



## **Análise de Sentimentos Aplicada a Vídeos Educacionais do Youtube em Busca de Padrões de Engajamento**

**Renan M. CICILIO<sup>1</sup>; Diego SAQUI<sup>2</sup>**

### **RESUMO**

O estudo investigou a relação entre a linguagem e a emoção demonstrada pelos professores em vídeo aulas educacionais e o engajamento dos alunos. Foi aplicado um algoritmo de Análise de Sentimento em legendas de vídeos do YouTube, relacionando a linguagem com a proporção de likes/views dos vídeos. Os resultados mostraram que a correlação entre o grau de positividade e o engajamento é mínima e não relevante para o envolvimento dos alunos. Outros fatores, como elementos visuais, técnicas linguísticas, piadas e conteúdo curioso, parecem influenciar mais o engajamento. A presença de linguagem negativa não teve um impacto considerável no engajamento geral. Conclui-se que a análise de sentimentos por si só não é suficiente para explicar o engajamento, sendo necessário explorar outros fatores, como técnicas criativas e interações com o público, para entender melhor o que motiva os espectadores a se envolverem com o conteúdo educacional em vídeos. Futuras pesquisas devem abordar esses elementos para uma compreensão mais completa do impacto da linguagem e de outros fatores no engajamento em vídeo aulas educacionais.

**Palavras-chave:** Vídeos; Engajamento; Youtube; Educação; Análise de Sentimentos

### **1. INTRODUÇÃO**

Vídeo aulas têm desempenhado diversos papéis na educação, sendo amplamente utilizadas em diferentes áreas de estudo (Sameer, 2020). Durante a pandemia de COVID-19, o aumento no uso dessas aulas destaca sua importância como ferramenta educacional (Rahmatika et al., 2021). No entanto, surgem questões sobre sua eficácia e como melhorar o engajamento entre alunos e professores no ambiente digital (Fidelis et al., 2016). Estudos indicam que certos tipos de vídeo aulas com técnicas interativas do instrutor são mais eficazes em capturar a atenção e gerar maior engajamento dos alunos (Kim et al., 2014).

Essa pesquisa avalia a linguagem e a emoção demonstrada pelo professor de um vídeo como possíveis fatores que contribuem para a eficácia e retenção da atenção do aluno. Estudos indicam que o estilo de abordagem frontal pode superar o estilo de abordagem lateral na efetividade das aulas (Beege, 2019). A investigação dessas características pode trazer insights valiosos para aprimorar as vídeo aulas e promover um engajamento mais eficaz na aprendizagem.

### **2. FUNDAMENTAÇÃO TEÓRICA**

O campo de "Análise de Sentimento" permite a identificação das emoções transmitidas pela linguagem em vídeos educacionais (ALLEN, 2003). O estudo aplicará um algoritmo de Análise de

<sup>1</sup> Renan M. Cicilio, Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais - *Campus Muzambinho*. Email: [renanarms11@gmail.com](mailto:renanarms11@gmail.com).

<sup>2</sup> Diego Saqui, Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais - *Campus Muzambinho*. Email: [diego.saqui@muz.ifsuldeminas.edu.br](mailto:diego.saqui@muz.ifsuldeminas.edu.br).

Sentimento em legendas extraídas de vídeos educacionais do YouTube, buscando correlacionar a linguagem utilizada com o engajamento dos alunos, medido pela proporção de likes/views do vídeo. Para essa análise, o Natural Language Toolkit (NLTK) é essencial.(NLTK, 2023).

O NLTK é amplamente utilizado para Processamento de Linguagem Natural (PLN) e a remoção de Stopwords melhora a eficácia do processamento de dados (Wilbur et al., s.d.; Kalaivani et al., 2021). Além disso, o NLTK oferece a ferramenta Sentiment Intensity Analyser que é usada para classificar a polaridade das frases (Mahreen et al., 2022).

A proporção de likes/views é uma métrica relevante para o estudo, pois está positivamente associada à duração da visualização do vídeo e ao sentimento dos comentários (PARK, et al., 2021).

### 3. MATERIAL E MÉTODOS

O conjunto de dados foi construído seguindo precauções específicas para garantir sua usabilidade, estas são: Os vídeos analisados são apenas da língua inglesa para facilitar análise, e nenhuma playlist foi usada no conjunto de dados para evitar viés. A biblioteca Pandas do Python foi utilizada para ler e manipular o conjunto de dados, focando na coluna de ID dos vídeos.

Em seguida, as legendas de cada vídeo foram extraídas em larga escala usando a API do YouTube Transcript, armazenando cada linha em um arquivo CSV com o nome do ID do vídeo.

Para a Análise de Sentimento, o NLTK foi empregado para remover as stopwords e deixar apenas o texto relevante (NLTK, 2023). O tratamento das legendas resultou em uma coluna "Treatment 1". (Figura 1)

A análise de sentimento foi realizada utilizando o módulo VADER do NLTK, categorizando cada frase em uma escala de -1 a 1. As pontuações foram salvas na coluna "Score" do arquivo CSV (Figura 1).

text	treatment_1	Score	Score_Untreated
hi everyone this is lecture 19 immune	hi everyone lecture 19 immune	{'neg': 0.0, 'neu': 0.645, 'pos': 0.355, 'compound': 0.296}	{'neg': 0.0, 'neu': 0.732, 'pos': 0.268, 'compound': 0.296}
system so today we're going to talk	system today we're going talk	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
about the nonspecific and adaptive	nonspecific adaptive	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
immune responses the nonspecific	immune responses nonspecific	{'neg': 0.0, 'neu': 0.476, 'pos': 0.524, 'compound': 0.296}	{'neg': 0.0, 'neu': 0.577, 'pos': 0.423, 'compound': 0.296}
immunity is also sometimes called innate	immunity also sometimes called innate	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}

**Figura 1: Exemplo de texto após análise de sentimentos**

A soma das pontuações compostas de todas as linhas do texto extraído forneceu a pontuação total de um vídeo, permitindo estabelecer a correlação entre o sentimento geral do vídeo

e seu engajamento.

