma opinião a respeito ou simplesmente contém o link para a notícia. Chama a atenção que dentre as palavras de menor impacto que figuram próximas à extremidade da nuvem estejam *desmonte*, *rasteira*, *sorradeira*, o *emoji de palhaço* (😂) e, mais ao centro, o símbolo de cruz (†). O conteúdo dos *tweets* que contém estas palavras é predominantemente pessimista.

⁴Fonte: https://www1.folha.uol.com.br/ciencia/2021/10/tuitaco-mobiliza-rede-social-contracorte-de-r-600-milhoes-na-ciencia.shtml?utm_source=whatsapp&utm_medium=social&utm_campaign=compwa

O método VADER pontua de acordo com o léxico e é possível que, dada a limitação do algoritmo em compreender o contexto, estes *tweets* tenham sido computados como majoritariamente neutros. Ademais, também considera algumas características próprias das mídias sociais como excesso de pontuação, por exemplo, mas não realiza a leitura de imagens e é possível que desconsidere alguns símbolos como o de cruz ou o associe a algo positivo. Salienta-se também que a palavra *desmante* apareceu em *tweets* que divulgaram o link de um artigo escrito por Denise Elias cujo título é “CNPQ e a genealogia de um desmante”, ou seja, não existem outras palavras que reforcem uma pontuação negativa para estes *tweets*. O mesmo ocorre com *rasteira*, os *tweets* associados a essa palavra reproduziram notícias sobre o pronunciamento do presidente do CNPq: “Corte de R\$ 600 mi da ciência foi ‘rasteira’ do governo, diz presidente do CNPq” (ELIAS, 2021).

5 CONSIDERAÇÕES FINAIS

Analisar as reações desses usuários é uma forma de compreender como determinado assunto é visto por este recorte específico da população. A presente pesquisa analisou as reações ao corte de recursos do CNPq no *Twitter* por meio da análise de sentimento utilizando o *software Orange Canvas* e o método VADER. Cada *tweet* foi classificado entre positivo, negativo ou neutro de acordo com o conteúdo da mensagem. Os resultados encontrados mostraram que há predominância de neutralidade a respeito do tema, muitos *tweets* assim classificados apenas replicaram informações e notícias. A mobilização de entidades em prol da ciência e contra o corte repercutiu e aumentou a visibilidade sobre o assunto. Apesar da neutralidade, os *tweets* classificados como positivos pela análise, em sua maioria, sequer abordavam o tema. Isso mostra que as reações exibidas na plataforma flutuaram entre neutras e negativas. Os *tweets* cujo teor foi classificado como negativo em sua maioria traziam preocupações e lamentações a respeito das consequências da medida.

Em suma, acredita-se que a análise de sentimento é um método de interessante aplicação ao conteúdo produzido em redes sociais, especialmente no *Twitter*. No entanto há algumas limitações ao aplicá-la utilizando ferramentas que independem de treinamento.

REFERÊNCIAS

ARAUJO, R. F. Aspectos teórico-metodológicos para análise de campanhas eleitorais online no twitter: do mapeamento das redes às conversações. **Liinc em Revista**, Rio de Janeiro, v.

10, n. 2, 2014. DOI: <https://doi.org/10.18617/liinc.v10i2.716>. Disponível em: <http://revista.ibict.br/liinc/article/view/3562>. Acesso em: 15 nov. 2021

DUTRA, L. F.; GOSLING, M. S. Analisando sentimentos no tripadvisor: a voz do consumidor como parâmetro decisório na medição da qualidade da informação em museus. **Informação & Sociedade: Estudos**, João Pessoa, v. 31, n. 1, p. 1-23, 2021. DOI: <https://doi.org/10.22478/ufpb.1809-4783.2021v31n1.54087>.

ELIAS, D. CNPq e a genealogia de um desmonte . **Revista Entre-Lugar**, Dourados, v. 12, n. 24, p. 389–394, 2021. DOI: <https://doi.org/10.30612/rel.v12i24.15319>. Disponível em: <https://ojs.ufgd.edu.br/index.php/entre-lugar/article/view/15319>. Acesso em: 12 nov. 2021.

FRANÇA *et al.* A ciência da informação e a covid-19: análise de sentimentos dos usuários no twitter. **Páginas A&B, Arquivos e Bibliotecas (Portugal)**, Porto, n. 14, p. 3-9, 2020. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/152820>. Acesso em: 10 nov. 2021.

HUTTO, C.; GILBERT, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. **Proceedings of the International AAAI Conference on Web and Social Media**, [S. l.], v. 8, n. 1, p. 216-225, 2014. Disponível em: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>. Acesso em: 15 nov. 2021.

MACHADO, G. B. *et al.* Perspectivas de pesquisa sobre inteligência estratégica antecipativa e coletiva (ieac) por meio da análise de sentimento: um cenário didático de uso. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 10, n. 1, p. 152-164, 2020. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/148563>. Acesso em: 18 nov. 2021.

MALINI, F. *et al.* O sentimento político em redes sociais: big data, algoritmos e as emoções nos tweets sobre o impeachment de Dilma Rousseff. **Liinc em Revista**, Rio de Janeiro, v. 13, n. 2, 2017. DOI: <https://doi.org/10.18617/liinc.v13i2.4089>. Disponível em: <http://revista.ibict.br/liinc/article/view/4089>. Acesso em: 13 nov. 2021.

MATOS, F. F. *et al.* Recuperação e classificação de sentimentos de usuários do Twitter em período eleitoral. **Informação & Informação**, Londrina, v. 25, n. 1, p. 92-114, abr. 2020. Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/35310>. Acesso em: 10 nov. 2021.

REIS, J. *et al.* Uma abordagem multilíngue para análise de sentimentos. *In*: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BRASNAM), 4. , 2015, Recife. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2015. ISSN 2595-6094. DOI: <https://doi.org/10.5753/brasnam.2015.6767>.

SANTOS, L. M. *et al.* Twitter, análise de sentimento e desenvolvimento de produtos: Quanto os usuários estão expressando suas opiniões? **Prisma.com**, Oporto, n. 13, p. 159-170, 2010. Disponível em: <https://ojs.letras.up.pt/ojs/index.php/prismacom/article/view/2005>. Acesso em: 10 nov. 2021.

SILVA, P. M. S. E.; SILVA, A. K. A. Aplicação das técnicas de análise de sentimento em redes sociais informais: revisão bibliográfica sistemática. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO - ENANCIB, 20., 2019, Florianópolis. **Anais [...]**. Florianópolis: [s.n.], 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/124085>. Acesso em: 11 nov. 2021.

SOUZA, R. R.; CAFÉ, L. M. A. Análise de sentimento aplicada ao estudo de letras de música. **Informação & Sociedade: Estudos**, João Pessoa, v. 28, n. 3, 2018. Disponível em: <https://periodicos.ufpb.br/index.php/ies/article/view/34884>. Acesso em: 11 nov. 2021.