Sentiment and Emotion Analysis of YouTube Comments Using NLP

Name:
Anmol Thapliyal

Course / Department:
B.Tech CSE – Department of Computer Science

Institution / Organization:
UPES, Dehradun, Uttarakhand, Bharat

Date:
27 November 2025

Introduction

This report is on a project that examines sentiment and emotions in YouTube comments on a large scale. The analysis uses a multi-model NLP pipeline that includes traditional methods, like VADER and TextBlob, as well as Transformer-based models, such as RoBERTa for sentiment and two Emotion models.The project aims to classify comments as positive or negative, identify emotions, and create emotional summaries for each video. This approach helps us understand how viewers react, their engagement patterns, and the content's impact.

Objectives:

- To analyze YouTube comments using multiple sentiment and emotion models.
- To classify each comment with sentiment and emotional labels.
- To generate per video emotion score for deeper content evaluation.
- To create a real world NLP pipeline capable of processing large datasets.
- To compare classical NLP and transformer based models for performance.

NOTE:The Emotion recognition models do not understand the concept od Sarcasm .

Methodology / Procedure

Tools & Libraries

- Python
- Pandas, NumPy
- NLTK (VADER Sentiment Analyzer)
- TextBlob
- HuggingFace Transformers
- Torch

Dataset

- 100k+ YouTube comments
- Columns: Video ID, Video Title, Comment, Author, Time
- Cleaned and processed into multiple CSVs
    - Clean dataset
    - Labeled dataset
    - Per-video summary

Steps Followed

- Loaded raw comments using Pandas
- Cleaned text using regular expressions and filters
- Applied sentiment models:
    - VADER
    - TextBlob
    - RoBERTa-Large Sentiment
- Applied emotion models:

    - Cardiff Emotion Model
    - Hartmann Emotion Model
- Calculated probabilities and top emotion labels
- Generated a per-comment labeled CSV
- Computed per-video emotion score using averages
- Exported processed datasets for analysis and reporting

Results / Findings

- All comments were successfully processed and labeled using five different models.
- VADER and TextBlob worked well for simple polarity detection.
- RoBERTa-Large gave more accurate deep-learning-based sentiment classification.
- Cardiff and Hartmann emotion models captured detailed emotional patterns.
- Hartman mostly gave neutral for all comments
- Per-video score file shows dominant emotions, average emotion scores, and comment counts.
- Common detected emotions included joy, anger, sadness, optimism, and neutral.
- The multi-model pipeline provided more reliable insights than any single model.

Discussion

The results show that combining classical NLP tools with transformer-based models gives a more complete understanding of user sentiment and emotion.
Traditional models are fast and simple but may miss deeper contextual meaning.
Transformer models captured complex language behaviors, emotional cues, and comment tone more accurately.
The multi-model structure reduces misclassification and increases reliability.
However, limitations include slow processing on CPU and challenges with sarcasm or context heavy comments.

Conclusion

The project successfully implemented a complete sentiment and emotion analysis pipeline for YouTube comments. Key outcomes include:

- A clean and processed dataset of over 100k comments
- Sentiment classification using three models
- Emotion classification using two advanced models
- A detailed per-comment labeled dataset
- A per-video emotion summary dataset
- Demonstration of an end-to-end NLP workflow

The project achieves its objective of analyzing viewer reactions and emotional patterns using multiple NLP models.

Recommendations

- Use GPU for faster model execution
- Add more datasets for better generalization
- Try additional models like BERT and DistilBERT
- Build a dashboard using Streamlit or Flask for real-time analysis
- Integrate topic modeling to understand recurring themes

References

- HuggingFace Transformers Documentation
- CardiffNLP Roberta Emotion Model
- Hartmann DistilRoBERTa Emotion Model
- NLTK VADER Documentation
- TextBlob Documentation
- Python & Pandas Documentation

Project Repository:

https://github.com/DrakonHellsworth/sem-3/tree/test/project