

# Uma análise da aprovação do presidente Jair Bolsonaro ao longo da pandemia de COVID-19 baseada nas transcrições de vídeos do YouTube

Paulo Henrique P. Marcolino<sup>1</sup>, Daniel Mendes Barbosa<sup>1</sup>

<sup>1</sup>Instituto de Ciências Exatas e Tecnológicas – Universidade Federal de Viçosa (UFV)  
Florestal – MG – Brazil

{paulo.marcolino,danielmendes}@ufv.br

**Abstract.** *The growing use of social networks opens up different fronts for research, given the large volume of data generated. In this study, an analysis of sentiments is carried out from the YouTube platform, on the subject of the approval of President Jair Messias Bolsonaro since the beginning of his term and during the pandemic of the COVID-19. We tried to establish a relationship between the feeling of the published video and the feelings generated by it, performing these measurements through extracted transcripts and user comments, respectively. The study identified patterns of channel behavior, indicating a polarization of the audience and a growing aversion and dissatisfaction of viewers towards media channels in general.*

**Resumo.** *A crescente utilização de redes sociais abre várias frentes de pesquisa dado o seu grande volume de dados. Neste estudo, é realizada uma análise de sentimentos tendo como base a plataforma YouTube, acerca da temática da aprovação do presidente Jair Messias Bolsonaro desde o início de seu mandato e durante a pandemia da COVID-19. Buscamos estabelecer uma relação entre o sentimento do vídeo publicado e os sentimentos gerados pelo mesmo, realizando essas medições por meio das transcrições extraídas e dos comentários dos usuários, respectivamente. O estudo identificou padrões de comportamento de canais, indicando uma polarização da audiência e uma crescente aversão e insatisfação por parte dos espectadores em canais de mídia em geral.*

## 1. Introdução

O YouTube é o segundo site mais acessado do Brasil e do mundo<sup>1</sup>, já oferecendo versões locais da plataforma em mais de 100 países ao redor do globo<sup>2</sup>. Sua plataforma atende cerca de 2 bilhões de usuários mensais<sup>3</sup>, sendo 83 milhões apenas no Brasil<sup>2</sup>.

Com base nesses números relevantes, quando se trata de redes sociais online, várias pesquisas e estudos passaram a buscar, analisar e interpretar os dados gerados dentro dessas plataformas. Na corrida presidencial do Brasil em 2018, por exemplo, foi possível perceber que essas mídias contemporâneas começaram a ganhar mais força e passaram a ser mais utilizadas quando comparadas às mídias tradicionais como televisão

<sup>1</sup><https://www.alexa.com/siteinfo/youtube.com>

<sup>2</sup><https://www.globalmediainsight.com/blog/youtube-users-statistics>

<sup>3</sup><https://www.statista.com/topics/2019/youtube/#dossierContents...outerWrapper>

e rádio, conforme abordado por [Silva and Barbosa 2019]. Essas novas plataformas digitais, em muitos casos, se tornaram os principais veículos de comunicação entre políticos e a população em geral, uma vez que o acesso à internet no país vem se tornando mais democrático e pervasivo ao longo do tempo<sup>4</sup>.

A partir da interação promovida nas redes sociais, um tipo de análise é capaz de determinar e aferir o sentimento transmitido pelo autor do conteúdo em questão. Com a aplicação de técnicas para análises de sentimentos, é possível estabelecer a aceitação, a recepção e o sentimento geral que determinado conteúdo gera no público-alvo, e assim prever resultados e interpretar o que está sendo discutido.

Em muitos casos, essas análises requerem normalizações e pré-processamentos dos mais diversos tipos, afim de promover uma melhor acurácia dos resultados. Tratamentos esses, que seriam impossíveis sem o auxílio da computação e da evolução e criação de algoritmos sofisticados. Nos últimos anos, redes neurais vêm ganhando muita relevância e a sua utilização para os mais diversos contextos possibilita o pré-processamento e as análises desses dados de forma automatizada e em um tempo viável.

Tomando como tema do estudo, a aceitação e aprovação do atual presidente do Brasil, Jair Messias Bolsonaro, desde o início do seu mandato em 2019, até o final do segundo trimestre de 2021, o presente trabalho tem como objetivo analisar os sentimentos dos vídeos disponíveis na plataforma YouTube, utilizando o conceito de rede neural para o pré-processamento de alguns dados e o método SentiStrength elaborado por [Thelwall et al. 2010] para a análise dos sentimentos. As análises realizadas tem como premissa, traçar um paralelo entre os sentimentos passados pelos vídeos (transcrições), dos principais canais que abordaram o assunto e a repercussão que esses geraram nos usuários, medida pelos sentimentos associados aos comentários dos mesmos ao longo deste período. Momento esse que foi marcado por uma alternância no governo e o surgimento do novo coronavírus (COVID-19), responsável por instaurar a maior pandemia da história, causando além das milhões de mortes, diversas instabilidades nas mais diversas vertentes sociais.

