

Explorando o sentimento no Twitter sobre a implementação do ensino remoto no Brasil em face à COVID-19

Felipe A. Leite¹, Paulo Alves da Silva¹, Douglas Vitória¹, Marcelo Iury S. Oliveira¹

¹Unidade Acadêmica de Serra Talhada – Universidade Federal Rural de Pernambuco (UAST - UFRPE) – Caixa Postal 063 – 56.900-000 – Serra Talhada – PE – Brasil

{lipe6330, alvespaulo737, douglas.alisson17, marceloiury}@gmail.com

Abstract. *The high contamination risk of COVID-19 has led to several countries taking social distance and isolation measures. As a result of governments' social isolation policies, remote learning was the alternative chosen by many schools and higher education institutions to continue classes and resume courses. However, such digitization induced by COVID-19 raised some issues about remote learning fitness for students expectations. This work carried out a study of sentiment analysis of messages from Twitter in order to understand the feeling of Brazilian users about the first months of introduction of remote education in Brazil during a pandemic of COVID-19.*

Resumo. *O alto risco de contaminação do COVID-19 tem levado vários países a tomar medidas de distanciamento social e isolamento. Como resultado dessas políticas, o ensino remoto foi a alternativa escolhida por muitas escolas e instituições de ensino superior para dar continuidade às aulas e retomar os cursos. No entanto, essa digitalização induzida pelo COVID-19 levantou algumas questões sobre a aptidão do aprendizado remoto em relação às expectativas dos alunos. Este trabalho realizou um estudo de análise de sentimento em mensagens do Twitter com o objetivo de compreender o sentimento dos usuários brasileiros sobre os primeiros meses de introdução da educação remoto emergencial no Brasil durante uma pandemia de COVID-19.*

1. Introdução

O surgimento do novo Coronavírus mudou completamente a rotina das pessoas em função do seu alto nível de contágio e taxa de mortalidade não desprezível. Como forma de contornar problemas causados pelas restrições e medidas de distanciamento social estabelecidas para mitigar os crescentes índices de contágio, a tecnologia da informação, junto aos seus mecanismos de utilização remota, foi empregada para o estabelecimento de telemedicina, trabalho remoto e ensino a distância [Watson et al. 2020].

Em relação à educação, esforços de governos e instituições terciárias em todo o mundo foram empregados para utilizar tecnologias de informação e comunicação com o intuito de permitir a aprendizagem dos estudantes de forma não presencial durante a pandemia [Ali 2020]. No Brasil, muitas redes de ensino e instituições de ensino superior têm optado pela modalidade de ensino remoto [Arruda 2020]. O ensino remoto difere da abordagem de educação a distância (EAD), já que o primeiro diz respeito às atividades de ensino mediadas por tecnologias, mas orientadas pelos princípios da educação presencial [Arruda 2020]. Por sua vez, a EAD, ainda que tecnicamente e conceitualmente refira-se

à mediação do ensino e da aprendizagem por meio de tecnologias, envolve planejamento anterior, considerações sobre o perfil de alunos e docentes e o desenvolvimento a médio e longo prazo de estratégias de ensino e aprendizagem que levem em consideração as dimensões síncronas e assíncronas [Moore et al. 2011, Arruda 2020].

Apesar de o ensino remoto ter permitido a retomada da aprendizagem de forma não-presencial, a realidade apresentada é bem diferente. Muitos alunos alegam dificuldade para acessar as aulas digitais ou afirmam que a experiência do ensino remoto está sendo ruim [Camacho et al. 2020]. Outro problema é a ineficácia das aulas a distância para quem não tem acesso à Internet [Camacho et al. 2020]. Ou seja, os impactos consistem tanto da dificuldade de adaptação à mudança do esquema escolar, quanto da vulnerabilidade social enfrentada pelos estudantes neste período de pandemia [Camacho et al. 2020, Galletta et al. 2020].

Isso leva os estudantes a expressarem seus sentimentos nas redes sociais, como é o caso do Twitter, que recebe milhares de opiniões acerca do mais diversos assuntos todos os dias. Inclusive, temas como “EAD” e “ensino remoto” foram classificados como uns dos tópicos mais populares, com um alto volume de postagens em mídia social¹.

Com o objetivo de compreender o sentimento dos usuários brasileiros sobre os primeiros meses de introdução do ensino remoto no Brasil durante a pandemia de COVID-19 e também analisar fatores que podem ter contribuído para um determinado sentimento, este trabalho realizou um estudo de análise de sentimentos das mensagens da rede social Twitter. A Análise de Sentimentos (AS), também chamada de Mineração de Opinião, é a área de estudo que analisa os sentimentos e opiniões das pessoas acerca de entidades [Liu and Zhang 2012]. Estas entidades podem ser produtos, serviços, organizações, indivíduos, problemas, eventos, tópicos e seus atributos; e as opiniões são expressas de forma não-estruturada, como em texto, por exemplo. Essa análise geralmente consiste em classificar uma opinião em um dentre três rótulos: *positiva*, *negativa* ou *neutra* [Liu and Zhang 2012, Yue et al. 2018, Alswaidan and Menai 2020].

