



Vinicius José Paes e Silva

Análise de Sentimentos em Tweets Relacionados ao Desmatamento da Floresta Amazônica

Recife

2021

Vinicius José Paes e Silva

Análise de Sentimentos em Tweets Relacionados ao Desmatamento da Floresta Amazônica

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

Recife

2021

Agradecimentos

Agradeço, primeiramente, a minha família, especialmente meus pais, Neildo e Verônica, meus irmãos Lucas e Victor, e minhas irmãs Luna, Laura e Isabel, e meus tios Júlio e Nadja, por sempre acreditarem em mim e sempre me incentivarem e me darem suporte em todos os momentos dessa jornada.

A Danielle Sales, pelo apoio incondicional, pelas noites em claro me dando apoio, por nunca me deixar desistir, por toda a parceria e por sempre me incentivar a ser melhor. Sem ela esse trabalho não seria possível. O seu apoio significa muito para mim.

Aos meus queridos companheiros, Davi, Emmanuel, Fernando e Gustavo, por todos os anos de amizade e por todo o apoio quase diário durante toda essa jornada.

A Ariana Lima, por ter me acompanhado em quase todos os desafios dessa e de tantas outras jornadas, por sempre estar disposta a me ajudar, me apoiar todos esses anos.

As pessoas que eu conheci na faculdade, Bianca, Brunna, Evele, Felipe Morais, Francisco, Gabriel, Hugo, Luiz, Manoel, Marta, Ricardo e tantos outros por dividirem essa luta comigo, por todos os ensinamentos e momentos compartilhados.

Aos meus amigos, Andressa, Cesar, Dandara, Dayse, Felipe Cavalcanti, Filipe Cabral, Giulia, Rodolfo, Sandy e Thiago, por todos os anos de amizade e por sempre me incentivarem a continuar.

E a tantos outros que conheci no caminho e que tornaram essa jornada e minha vida um pouco melhor.

Análise de Sentimentos em Tweets Relacionados ao Desmatamento da Floresta Amazônica

Vinicius J. Paes¹, Ermeson Andrade²

¹Departamento de Estatística e Informática – Universidade Federal Rural de Pernambuco (UFRPE)
Recife – PE – Brazil

²Departamento de Computação – Universidade Federal Rural de Pernambuco (UFRPE)
Recife – PE – Brasil

{vinicius.paes, ermeson.andrade}@ufrpe.br

Abstract. *The Amazon Forest is being devastated at the fastest pace in recent years. In 2021, the Amazon rainforest registers the largest accumulation of deforestation in 5 years, increasing from 13 thousand km² between August 2020 and July 2021. An increase of 22% compared to the same period in the previous year, the highest number since 2006. Although many works address the issue of deforestation, none of them focus on analyzing the sentiments of the Brazilian population regarding the issue. This work presents an analysis of the sentiments of the Brazilian population related to the deforestation of the Amazon rainforest through the text mining of Twitter and aims to understand how Brazilian users opine and dialogue about the deforestation of the Amazon rainforest. The results reveal that Brazilian users tend to react to events related to deforestation in the Amazon forest on Twitter and that most users have a negative sentiment about the topic, reaching peaks of approximately 60% of tweets in a given time.*

Resumo. *A Floresta Amazônica está sendo devastada no maior ritmo dos últimos anos. Em 2021, Amazônia registra o maior acúmulo de desmatamento em 5 anos, passando de 13 mil km² entre agosto de 2020 a julho de 2021. Um aumento de 22% no desmatamento, em relação ao mesmo período do ano anterior, sendo o maior número desde 2006. Embora muitos trabalhos abordem o tema de desmatamento, nenhum deles foca em analisar os sentimentos da população brasileira com relação ao tema. Este trabalho apresenta uma análise dos sentimentos dos usuários brasileiros do Twitter relacionados ao desmatamento da Floresta Amazônica através da mineração de texto do Twitter e busca entender como os usuários brasileiros opinam e dialogam sobre a desmatamento da Floresta Amazônica. Os resultados revelam que os usuários brasileiros tendem a reagir a acontecimentos relacionados ao desmatamento da floresta Amazônica no Twitter e, que em sua maioria, os usuários apresentam sentimento negativo sobre o tema, alcançando picos de aproximadamente 60% dos tweets em determinado momento.*

1. Introdução

A preservação ambiental tem sido um tema recorrente entre os principais polos de discurso mundial, seja no âmbito político, acadêmico, ou em conversas do cotidiano. É

praticamente impossível passar despercebido pelas mudanças climáticas que vêm assolando o nosso planeta. Com isso, o aquecimento global vem se tornando um dos assuntos mais discutidos mundialmente, e a preservação das florestas andam lado a lado com esse tema. O Brasil possui a maior floresta tropical do mundo, a Floresta Amazônica, a qual vem sendo alvo de ações ambientais e programas de preservação, ao mesmo tempo que sofre ações de desmatamento ilegais como queimadas e grilagem. Segundo o INPE (Instituto Nacional de Pesquisas Espaciais) [INPE 2020], estima-se que cerca de 17% do total do bioma amazônico tenha sido desmatado até o ano de 2020. Esses fatos tem levantado debates acirrados sobre a importância da floresta, como é a forma correta de defender e sobre a eficácia de políticas públicas de prevenção ao desmatamento.

Durante os últimos anos, diversos eventos ocorreram em torno do desmatamento da Amazônia, os quais vêm ascendendo bastante debate nas redes sociais sobre esse tema, como novas políticas públicas, queimadas ilegais, substituição do ministro do meio ambiente, pressão da comunidade internacional, etc. Alinhado ao aumento do uso das redes sociais durante a pandemia do COVID-19 [Moreira and Pinheiro 2020], esses debates têm ganhado cada vez mais espaço e visibilidade. Conseqüentemente, muitos dados são gerados, em formato de texto, vídeo e áudio através das redes sociais, como o Twitter. No entanto esses dados, em sua grande maioria, não são estruturados. Isto é, são textos de opinião e não classificados.