Apesar da existência de diversos estudos presentes na literatura explorando a plataforma YouTube, existe a escassez de trabalhos relacionando transcrições de vídeos com comentários no idioma Português. Diante disso, essa oportunidade de pesquisa foi a maior motivação para este estudo. O presente trabalho está organizado da seguinte forma: a seção 2 contempla a fundamentação teórica; os trabalhos relacionados são abordados na seção 3; a metodologia é apresentada na seção 4; na seção 5 são discutidos os experimentos e resultados; por fim, na seção 6 são apresentadas as considerações finais e trabalhos futuros.

## **2. Fundamentação Teórica**

Nessa seção serão abordados dois conceitos que foram essenciais para o desenvolvimento do estudo: a adoção de uma estratégia para a realização da análise de sentimentos e a definição da arquitetura e implementação da rede neural utilizada.

---

<sup>4</sup><https://agenciabrasil.ebc.com.br/geral/noticia/2021-08/brasil-tem-152-milhoes-de-pessoas-com-acesso-internet>

## 2.1. Análise de Sentimentos

A pesquisa abordada por [Araújo et al. 2013], contempla o estudo individual acerca da performance dos métodos: Emoticons, LIWC, SentiStrength, SentiWordNet, SenticNet, SASA, Happiness Index e PANAS-t. Diante dos resultados apresentados por esse trabalho, foi definido como método de análise para o presente estudo, a utilização do *SentiStrength*, uma metodologia baseada em aprendizado de máquina e na utilização de dicionários léxicos, que demonstrou um bom desempenho e precisão em análises de dados de diversas plataformas digitais, incluindo o YouTube.

## 2.2. Rede neural

Algoritmos para reconhecimento automático de fala, ou *Automatic Speech Recognition* (ASR), como o utilizado pelo YouTube, geram palavras sem nenhum tipo de tratamento, ou seja, sem pontuação ou letras maiúsculas. Tendo como premissa para a análise de sentimentos a utilização do *SentiStrength*, um método eficiente para a classificação de sentenças curtas, se tratando de transcrições o ideal seria que as mesmas possuíssem pontuação, o que possibilitaria a divisão do texto em frases. Diante disso, como estratégia para recuperar a pontuação das transcrições obtidas, foi implementada uma rede neural.

Existem diversas arquiteturas para redes neurais, projetadas para os mais diversos contextos. Tendo o objetivo mencionando, a rede a ser escolhida deveria seguir a premissa de ter como entrada um fluxo de palavras e como saída, também um fluxo de palavras, porém esse estaria capitalizado e pontuado. Modelos desse tipo são conhecidos como *sequence-to-sequence* (Seq2seq).

Definido o contexto da rede, foi escolhida a arquitetura Transformer, proposta em 2017. Esse modelo vem se tornando bastante relevante na área, considerada o atual estado da arte se tratando de modelos Seq2seq e processamento de linguagem natural (PLN) conforme proposto por [Vaswani et al. 2017]. A figura 1 ilustra ao lado esquerdo, a arquitetura em uma versão mais alto nível para facilitar o entendimento de seus componentes. Ao lado direito dessa mesma figura, o modelo original.

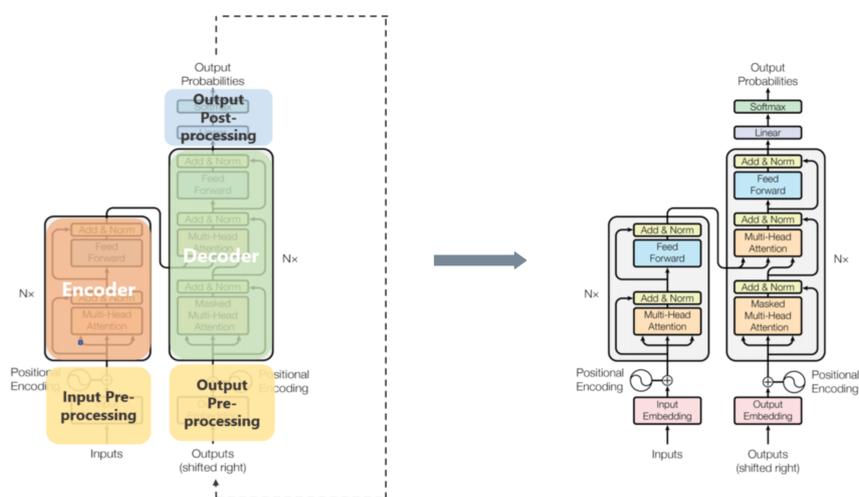


Figura 1. Arquitetura do Transformer. Adaptado de [Kulshrestha 2020]