Neste estudo, foram analisados fatores econômicos, sociais e geográficos a fim de compreender a visão das pessoas sobre a adoção do ensino remoto durante a pandemia de COVID-19. Desta forma, podemos identificar os problemas enfrentados pelas pessoas e questionar se as técnicas adotadas estão contribuindo para o ensino neste período.

O restante deste trabalho está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 detalha a metodologia utilizada nesta pesquisa; enquanto que os resultados alcançados e suas discussões podem ser encontrados na Seção 4; dificuldades encontradas durante a pesquisa são apresentadas na Seção 5; e, por fim, a Seção 6 traz as considerações finais.

2. Trabalhos relacionados

A análise de sentimentos nas redes sociais tornou-se um grande tema de estudo [Yue et al. 2018, Alswaidan and Menai 2020]. Um número crescente de trabalhos tem apresentado novas técnicas de análise de sentimentos, além de estudos que analisam o sentimento das pessoas a respeito de temas específicos, a exemplo de educação e ensino [Yue et al. 2018, Alswaidan and Menai 2020]. Inclusive, o emprego da análise de

¹<https://bit.ly/3IHCmpQ>

sentimentos a respeito da educação durante período pandêmico já foi abordado também pela literatura, tais como os trabalhos de [AL-Rubaiee et al. 2016, Duong et al. 2020, Wang and Cruz 2020, Pastor 2020, Mostafa 2021].

Um destes trabalhos foi realizado por [AL-Rubaiee et al. 2016], os quais fizeram uso dos algoritmos *Support Vector Machine* (SVM) e *Naïve Bayes* (NB) para classificar a polaridade de *tweets* de estudantes da universidade King Abdulaziz, na Arábia Saudita. Os autores buscaram compreender a opinião dos universitários a respeito da EAD e consequentemente melhorar seus processos de ensino. Com os resultados, eles perceberam que o SVM alcançou uma maior precisão quando comparado ao NB. O trabalho em si apresentou uma técnica para classificar as opiniões, contudo não foram evidenciados os sentimentos que os estudantes possuíam a respeito da educação a distância.

No trabalho de [Duong et al. 2020], foram verificadas as implicações sociais que a pandemia de COVID-19 provocou em relação ao público em geral e aos estudantes do ensino superior nos EUA, a partir da extração de opiniões do Twitter. Os autores analisaram um período de dois meses, quando o distanciamento social em todo o país e o fechamento de escolas atraíram grandes preocupações. Utilizando técnicas de aprendizagem profunda e o modelo RoBERTa para classificar as opiniões, foi identificado que os estudantes publicaram mais *tweets* com teor negativo relacionados ao fechamento das escolas. Entretanto, a pesquisa, apesar de destacar a preocupação dos jovens e o sentimento de desaprovação em relação ao ensino remoto, não evidenciou os fatores que levaram os estudantes a terem esse posicionamento.

A fim de ter um entendimento mais completo em torno do cenário que a pandemia de COVID-19 ocasionou na educação dos EUA, [Wang and Cruz 2020] conduziram um estudo em larga escala sobre o assunto. Eles utilizaram um conjunto de dados georreferenciados do Twitter, abrangendo todo o país, durante partes da primavera, verão e outono de 2020. No que diz respeito ao processo metodológico, foi utilizado o dicionário labMT para análise de sentimentos, o qual indicou uma predominância de sentimentos positivos. Contudo, os autores afirmam que esses resultados não são totalmente confiáveis, visto que a palavra “*positive*” (positivo), encontrada em muitos *tweets*, contribuiu bastante para classificação dos dados, embora o contexto da mensagem em alguns casos indicasse um sentimento contrário. Para a análise de conteúdo, os autores fizeram uso do *Linguistic Inquiry and Word Count* (LIWC), o qual mostrou os tópicos de maior preocupação entre os universitários, como financiamento e mensalidades. Porém, cabe ressaltar que a utilização de algoritmos mais robustos para classificação dos dados poderia ter dado uma maior confiança nos resultados obtidos.

Ainda, o trabalho de [Pastor 2020] teve como objetivo revelar o sentimento de estudantes da Universidade Estadual de Pangasinan, nas Filipinas, em relação à introdução do paradigma de ensino assíncrono em função da COVID-19. Diferentemente dos outros trabalhos apresentados nesta seção, a autora realizou a coleta de dados com base em um formulário contendo uma pergunta aberta, o qual foi enviado para os universitários via *messenger*. A pesquisa apontou alguns fatores que prejudicam a adoção do ensino remoto, como a conectividade à Internet, que, embora não seja tão rápida comparada a outros países, serve para lembrar a realidade vivenciada por vários estudantes durante este período de pandemia. Porém, vale ressaltar que a autora não especificou quais técnicas foram utilizadas para a análise de sentimentos.