Uma das técnicas utilizadas em *Big Data* é a análise de sentimentos. Essa técnica utiliza a mineração de texto e processamento de linguagem natural para extrair informações no âmbito de sentimentos e emoções. É possível classificar os textos em positivos, negativos e neutros e também determinar as emoções encontradas, como raiva, medo, confiança, entre outras. Esse tipo de análise permite uma melhor classificação de críticas, opiniões e posicionamentos sobre um determinado assunto. Se utilizando dessa técnica, é possível extrair informações sobre os debates acerca do desmatamento da floresta e das políticas que foram e estão sendo implementadas para prevenir o desmatamento. Alinhando técnicas de mineração de textos em redes sociais à análise de sentimento, é possível extrair opiniões com o objetivo de identificar o sentimento dos usuários brasileiros do Twitter sobre esse tema. Assim, este trabalho tem como objetivo identificar o sentimento dos usuários brasileiros do Twitter sobre o desmatamento através dos conteúdos de *tweets*, entre 1 Janeiro de 2019 até 30 de Agosto de 2021. Para alcançar esse objetivo, foram coletados aproximadamente 252 mil *tweets*, contendo o termo “Amazônia” simultaneamente com o termo “queimadas” ou “desmatamento”. Foi possível observar que o sentimento predominante durante a maior parte da duração pesquisa foi o negativo, representando 49,7% dos *tweets* analisados, em segundo lugar os *tweets* neutros, com 28,97% dos *tweets* e, por último, o sentimento menos presente foi o positivo, com 21,34% dos *tweets*. Durante a análise se observou que eventos como queimadas ou mudanças em políticas públicas ocasionam um aumento no numero dos *tweets*, os quais geralmente são seguidos por um aumento na presença do sentimento negativo, assim como de emoções negativas.

O trabalho está organizado da seguinte maneira: a Seção 2 apresenta informações e conceitos básicos sobre técnicas e teorias utilizadas na elaboração e execução deste trabalho. A Seção 3 detalha os trabalhos relacionados e referências bibliográficas. A Seção 4 demonstra a metodologia utilizada para o desenvolvimento e obtenção dos resultados

deste trabalho. A Seção 5 detalha os resultados obtidos. Por fim, a Seção 6 apresenta as conclusões e sugere os trabalhos futuros.

2. Fundamentação

Nesta seção serão apresentadas definições e breves explicações sobre conceitos básicos explorados neste trabalho.

2.1. Big Data

Big data é o termo utilizado para descrever volumes de dados enormes e que, normalmente, não são estruturados. Esses dados podem vir das mais diversas fontes, como planilhas, sites, redes sociais e dispositivos inteligentes [Sagiroglu and Sinanc 2013]. Se existe uma forma de coletar e distribuir dados em grande escala, é possível classificar como *Big Data*. Um dos maiores desafios desse cenário é a classificação e o tratamento desses dados, pois muitas vezes eles vêm de contextos onde se usa linguagem natural ou informações não estruturadas.

No contexto deste trabalho, os dados são obtidos de uma rede social através de textos apresentados em formato de *microblog* e escritos em linguagem natural, os quais, frequentemente utilizam de figuras, *emojis*, e, muitas vezes, contendo *links* misturados ao conteúdo dos textos, resultando na necessidade de um processo delicado de limpeza dos dados. É necessário, para a padronização dos textos, a extração dos *links*, a remoção de figuras e *emojis* e a remoção de palavras que não trazem conteúdo ao texto, conhecidas como *stopwords*, como, por exemplo, artigos e preposições. Consequentemente, o uso de técnicas de *Big Data* como mineração de texto são imprescindíveis.

2.2. Mineração de texto

Mineração de texto é uma técnica dentro da área de mineração de dados que tem como objetivo extrair informação a partir de textos diversos em linguagem natural. É o processo de transformar textos sem estrutura em padrões de informações [IBM 2020]. Essa técnica possui diversas aplicações, como categorização, detecção de padrões em literatura, identificação de temas, sumarização e análise de sentimentos. O texto pode ser extraído de diversas fontes, como documentos, planilhas, livros, redes sociais e diversos outros materiais de literatura que haja dados passíveis de padronização. E, para realizar essa extração, existem diversas técnicas como *web scraping*, que percorre páginas *web* recolhendo textos de determinado filtro, ou a simples leitura e extração organizada de uma planilha de inventários de uma organização.

Uma técnica comumente usada na área de mineração de textos e utilizada nesse trabalho é a *tokenização*. Essa técnica consiste em dividir pedaços de texto em sentenças individualizadas para facilitar e refinar o trabalho de análise [Gupta and Lehal 2009].

2.3. Análise de sentimentos

A análise de sentimento tem como objetivo estudar, extrair e quantificar informações sobre emoções subjetivas e sentimentos contidos em um documento ou texto para melhor entender seu impacto dentro de uma determinada comunidade.

O processo de análise de sentimento consiste, primeiramente, na extração do texto, sendo seguido pelo processo de *tokenização*, citado na Seção 2.2, no qual as sentenças do

texto são separadas e essas sentenças, ou *tokens*, recebem uma pontuação [Pozzi et al. 2016]. Essa pontuação pode ser estabelecida a partir de algumas técnicas distintas. Duas comumente utilizadas são o aprendizado de máquina e a classificação léxica. Para fins desse trabalho, utilizamos a classificação léxica, que consiste em utilizar um dicionário de palavras previamente pontuadas para definir o sentimento do *token*, ou sentença, e detectar a presença de emoções dentro do texto. Uma vez que a pontuação foi definida, é possível determinar qual sentimento é predominante dentro de uma sentença. E, assim, somando a pontuação de todas as sentenças presentes num texto, pode-se determinar seu sentimento predominante, assim como as emoções presentes no mesmo.

