

Sentiment Analysis of Tweets Related to SUS Before and During COVID-19 Pandemic

Henrique Silva, Danilo Araújo, Jamilson Dantas, and Ermeson Andrade

Abstract—The COVID-19 pandemic has affected the whole world since the beginning of 2020. In Brazil, over 70% of the population rely on Brazil’s Unified Health System (SUS). Knowing public opinion related to SUS is very important for improving services and assistance provided by such an entity. Sentiment analysis has been used in several applications, including social networks and blogs, to extract public opinion. Even though other papers have already worked with sentiment analysis, none of them have focused on SUS. Therefore, this paper aims to analyze the sentiments shown by Brazilian Twitter users about SUS before and during the COVID-19 pandemic. A database of Portuguese tweets regarding SUS posted between December 2019 and October 2020 was created to reach this goal. The tweets were pre-processed, classified, and then analyzed. The results show that, in most cases (36% of the total tweets collected), users favor SUS.

Index Terms—SUS, Brazilian Health System, Sentiment Analysis, Social Media.

I. INTRODUÇÃO

A pandemia da COVID-19, classificada assim pela Organização Mundial da Saúde (OMS) em 12 de Março de 2020 [33], paralisou o mundo no ano de 2020, forçando todos a se ajustarem à nova realidade com regras de circulação e conduta social [4]. A pandemia também causou uma grande exaustão nos sistemas de saúde de muitos países [29], colocando profissionais de saúde na linha de frente do combate à COVID-19 e arriscando suas vidas para salvar outras. No Brasil, além das novas regras sociais, foram implantados diversos hospitais de campanha para minimizar a eventual sobrecarga de leitos no Sistema Único de Saúde Brasileiro (SUS). Dados da Agência Nacional de Saúde (ANS) mostram que, em 2018, apenas 47 milhões de brasileiros, aproximadamente 22,6% da população, tinha acesso a algum plano de saúde particular [2]. Para quem não possui um plano de saúde ou condições financeiras para atendimento em hospitais particulares, a rede de hospitais públicos do SUS é a única possibilidade para busca por atendimento.

A percepção da importância do SUS sempre foi muito polarizada no Brasil. De um lado, existem reclamações sobre longas filas de espera para agendamento de consultas, hospitais superlotados e a falta de equipamentos para realização de exames [26]. Por outro lado, existem pessoas que apoiam a

atuação do SUS. Essas pessoas põem em evidência os impactos positivos do SUS na vida dos brasileiros, como controle de vigilância sanitária e acesso gratuito a medicamentos [7]. O fato é que existência do SUS afeta a vida de todos os brasileiros e a experiência que cada um teve utilizando um dos vários serviços ofertados pelo SUS pode variar bastante.

Durante a pandemia, ocorreu um forte movimento de conscientização mundial para as pessoas permanecerem em casa, para evitar a contaminação e disseminação do vírus [13]. Com cada vez mais brasileiros sendo obrigados a ficar em suas casas, foi necessário encontrar novas formas de entretenimento. Neste mesmo período, algumas mídias sociais, como o Twitter, registraram um crescimento de 40%, indicando que as pessoas começaram a passar mais tempo nesses aplicativos [28]. Nos anos recentes, a popularização do acesso a Internet e de mídias sociais como o Twitter permitiu que cada vez mais pessoas gerassem dados na Internet, e em sua grande maioria, dados não estruturados. Isto é, dados em formato de vídeos, imagens, áudios e textos, como *tweets*. A área do conhecimento que estuda formas de tratar e obter informação relevante de grandes volumes de dados em formatos variados ficou conhecido como *Big Data* [31].

Uma das técnicas utilizadas por analistas de dados para tratar da questão dos textos em *Big Data* é a Análise de Sentimentos (AS), que é uma técnica da área de mineração de textos. A AS tem como principal objetivo “extrair informações úteis referentes a sentimentos em textos, informações tais como conhecimentos, críticas e opiniões sobre determinado tema” [10]. Um dos trabalhos que utilizou essa técnica foi [12], com o objetivo mostrar a opinião de usuários do Twitter sobre diferentes acontecimentos da Copa do Mundo FIFA de 2014, gerando mapas de calor e mostrando termos mais frequentes. Outro trabalho que utilizou a AS foi [23], no qual os autores utilizaram-se da AS para realizar predições dos resultados das eleições presidenciais de 2018 no Brasil. Para isso, foi realizada a classificação dos sentimentos em *tweets* sobre os principais candidatos a presidência. Apesar de terem construído uma metodologia robusta para a análise, o objetivo de prever corretamente o resultado das eleições não foi atingido.