Em seu trabalho, [Mostafa 2021] realizou análise de sentimentos em meio à pandemia de COVID-19 visando entender o sentimento dos alunos egípcios em relação ao processo de aprendizagem durante este período. Para o trabalho, os dados foram coletados através de um questionário realizado por meio de planilhas eletrônicas. No que diz respeito ao processo metodológico para classificação dos sentimentos, foi utilizado o algoritmo Word2vec combinado com os classificadores *Naïve Bayes*, SVM e Árvore de Decisão. Dessa forma, a autora pôde comparar o desempenho dos algoritmos, chegando à conclusão de que o *Naïve Bayes* foi superior no cenário avaliado. Contudo, ela reconhece que o tamanho da amostra utilizada foi relativamente pequeno.

Convém destacar que não foram encontrados trabalhos de pesquisa que abordassem o tema no contexto brasileiro. O presente trabalho busca, assim, contribuir com o estado da arte ao apresentar uma análise de sentimentos sob a perspectiva de estudantes brasileiros por meio da análise de *tweets* em português. Outra contribuição foi a construção de um conjunto de dados anotados sobre sentimentos e emoções relacionadas ao ensino remoto e à COVID-19.

3. Metodologia

Este trabalho, em sua perspectiva objetiva, é exploratório na medida em que busca proporcionar maior entendimento sobre um problema e também permitir a construção de hipóteses [Yin 2013]. Quanto a seu paradigma de pesquisa, ele pode ser definido como pragmático que visa, ao invés de encontrar a verdade ou a realidade, facilitar a resolução de problemas humanos, ao buscar a aplicação de “tudo que funciona” para resolver o problema e inclui uma combinação de diferentes estratégias de pesquisa [Creswell 2010]. Consequentemente, o pragmatismo tem uma forte base filosófica nos métodos mistos ou campos de pluralismo metodológico.

Com base em nossos objetivos e no paradigma pragmático, a metodologia utilizada consistiu em 3 etapas de desenvolvimento: 1) coleta de dados e anotação do corpus; 2) Pré-processamento dos dados; 3) construção do modelo e classificação; 4) análise dos dados.

3.1. Coleta de dados e Anotação do corpus

Os *tweets*, isto é, os dados do Twitter, foram extraídos durante os períodos de 28/04/2020 a 02/06/2020 e 22/06/2020 a 31/07/2020 por meio de um software extrator construído através da biblioteca Python Tweepy². Para extrair *tweets* relacionados, palavras-chave e *hashtags* como “coronavirus”, “COVID”, “COVID-19”, “COVID19”, “ensino remoto” e “EAD”³ foram usadas. Ao todo, foram coletados 162.921 *tweets* escritos em Português.

Após a extração, 2.100 *tweets* foram rotulados manualmente pelos autores deste trabalho classificando a polaridade da opinião contida no *tweet* acerca do ensino remoto, ou seja, classificando-os em três classes: *positiva*, *negativa* e *neutra*. Optou-se por uma anotação balanceada, garantindo que a base dos dados rotulados contivesse 700 *tweets* para cada uma das três classes. A anotação balanceada foi escolhida como forma de não enviesar a classificação. Além disso, cada documento foi classificado por dois anotadores,

²<https://www.tweepy.org>

³Apesar de EAD e ensino remoto serem modalidades distintas, constatamos que os usuários do Twitter utilizavam com mais frequência o termo EAD em referência ao ensino remoto

tendo as divergências sido resolvidas, quando ocorreram, como forma de mitigar possíveis vieses também na anotação.

3.2. Pré-processamento dos dados

Para a posterior utilização dos dados coletados, foram necessárias técnicas de pré-processamento. Tendo as bibliotecas NLTK⁴ e *scikit-learn*⁵ sido utilizadas para esse fim.

Primeiramente, todos os *tweets* foram colocados em minúsculo e realizou-se a tokenização, isto é, a divisão dos textos em *tokens*, com o auxílio do *TweetTokenizer* da biblioteca NLTK. Então os *tokens* passaram por um processo de filtragem, no qual as URLs e os identificadores de usuários foram removidos. Optou-se por manter as *hashtags*, pois considerou-se que elas poderiam conter informações importantes para a classificação de opinião. Por fim, utilizou-se a técnica TF-IDF para transformar os documentos em vetores de características, etapa necessária para a utilização dos classificadores.

3.3. Construção do Modelo e Classificação

Para a realização da análise de sentimentos, foi desenvolvido um modelo para classificação de sentenças de texto em 3 classes de sentimentos (negativo, neutro e positivo). Para tanto, foi construída uma aplicação de em Python, com o auxílio da biblioteca *scikit-learn*. Cinco classificadores de Aprendizado de Máquina foram utilizados: *Multinomial Naïve Bayes*, SVM, *Random Forest*, uma rede neural MLP e o classificador de Regressão Logística. A escolha dos classificadores se deu pelo fato de os quatro primeiros serem os mais utilizados para mineração de texto com a Língua Portuguesa [Souza et al. 2018], ao passo que o de Regressão Logística também é considerado bastante útil para classificação de texto [Yu et al. 2011].