2.4. Twitter

O Twitter é uma rede social focada na facilitação da comunicação e conexão entre pessoas através da troca de mensagens rápidas e frequentes [Twitter a]. Nessa rede, as pessoas possuem perfis que funcionam como um *microblog*, onde elas podem expressar opiniões e ideias através de textos curtos, limitados a 280 caracteres e chamados de *tweets*. Esses *tweets*, segundo a política de dados e privacidade do Twitter, são públicos, e podem ser acessados por qualquer pessoa do mundo, sem a necessidade de possuir uma conta na plataforma. Cada usuário, por sua vez, tem a possibilidade de configurar seus próprios *tweets* como privados, e possui um nome único, identificável e pesquisável dentro da rede social.

O Twitter permite, ainda que os assuntos de seus *tweets* possam ser classificados através do símbolo “#”, chamado de *hashtag*. As *hashtags* são utilizadas pelo Twitter como uma forma de indexação e classificação das mensagens para facilitar sua busca. Porém, é possível utilizar qualquer elemento presente em um *tweet*, como nome de usuário, *hashtag* ou até mesmo uma palavra ou frase contida no texto para realizar uma busca. Segundo a última divulgação [Twitter b], o Twitter possui aproximadamente 211 milhões de usuários ativos em diversos países e suporta mais de 37 línguas, incluindo o português brasileiro.

3. Revisão Bibliográfica

A análise de sentimento é uma área complexa e que vem sendo cada vez mais utilizada e incrementada com o passar dos anos. Vários trabalhos vêm explorando não só as diversas técnicas que podem ser utilizadas nessa área, mas também os diversos propósitos empregados a ela. Embora existam muitos trabalhos que focam na análise de sentimentos, nenhum deles se concentra em analisar os sentimentos dos brasileiros em relação ao desmatamento na floresta amazônica.

Em [Pozzi et al. 2016], os autores apresentaram um estudo sobre as diversas técnicas aplicáveis em análise de sentimento dedicadas a redes sociais. São explorados os diversos desafios encontrados na área e também é levantando as melhores técnicas para tratamento e limpeza de dados, classificação léxica e processamento de linguagem natural, sempre com o foco no contexto do tipo de linguagem utilizada nas redes sociais. O trabalho traz experiência para o processo de mineração e classificação, podendo ser utilizado como um material de apoio nesse processo.

Em [Brum et al. 2020], os autores apresentam um estudo sobre o processo de caracterização de textos, onde técnicas de análise de sentimento na rede social *Twitter*

são usadas com o intuito de classificar os *tweets* relacionados a pandemia da COVID-19. Além disso, é elaborado um entendimento do sentimento geral da população sobre os acontecimentos marcados durante a pandemia. O estudo mostra que os temas mais discutidos no período foram relacionados a quarentena, hidroxiclороquina, aglomeração e distância social. Adicionalmente, essas discussões foram focadas em questões políticas.

Em [Euzebio et al. 2020], os autores apresentam uma análise estatística em textos do *Twitter* a fim de identificar surtos de dengue. O trabalho reúne técnicas de classificação de textos e análise estatística no processo de identificação de dengue para cidades de médio porte. Em [Cristiani et al. 2020], os autores aplicam o processo de análise de sentimentos no *Twitter* a fim de classificar e entender as opiniões e os sentimentos da população em relação as eleições de 2018 no Brasil. Embora o trabalho não apresente nenhuma análise estatística, os autores afirmam que a análise dos *tweets* positivos para os candidatos possuem uma forte relação com as pesquisas oficiais realizadas por empresas de pesquisa de opinião.

Em [Silva et al. 2021], os autores aplicaram o processo de análise de sentimentos na rede social *Twitter* a fim de entender as opiniões e os sentimentos da população em relação a atuação do SUS durante a pandemia de COVID-19. O trabalho relata os desafios da metodologia adotada, assim como diversas informações sobre o processo de classificação e análise de *tweets*. Além disso, o trabalho declara que o mesmo traz informações de valor sobre o SUS tanto para grupos políticos como para a própria população, que podem utilizar-se dos dados do trabalho para entender melhor o que outras pessoas acham sobre diferentes pontos relacionados ao tema SUS.

4. Metodologia

Este trabalho almeja identificar o sentimento dos usuários brasileiros do Twitter sobre o desmatamento na Amazônia através da rede social Twitter. Para tal, foram realizadas uma série de etapas que serão apresentadas a seguir.

4.1. Coleta e tratamento dos dados

Os dados foram coletados em Agosto de 2021 e foram reunidos aproximadamente 252 mil *tweets*, datados entre 1 de Janeiro de 2019 a 21 de Agosto de 2021. A coleta foi realizada utilizando a biblioteca para Python *Twint* [Twint 2021], juntamente com a biblioteca *Jupyter Notebook* [Jupyter 2021]. Para a coleta, foram filtrados os *tweets* que contivessem os termos “Amazônia” em conjunto com os termos “desmatamento” ou “queimadas”. Essa decisão foi tomada para filtrar *tweets* não relacionados a floresta Amazônica, uma vez que o nome Amazônia está presente em produtos e marcas diversas. Essa regra restringe a pesquisa, porém garante que os resultados estão de acordo com o esperado para o escopo.

Devido a limitações impostas pelo próprio *Twitter*, as coletas foram realizadas em períodos de duas semanas com uma pausa de alguns minutos no tempo na execução do código, evitando o bloqueio do Twitter. Uma vez coletados, os *tweets* passaram por um processo de limpeza, que consistiu em: remoção de acentuações, remoção de *links*, remoção de *emojis* e caracteres especiais e, por fim, todas as palavras foram passadas para caracteres minúsculos.

4.2. Classificação

Para a classificação dos sentimentos dos *tweets*, escolheu-se utilizar a classificação com *NRC sentiment* da biblioteca *Syuzhet* [Syuzhet 2021] em R. Essa biblioteca foi escolhida porque é uma das poucas que pode ser utilizada em português sem a necessidade de ferramentas adicionais, além de possibilitar a análise de emoções mais complexas como “medo” e “alegria”. Um dos maiores desafios durante essa fase foi encontrar uma biblioteca capaz de trabalhar com análise léxica de sentimentos sem a necessidade de ferramentas de tradução e durante a pesquisa apenas a *Syuzhet* se mostrou eficaz. Além disso, de acordo com [Misuraca et al. 2020], *Syuzhet* possui uma das melhores acurácias em termo de predição dos sentimentos.