No Brasil, a pandemia deixou marcas irreparáveis, atingindo o número de 200 mil óbitos relacionados a pacientes infectados com a COVID-19 até janeiro de 2021, 10 meses após a confirmação do primeiro caso de contágio registrado no país [20]. Neste período, muito se falou sobre a participação do SUS no combate à COVID-19 e de todas as medidas implementadas para ajudar nesse combate. Assim, este trabalho tem como objetivo identificar o sentimento da população brasileira

Henrique Silva, Universidade Federal Rural de Pernambuco, hsilva.fps@gmail.com

Danilo Araújo, Doutor em Engenharia Elétrica, UFPE - Universidade Federal de Pernambuco, danilo.araujo@ufrpe.br

Jamilson Dantas, Ciência da Computação, Universidade Federal de Pernambuco, jrd@cin.ufpe.br

Ermeson Andrade, Ciências da Computação, Universidade Federal de Pernambuco, ermeson.andrade@ufrpe.br

sobre o SUS através dos conteúdos de *tweets*, aplicado no contexto da pandemia da COVID-19. Também almeja-se realizar comparações dos sentimentos apresentados em diferentes momentos antes e durante a pandemia. Para chegar neste objetivo, foram coletados 27,5 mil *tweets* mencionando as palavras “Saúde” e “SUS” no período de 1 de dezembro de 2019 a 31 outubro de 2020. Os *tweets* coletados foram analisados e os resultados foram interpretados. Os resultados encontrados mostraram que, durante todo o período de pesquisa, 36,4% dos *tweets* coletados demonstraram apoio a manutenção do serviço do SUS, sendo que essa classe de sentimentos positivos foi a classe com mais *tweets* associados. A classe neutra aparece em segundo lugar com 32,5% dos *tweets* e por último ficou a classe negativa, com 31,1%. Os momentos nos quais os sentimentos encontrados foram de maioria negativa estão relacionados com o medo das consequências negativas da pandemia na atuação do SUS.

Este trabalho está organizado conforme segue. A Seção II apresenta informações básicas para um melhor entendimento deste trabalho. A Seção III detalha os trabalhos relacionados. A Seção IV explica a metodologia utilizada para chegar nos resultados obtidos. A Seção V detalha os resultados das análises realizadas. Finalmente, a Seção VI apresenta as considerações finais e descreve os possíveis trabalhos futuros.

II. FUNDAMENTAÇÃO

Nesta seção, serão apresentadas explicações sobre alguns dos principais conceitos e entidades exploradas neste trabalho.

A. Twitter

O Twitter é uma mídia social focada em assuntos do momento, na qual os usuários podem se expressar em textos curtos limitados a 280 caracteres chamados de *tweets*. Essa mídia social possui um alto nível de engajamento por partes dos usuários, tendo registrado em 2019 uma média 140 milhões de postagens diárias e 330 milhões de usuários ativos mensalmente [11]. De acordo com a política de dados do Twitter, que é um contrato aplicado a todos os usuários da mídia social, “o Twitter é público e os *tweets* são imediatamente visíveis e pesquisáveis por qualquer pessoa em todo o mundo” [32]. Para os usuários que não concordam com essa regra, existe a possibilidade de limitar quem tem permissão de visualizar seus *tweets*.

No Twitter, cada usuário possui um nome único precedido pelo caractere “@” para identificar cada usuário da plataforma e assuntos podem ser marcados com uma palavra precedida do caractere “#”. As palavras que identificam os assuntos são chamadas de *hashtags*. Os assuntos mais comentados no momento dentro da plataforma ganham destaque para atrair outros usuários para a conversa ou visualização. Vale ser ressaltado que todas as informações contidas nos *tweets* podem ser utilizadas para filtrar a pesquisa por parte dos usuários. Existem algumas aplicações que realizam tal pesquisa, de modo a facilitar a coleta das informações para fins de análise, como a Twint [1]. O Twint é uma aplicação desenvolvida em Python que recebe parâmetros para buscar *tweets* relacionados à pesquisa e retorna os dados em formato JSON ou CSV.

B. Sistema Único de Saúde - SUS

Em 2020, o SUS completou 30 anos de existência. Sua criação se deu a partir da Lei 8080/1990, a qual oficializou a criação de um sistema de saúde de acesso universal e gratuito, independente de cidadania, renda, classe social ou titularização de plano privado de assistência médica [22]. O objetivo do SUS é cuidar da saúde em geral, atendendo todas as áreas relacionadas ao bem estar da população, incluindo saneamento e segurança no trabalho [24]. O SUS também é responsável por 98% do mercado de vacinas no Brasil, oferecendo gratuitamente todas as vacinas recomendadas pela OMS, além de mais de 10 vacinas para pessoas em grupos com condições específicas, como portadores de HIV [5].