Os cinco classificadores foram combinados utilizando um Sistema de Múltiplos Classificadores, também conhecido como *ensemble*, o qual combina classificadores diferentes visando melhorar o desempenho da classificação [Woźniak et al. 2014]. Geralmente, essa combinação alcança resultados melhores que classificadores individuais [Vitório et al. 2019b], então optou-se por utilizá-la como forma de prover uma precisão maior na classificação dos *tweets*.

Com o mesmo objetivo apontado acima, também foi utilizada a técnica de *Active Learning* (AL). Esta é uma técnica de Aprendizado de Máquina semi-supervisionado que consiste na anotação de dados no decorrer da classificação, os quais são utilizados para retreinar o classificador. O método AL usa um conjunto inicial de dados rotulados para o primeiro treinamento do modelo. Então, uma estratégia de seleção é utilizada para solicitar o rótulo de novos dados a fim de atualizar o modelo de aprendizagem. Estes novos dados são aqueles considerados pela estratégia de seleção mais adequadas para o problema [Zimmermann et al. 2015].

Sendo assim, o *ensemble* de classificadores, tendo sido treinado com os 2.100 *tweets* anotados manualmente, foi utilizado para classificar os *tweets* restantes. O algoritmo, então, solicitava ao usuário o rótulo daqueles *tweets* para os quais os classificadores tiveram menos certeza da sua classe. Esta estratégia de seleção, que leva em conta a incerteza do classificador, é denominada *Uncertainty* [Vitório et al. 2019a]. E os *tweets* cujos

⁴<https://www.nltk.org>

⁵<https://scikit-learn.org/stable/>

rótulos foram fornecidos eram adicionados ao modelo, melhorando-o. Neste trabalho, a biblioteca modAL [Danka and Horvath 2018] foi utilizada para implementação da técnica de AL.

Como, nesta pesquisa, lidou-se com um conjunto grande de dados distribuídos temporalmente, os quais são conhecidos como “fluxos contínuos de dados”, optou-se pela utilização das duas técnicas supracitadas (i.e., *Active Learning* e *ensembles*). Ao se trabalhar com fluxos de dados contendo opiniões, é possível se deparar com os chamados *opinion drifts*: mudanças que podem ocorrer nas opiniões das pessoas acerca de alguma entidade com o passar do tempo [Wang et al. 2013]. E ambas as técnicas utilizadas auxiliam no trato desse problema, melhorando a classificação.

Além disso, como forma de garantir um maior grau de confiança na classificação, foram selecionados apenas os *tweets* para os quais o *ensemble* teve uma certeza de pelo menos 60% para sua classe, além daqueles selecionados pela estratégia de AL. Dessa forma, 66.890 *tweets* foram considerados nesta pesquisa. Após a anotação automática de todos os *tweets*, 200 deles foram escolhidos aleatoriamente e anotados manualmente por um dos autores. Com essa anotação, pôde-se estimar a acurácia do classificador: 87%.

Por fim, foi realizada a identificação dos estados brasileiros aos quais pertenciam os autores dos *tweets* classificados. Sabe-se que grande parte dos usuários do Twitter não fornece uma localização válida em seu perfil. Sendo assim, montou-se uma aplicação Python com a ajuda da biblioteca pyUFbr⁶, a qual contém a lista de municípios brasileiros por estado, para identificar automaticamente as localidades.

3.4. Análise dos Dados

O objetivo deste estudo é fornecer um entendimento sobre os sentimentos de brasileiros em relação aos primeiros meses de introdução do ensino remoto emergencial durante a pandemia, bem como quais fatores contribuíram para este sentimento. Para isso, foi definida a seguinte questão geral de pesquisa: *Qual a visão que a maioria dos brasileiros usuários do twitter possuem em relação aos primeiros meses de implantação do ensino remoto emergencial?*

Também definimos outras questões de pesquisa específicas para orientar e estruturar a análise e síntese de dados e evidências, a exemplo de:

- O estado ou a cidade em que a pessoa está situada pode ser um fator que influencia o seu sentimento?
- Habitar em cidades que possuem maior ou menor riqueza fazem as pessoas apresentarem uma opinião positiva ou negativa acerca do ensino remoto?
- De maneira geral, quais são os fatores associados ao sentimento dos alunos em relação ao ensino remoto?