Na classificação, cada *tweet* é separado em sentenças chamadas de *tokens*, e cada *token* recebe uma pontuação para cada sentimento e emoção encontrada. Uma vez pontuadas as frases, é gerada uma pontuação geral para o *tweet*, que determina sua polaridade, entre “positivo”, “negativo” e “neutro”. A polaridade “neutra” acontece somente nos casos em que existe um empate de pontuações entre os *tokens* “positivos” e “negativos” no *tweet*. Os *tweets* em que a pontuação negativa foi encontrada em maior quantidade foram classificados como “negativos” e os *tweets* onde a pontuação positiva foi encontrada em maior quantidade foram classificados como “positivos”. Além da polaridade dos *tweets*, foi possível identificar emoções. Das 10 emoções disponíveis na biblioteca *Syuzhet*, foram escolhidas 8 para serem utilizadas nessa pesquisa, baseado em sua relevância, que foram: “raiva”, “antecipação”, “nojo”, “medo”, “alegria”, “tristeza”, “surpresa” e “confiança”. Um mesmo *tweet* pode possuir mais de uma das emoções descritas, por este motivo, durante a classificação e análise, optou-se por utilizar como parâmetro a frequência em que essas emoções aparecem nos *tweets*.

4.3. Análise

Para realizar a análise, o primeiro passo foi organizar os resultados das análises de sentimento e polaridade em gráficos. Com esses resultados organizados, foram definidos e identificados períodos em que existiram um aumento significativo na frequência de *tweets* sobre o assunto em foco. Esses períodos foram Janeiro de 2019, Agosto de 2019, Abril a Maio de 2020 e Abril a Maio de 2021. Esses agrupamentos foram utilizados para definir em quais períodos seriam geradas as nuvens de palavras e os gráficos com emoções. Para a geração de nuvens de palavras, foram identificados os termos mais frequentes dentro do período determinado. Termos são considerados até duas palavras que aparecem frequentemente em conjunto. Um exemplo é o termo “impedir desmatamento”, o qual aparece como um dos termos mais frequentes em todos os períodos analisados. Para essa análise e criação das nuvens de palavras foram utilizadas as bibliotecas *TextBlob* [TextBlob 2021] para processar a frequência dos termos dentro dos textos e *wordcloud* [WordCloud 2021] para gerar as nuvens de palavras, ambas bibliotecas *Python*. Com os gráficos e as nuvens de palavras geradas, foi realizada uma pesquisa traçando paralelos entre os resultados demonstrados nos gráficos com notícias e acontecimentos reportados pela mídia. De forma a comprovar a relação entre o aumento da frequência dos *tweets* e o aumento de emoções específicas com as notícias e os eventos.

5. Resultados

Após a coleta e processamento dos dados definidos na **Seção 4.1**, foi iniciado o processo de análise dos mesmos. Durante o decorrer desta pesquisa foi possível observar que a movimentação ou a comoção social sobre o tema Amazônia decorre majoritariamente em consequência de acontecimentos marcantes nesta região. É possível observar nas Figuras 1 e 2 que a frequência de *tweets* e a presença de sentimentos nos *tweets* cresce de forma significativa em datas e momentos específicos ao longo dos anos.

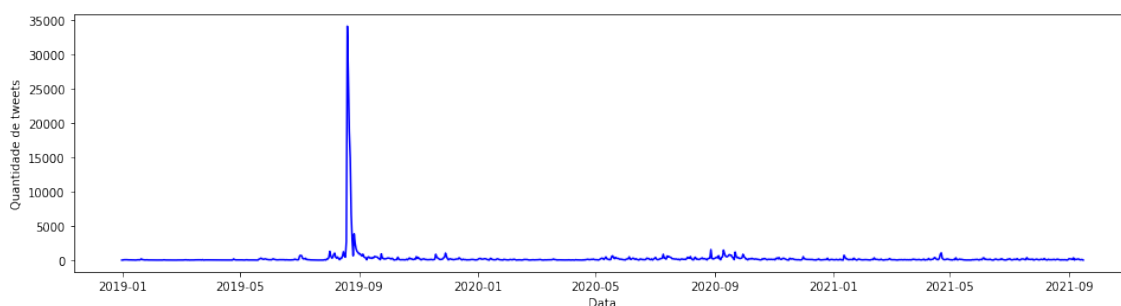


Figura 1. Frequência de *tweets* ao longo do período pesquisado de Janeiro de 2019 a Setembro de 2021.

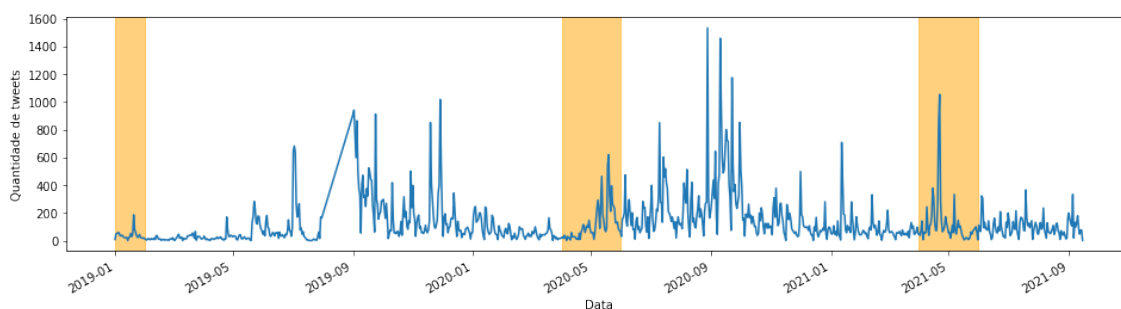


Figura 2. Frequência de *tweets* ao longo do período pesquisado de Janeiro de 2019 a Setembro de 2021, sem considerar o pico de Agosto de 2019.