Outro ponto de atuação do SUS é na vigilância sanitária. No Brasil, a Agência Nacional de Vigilância Sanitária (Anvisa), que é um órgão do SUS, é responsável por normatizar, controlar e fiscalizar produtos, substâncias e serviços de interesse para a saúde. Ou seja, para ser vendido no Brasil, qualquer medicamento precisa ter autorização da Anvisa, que garante a segurança e eficácia do medicamento para pacientes [6], [7]. Esses medicamentos incluem as vacinas desenvolvidas para a prevenção da COVID-19. Em 17 de janeiro de 2021 a Anvisa aprovou o uso emergencial de duas vacinas, garantindo uma sua segurança e eficácia para uso na população [27].

C. Mineração de Texto

A área de mineração de textos é uma subárea da mineração de dados. Essa área tem como objetivo obter informações de textos, que podem estar armazenados em formatos estruturados, como planilhas e documentos *HTML*, ou não estruturados, como texto plano. Dados não estruturados são os tipos de dados mais comuns na Internet e representam pelo menos 80% dos dados textuais armazenados. Entre as aplicações dessa área estão a categorização, para identificação dos temas chave abordados no documento analisado para definir qual o seu tópico principal, sumarização, para obter um resumo das principais informações contidas em um texto [21], e análise de sentimentos, que será explorada melhor na próxima subseção.

1) *Análise de Sentimentos*: A AS, tem como principal objetivo extrair informações sobre os sentimentos apresentados em documentos de texto. Conforme abordado no trabalho [10], dentre as aplicações dessa técnica estão a identificação de sarcasmo ou ironia, adaptação de domínios e o principal deles, a classificação de polaridade. Para realizar uma AS, após definir qual o problema a ser resolvido, é necessário realizar um tratamento inicial dos dados, que entre outros processos, inclui a separação do texto em blocos para análise. As próximas etapas são para obter os *tokens*¹, fazendo uma filtragem e um refinamento necessário para a análise. Por fim, é necessário dar uma pontuação de sentimento para cada bloco de acordo com os *tokens* que cada um contém [3].

Uma boa análise de sentimentos considera muito mais do que apenas a frequência das palavras. Em [3] os autores apresentaram vários aspectos importantes para a AS, como a marcação de classe gramaticais das palavras em frases,

¹*Tokens* são unidades básicas individuais que trazem valor para a análise a realizar [25].

conhecida como *part of speech*. Também são evidenciados alguns dos principais desafios ao executar a tarefa de AS em mídias sociais, como a identificação de *spam* e as dificuldades inerentes a natureza do problema de processamento de linguagem natural (PLN), já que são dados subjetivos e o computador não consegue identificar a semântica de todas as palavras dos textos. Por fim, os autores apresentaram possíveis técnicas para AS, como considerar *emoticons* e utilizar bases de conhecimento como a SenticNet. Essas bases de conhecimento possuem informações de semântica de palavras em diferentes contextos.

III. REVISÃO BIBLIOGRÁFICA

Existem algumas dificuldades ao trabalhar com o mineração de textos utilizando o Twitter como fonte de dados. Um dos principais problemas é que os textos do Twitter contém bastante ruído, devido ao fato dos *tweets* serem constantemente informais e de vocabulário inconsistente [9]. Uma das aplicações mais comuns da AS em trabalhos disponíveis na literatura consiste em analisar o sentimento de um determinado grupo em contraponto aos acontecimentos reais. Outra aplicação é tentar agrupar usuários de acordo com suas opiniões. Os trabalhos a seguir realizaram AS para endereçar tais objetivos.

Em [12], o autor coletou *tweets* marcados com a *hashtag* “#worldcup” durante o período da Copa do Mundo Fifa de 2014. Os *tweets* foram separados em conjunto de treino, teste e classificação. Os *tweets* de testes foram classificados manualmente pelo autor e utilizados para validar a qualidade do modelo *Naive Bayes* para AS gerado durante o treinamento. O modelo gerado é capaz de classificar os *tweets* em Negativo, Neutro, Positivo e Ambíguo. Os dados processados e analisados foram utilizados para mostrar a opinião de usuários do Twitter correlacionando com os fatos acontecidos durante a copa, gerando mapas de calor e mostrando termos mais frequentes.

Com o objetivo de avaliar a opinião de usuários do Twitter sobre um novo produto, em [34] o autor realizou uma análise de sentimentos sobre *tweets* relacionados ao sistema operacional Android 10. Foram coletados 7 mil *tweets* em inglês com a *hashtag* “#Android10” para serem analisados com uma implementação de AS da biblioteca *Pattern* de Python. Os resultados obtidos foram apresentados em gráficos, e concluiu-se que apesar das limitações da biblioteca de classificação, os sentimentos apresentados pelos usuários foram de maioria positiva.

Para realizar uma análise sobre os protestos que ocorreram no Brasil em 2013, os autores de [17] criaram um modelo de classificação de polaridade baseado no algoritmo *Naive Bayes* para classificar *tweets* coletados em *hashtags* relacionadas aos protestos de 2013. Após as classificações, foram exibidos dados como a variação do percentual positivo e negativo dos *tweets* ao passar das semanas e também os fatos acontecidos em cada semana. Adicionalmente, a localização dos *tweets* foram coletadas de modo a destacar onde o tema foi mais comentado e qual foi o sentimento foi apresentado nas regiões do Brasil.