O modelo basicamente se divide em dois blocos, sendo o primeiro conhecido como *encoder* e o segundo como *decoder*. Em linhas gerais, o *encoder* é a estrutura responsável por codificar a entrada, composta por sequências de palavras, em uma representação numérica que será a entrada para o *decoder*. A codificação aqui se trata de uma representação contextual e não apenas um mapeamento de palavras em números(*tokens*). O modelo leva em consideração também a ordem das palavras, ou seja, a sua posição em uma frase. Esse é um fator característico e importante, visto que, por se tratar de *Seq2seq*, duas frases compostas exatamente pelas mesmas palavras em ordens diferentes podem ter sentidos e significados distintos.

Adentrando a estrutura do *encoder*, o mesmo é composto basicamente por dois elementos, sendo uma rede neural do tipo *Feed Forward* e uma camada conhecida como *Self-Attention*, o elemento mais importante e o coração do modelo. Esse mecanismo vai agregando informação à representação contextual, de modo a identificar padrões e direcionar o foco e a atenção para determinados *tokens*, uma vez que em uma frase existem palavras que são mais importantes do que outras, palavras que remetem a outras e/ou palavras que só fazem sentido na presença de outras. Em suma, para cada palavra existe um peso que a conecta a uma outra palavra, porém esse peso depende do contexto.

O *decoder* é composto por três elementos. Além de uma rede *Feed Forward* e o mecanismo *Self-Attention*, existe uma camada intermediária, a *Encoder-Decoder Attention*, que, além de levar em conta a representação das palavras que já foram analisadas, também leva em consideração a representação das palavras que serão a sua entrada.

### 3. Trabalhos Relacionados

Apesar do campo ser muito explorado, ainda existem inúmeros desafios a serem contornados e inúmeras situações a serem investigadas. Especificamente sobre a plataforma YouTube, mesmo com os vários mecanismos que a rede oferece como número de *likes* e *dislikes*, número de *views*, número de vezes que um vídeo foi favoritado, entre outros, grande parte das análises são feitas a partir dos comentários dos usuários. As pesquisas de [Bhuiyan et al. 2017] e [Filippova and Hall 2011], buscaram estabelecer uma relação entre os metadados supracitados e os comentários como um maneira de entender melhor e aprimorar a acurácia dos sentimentos extraídos dos comentários.

Na pesquisa de [Asghar et al. 2015], são explorados, além dos comentários, alguns outros metadados como títulos e descrições dos vídeos, a partir de análises usando várias técnicas presentes na literatura, como: métodos léxicos, algoritmos de aprendizado de máquina supervisionados e não-supervisionados.

Explorando outros contextos e aplicações relacionados a essas análises, o trabalho de [Silva and Barbosa 2019] buscou traçar a relação entre o âmbito político e o comportamento dos usuários na plataforma. Investigando apenas comentários, foi abordada a ideia de polaridades de modo a determinar a aceitação dos candidatos presentes na corrida eleitoral para a presidência do país. Foi adotado neste estudo algumas medidas e técnicas abordadas pelo estudo citado anteriormente, como, por exemplo, o fluxo estabelecido para o pré-processamento de comentários.

Em estudos como o de [Kaushik et al. 2013], são propostos algoritmos para reconhecimento automático de fala devido as dificuldades em se extrair transcrições que

possuem uma boa acurácia de forma que pudessem ser utilizadas para análise de sentimentos. Dificuldades essas que podem ter como causa diversos fatores, como por exemplo: ruídos no vídeo, condições de gravação, erros na pronúncia, sotaques diversos, entre outros fatores.

Dada a quantidade de dados que a instituição Google(Alphabet Inc.), atual proprietária do YouTube, processa e administra diariamente, o algoritmo de geração de transcrições da plataforma possui uma melhoria contínua<sup>5</sup>, fazendo com que as transcrições possuam uma significativa acurácia conforme constatado por [Kim et al. 2019]. Diante disso, o presente trabalho se baseou nas transcrições geradas pela plataforma em questão.

## 4. Metodologia

A metodologia foi dividida em quatro partes e cada uma dessas será melhor abordada nas subseções a seguir. A divisão dessa seção foi feita da seguinte forma: (i) coleta de dados do YouTube, (ii) transformação da transcrição, (iii) análise de sentimentos das transcrições e comentários e por fim, (iv) normalização dos resultados da análise de sentimentos. A figura 2 ilustra as etapas do processo.