É possível observar na Figura 3, que durante o período de Janeiro a Junho de 2019, o sentimento predominante era o neutro. Esse sentimento corresponde a *tweets* com notícias de veículos jornalísticos ou usuários que *retweetam* esses conteúdos, como é visto na Figura 5. Entretanto, em Agosto de 2019 pode-se ver uma grande mudança no padrão, com um aumento significativo na frequência de *tweets* sobre o tema, alcançando um pico recorde de 35000 *tweets* em um único mês (ver Figura 1). Além de mudanças na distribuição de sentimentos, os *tweets* negativos se tornaram predominantes e houve um aumento considerável nos *tweets* que apresentam a emoção raiva. Observando o gráfico presente na Figura 4, é possível notar que o aumento da emoção "raiva" coincide com o aumento da frequência de *tweets* presentes na Figura 1.

No dia 10 de Agosto de 2019 ocorreu o chamado "dia do fogo" [Machado 2019], quando se iniciou uma série de focos de incêndios na Amazônia, que chegaram a ter consequências em São Paulo, onde no dia 19 de Agosto [G1 2019] escureceu no início da tarde, graças à fumaça das queimadas na região amazônica. Nesse mês se observou uma mudança no padrão de comportamento que havia anteriormente, não ape-

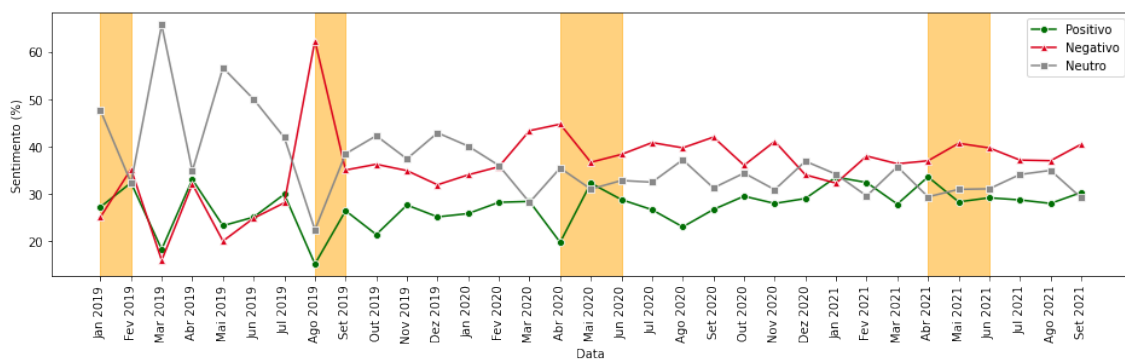


Figura 3. Proporção dos sentimentos em tweets ao longo do período pesquisado de Janeiro de 2019 a Setembro de 2021.

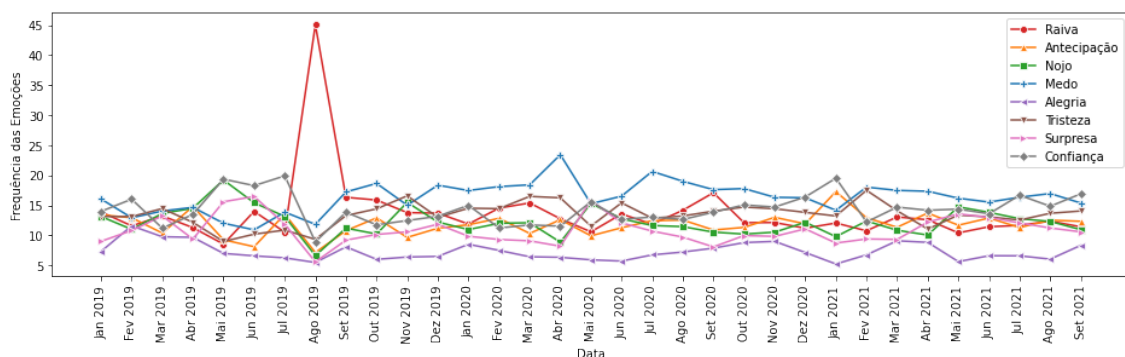


Figura 4. Emoções presente nos tweets durante o período pesquisado de Janeiro de 2019 a Setembro de 2021.

TWEET: O governo cortou em 58% a verba para contratação de profissionais para prevenção e controle de incêndios florestais entre 2019 e 2020, apesar do aumento das queimadas na Amazônia e no Pantanal. <https://t.co/iHe6pb302A>

Figura 5. Exemplo de tweet com alta repercussão.

nas pelo número de *tweets*, mas também pela atitude dos usuários que passou de uma observação passiva, apenas compartilhando notícias sobre o tema, para cobranças e debates sobre o assunto. É possível observar dentre os *tweets*, organizações sobre abaixo assinados, manifestações e outras movimentações organizadas que aumentaram a quantidade de mensagens sobre o tema. Como exemplo, o *#ActForAmazon* que movimentou usuários brasileiros e internacionais durante o período de Agosto de 2019.

Essas mudanças ocasionaram um aumento do fluxo de *tweets* sobre o tema que persiste até o fim do período analisado. Esse ponto pode ser considerado um marco dentro da pesquisa. Até esse momento, a maior parte dos *tweets* não emitiam opiniões, apenas compartilhavam *links* e notícias relativas ao assunto. Porém, a partir de Agosto de 2019, com todos esses acontecimentos, é possível notar que não existe mais um pico de *tweets* neutros como havia anteriormente. Nos meses subsequentes, os usuários tenderam a se posicionar, de fato, sobre o tema, especialmente quando ocorre algum evento relevante relacionado ao desmatamento da Amazônia.

A partir da análise anterior e ao analisar *tweets* mais populares, com grande número de *retweets* e *likes*, é possível observar que a raiva contida nos *tweets* não é direcio-

nada à Amazônia, mas sim aos acontecimentos e políticas públicas relacionadas à floresta. O que é um resultado esperado baseado nos termos da pesquisa, e mostra com clareza a insatisfação dos usuários com relação aos acontecimentos relatados. Muitos dos *tweets* mais curtidos e *retweets* reclamam de como os diversos momento de crises ambientais presentes no decorrer dessa pesquisa são abordados, como os exemplos das Figuras 5 e 6.