No contexto político internacional, em [8] foi realizado um estudo sobre o comportamento de usuários americanos

no Twitter. Nesse estudo, foi analisado como utilizar as informações do usuário para identificar qual o partido político que cada usuário apoia. As informações utilizadas foram a identidade social dos usuários, as notícias que eles compartilham e quais as *hashtags* usadas pelos mesmos. O principal objetivo desse trabalho foi criar uma base de dados com as notícias dos principais sites americanos e as atividades de compartilhamento dessas notícias no Twitter. Os dados foram coletados para o mês de setembro de 2018. Após a análise, os usuários foram organizados pelos seus estados de origem e uma análise comparativa foi realizada de estado a estado com o os resultados reais da eleições de meio mandato dos Estados Unidos, obtendo uma acurácia de 90% na classificação de quais partidos ganharam em quais estados.

A análise de sentimentos já foi aplicada em contextos políticos no Brasil, como demonstrado em [23]. Nesse trabalho os autores objetivaram prever o resultado das eleições presidenciais de 2018 de acordo com os sentimentos apresentados pelos usuários do Twitter em suas postagens. Foram coletados 1000 *tweets* sobre os principais candidatos para gerar nuvens de palavras com os termos mais frequentes, classificação de polaridade e a análise de outros sentimentos, como surpresa, alegria e medo. Para realizar a classificação, foi utilizada uma plataforma chamada *Orange Canvas*. O objetivo de prever o resultado não foi atingido, porém as análises e metodologia apresentadas pelos autores agregam bastante valor ao trabalho.

Dentre os trabalhos citados, alguns fazem análises sobre questões de políticas públicas, como protestos e eleições, porém nenhum deles aborda o tema do SUS, que afeta uma grande parcela da população brasileira, e está em alta devido à pandemia da COVID-19. Além disso, diferentemente desses trabalhos, o trabalho proposto aqui objetiva analisar como acontecimentos relacionados ao SUS impactaram a opinião dos brasileiros usuários do Twitter no contexto da pandemia da COVID-19. Vale ressaltar que foram utilizados 27,5 mil *tweets* para classificação de sentimentos, em contrapartida com os trabalhos similares que utilizaram quantidades inferiores a 10 mil. A AS sobre o tema do SUS traz informações de valor tanto para grupos políticos como para a própria população, que podem utilizar-se desses dados para entender melhor o que outras pessoas acham sobre diferentes pontos relacionados ao SUS.

IV. METODOLOGIA

O trabalho proposto almeja identificar o sentimento da população brasileira sobre o SUS por meio dos conteúdos de *tweets*, aplicado no contexto da pandemia da COVID-19. Para atingir esse objetivo, foram realizadas uma série de etapas que estão descritas na Fig. 1. Em primeiro lugar, foram apresentadas as técnicas utilizadas para realizar a coleta e tratamento dos *tweets*. Em seguida, foram detalhadas as técnicas de classificação de sentimentos utilizadas. Por fim, foi descrito o processo de análise dos resultados. Abaixo, as macro atividades da metodologia são detalhadas.

A. Coleta e Tratamento dos Dados

Os dados foram coletados em outubro de 2020, com a utilização da API (*Application Programming Interface*) Twint.

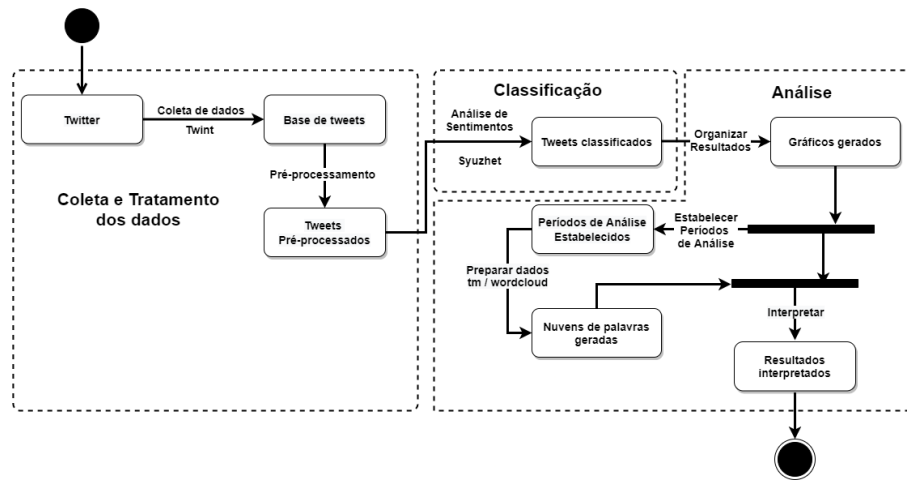


Fig. 1. Fluxograma da metodologia do trabalho.