4. Resultados e Discussão

O gráfico da Figura 1 apresenta a variação, ao longo do tempo, do sentimento da população brasileira usuária do Twitter e que manifestou opiniões sobre a substituição das aulas presenciais pelo ensino remoto durante a pandemia de COVID-19. O resultado contempla *tweets* oriundos das 27 unidades federativas do país. Para essa análise não

⁶<https://github.com/Sidon/py-ufbr>

foram considerados *tweets* neutros, os quais representavam 42% do total coletado, sendo assim, do montante de 66.890 *tweets* classificados, 38.933 *tweets* foram utilizados para esta análise. Ao observar o gráfico, fica evidente que o sentimento negativo predominou durante todo o período da coleta. De modo geral, analisando todo o período, pode-se apontar que a rejeição ao ensino remoto pelos usuários do Twitter foi de 74%, enquanto que apenas 26% dos *tweets* tinham cunho positivo.

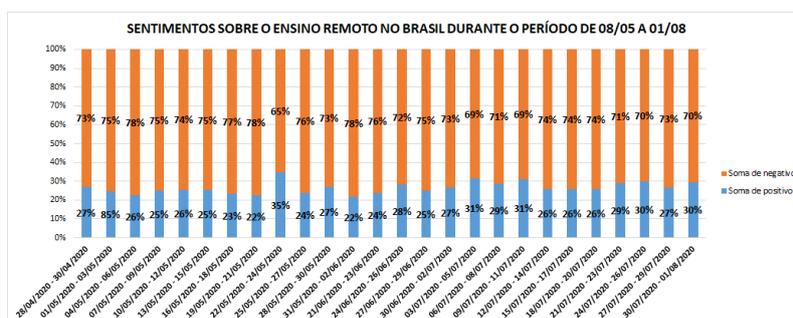


Figura 1. Distribuição dos sentimentos ao longo do tempo no Brasil.

A Figura 2 apresenta as nuvens de palavras obtidas considerando os *tweets* positivos, 2(a), e negativos, 2(b). Primeiramente analisando as palavras da nuvem positiva, podemos observar a presença de palavras de apoio como “*obrigada ead*” e “*parte boa*”, as quais devem se referir principalmente à metodologia de avaliação utilizada pelos professores, considerada mais confortável do que a metodologia presencial, na visão dos alunos usuários do Twitter. As palavras “*melhor*”, “*bom*”, “*feliz*”, “*amo*” e “*fácil*” também apareceram relacionadas à forma de condução das atividades durante a aplicação do ensino remoto. As palavras “*faculdade*” e “*curso*” dizem respeito à existência da educação a distância em modalidades de ensino técnico e superior, nas quais os usuários afirmam que utilizaram essa modalidade. Enquanto a palavra “*semestre*” representa a aceitação da comunidade ao ensino remoto como um meio de evitar a perda total do semestre letivo.



Figura 2. (a) Nuvem de palavras dos *tweets* positivos. (b) Nuvem de palavras dos *tweets* negativos.

Já analisando a nuvem de palavras negativas, pode-se enxergar uma predominância de ofensas e xingamentos. Além disso, outras palavras da nuvem nos ajudam a entender o motivo do descontentamento, como “*atividade*”, que se refere à quantidade exagerada de atividades passadas pelos professores como uma tentativa de suprir o

acompanhamento presencial e garantir que o conteúdo programado seja concluído. Esse tipo de metodologia empregada pelos professores, contudo, trouxe como consequência a indignação por partes das pessoas, que se veem sobrecarregados e desesperados, como pode-se notar pelo uso das palavras “*surtando*”, “*surto*” e “*psicológico*”, as quais se referem ao desgaste mental sofrido pelas pessoas nesta forma de ensino.

Na Figura 2(b), também pode-se notar o uso da palavra “*aguento*” (geralmente associada a ao advérbio “*não*”), referindo-se ao descontentamento das pessoas em relação ao ensino remoto ou às atividades relacionadas a esta modalidade de ensino. A palavra “*desistir*” também se destacou, representando os alunos que não aceitaram este meio de aprendizagem e desejavam abandonar ou trancar o curso. Outras palavras também expressaram sentimentos negativos declarados como “*ódio*” e “*odeio*”. Por fim, é interessante notar o uso do termo “*prestar atenção*” para se referir negativamente ao ensino remoto. Este termo está relacionado ao fato de que muitos alunos não conseguem se concentrar nas aulas online da mesma forma que faziam com as aulas presenciais. Sabe-se que quando o aluno está em casa, ele tende a procrastinar mais, ou ainda, não consegue encontrar um ambiente propício para o estudo.

Pode-se concluir que vários problemas surgiram nessa transição acelerada para o ensino remoto: falta de materiais didáticos preparados antecipadamente para esta modalidade de ensino, carência de tecnologia nas instituições, acesso limitado à Internet em algumas regiões e o fato de alguns alunos e professores não possuírem notebooks.