TWEET: Latifundiários, donos de gigantescas áreas desmatadas na Amazônia, querem ampliar suas propriedades roubando terras públicas para promover mais desmatamento. Não podemos aceitar! Grilagem é crime e promove a destruição de nossas florestas. #PL510N ao #grilagemnao <https://t.co/gzYcLd1c1r>

Figura 6. Exemplo de *Tweet* com alta repercussão.

Ainda é possível notar que os *tweets* com sentimentos positivos representam usuários cujo foco principal é mostrar a importância da Amazônia para o ecossistema brasileiro e global. Esses usuários se portam de forma didática nas redes dando preferência a conteúdos com âmbito educacional e interações que visam sentimentos mais positivos, como confiança.

Como é possível ver na Figura 7, os termos mais presentes durante todo o período da coleta, na sua maioria, são termos de denúncia de crimes cometidos na Amazônia, como os termos negativos “desmatamento”, “exploração”, “queimadas” em conjunto com os termos de ação “impedir”, “petição”, o que nos indica que boa parte do Twitter interage com esse tema em tom de denúncia e cobrança. É possível, ainda, observar o termo “changebr” aparecendo na nuvem de palavras, que é um conhecido site de abaixo assinados.



Figura 7. Nuvem de palavras representando os termos mais utilizados por usuários durante o período de janeiro de 2019 a agosto de 2021.

A fim de identificar outros períodos relevantes para a análise, retiramos o pico de Agosto de 2019, conforme apresentado na Figura 2. Baseado nessa figura e na Figura 1,

escolhemos para estas análises os seguintes períodos para um melhor entendimento dos sentimentos dos brasileiros relacionados ao desmatamento: Janeiro de 2019, Agosto de 2019, Abril a Maio de 2020 e Abril a Maio de 2021. As análises desses períodos são detalhados a seguir.

5.1. Janeiro de 2019

Nesse período, é possível observar na Figura 3 que o sentimento majoritário é o neutro e, por mais que seja possível identificar diversas emoções, como demonstrado na Figura 4, nenhuma se destaca sobre as outras. Esse período é caracterizado por uma baixa quantidade de *tweets* sobre o assunto Amazônia, e é o início do nosso período de estudo. Os *tweets* desse período consistem, majoritariamente em notícias e *retweets* de notícias sem uma opinião anexa. A maioria das interações nesse período foram feitas por veículos jornalísticos e relatavam dados sobre desmatamento mensal e queimadas. Sem muita interação dos usuários sobre os assuntos. Nesse período, a polaridade negativa e positiva dos usuários era extremamente similar, alcançando máxima de 33% de *tweets* com sentimentos negativos e 31% de positivos, e mínima de 18% de *tweets* sentimentos negativos e 16% com positivos.

5.2. Agosto de 2019

O período de Agosto de 2019 foi um ponto crítico observado durante esse trabalho, em que é possível observar um aumento significativo na frequência de *tweets*, tendo sido registrado um pico recorde de 35000 em um único mês, o maior registrado durante o período analisado. Além disso, houve mudanças na distribuição de sentimentos que passaram de maioria neutros com cerca de 65%, para 60% de *tweets* negativos. Ainda foi registrada a maior baixa de sentimentos positivos, com apenas 12% dos *tweets* nessa categoria.

No dia 10 de Agosto de 2019 foi registrada uma série de incêndios na Floresta Amazônica, iniciando um período recorde de queimadas. Veículos de mídia do mundo inteiro noticiaram os acontecimentos e, nas redes, os usuários acompanharam e dissertaram diversas opiniões sobre os acontecimentos. Muitos usuários e organizações cobraram ação do governo e ajuda de entidades responsáveis, várias petições foram criadas para tentar trazer visibilidade ao assunto e arrecadar fundos. Esse movimento fica visível quando observamos a nuvem de palavras do período na Figura 8.

Além do aumento do sentimento negativo, o padrão de emoções encontrado nos *tweets* também mudou. Nos períodos anteriores era possível encontrar um equilíbrio entre as emoções, com a emoção de “confiança” sendo a de maior ocorrência. Agora, depois do dia do fogo, a emoção mais presente nos *tweets* do período se tornou “raiva” com aproximadamente 56,81% de presença nos *tweets* do período. Além disso, foi possível notar uma baixa histórica na presença da emoção “alegria”, com aproximadamente 9% de aparições como mostra a Figura 4, indicando não só um descontentamento, mas também uma comoção dos usuários.

Ao analisar os *tweets* desses usuários, é possível perceber que o sentimento de raiva e o descontentamento é direcionado a falta de políticas públicas para lidar com os acontecimentos registrados no dia 10. Também é possível observar a presença de discussões políticas sobre o posicionamento do atual governo e oposição, como demonstra a Figura 10 com os 5 *tweets* com mais *retweets* do período. Essas discussões demonstram



Figura 8. Nuvem de palavras representando os termos mais utilizados por usuários durante o mês de agosto de 2019.

um tom mais agressivo de acusação e questionamentos sobre quem seriam os culpados pelos acontecimentos reportados.