Com essa API, foram coletados 27,5 mil *tweets* em português, postados entre 1 de dezembro de 2019 e 31 outubro de 2020, sendo 2,5 mil *tweets* de cada mês. O código de coleta foi executado no ambiente virtual Google Colab e após coletados, os *tweets* foram exportados para formato CSV, separados por mês de publicação.

Para a coleta, escolheu-se pesquisar por *tweets* que contém as palavras “Saúde” e “SUS”. As duas palavras não precisam aparecer em sequência, porém devem apenas estar contidas no mesmo *tweet*. Um problema encontrado quando usado somente a hashtag “SUS” foi a quantidade de *tweets* em espanhol não relacionadas ao tema do SUS. Isso ocorreu mesmo filtrando apenas por *tweets* em português “pt-BR”, visto que o próprio Twitter tem dificuldade em classificar os *tweets* em português ou espanhol devido às semelhanças desses idiomas. Além disso, a palavra “SUS” é bastante usada no espanhol. Vale ressaltar que a escolha dos termos “Saúde” e “SUS” restringe um pouco a pesquisa, porém garante que irá retornar *tweets* nos quais as duas palavras aparecem e que o que será retornado tem relação com o tema do trabalho, sem acrescentar viés na busca.

Os *tweets* coletados precisaram ser pré-processados para limpar informações irrelevantes para a análise, como *links*, *hashtags* e marcações de usuários. Todo o texto foi passado para caracteres minúsculos. Caracteres especiais como os *emojis* e os parêntesis foram removidos. Foram apenas deixados alguns símbolos, como o de %, para contextualizar alguns números que aparecem nos *tweets*. A Fig. 2 apresenta alguns exemplos de *tweets* antes e depois do pré-processamento.

B. Classificação

Para realizar a classificação dos sentimentos dos *tweets*, escolheu-se utilizar a classificação com o *NRC sentiment* da biblioteca *Syuzhet* em R. Essa biblioteca foi escolhida pois ela consegue classificar textos em português e fornece uma análise de sentimento por cada frase do texto. A biblioteca retorna 12 sentimentos diferentes, incluindo “Raiva”, “Alegria”, “Tristeza” e “Nojo”, além da classificação de polaridade. Cada frase recebe 1 ponto para cada sentimento apresentado, caso

o sentimento esteja presente, e a pontuação de sentimentos do texto completo consiste na soma da pontuação de todas as frases.

Para a classificação de polaridade, decidiu-se considerar as classes “Positivo”, “Negativo” e “Neutro”, sendo a classe neutra associada a um *tweet* no qual a classificação de polaridade resultou em um empate entre as classes “Positivo” e “Negativo”. Nos *tweets* em que a pontuação negativa foi maior que a positiva foram associados à classe “Negativo”, e os *tweets* nos quais a pontuação positiva foi superior, à classe “Positivo”. Dentre os 10 sentimentos que a biblioteca *Syuzhet* consegue classificar, foram escolhidos 6 que têm mais relevância para o contexto deste trabalho, são eles: Raiva, Nojo, Medo, Alegria, Tristeza e Confiança. Para esses sentimentos, optou-se por exibir nos gráficos a frequência com que esses sentimentos apareceram nos *tweets* coletados, enquanto no gráfico de polaridade, foi exibido o percentual de cada classe em cada mês para os *tweets* coletados.

C. 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 períodos de análise de acordo com as semelhanças encontradas entre os meses de análise. Os períodos estabelecidos foram de dezembro de 2019 até fevereiro de 2020, de março de 2020 até abril de 2020, de maio de 2020 até agosto de 2020, o mês de setembro de 2020 e mês de outubro de 2020. Esses agrupamentos foram utilizados para definir para quais períodos seriam geradas nuvens de palavras. Para a geração das nuvens de palavras, foi realizada a identificação dos bigramas mais frequentes de cada conjunto de texto. Bigramas podem ser analisados ao nível de palavras ou letras. A nível de palavras, bigramas são qualquer sequência de duas palavras consecutivas em um texto [16], [15]. Os termos mais frequentes em cada nuvem de palavra são os bigramas mais frequentes em cada período analisado. Neste trabalho, foi utilizada a biblioteca *tm* [18] para identificação dos bigramas e a biblioteca *wordcloud* para gerar as nuvens de palavras. Além disso, também foi real-