Embora uma minoria tenha se adaptado à implantação desse ensino remoto acelerado, seja por conseguirem melhores notas ou devido a terem uma maior facilidade para com a forma como os conteúdos são apresentados, pôde-se notar que a maioria dos alunos não enxerga da mesma forma. Tendo como base as nuvens de palavras, é notório o uso de palavras ofensivas, em muitas situações empregadas devido ao desgaste mental dos alunos, a dificuldade de focar nas aulas e principalmente a questão do conhecimento que, de certo modo, não é adquirido com a mesma qualidade que o ensino presencial permite.

A fim de compreender se fatores geográficos exercem algum tipo de influência na aceitação/rejeição do ensino remoto, também foi realizada uma análise da distribuição dos sentimentos por região do Brasil (Figura 3).

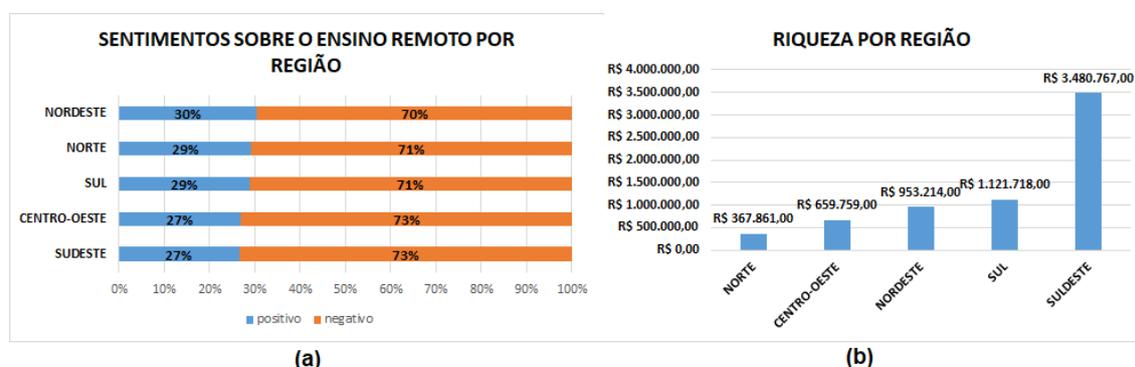


Figura 3. (a) Proporção dos sentimentos por região. (b) Produto Interno Bruto (PIB) por região.

Ao visualizar os resultados, torna-se claro que todas as regiões apresentaram uma

maior proporção de sentimentos negativos. Além disso, pode-se considerar que em todas elas a proporção de *tweets* negativos foi semelhante: em torno de 70%. Sendo assim, não é possível estabelecer uma relação entre a localização dos estudantes, de maneira isolada, com o nível de insatisfação com o ensino remoto.

Sabe-se, contudo, que em regiões como a Norte e a Nordeste, principalmente no interior, há uma escassez de recursos tecnológicos, o que poderia levar a uma insatisfação ainda maior nessas áreas. Porém, no recorte apresentado neste trabalho, este não foi um fator a se levar em consideração, pelo fato de que foram utilizados dados do Twitter, portanto todos os usuários cujas opiniões embasaram esta pesquisa possuíam algum tipo de acesso à Internet.

Ao se analisar o fator econômico, foram tomados como métrica os últimos dados emitidos pelo IBGE sobre o Produto Interno Bruto (PIB) de cada estado, dessa forma pôde-se calcular a riqueza por região (Figura 3(b)). Mediante a comparação dessa métrica com a proporção dos sentimentos por região, percebeu-se que também não existe uma relação entre eles, já que a região Sudeste possui o maior PIB e apresentou o maior nível de insatisfação. Porém, isso também pode decorrer do fato apresentado anteriormente, de que todas as opiniões utilizadas foram de usuários que possuem acesso à Internet.

5. Dificuldades e Limitações

No que tange à condução do trabalho, e principalmente aos recursos disponíveis para sua efetivação, é de fundamental importância tornar evidente os fatores que, de alguma forma, inibiram a ampliação do cenário apresentado. Por esse motivo, nesta seção são descritas as dificuldades e limitações encontradas no decorrer da pesquisa.

De início, pode-se apontar a falta de recursos para a Língua Portuguesa como uma das dificuldades enfrentadas. Como pôde-se observar, alguns dos trabalhos relacionados utilizaram técnicas mais robustas, apresentando uma confiança maior para a classificação. Essas técnicas não puderam ser utilizadas nesta pesquisa pelo fato de muitas delas serem dependentes da língua e não estarem disponíveis para o Português. Pois ainda existem várias lacunas para a mineração de texto, e, conseqüentemente, para a Análise de Sentimentos, utilizando dados escritos na Língua Portuguesa [Souza et al. 2018].

Também cabe destacar que a rede social utilizada para o estudo, o Twitter, impõe em sua estrutura uma série de bloqueios que limitaram o acesso a alguns recursos da plataforma. A existência de perfis privados, bem como a possibilidade de ocultar campos informativos sobre o usuário impossibilitaram a realização de análises mais minuciosas. Neste sentido, a não divulgação de campos como “data de nascimento” e “gênero” inviabilizam a demonstração de, por exemplo, como ocorreu a aceitação da adoção do ensino remoto por faixa-etária e, ocasionalmente, pelo gênero dos usuários. Por conta disso, não foi possível realizar um análise parecida com a do trabalho de [Duong et al. 2020], os quais necessitaram utilizar técnicas bem mais complexas para extrair essas informações.