Figura 2. Etapas da metodologia

### 4.1. Coleta de dados

A coleta e composição da base de dados foi dividida em duas etapas. A primeira se deu a partir da utilização da YouTube Data API<sup>6</sup>, onde foram coletados os vídeos postados em 2019, 2020 e 2021(até Agosto), tendo como parâmetros para os termos de busca: "Bolsonaro" e "Jair Messias Bolsonaro".

A segunda etapa, diz respeito às transcrições dos vídeos, contudo a API oficial do YouTube não disponibiliza uma forma de coletar esse dado. As transcrições então, foram coletadas com o uso de um pacote da linguagem Python, o YouTube Transcript API<sup>7</sup>. Com esta API e já possuindo os códigos de identificação dos vídeos adquiridos na primeira etapa, as transcrições disponíveis foram coletadas. A figura 3 ilustra todo o processo de composição da base de dados.

### 4.2. Transformação da transcrição

Para implementação do modelo Transformer, foi utilizado o *framework* voltado à *machine learning* de nome TensorFlow<sup>8</sup> e devido a sua complexidade e a limitação de hardware, o treinamento foi feito inteiramente pela plataforma Google Colaboratory<sup>9</sup>. Foi utilizado

<sup>5</sup>Rede neural - <https://ai.googleblog.com/2015/08/the-neural-networks-behind-google-voice.html>

<sup>6</sup>YouTube Data API - <https://developers.google.com/youtube/v3>

<sup>7</sup>YouTube Transcript API - <https://pypi.org/project/youtube-transcript-api/>

<sup>8</sup>Framework TensorFlow - <https://www.tensorflow.org/>

<sup>9</sup>Google Colaboratory - <https://colab.research.google.com>



**Figura 3. Fluxo de coleta de dados**

uma máquina de processador Intel(R) Xeon(R) CPU @ 2.30GHz, GPU Tesla K80 e 12GB de memória RAM.

#### 4.2.1. Pré-processamento do Corpus e Treinamento

Para o treinamento do modelo de pontuação, foi utilizado o corpus CHAVEFolha<sup>10</sup>. Essa base de dados é formada por textos completos (edição diária) do jornal Folha de São Paulo, publicados nos anos de 1994 e 1995, totalizando cerca de 730 edições e aproximadamente 35,7 milhões de palavras. A escolha do corpus se deu pelo fato do tamanho dos textos ser relevante para o treinamento do modelo, uma vez que em um texto do conjunto existem diversas frases, possibilitando assim que o modelo aprenda e identifique a existências de pontuações entre as palavras e não apenas ao final de um frase.

Conforme mencionado na seção 2, tanto o *encoder* quanto o *decoder* possuem a sua respectiva entrada. Assim sendo, para a utilização do CHAVEFolha, foi necessário um tratamento prévio dos textos. A entrada do *encoder* foi constituída do texto pré-processado e limpo (sem pontuação e/ou capitalização), similar a um texto gerado por um ASR. Já a entrada do *decoder*, foi constituída do texto original, capitalizado e pontuado.

Para a realização do treinamento, os 730 arquivos foram divididos em cinco grupos de mesmo tamanho aproximadamente, de modo a aplicar uma ideia de validação cruzada como estratégia para contornar o *overfitting* ou sobre-ajuste do modelo, situação essa na qual a rede neural apenas memoriza a base e não é capaz de generalizar suas predições. Após a construção e treinamento da rede, por meio do modo de inferência do modelo, todas as transcrições foram pontuadas e capitalizadas.

#### 4.3. Análise de sentimentos

Tendo as transcrições capitalizadas e pontuadas e os comentários coletados, cada um desses elementos passou por um pré-processamento fundamental para a aplicação do método de análise de sentimentos, conforme ilustrado na figura 4. Após essa etapa os elementos foram submetidos ao método de análise *SentiStrength*.

<sup>10</sup>CHAVEFolha - <https://www.linguateca.pt/chave/>



Figura 4. Etapas do pré-processamento. Adaptado de [Silva and Barbosa 2019]

#### 4.4. Normalização dos sentimentos

Em relação às transcrições, vídeos longos possuem mais sentenças, logo uma soma direta de resultados da análise de sentimentos, resultaria em um valor desbalanceado já que o mesmo estaria enviesado pela duração vídeo, e não pelo sentimento expressado pelo vídeo propriamente dito.