- TWEET: BH amanheceu doente pela fumaça dos incêndios em Nova Lima, Sabará e na Serra da Moeda. Quando falo da urgência da #Grev eGlobalPeloclima, não é só sobre a Amazônia. Contra o desmatamento q asfixia o Brasil, amanhã estaremos nas ruas! Na Praça Sete, às 17h. <https://t.co/RBS3hqnf9X>
- TWEET: Bolsonaro sugeriu que as mesmas pessoas que têm filmado queimadas na Amazônia estariam por trás dos incêndios. <https://t.co/D33jmC4ML4F>
- TWEET: A esquerda não entendeu que perdeu às eleições, em oito meses de governo, deixaram estragos com ações terroristas: 1) Lula livre; 2) educação; 3) queimadas na Amazônia. Qual será a próxima?
- TWEET: Quem riscou o fósforo? As queimadas na Amazônia já estão no centro de uma crise internacional e Bolsonaro é responsável, sim. O governo pode se esquivar o quanto quiser, mas há fuligem nas mãos do presidente e seus ministros. Como? Segura que vem thread. <https://t.co/LckQaWeQ8d>
- TWEET: A Amazônia sofre o maior ataque orquestrado por organizações internacionais da história. As queimadas são o pano de fundo para o real interesse de quem quer a biodiversidade e a riqueza mineral da região. É triste ver brasileiros a favor Macron e contra o Brasil. #MacronLies

Figura 9. Tweets com maior repercussão durante o período de agosto de 2019.

5.3. Abril a Maio de 2020

Outro período observado durante o trabalho foram os meses de Abril a Maio de 2020, sendo Abril o primeiro mês estudado durante a pandemia do COVID-19. Neste período, os veículos de mídia questionavam o impacto da pandemia no que diz respeito ao desmatamento da Floresta Amazônica. Simultaneamente com as discussões sobre os impactos da pandemia, em Maio foi reportado o maior recorde de desmatamento até aquele mês, com um índice de alerta de desmatamento 78% maior que o mesmo período no ano anterior [Oliveira 2020]. Em paralelo, o número de multas a municípios com altos índices de desmatamento alcançou recordes de baixa, cerca de 40% a menos que o mesmo período no ano anterior [Muniz et al. 2020].

O sentimento dos usuários acompanhou a tendência reportada em agosto de 2019, que reagiu com um aumento significativo na frequência de tweets, demonstrado nas Fi-

guras 1 e 2, e um aumento no sentimento “negativo” de 30% para 45%, demonstrado na Figura 3. Apesar de menor do que o pico apresentado em Agosto de 2019, ainda representa um aumento, comparado com os períodos apresentados entre as duas datas. Essa nova média do sentimento negativo se torna, daqui pra frente, o novo padrão durante o resto do estudo, se mantendo entre 38% e 45%. Vale ser destacado que nesse período é apresentada a última grande baixa no sentimento “positivo” durante o estudo, com um valor de 22%. Nesse período as emoções que mais se destacaram foram “medo”, “tristeza” e “confiança”, como demonstrado na Figura 4. É possível que durante o período de pandemia o medo se torna a emoção mais predominante entre os usuários brasileiros no que se refere ao desmatamento da Amazônia.

É possível observar uma mudança nos termos mais comentados do período, conforme mostrado na nuvem de palavras presente na Figura 11. Assim como nos cinco *tweets* com mais *retweets* do período detalhado na Figura 10. As discussões na rede tomam um foco mais político, é possível observar uma diminuição em ações, como petições e abaixo assinados, assim como movimentos ambientalistas. Esse período, assim como o dia do fogo em Maio de 2019, repercutiu mundialmente, trazendo questionamentos sobre os esforços prestados pelo governo vigente no combate aos desmatamento e as queimadas durante o período da pandemia do COVID-19. Essa discussão trazida pela mídia e pelos órgãos internacionais é o tema mais discutido no momento, associado a queimadas na Amazônia. Os usuários demonstram insatisfação com as medidas tomadas desde o período passado e com os novos números recordes. O medo das consequências do desmatamento também é um assunto recorrente entre os usuários. É possível observar uma amenização no tom dos debates, o clima de guerra e troca de acusações é substituído por cobrança de ações mais imediatas e direcionadas ao governo vigente.

TWEET: Mesmo quando a #pandemia freia a economia, o garimpo e o #desmatamento ilegal em terras indígenas na #Amazônia permanece m a todo vapor. Artigo de @deiarauscher e @mlcanineu da @hrw_brasil para @elpais_brasil <https://t.co/3t1SyHZ2qj>
TWEET: Importante ação para a obtenção de recursos para o combate ao desmatamento da #Amazônia. É a atuação do @meioambiente no #ConselhodaAmazoniaLegal <https://t.co/6PxIMkesc7>
TWEET: A operação “Verde Brasil 2” foi lançada pelo governo Jair Bolsonaro para combater o desmatamento ilegal e focos de incêndio na #Amazônia. <https://t.co/ZNbP1U1dzq>
TWEET: Enquanto as emissões de gases de efeito estufa desabam no mundo com a pandemia, no Brasil elas explodem: resultado do estímulo governamental à invasão de terras, ou seja, ao desmatamento na Amazônia [@tassoazevedo](https://t.co/CVkfFTLWck) explica os números hoje às 14:00 @obsclima
TWEET: “500 dias de negacionismo do aquecimento global, corrupção e má gestão no ministério do meio ambiente. Salles é um presente do partido NOVO.” Desmatamento da Amazônia cresce 171% em abril e é o maior em 10 anos - 19/05/2020 - UOL Notícias <http://t.co/Fh8XGs86Xo> @UOLNoticias #UOL

Figura 10. Tweets com maior repercussão durante o período de abril a agosto de 2020.

5.4. Abril a Maio de 2021

O período de Abril a Maio de 2021, o último estudado por este trabalho, apresenta muitas similaridades com o que foi demonstrado no período de Abril a Maio de 2020 (ver **Seção 5.3**). Mais uma alta histórica de desmatamento, o pior da série histórica de Abril desde 2015 [Niklas and Grandelle 2021], gerou um aumento na frequência de *tweets*, porém menor do que os aumentos registrados no mesmo período dos 2 anos anteriores. A polaridade de sentimentos permaneceu próxima à média obtida nos períodos anteriores, como é possível ver na Figura 3. Os *tweets* negativos representam 37% do total dos *tweets* e os positivos representam 32%, sendo esse o maior valor de *tweets* positivos desde Abril de 2019.

da rede, a floresta representa um tesouro natural e do povo brasileiro e é de responsabilidade do governo brasileiro preservá-la. Nos meses onde o sentimento se mostrou mais negativo, esse sentimento vinha após alguma notícia sobre crescimento do desmatamento, queimadas ou presença de grileiros na região da floresta amazônica, trazendo uma alta de emoções como “raiva” e “medo”. A partir disso, é possível entender que o sentimento negativo vem de um medo e desejo de preservação da floresta, assim como um desejo de mudança da situação atual.