Além disso, sabe-se que os textos de redes sociais, principalmente textos curtos como os do Twitter, o qual possui um limite de 280 caracteres por publicação, trazem inerentemente dificuldades para a classificação. Textos de redes sociais costumam ser escritos de maneira informal, com gírias, abreviações, erros gramaticais e, muitas vezes, de forma ambígua e/ou irônica [Souza et al. 2016].

Por fim, tendo em vista que foi realizada uma análise geográfica nesta pesquisa, é importante ressaltar a existência de dificuldades em obter a localização precisa dos usuários. O Twitter permite que o usuário insira sua localização manualmente, podendo ele criar derivações de locais já existentes ou até mesmo preencher o campo com informações que não correspondem a nenhuma localização geográfica. Dessa forma, é comum encontrar estados, cidades e até países com outra nomenclatura, além de qualquer outro tipo de informação que não é útil para deduzir a residência do usuário.

Vale ressaltar que a coleta dos *tweets* foi realizada de forma que todos eles continham o campo de localização preenchido. Ainda assim, de todos os *tweets* classificados e utilizados nesta pesquisa, só foi possível identificar uma localização válida em cerca de 62,6% deles, mesmo utilizando alguns filtros e palavras-chave para extração da localidade. Como critério, palavras abreviadas de cidades e estados foram utilizadas. Por exemplo, fez-se uso das siglas dos aeroportos de algumas capitais brasileiras, como “SSA” (Salvador, na Bahia), “BSB” (Brasília), “SLZ” (São Luís do Maranhão) e “CWB” (Curitiba-PR). Também fez-se uso de abreviações, como “Jampa”, para se referir a João Pessoa, na Paraíba, “Sampa” para São Paulo capital e “BH” para Belo Horizonte, em Minas Gerais. Além de nomes populares de cidades como “Cidade Maravilhosa”, para se referir à cidade do Rio de Janeiro, e “Capital do Forró” para Caruaru-PE, os quais também auxiliaram no processo de identificação. Contudo, a dificuldade em obter a geolocalização, bem como a existência de cidades com nomes iguais em estados distintos ocasionaram uma redução do conjunto de dados utilizado para as análises. A Tabela 1 traz exemplos do preenchimento do campo de localização dos usuários do Twitter.

Tabela 1. Exemplos do preenchimento do campo de localização e sua localização geográfica correspondente.

Campo localização	Cidade correspondente	UF
“Blumenau, Brasil”	Blumenau	SC
“Brasília”	Brasília	DF
“ela/dela”	-	-
“em casa”	-	-
“Fortal city”	Fortaleza	CE
“jf”	Juiz de Fora	MG
“no meio do mato”	-	-
“Nova Iorque, mas a do Maranhão”	Nova Iorque	MA
“praia de itapuã, vv”	Vila Velha	ES
“recife”	Recife	PE
“Rio Grande do Sul, Brasil”	-	RS
“rua dos bobos nº 0”	-	-
“São Bernardo do Campo, Brasil”	São Bernardo do Campo	SP
“sp”	-	SP
“Ssa-Ba”	Salvador	BA

6. Considerações finais

Nesta pesquisa, realizou-se uma Análise de Sentimentos de usuários do Twitter acerca dos primeiros meses da adoção do ensino remoto nas instituições de ensino brasileiras durante a pandemia de COVID-19. A partir da classificação das opiniões, pôde-se perceber que

a maior parte delas (74%) foi negativa, mostrando uma grande insatisfação com essa modalidade de ensino, ou pelo menos com sua adoção devido ao distanciamento social.

Também foi possível identificar fatores que levaram a essa rejeição ao ensino remoto, como a quantidade excessiva de atividades passadas pelos professores, o que ocasionou um desgaste mental nos alunos, e a incapacidade de muitos alunos de se concentrar no estudo fora da instituição de ensino. Já outros fatores conhecidos por ocasionarem uma rejeição a essa modalidade, como a vulnerabilidade social e a falta de acesso a tecnologias [Camacho et al. 2020], não foram identificados neste trabalho. Porém, isso se deve ao fato do tipo de dado utilizado: mensagens expressas em redes sociais, garantindo que os alunos cujas opiniões foram consideradas possuíam acesso à Internet.

Como trabalhos futuros, pretende-se realizar uma nova Análise de Sentimentos para verificar como se encontra o sentimento dos alunos, agora que estes podem estar mais adaptados ao ensino remoto, assim como analisar novamente os *tweets* para ter um comparativo temporal. Além disso, pretende-se realizar a análise de outros fatores como faixa etária (embora difícil de ser analisada no Twitter) e gênero dos usuários.