Como a ideia proposta é analisar o vídeo como um todo, o valor de cada sentença foi somado e posteriormente, foi feita uma média aritmética. Sendo  $n$  o número de sentenças de um vídeo e  $s$  representando o sentimento extraído da sentença, conforme a eq. 1.

$$\text{Sentimento final de um vídeo} = \left( \sum_{n=1}^n s_n \right) / n \quad (1)$$

Em relação aos comentários, buscando seguir a mesma linha de raciocínio das transcrições, os valores dos comentários foram somados e assim como anteriormente, foi feito uma média aritmética desse valor, normalizando também a questão da quantidade de comentários por vídeos. Sendo  $n$  o número de comentários associados a um vídeo e  $c$  o sentimento extraído daquele comentário, conforme a eq. 2. Se tratando de comentários, vale ressaltar a situação na qual um comentário possui várias sentenças. Nesse caso, as sentenças não foram divididas para análise de sentimentos nem para o cálculo do sentimento final do mesmo.

$$\text{Sentimento final dos comentários} = \left( \sum_{n=1}^n c_n \right) / n \quad (2)$$

## 5. Experimentos e Resultados

Nesta seção, serão apresentados os resultados obtidos pela rede neural no que tange a recuperação da pontuação das transcrições e, em seguida, os resultados finais referentes ao objetivo deste estudo.

### 5.1. Resultados da transformação

O número de *epochs*, ou seja, o número de vezes que a base de treinamento passou completamente pela rede, foi 14. Como hiper-parâmetros do modelo, levando em conta a limitação de recursos computacionais, foram utilizados parâmetros reduzidos se comparados aos do estudo original proposto por [Vaswani et al. 2017]. Especificações aplica-

das: 4 *layers*, 8 *heads*, 0.1 para *dropout rate*, 128 para *dimension model*( $d_{model}$ ) e 512 para *inner layer dimension*( $d_{ff}$ ).

Ao final de todo o treinamento, duas métricas foram utilizadas para avaliar a precisão do modelo: perda(*loss*) e acurácia(*accuracy*). A taxa de perda representa de forma direta a qualidade do modelo, ou seja, caso o número de erros seja alto, a perda será alta, denotando um modelo incorreto e ineficiente. Em outras palavras, expressa de forma mais pessimista, o quão ruim é o modelo. Para o cálculo dessa taxa foi utilizado o método de *Cross Entropy*. Inicialmente a taxa foi de 9.59, já ao final do treinamento a taxa ficou em 0.376. Um resultado considerado positivo e satisfatório uma vez que a distância entre a previsão e o resultado correto é pequeno, próximo a zero.

Se tratando da acurácia do modelo, essa taxa mede em percentual a performance do modelo, comparando as previsões do modelo com os valores corretos. Para essa métrica foi atingido o valor 0.9092. Denotando uma boa acurácia, considerada suficiente para o prosseguimento do estudo e transformação das transcrições.