Os dados de estatísticas do YouTube, como número de likes e visualizações, foram obtidos usando a YouTube Data API, e a proporção de likes/views foi calculada como a métrica para testar e avaliar a hipótese proposta pelo estudo.

#### 4. RESULTADOS E DISCUSSÃO

	Likes	Views	Like/View	Neutral Score	Positive Score	Negative Score	Total Score
Likes	1.0	0.9	0.2	-0.1	-0.0	0.1	-0.1
Views	0.9	1.0	-0.1	-0.1	0.0	0.1	-0.1
Like/View	0.2	-0.1	1.0	-0.3	-0.1	-0.2	-0.1
Neutral Score	-0.1	-0.1	-0.3	1.0	0.9	0.8	0.6
Positive Score	-0.0	0.0	-0.1	0.9	1.0	0.6	0.8
Negative Score	0.1	0.1	-0.2	0.8	0.6	1.0	0.1
Total Score	-0.1	-0.1	-0.1	0.6	0.8	0.1	1.0

**Figura 2:** Tabela de correlações entre os valores numéricos do *dataset*.

Após o tratamento e análise, foi estabelecido uma tabela de correlações (Figura 2) entre a pontuação total e a proporção de likes/views de cada vídeo. Os dados mostram que a correlação entre o grau de positividade e o engajamento é mínima e praticamente irrelevante para o engajamento, pelo menos considerando a métrica utilizada. Parece que outros fatores, como o uso de elementos visuais, técnicas linguísticas, piadas e conteúdo curioso, podem ser mais relevantes para o engajamento.

Exemplos de vídeos que ilustram isso são aqueles que ensinam sobre termodinâmica. Enquanto um vídeo com uma abordagem padrão pode ser monótono e não estimular o espectador, um vídeo mais elaborado, com gráficos interessantes, trocadilhos e informações curiosas, pode ser mais envolvente, independentemente de sua pontuação de análise de sentimentos.

É importante notar que é raro encontrar vídeos com pontuação total negativa, o que sugere que a linguagem negativa tem pouco ou nenhum impacto no envolvimento geral dos alunos. No entanto, existe um limite para quão baixa a pontuação pode chegar antes que comece a afetar negativamente o envolvimento.

#### 5. CONCLUSÃO

A análise de sentimentos revelou que a correlação entre o grau de positividade e o engajamento em vídeos educacionais é mínima e praticamente irrelevante para a proporção de likes/views. Outros fatores, como o uso de elementos visuais, técnicas linguísticas, piadas e conteúdo curioso, parecem ter um papel mais significativo no envolvimento dos alunos. A análise

de sentimentos por si só não é suficiente para explicar o engajamento; fatores não abordados, como técnicas criativas e interações com o público, podem ser cruciais para compreender o que realmente motiva os espectadores a se envolverem com o conteúdo. Futuras pesquisas devem explorar esses elementos para obter uma visão mais abrangente do impacto da linguagem e de outros fatores no engajamento em vídeos educacionais.

## REFERÊNCIAS

ALLEN, James F. **Natural language processing**. [S.l.]: John Wiley and Sons Ltd., 2003.

MAIK BEEGE e colab. Social entities in educational videos: Combining the effects of addressing and professionalism. *Computers in Human Behavior*, v. 93, p. 40–52, 2019. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0747563218305867>>.

KIM, Juho e colab. Data-driven interaction techniques for improving navigation of educational videos: Proceedings of the 27th Annual ACM symposium on user interface software and technology. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/2642918.2647389>>.

KALAIVANI, RAMA E. **The effect of stop word removal and stemming in ... - proquest.com**. Disponível em: <<https://www.proquest.com/openview/7d09e95c5e0d540c8765cf0d41a7ce39/1?pq-origsite=gscholar&cbl=2031963>>.

MAHREEN, Juveriya. **Sentiment analysis using vader**. Disponível em: <<https://www.analyticsvidhya.com/blog/2022/10/sentiment-analysis-using-vader/#:~:text=It%20is%20a%20rule%2Dbased,method%20to%20determine%20the%20sentiment.>>>.

PARK, Minsu e NAAMAN, Mor e BERGER, Jonah. **A data-driven study of view duration on YouTube**. Proceedings of the International AAAI Conference on Web and Social Media, v. 10, p. 651–654, 2021. Disponível em: <<https://ojs.aaai.org/index.php/ICWSM/article/view/14781>>.

RAHMATIKA, Rahmatika e YUSUF, Munawir e AGUNG, Leo. **The effectiveness of YouTube as an online learning media**. Disponível em: <<https://ejournal.undiksha.ac.id/index.php/JET/article/view/33628>>.

SAMEER, Ahmed. **Videoconferencing Classes: A solution to the social distance caused by COVID-19 or a big problem?** SciELO Preprints, 2020. Disponível em: <<https://preprints.scielo.org/index.php/scielo/preprint/view/787>>.

**What is natural language processing?** Disponível em:

<<https://www.ibm.com/topics/natural-language-processing>>.

Disponível em: <<https://www.nltk.org/>>.