



DEEP LEARNING PARA CLASSIFICAÇÃO DE SENTIMENTO EM VÍDEOS UTILIZANDO LEGENDAS E COMENTÁRIOS

Jean Ferreira DIAS¹; **Diego SAQUI²**; **Heber Rocha MOREIRA³**

RESUMO

Este relato de pesquisa explora a aplicação de técnicas de Deep Learning para a classificação de sentimento em vídeos do YouTube, utilizando legendas e comentários como fontes de dados. Desenvolveu-se um modelo híbrido de Deep Learning, combinando camadas de embedding e LSTM bidirecional com entradas TF-IDF, para capturar nuances contextuais e estatísticas da linguagem. O modelo foi treinado e avaliado em 5 rodadas, demonstrando desempenho superior (Acurácia: 0.8857 ± 0.0653 ; F1-Score: 0.8566 ± 0.0828) em comparação com abordagens léxicas (VADER) e pré-treinadas (RoBERTa). A análise de um dataset completo revelou predominância de sentimentos negativos e uma correlação fraca entre sentimento e métricas de engajamento, sugerindo que o sentimento expresso não influencia diretamente o engajamento do vídeo. Os resultados indicam a eficácia da abordagem híbrida para análise de sentimento em conteúdo audiovisual.

Palavras-chave: Processamento de Linguagem Natural; YouTube; Engajamento; Legendas

1. INTRODUÇÃO

Desde sua consolidação a internet se tornou a maior fonte de dados para estudos. Disponíveis em diversas fontes e formatos, como texto, imagens, áudios e vídeos, o maior desafio passou de onde conseguir dados para como lidar com esses tipos de dados desestruturados (FUHR e FRANZEN, 2020).

O avanço das redes sociais ampliou ainda mais a geração de dados não estruturados, especialmente nas plataformas de compartilhamento de vídeos como o YouTube, onde usuários manifestam opiniões em comentários e reações. Essas informações podem ser exploradas por técnicas de análise de sentimentos para apoiar estratégias de marketing digital e compreensão do comportamento do público (SILVA, 2021).

O presente trabalho explora a eficácia de técnicas avançadas de aprendizado profundo, aplicando-as para classificar sentimentos em vídeos do YouTube, com foco nas legendas e comentários. O objetivo é entender a relação entre o sentimento expresso pelo público e as métricas de engajamento dos vídeos, oferecendo insights que possam guiar estratégias de criação e promoção de conteúdo na plataforma.

¹Jean Ferreira Dias, Bacharelado em Ciência da Computação, IFSULDEMINAS – Campus Muzambinho. E-mail: jean.dias@alunos.ifsuldeminas.edu.br.

²Diego Saqui, Universidade Federal de Lavras (UFLA) - Campus São Sebastião do Paraíso. E-mail: diego.saqui@ufla.br.

³Heber Rocha Moreira, IFSULDEMINAS – Campus Muzambinho. E-mail: heber.moreira@ifsuldeminas.edu.br.

2. FUNDAMENTAÇÃO TEÓRICA

O processamento de linguagem natural (PLN) tem avançado significativamente com técnicas de aprendizado profundo, permitindo converter textos em representações numéricas compreensíveis por algoritmos. Isso facilita a identificação de relações sintáticas e semânticas entre palavras, aprimorando tarefas como a análise de sentimentos (ARRUDA, 2019).

A análise de sentimentos visa classificar textos em positivos, negativos ou neutros, avaliando emoções e opiniões. Essa técnica é especialmente relevante em mídias sociais, onde usuários expressam suas percepções sobre produtos e serviços, influenciando diretamente estratégias de marketing e engajamento (SILVA, 2021).

Em contraste, métodos baseados em aprendizado profundo, particularmente redes neurais recorrentes como LSTM, têm se mostrado mais eficazes na captura de contextos complexos e nuances das linguagens naturais. Esses métodos possibilitam representações contextuais que levam em conta diferentes significados dependendo do contexto das palavras, melhorando a análise de sentimentos (ARRUDA, 2019).

Por fim, a integração de técnicas estatísticas com aprendizado profundo tem resultado em modelos híbridos mais precisos e robustos, especialmente para análise de sentimentos e classificação textual. Essas abordagens oferecem grande potencial para automatizar e aprimorar a compreensão de grandes volumes de dados textuais (TREVILATO, 2025).

3. MATERIAL E MÉTODOS

Para alcançar os objetivos propostos nesta pesquisa, foram seguidos alguns passos metodológicos.

Começando pela criação do Dataset utilizado para o treinamento do modelo. Este dataset foi criado a partir de legendas de vídeos do youtube, que foram recuperadas através da biblioteca youtube-transcript-api. Tais legendas foram separadas em sentenças e classificadas manualmente entre positivas e negativas.

Foi implementado um modelo de *Deep Learning* híbrido utilizando a biblioteca Keras (TensorFlow). Este modelo combina dois métodos de entradas diferentes: uma sequencial processada por uma camada de embedding seguida por uma camada bidirecional LSTM, para melhor compreensão de contextos; e uma entrada TF-IDF utilizada para capturar a importância estatística de cada palavra. As saídas desses dois métodos são concatenadas e passadas por camadas de regularização, chegando em uma camada de saída Sigmoid para classificação binária entre sentimentos positivos e negativos.