Tweet Original	Tweet Formatado
Parte dos recursos (45%) é direcionada para o Sistema Único de Saúde (SUS). Toffoli acrescentou que o objetivo da Medida Provisória suspensa pela Corte era "semelhante" ao que foi pretendido com a edição da norma com os novos valores +	parte dos recursos 45% é direcionada para o sistema único de saúde sus toffoli acrescentou que o objetivo da medida provisória suspensa pela corte era semelhante ao que foi pretendido com a edição da norma com os novos valores
Saúde pra dar e vender?! VENDER? Saúde é um direito e deve ser pública e de qualidade. Defenda o SUS <U+0001F44D><U+0001F3FB>!	saúde pra dar e vender?! vender? saúde é um direito e deve ser pública e de qualidade defenda o sus !
@ @ @ Depende muito. Sucesso não é só financeiro. Por exemplo o país mais potente do mundo tornou um tratamento medico tao caro que muitos não tem acesso. O SUS(apesar de sucateado) é um sistema de saúde infinitamente mais bem bolado que aquele que os americanos tem acesso.	depende muito sucesso não é só financeiro por exemplo o país mais potente do mundo tornou um tratamento medico tao caro que muitos não tem acesso o sus apesar de sucateado é um sistema de saúde infinitamente mais bem bolado que aquele que os americanos tem acesso
@ @ @ ... respeito a toda a sociedade. Tem impacto na saúde pública. Está relacionado a inúmeros problemas de saúde (hipertensão, diabetes, infarto, problemas osteoarticulares, vários tipos de câncer). Tem impacto no SUS. É meu dinheiro (seu também) gasto por milhares de pacientes com...	respeito a toda a sociedade tem impacto na saúde pública está relacionado a inúmeros problemas de saúde hipertensão diabetes infarto problemas osteoarticulares vários tipos de câncer tem impacto no sus é meu dinheiro seu também gasto por milhares de pacientes com
adeus ano velho feliz ano novo que tudo se realize no ano que vai nascer muito dinheiro no bolso saúde pra dar e só dar mesmo pq vender saúde é coisa de arrombado viva o sus	adeus ano velho feliz ano novo que tudo se realize no ano que vai nascer muito dinheiro no bolso saúde pra dar e só dar mesmo pq vender saúde é coisa de arrombado viva o sus

Fig. 2. Tweets antes e depois do processamento.

izada uma comparação entre os acontecimentos e os assuntos mais comentados em cada período analisado, para justificar os resultados apontados nos gráficos.

V. RESULTADOS

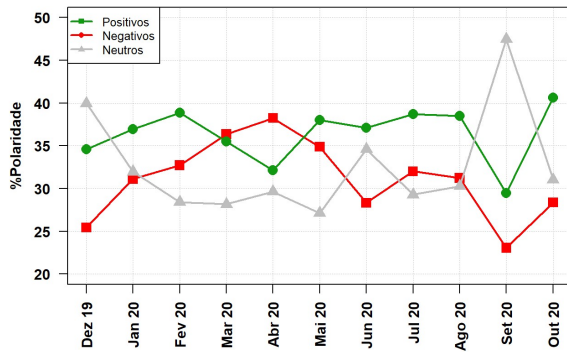


Fig. 3. Polaridade em tweets sobre o SUS por mês.

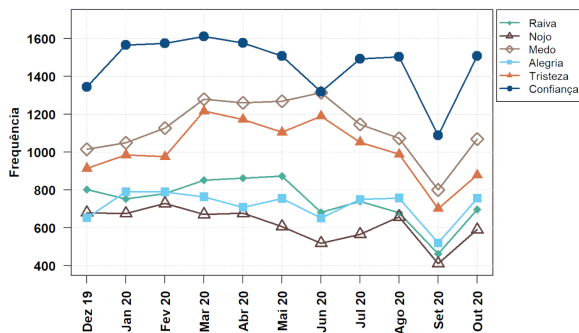


Fig. 4. Frequência de sentimentos em tweets sobre o SUS por mês.

Após executar os passos definidos na Seção 4, foi iniciado o processo de análise dos resultados obtidos. Na Fig. 3, o percentual polaridade dos comentários no Twitter sobre o SUS foram organizados por meses, em que cada mês contém tweets coletados entre o primeiro e o último dia de cada mês. Ao observar a Fig. 3, nota-se que no período anterior a pandemia

da COVID-19, que é representado pelos meses de dezembro de 2019 a fevereiro de 2020, os sentimentos positivos foram maioria e apresentaram uma tendência de crescimento. Entre os termos que se destacam na categoria positiva da Fig. 5 estão “incluir milhoes” e “brasileiros sus”. Esses termos, junto com outros na categoria positiva, fazem menção ao programa Previne Brasil, que tem a intenção de incluir no SUS mais 50 milhões de brasileiros que não têm acompanhamento médico [14]. Os tweets sobre o esse tema foram em grande maioria classificados como positivo. Os termos “sus novas”, “regras visam” e “estrangular maior” estão relacionados com uma notícia compartilhada na mídia social, intitulada: “SUS: novas regras visam estrangular o maior sistema público de saúde mundial”. Todos esses tweets contêm esse artigo, sem expor nenhuma opinião sobre o mesmo, por isso foram classificados como neutro. Nesse período, outro tema que foi bastante discutido foi plano de saúde. Nesse mês aconteceu a extensão da faixa etária do benefício de plano de saúde para filhos de senadores, filhos com até 33 anos. A reação dos usuários a tal acontecimento justifica a subida na incidência de tweets negativos no mês fevereiro.