A metodologia utilizada para coleta e classificação dos *tweets* neste trabalho pode ser utilizada em outros contextos e para outros temas dentro da plataforma *Twitter*, de modo que estudos similares podem ser conduzidos sobre diversos temas. Apesar da limitação das tecnologias para lidar com algumas particularidades da linguagem natural, ou a necessidade de remoção de *emojis* e outros símbolos, tais limitações podem ser abordadas em trabalhos futuros para melhorar a classificação e a análise dos sentimentos. Os *emojis* são um exemplo de simbologia de linguagem que carrega significado, mas que é extremamente difícil de se trabalhar, podendo agregar valor futuro a área de análise de sentimentos. Outro ponto para abordagens futuras seriam a coleta, classificação e análise de sentimentos em outras redes sociais como o Facebook, que é hoje a rede social com o maior número de usuários ativos no mundo. Pode-se também investigar o impacto de *bots* nas redes sociais e como essas entidades impactam a análise de sentimento.

Referências

- Brum, P., Cândido Teixeira, M., Miranda, R., Vimieiro, R., Meira Jr, W., and Pappa, G. (2020). A Characterization of Portuguese Tweets Regarding the Covid-19 Pandemic. pages 177–184.
- Cristiani, A., Lieira, D., and Camargo, H. (2020). Sentiment Analysis of Brazilian Elections Tweets. In *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*, pages 153–160, Porto Alegre, RS, Brasil. SBC.
- Euzebio, C., Agy, S., Jr., C. B., Porto, L., Alcarás, J. R., Martinez, A., and Ruiz, E. (2020). Statistical analysis of small twitter data collection to identify dengue outbreaks. In *Anais do VIII Symposium on Knowledge Discovery, Mining and Learning*, pages 17–24, Porto Alegre, RS, Brasil. SBC.
- G1 (2019). Dia vira ‘noite’ em SP com frente fria e fumaça vinda de queimadas na região da Amazônia. Disponível em: <https://glo.bo/3nZU2Bu>. Acesso em: 28 de novembro de 2021.
- Gupta, V. and Lehal, G. (2009). A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence*, 1.
- IBM (2020). IBM Cloud Education: What is Text Mining? Disponível em: <https://www.ibm.com/cloud/learn/text-mining>. Acesso em: 28 de novembro de 2021.
- INPE (2020). Monitoramento do território: Florestas. Disponível em: <http://www.inpe.br/faq/index.php?pai=6>. Acesso em: 28 de novembro de 2021.
- Jupyter (2021). Project Jupyter. Disponível em: <https://jupyter.org/>. Acesso em: 04 de dezembro de 2021.

- Machado, L. (2019). O que se sabe sobre o ‘Dia do Fogo’, momento-chave das queimadas na Amazônia. Disponível em: <https://www.bbc.com/portuguese/brasil-49453037>. Acesso em: 28 de novembro de 2021.
- Misuraca, M., Forciniti, A., Scepi, G., and Spano, M. (2020). Sentiment Analysis for Education with R: packages, methods and practical applications. *arXiv preprint arXiv:2005.12840*.
- Moreira, A. and Pinheiro, L. (2020). OMS declara pandemia de coronavírus. Disponível em: <https://g1.globo.com/bemestar/coronavirus/noticia/2020/03/11/oms-declara-pandemia-de-coronavirus.ghtml>. Acesso em: 28 de novembro de 2021.
- Muniz, B., Fonseca, B., and Ribeiro, R. (2020). Governo Bolsonaro reduz multas em municípios onde desmatamento cresce. Disponível em: <https://bit.ly/3HV65I1>. Acesso em: 28 de novembro de 2021.
- Niklas, J. and Grandelle, R. (2021). Desmatamento recorde na Amazônia: 2021 teve pior abril da série histórica, mostra Inpe. Disponível em: <https://glo.bo/3rBPL9y>. Acesso em: 04 de dezembro de 2021.
- Oliveira, E. (2020). Com recorde em maio, alertas de desmatamento na amazônia indicam que temporada pode ter devastação maior que a anterior. Disponível em: <https://glo.bo/3D136KA>. Acesso em: 28 de novembro de 2021.
- Pozzi, F., Fersini, E., Messina, E., and Liu, B. (2016). *Sentiment analysis in social networks*. Morgan Kaufmann.
- Sagiroglu, S. and Sinanc, D. (2013). Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)*, pages 42–47. IEEE.
- Silva, H., Andrade, E., Araújo, D., and Dantas, J. (2021). Análise de Sentimentos de Tweets Relacionados ao SUS Antes e Durante a Pandemia do COVID-19. *IEEE Latin America Transactions*, 100(XXX).
- Syuzhet (2021). Syuzhet. Disponível em: <https://github.com/mjockers/syuzhet>. Acesso em: 04 de dezembro de 2021.
- TextBlob (2021). TextBlob: Simplified Text Processing. Disponível em: <https://textblob.readthedocs.io/en/dev/>. Acesso em: 04 de dezembro de 2021.
- Twint (2021). TWINT - Twitter Intelligence Tool. Disponível em: <https://github.com/twintproject/twint>. Acesso em: 04 de dezembro de 2021.
- Twitter. About Twitter — Our company and priorities. Disponível em: <https://about.twitter.com/>. Acesso em: 28 de novembro de 2021.
- Twitter. Q3’ 2021 Shareholder Letter. Disponível em: https://s22.q4cdn.com/826641620/files/doc_financials/2021/q3/Final-Q3’21-Shareholder-letter.pdf. Acesso em: 28 de novembro de 2021.
- WordCloud (2021). WordCloud for Python. Disponível em: https://amueller.github.io/word_cloud/. Acesso em: 04 de dezembro de 2021.