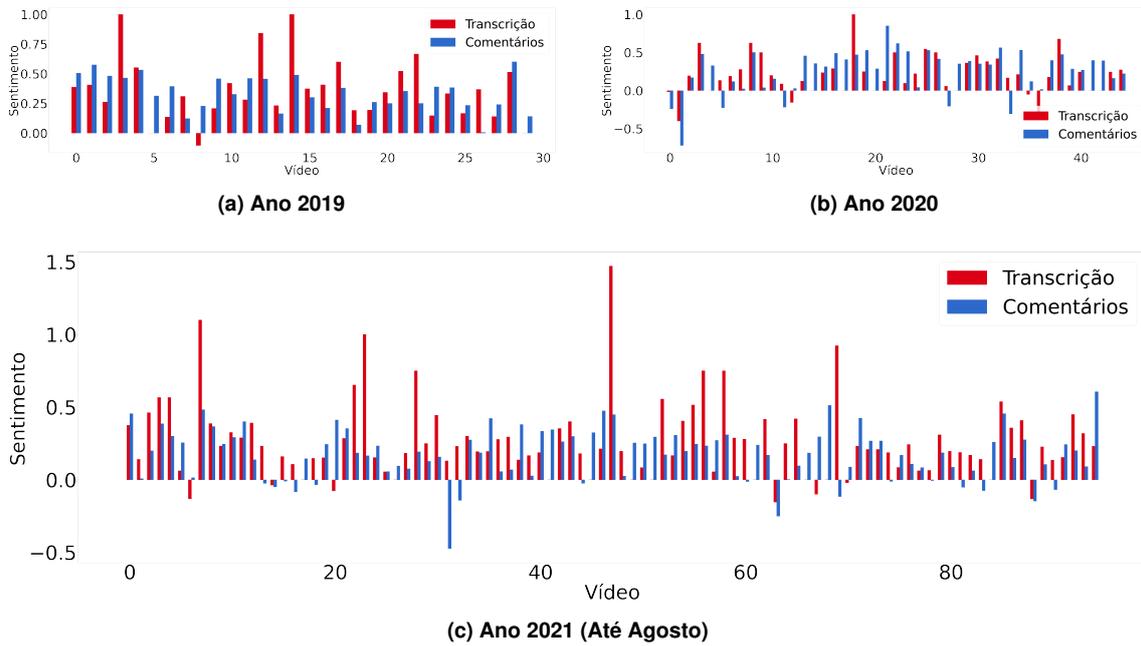
## 5.2. Resultados da análise

Diante desses resultados das coletas, foram identificados dois grandes grupos de canais: canais de opinião(pessoas comuns) e canais de veículos de comunicação(órgãos governamentais, imprensa, entre outros). Em termos percentuais, canais que publicam entre 1 e 5 vídeos, representam cerca de 86% de toda a base. O segundo grupo, apesar de possuir um menor número absoluto de canais, é o conjunto que detém o maior volume de publicação.

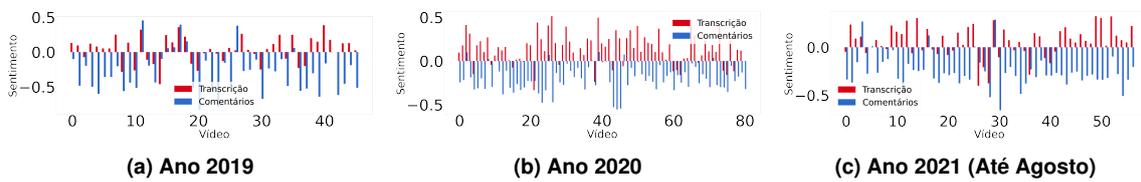
Como demonstração de resultados, serão apresentados apenas alguns gráficos que ilustram os padrões de comportamento de alguns canais e a repercussão dos usuários, de modo a possibilitar uma melhor visualização e entendimento acerca dos resultados obtidos. Em todos os gráficos, o eixo horizontal representa o número simbólico do vídeo e o eixo vertical representa a média de sentimentos, tanto da transcrição quanto dos comentários.

O primeiro canal abordado e evidenciado corresponde ao canal oficial do presidente Jair Messias Bolsonaro. A figura 5 ilustra o posicionamento tanto do autor quanto dos usuários que assistem e acompanham o conteúdo. É possível perceber rapidamente que ao longo dos anos, o número de vídeos publicados aumentou consideravelmente, principalmente levando em consideração o fato de as coletas do ano de 2021 terem sido feitas até o mês de Agosto. Quanto a relação entre o sentimento expressado pelo vídeo e o sentimento expressado pelos usuários, percebe-se que o padrão aqui é uma positividade em ambos, ou seja, o sentimento dos espectadores é o mesmo dos vídeos.

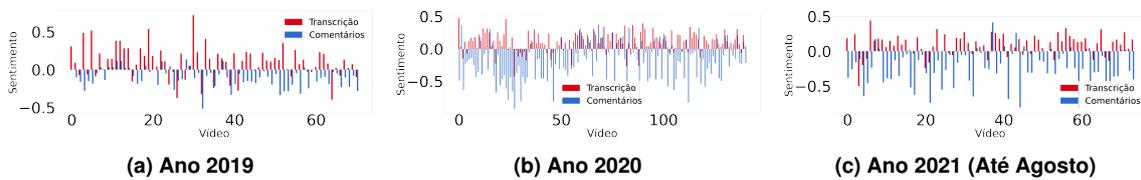
As figuras 6, 7 e 8 representam uma visualização geral de alguns dos principais veículos de mídias presentes na base de dados coletada. Dentre esses, estão nas respectivas imagens os canais: Morning Show, Os Pingos nos Is e Record News. O comportamento demonstrado nesses se propaga por todos os canais desse nicho, ou seja, canais de veículos de imprensa e notícias em geral. Eles apresentam majoritariamente uma linguagem positiva e/ou neutra, mas, em contrapartida, os espectadores tendem a ser negativos, sentimento esse que se fez mais presente conforme o passar dos anos. Um comportamento completamente oposto ao visto no canal do presidente(figura 5), denotando possíveis indícios de uma polarização.