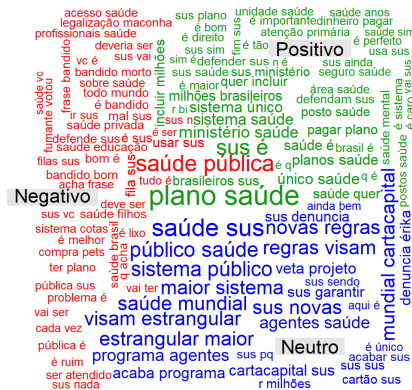


Fig. 5. Nuvem de palavras para Dezembro de 2019 até Fevereiro de 2020. “plano saúde” tem 548 ocorrências positivas, “saúde pública” tem 290 ocorrências negativas e “saúde sus” tem 196 ocorrências neutras.

No mês de março, primeiro mês de pandemia, os sentimen-

- [7] Blog da Saúde - Ministério da Saúde. Quando você usa o sus? Disponível em: <http://www.blog.saude.gov.br/geral/35494-quando-voce-usa-o-sus>, Acesso em: 26 jan. 2021, 2021.
- [8] G. Brena, M. Brambilla, S. Ceri, M. Di Giovanni, F. Pierri, and G. Ramponi. News sharing user behaviour on twitter: A comprehensive data collection of news articles and social interactions. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 592–597, 2019.
- [9] A. Brew, D. Greene, D. Archambault, et al. Deriving insights from national happiness indices. In *2011 IEEE 11th International Conference on Data Mining Workshops*, pages 53–60. IEEE, 2011.
- [10] L. Britto and L. Pacífico. Análise de sentimentos para revisões de aplicativos mobile em português brasileiro. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 1080–1090. SBC, 2019.
- [11] Broadband Search. Average time spent daily on social media (latest 2020 data). Disponível em: <https://www.broadbandsearch.net/blog/average-daily-time-on-social-media#post-navigation-2>, Acesso em: 22 set. 2020, 2020.
- [12] J. A. CARVALHO FILHO. Mineração de textos: Análise de sentimento utilizando tweets referentes à copa do mundo 2014. 2014.
- [13] Centers for Disease Control and Prevention(CDC). Social distancing. Disponível em: <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html>, Acesso em: 22 set. 2020, 2020.
- [14] A. Coifman. Programa previne brasil: Como impacta a gestão da saúde?, blog hygia. Disponível em: <https://blog.hygia.com.br/programa-previne-brasil>, Acesso em: 22 set. 2020, 2020.
- [15] M. J. Collins. A new statistical parser based on bigram lexical dependencies. In *34th Annual Meeting of the Association for Computational Linguistics*, pages 184–191, Santa Cruz, California, USA, June 1996. Association for Computational Linguistics.
- [16] K. Corbin. Double, triple, and quadruple bigrams. *Word Ways*, 22(3):8, 1989.
- [17] T. C. de França and J. Oliveira. Análise de sentimento de tweets relacionados aos protestos que ocorreram no brasil entre junho e agosto de 2013. In *Anais do III Brazilian Workshop on Social Network Analysis and Mining*, pages 128–139. SBC, 2014.
- [18] I. Feinerer, K. Hornik, and D. Meyer. Text mining infrastructure in r. *Journal of Statistical Software, Articles*, 25(5):1–54, 2008.
- [19] J. Fucs. Em salvador, um hospital público que parece privado, Época. Disponível em: <https://epoca.globo.com/tempo/noticia/2014/04/em-salvador-um-bhospital-publico-que-parece-privado.html>, Acesso em: 22 set. 2020, 2020.
- [20] G1. Brasil chega a 200 mil mortes por covid-19 sem vacina e sob risco de repetir piores momentos da pandemia. Disponível em: <https://g1.globo.com/bemestar/coronavirus/noticia/2021/01/07/brasil-chega-a-200-mil-mortes-por-covid-19-sem-vacina-e-sob-risco-de-repetir-piores-momentos-da-pandemia.ghtml>, Acesso em: 26 jan. 2021, 2021.
- [21] V. Gupta, G. S. Lehal, et al. A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1):60–76, 2009.
- [22] Justiça Federal - Brasil. Acesso ao sus é universal e destinado ao atendimento de todas as classes sociais. Disponível em: <https://www.cjf.jus.br/cjf/outras-noticias/2013/julho/acesso-ao-sus-e-universal-e-destinado-ao-atendimento-de-todas-as-classes-sociais>, Acesso em: 22 set. 2020, 2020.
- [23] F. F. Matos, L. H. de Magalhães, and R. R. Souza. Recuperação e classificação de sentimentos de usuários do twitter em período eleitoral. *Informação & Informação*, 25(1):92–114, 2020.
- [24] Ministério da Saúde - Brasil. Brasil comemora a lei dos 30 anos da criação do sus. Disponível em: <https://antigo.saude.gov.br/noticias/agencia-saude/47491-brasil-comemora-a-lei-dos-30-anos-da-criacao-do-sus>, Acesso em: 22 set. 2020, 2020.
- [25] T. Mohler. The 7 basic functions of text analytics & text mining, lexalytics. Disponível em: <https://www.lexalytics.com/lexablog/text-analytics-functions-explained>, Acesso em: 2 fev. 2021, 2020.
- [26] J. A. Morsch. Superlotação nos hospitais: Causas, consequências e possíveis soluções. Disponível em: <https://telemedicinamorsch.com.br/blog/superlotacao-nos-hospitais>, Acesso em: 22 set. 2020, 2020.
- [27] W. Máximo. Por unanimidade, anvisa aprova uso emergencial de vacinas contra covid, agência brasil. Disponível em: <https://agenciabrasil.ebc.com.br/saude/noticia/2021-01/por-unanimidade-anvisa-aprova-uso-emergencial-de-vacinas-contra-covid>, Acesso em: 26 jan. 2021, 2021.
- [28] Negócios da Comunicação. Redes sociais e influenciadores crescem na pandemia, negócios da comunicação. Disponível em: <https://portaldacomunicacao.com.br/2020/04/redes-sociais-e-influenciadores-crescem-na-pandemia>, Acesso em: 22 set. 2020, 2020.
- [29] Portal Europeu de Dados. Pressure on healthcare systems: Coping with demand for icu and hospital beds. Disponível em: <https://www.europeandataportal.eu/pt/impact-studies/covid-19/pressure-healthcare-systems-coping-demand-icu-and-hospital-beds>, Acesso em: 22 set. 2020, 2020.
- [30] SESAB. Parcerias público-privadas (ppps). Disponível em: <http://www.saude.ba.gov.br/atencao-a-saude/comofuncionaosus/ppp/>, Acesso em: 23 jan. 2021.
- [31] C. Taurion. *Big data*. Brasport, 2013.
- [32] Twitter Privacidade. Política de privacidade twitter. Disponível em: <https://twitter.com/pt/privacy>, Acesso em: 22 set. 2020, 2020.
- [33] World Health Organization. Who announces covid-19 outbreak a pandemic. Disponível em: <https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic>, Acesso em: 22 set. 2020, 2020.
- [34] J. XAVIER. O uso da análise de sentimentos no twitter para avaliar a opinião do público consumidor a respeito do sistema operacional mobile android 10. 2019.