Referências

- AL-Rubaiee, H. S., Qiu, R., Alomar, K., and Li, D. (2016). Sentiment analysis of arabic tweets in e-learning. *Journal of Computer Science*.
- Ali, W. (2020). Online and remote learning in higher education institutes: A necessity in light of covid-19 pandemic. *Higher Education Studies*, 10(3):16–25.
- Alswaidan, N. and Menai, M. E. B. (2020). A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, 62(8):2937–2987.
- Arruda, E. P. (2020). Educação remota emergencial: elementos para políticas públicas na educação brasileira em tempos de covid-19. *EmRede-Revista de Educação a Distância*, 7(1):257–275.
- Camacho, A. C. L. F., Fuly, P., Santos, M., and Menezes, H. (2020). Alunos em vulnerabilidade social em disciplinas de educação à distância em tempos de covid-19. *Research, Society and Development*, 9:275973979.
- Creswell, J. W. (2010). Projeto de pesquisa métodos qualitativo, quantitativo e misto. In *Projeto de pesquisa métodos qualitativo, quantitativo e misto*. Artmed.
- Danka, T. and Horvath, P. (2018). modAL: A modular active learning framework for Python. available on arXiv at <https://arxiv.org/abs/1805.00979>.
- Duong, V., Pham, P., Yang, T., Wang, Y., and Luo, J. (2020). The ivory tower lost: How college students respond differently than the general public to the covid-19 pandemic. *arXiv preprint arXiv:2004.09968*.
- Galletta, D. F., Gaskin, J. E., Koch, H., Anderson, G., Jessup, L., King, J. L., Lowry, P. B., and Wetherbe, J. (2020). Educational disruption & rising faculty expectations. In *Proceedings of AMCIS 2020*.
- Liu, B. and Zhang, L. (2012). *A Survey of Opinion Mining and Sentiment Analysis*, pages 415–463. Springer US, Boston, MA, EUA.

- Moore, J. L., Dickson-Deane, C., and Galyen, K. (2011). e-learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education*, 14(2):129–135.
- Mostafa, L. (2021). Egyptian student sentiment analysis using word2vec during the coronavirus (covid-19) pandemic. In Hassanien, A. E., Slowik, A., Snášel, V., El-Deeb, H., and Tolba, F. M., editors, *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2020*, pages 195–203, Cham. Springer International Publishing.
- Pastor, C. K. L. (2020). Sentiment analysis on synchronous online delivery of instruction due to extreme community quarantine in the philippines caused by covid-19 pandemic. *Asian Journal of Multidisciplinary Studies*, 3(1):1–6.
- Souza, E., Castro, D., Vitório, D., Teles, I., Oliveira, A. L. I., and Gusmão, C. (2016). Characterizing user-generated text content mining: A systematic mapping study of the portuguese language. In *New Advances in Information Systems and Technologies*, pages 1015–1024, Cham. Springer International Publishing.
- Souza, E., Costa, D., Castro, D. W., Vitório, D., Teles, I., Almeida, R., Alves, T., Oliveira, A. L. I., and Gusmão, C. (2018). Characterising text mining: a systematic mapping review of the portuguese language. *IET Software*, 12(2):49–75.
- Vitório, D., Souza, E., and Oliveira, A. L. I. (2019a). Evaluating active learning sampling strategies for opinion mining in brazilian politics corpora. In *19th EPIA Conference on Artificial Intelligence*, pages 695–707, Cham. Springer.
- Vitório, D., Souza, E., and Oliveira, A. L. I. (2019b). Using active learning sampling strategies for ensemble generation on opinion mining. In *8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 114–119. IEEE.
- Wang, D., Feng, S., Wang, D., and Yu, G. (2013). Detecting opinion drift from chinese web comments based on sentiment distribution computing. In *Web Information Systems Engineering – WISE 2013*, pages 72–81.
- Wang, Z. and Cruz, I. F. (2020). Analysis of the impact of covid-19 on education based on geotagged twitter. COVID-19, page 15–23, New York, NY, USA. Association for Computing Machinery.
- Watson, R., Corbett, J., Galletta, D. F., Ives, B., Mandviwalla, M., and Tremblay, M. (2020). Covid-19 and is: Challenges and opportunities for people, careers, and institutions. In *Proceedings of AMCIS 2020*.
- Woźniak, M., Graña, M., and Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16:3 – 17.
- Yin, R. K. (2013). *Case study research: Design and methods*. Sage publications.
- Yu, H.-F., Huang, F.-L., and Lin, C.-J. (2011). Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning*, 85(1-2):41–75.
- Yue, L., Chen, W., Li, X., Zuo, W., and Yin, M. (2018). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60(2):617–663.
- Zimmermann, M., Ntoutsis, E., and Spiliopoulou, M. (2015). Incremental active opinion learning over a stream of opinionated documents. *arXiv preprint arXiv:1509.01288*.