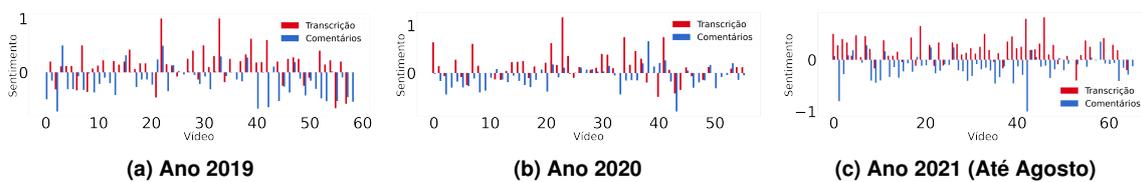
**Figura 5. Relação entre transcrições e comentários do canal Jair Bolsonaro**



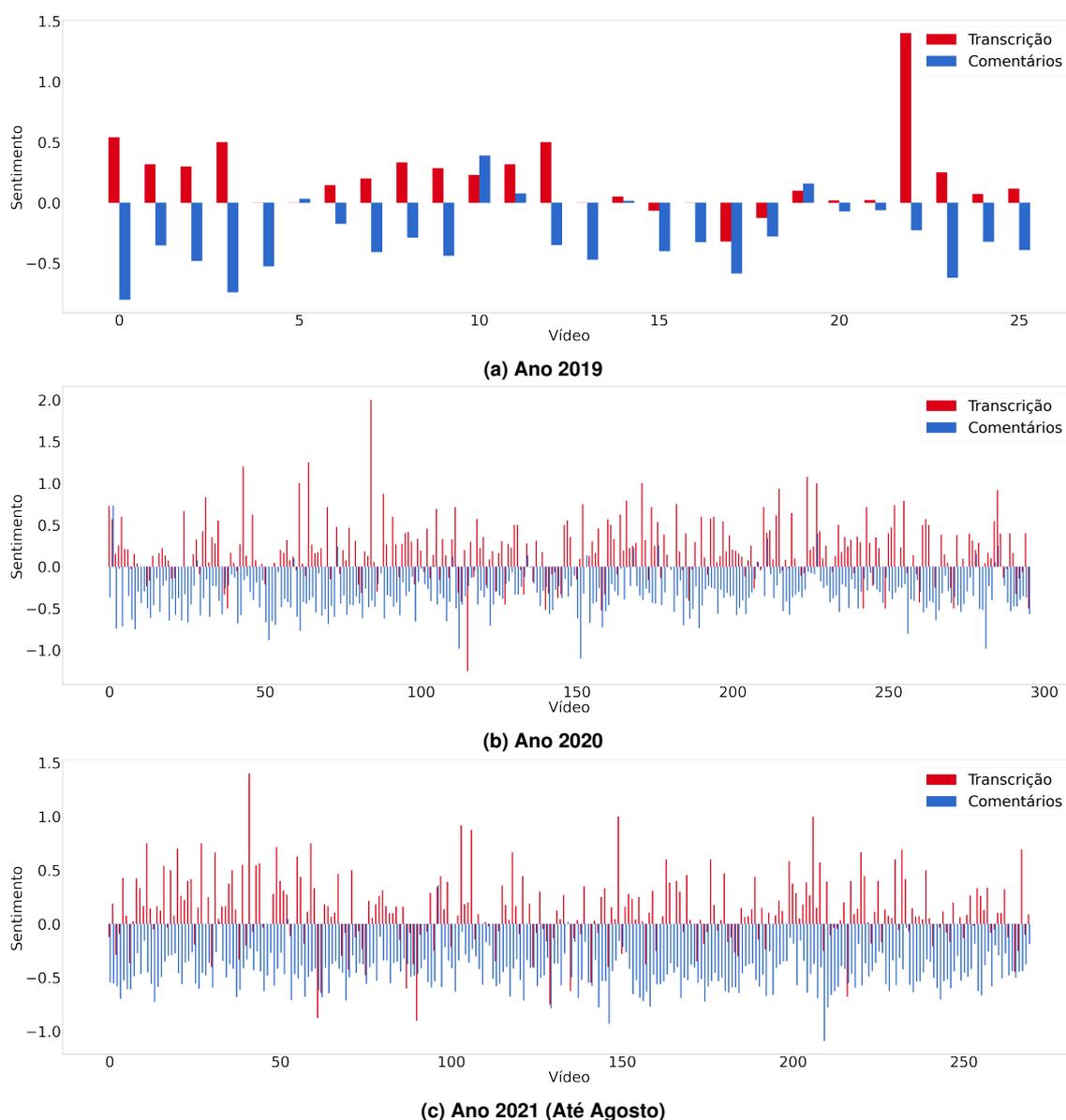
**Figura 6. Relação entre transcrições e comentários dos canal Morning Show**



**Figura 7. Relação entre transcrições e comentários do canal Os Pingos nos Is**



**Figura 8. Relação entre transcrições e comentários do canal Record News**



**Figura 9. Relação entre transcrições e comentários do canal UOL**

Um canal que se destacou na base de dados foi o canal UOL, que mudou significativamente o seu comportamento ao longo dos anos no que tange a volume de publicação conforme pode ser observado na figura 9, mas como mencionado anteriormente, o padrão de sentimento para canais desse nicho se manteve.