Henrique Silva Graduando do curso de Ciência da Computação pela Universidade Federal Rural de Pernambuco. Tem experiência na área de Ciência da Computação e Sistema de Informação, com ênfase na desenvolvimento de software. Tem interesse em explorar tópicos como Ciência de Dados.



Danilo Ricardo Barbosa de Araújo recebeu o título de Doutor em Engenharia Elétrica pela UFPE - Universidade Federal de Pernambuco em 2015; recebeu o título de Mestre em Engenharia da Computação pela UPE - Universidade de Pernambuco em 2011; e recebeu o título de Bacharel em Ciência da Computação pela Unicap - Universidade Católica de Pernambuco em 2002. Ingressou como docente do Departamento de Computação da UFRPE - Universidade Federal Rural de Pernambuco em 2013 e é atualmente Professor Adjunto

III. Possui interesse por temas de pesquisa relacionados a Redes de Alta Capacidade, Inteligência Artificial e Redes Complexas.



Jamilson Dantas Possui graduação em Sistemas de Informação pelo Centro Universitário do Rio São Francisco - UNIRIOS (2009), mestrado (2013) e doutorado (2018) em Ciência da Computação pela Universidade Federal de Pernambuco - UFPE. É professor no Centro de Informática da UFPE. Atua nas áreas de avaliação de desempenho de sistemas, redes de computadores e sistemas distribuídos.



Ermeson Andrade Possui graduação em Ciência da Computação pela Universidade Católica de Pernambuco (2006), mestrado (2009) e doutorado (2014) em Ciências da Computação, ambos pela Universidade Federal de Pernambuco. Entre 2010 e 2011, fez doutorado sanduíche na Universidade de Duke (Estados Unidos). Atualmente, é professor adjunto da Universidade Federal Rural de Pernambuco. Tem experiência na área de Ciência da Computação, com ênfase em Engenharia de Software e Avaliação de Desempenho, atuando principalmente nos seguintes

temas: Modelagem de Desempenho e Dependabilidade, Avaliação de Modelos, Teste e Qualidade de Software.