O treinamento do modelo foi feito em 5 rodadas, usando diferentes random states para os splits de treino/teste, o que garante uma avaliação mais robusta. O modelo foi comparado com

outras duas abordagens de análise de sentimento: VADER (*Valence Aware Dictionary and sEntiment Reasoner*), um modelo baseado em análise léxica; e RoBERTa (*Robustly Optimized BERT Pretraining Approach*), um modelo transformer pré-treinado. A comparação foi realizada em 5 rodadas de teste sobre o dataset rotulado manualmente.

Por fim, o modelo treinado foi aplicado para analisar um novo dataset mais completo, feito a partir da API do YouTube e biblioteca youtube-transcript-api. Tal dataset apresenta informações como IDs de vídeos, metadados como título, número de visualizações, curtidas e contagem de comentários, além da legenda do vídeo e os principais comentários. Para cada vídeo, o sentimento foi calculado a partir da legenda (média das classificações de cada sentença) e para os comentários (média das classificações de todos os comentários). Com todas as informações necessárias prontas foram gerados os dados sobre a correlação entre as métricas de engajamento (visualizações, curtidas, contagem de comentários) e os sentimentos apresentados, representados por uma matriz de correlação.

4. RESULTADOS E DISCUSSÃO

Na **tabela 1** temos a análise comparativa entre o modelo proposto e os modelos já existentes VADER e RoBERTa.

Tabela 1:

Modelo	Acurácia (Média ± DP)	F1-Score (Média ± DP)	Precisão (Média ± DP)	Recall (Média ± DP)
Híbrido	0.8857 ± 0.0653	0.8566 ± 0.0828	0.9167 ± 0.0846	0.8040 ± 0.0812
VADER	0.6608 ± 0.0067	0.5851 ± 0.0121	0.6105 ± 0.0078	0.5619 ± 0.0181
RoBERTa	0.7323 ± 0.0064	0.5760 ± 0.0122	0.8849 ± 0.0103	0.4270 ± 0.0114

Fonte: do autor

Analizando os resultados podemos observar pelas métricas avaliadas que o modelo híbrido apresentou um desempenho superior. Esses resultados indicam que a combinação de embeddings contextuais com representações estatística (TF-IDF) foi eficaz para captar as nuances das frases, resultando em uma classificação mais precisa dos sentimentos. No entanto, é importante ressaltar que o modelo proposto foi treinado no mesmo dataset utilizado para sua avaliação, o que pode ter influenciado em seu desempenho superior em relação ao modelo RoBERTa.

Ao aplicar o modelo híbrido ao dataset completo de 938 vídeos, observou-se uma predominância geral de sentimentos negativos, tanto nas legendas quanto nos comentários. A correlação entre os dados analisados foi representada em uma matriz de correlação (**tabela 2**), através dela podemos ver que a correlação entre as métricas de engajamento e os sentimentos expressos nos vídeos é fraca.

A correlação mais forte observada foi entre o sentimento das legendas e o sentimento dos comentários (**0,2089**), indicando que vídeos mais positivos tendem a ter comentários também mais positivos. As correlações entre sentimentos e métricas de engajamento foram todas próximas de zero, sugerindo que o sentimento expresso não influencia diretamente no engajamento do vídeo.

Tabela 2:

	Visualizações	Curtidas	Quantidade de Comentários	Sentimento Legendas	Sentimento Comentários
Visualizações	1,0000	0,5589	0,3118	0,0441	0,0604
Curtidas	0,5589	1,0000	0,6569	-0,0171	0,0568
Quantidade de Comentários	0,3118	0,6569	1,0000	-0,0634	-0,1275
Sentimento Legendas	0,0441	-0,0171	-0,0634	1,0000	0,2089
Sentimento Comentários	0,0604	0,0568	-0,1275	0,2089	1,0000

Fonte: do autor

5. CONCLUSÃO

Este estudo demonstrou a eficácia de uma abordagem híbrida de Deep Learning para a classificação de sentimento em vídeos do YouTube, utilizando legendas e comentários como fonte de dados. Os resultados obtidos com o modelo proposto superaram os de abordagens léxicas (VADER) e pré-treinadas (RoBERTa), evidenciando a capacidade da combinação de embeddings contextuais (LSTM) com representações estatísticas (TF-IDF) em capturar nuances de linguagens e a polaridade do sentimento de forma mais precisa. A análise do dataset completo revelou uma predominância de sentimentos negativos, tanto nas legendas quanto nos comentários. Também foi identificada uma correlação fraca entre o sentimento expresso e as métricas de engajamento (visualizações curtidas e quantidade de comentários). Isso sugere que apenas o sentimento não seja o principal impulsionador do engajamento em vídeos do YouTube.

REFERÊNCIAS

ARRUDA, Adriano Souza et al. Representação de palavras para aplicações de aprendizado profundo no processamento de linguagem natural. 2019.

FÜHR, Bruno Edgar; FRANZEN, Evandro. desenvolvimento de uma ferramenta de coleta e armazenamento de dados disponíveis em redes sociais. 2014.

SILVA, Geovane do Nascimento et al. Análise de correlação entre comentários e curtidas/descurtidas de vídeos do youtube por meio de análise de sentimentos. 2021.

TREVILATO, Larissa Alves. Análise Comparativa de Modelos de Representação de Texto e Métodos de Aprendizado de Máquina na Classificação de Notícias Falsas em Português. 2025.