Analisando agora a base de dados de um modo geral, cerca de 66% dos vídeos possuem um caráter positivo, 27.15% negativo e 6.95% totalmente neutro. Analisando a média da negatividade de comentários ao longo dos anos, em 2019, 61% eram comentários negativos. Em 2020, esse número cresceu para 66%, e no ano de 2021 (Até Agosto) a quantidade de comentários negativos representa cerca de 81% da base de dados desse ano. Avaliando o sentimento do vídeo e o sentimento que o mesmo propaga, é possível constatar uma crescente na insatisfação em relação ao atual governo.

Grande parte do volume de vídeos publicados relacionados ao tema, são de canais

de comunicação, notícias e jornalismo no geral. Vale destacar que a linguagem utilizada por esses, ainda que classificados como positivos, tende a possuir um caráter mais imparcial e neutro, fato que pode justificar o grande volume de vídeos positivos e/ou neutros. Foi constatado que em média, os vídeos desses canais classificados como positivos, apresentam cerca de 70% de neutralidade, ou seja, apenas em alguns trechos do vídeo é expressado uma positividade, porém essa positividade não se propaga para os usuários. A reação é justamente oposta, como pôde ser observado.

## 6. Considerações Finais e Trabalhos Futuros

Neste trabalho foi explorada a relação entre os sentimentos passados pelos vídeos, a partir de suas transcrições e a reação dos usuários, medida por meio dos comentários. Acerca da aprovação do presidente ao longo do tempo, é possível perceber que o volume de vídeos relacionados ao tema aumentou e diante dos resultados foi possível abordar duas análises distintas.

A primeira análise diz respeito a polarização, como houve no canal do presidente. Não importando as entidades envolvidas nos conteúdos passados pelos vídeos, os comentários sempre tenderam a seguir o sentimento transmitido pelos vídeos, denotando assim uma convergência de ideais. A segunda, se refere ao comportamento dos canais de veículos de mídia, onde mesmo os vídeos transmitindo neutralidade e/ou positividade, os comentários negativos foram aumentando durante o intervalo de tempo analisado, denotando uma crescente insatisfação.

Acerca da plataforma e do crescimento de usuários e canais, foi constatado também e vale ressaltar que vêm havendo uma profissionalização do conteúdo produzido para o YouTube, com a crescente inserção dos veículos de mídias tradicionais na plataforma.

Em trabalhos futuros, considera-se o estudo da aplicação de outras técnicas para análise de sentimentos como análise em nível de aspectos aliada a técnicas de detecção de entidades, bem como a utilização de outros metadados disponibilizados pelo YouTube, como dados estatísticos dos vídeos e *threads* de comentários de modo a determinar quais vídeos e canais promovem mais discussões e geram mais engajamento entre os usuários.

## Referências

- Araújo, M., Gonçalves, P., Benevenuto, F., and Cha, M. (2013). Métodos para análise de sentimentos no twitter. In *Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)*, page 19. sn.
- Asghar, M. Z., Ahmad, S., Marwat, A., and Kundi, F. M. (2015). Sentiment analysis on youtube: A brief survey. *arXiv preprint arXiv:1511.09142*.
- Bhuiyan, H., Ara, J., Bardhan, R., and Islam, M. R. (2017). Retrieving youtube video by sentiment analysis on user comment. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 474–478. IEEE.
- Filippova, K. and Hall, K. B. (2011). Improved video categorization from text metadata and user comments. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 835–842.

- Kaushik, L., Sangwan, A., and Hansen, J. H. (2013). Sentiment extraction from natural audio streams. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8485–8489. IEEE.
- Kim, J. Y., Liu, C., Calvo, R. A., McCabe, K., Taylor, S. C., Schuller, B. W., and Wu, K. (2019). A comparison of online automatic speech recognition systems and the nonverbal responses to unintelligible speech. *arXiv preprint arXiv:1904.12403*.
- Kulshrestha, R. (2020). Transformers. <https://towardsdatascience.com/transformers-89034557de14>.
- Silva, C. A. and Barbosa, D. M. (2019). Analyzing the acceptance of the 2018 brazilian presidential election's main candidates based on youtube comments. In *Proceedings of the 25th Brazilian Symposium on Multimedia and the Web*, pages 377–384.
- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12):2544–2558.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.