"Unveiling Audience Insights"

VIPS

योगः कर्मसु कौशलम्

IN PURSUIT OF PERFECTION

Under the supervision
of
Ms. Vidushi

By
Manas Saluja , 09117711921
Shruti Thareja ,10417711921

School of Engineering & Technology
Vivekananda Institute of Professional Studies – Technical Campus
Pitampura, Delhi - 110034

Table of Contents

INTRODUCTION*:*

In today’s digital world, platforms like YouTube have become a major hub for user-generated content, ranging from entertainment, education, to product reviews. Every day, millions of videos are uploaded, and an even larger number of comments are generated by viewers. These comments represent the opinions, feelings, and reactions of audiences towards the content they consume, providing valuable insights into public sentiment. Analyzing these comments can help content creators, marketers, and businesses understand their audience better, improving both their content strategy and business decisions.

Despite the potential insights embedded within these comments, manually analyzing thousands of comments on popular videos is impractical and time-consuming. This project addresses the need for an automated system capable of collecting, processing, and analyzing these comments. By utilizing Natural Language Processing (NLP) techniques, this system can perform sentiment analysis on a large scale, providing users with actionable insights into how their content is being received.

Through web scraping techniques and advanced machine learning models, this project allows users to input YouTube video URLs, extract the comments, and analyze the sentiment expressed within them, whether it is positive, negative, or neutral. This automated solution will save time, reduce bias in sentiment interpretation, and offer a real-time view of viewer sentiment. Beyond simple sentiment analysis, the system can also predict trends in viewer sentiment over time, enabling businesses and content creators to adjust their strategies accordingly.

This project not only showcases the power of web scraping and NLP but also emphasizes the growing importance of data-driven insights in today's content-driven world. By applying advanced machine learning models and deploying this analysis via an easy-to-use web interface, the project bridges the gap between technology and business intelligence.

Problem Statement :

YouTubers and content creators often struggle to manually sift through thousands of comments on their videos to understand audience feedback. There is a lack of tools that can automatically analyze and categorize viewer sentiments, making it difficult for creators to make data-driven decisions about their content. This project addresses this issue by automating the extraction and sentiment analysis of YouTube comments, providing YouTubers with quick insights into their audience's opinions.

Currently, there is no efficient, scalable, and automated way for users to extract and analyze YouTube comments to identify patterns in sentiment (positive, negative, or neutral) or derive meaningful conclusions about audience reactions. Without such a system, content creators are left guessing or relying on rudimentary methods to understand how their audience feels about their content, which can lead to missed opportunities for improvement and engagement.

Feasibility Study:

1. Technical Feasibility

This project is technically feasible due to the availability of well-established tools. Web scraping can be done efficiently using the YouTube Data API, while sentiment analysis can be implemented using Python libraries such as TextBlob or VADER. Advanced models like BERT (via transformers) can also be used if needed. For web development, Flask provides a simple backend, and the project can be deployed to platforms like Heroku or AWS.

2. Operational Feasibility

The project is operationally viable, with a simple user interface where users input a video URL to obtain a sentiment analysis report. The application can be scaled as needed, and maintaining it will mostly involve keeping the scraping methods up to date with any API changes.

3. Economic Feasibility

The project is cost-effective, relying on free tools and open-source technologies. Cloud platforms like Heroku or AWS offer free tiers for small-scale applications, allowing the project to operate at minimal cost initially, with potential for scalability based on user demand.

4. Time Feasibility:

The project can be completed within the typical time frame of a minor project, considering the complexity of web scraping and sentiment analysis, with ample time for testing and refinement.

Hardware and Software Requirements:

Hardware Requirements:

Processor:	11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz 2.40 GHz
Installed RAM :	8.00 GB (7.65 GB usable)
System type:	64-bit operating system, x64-based processor
HDD Capacity:	512GB SSD
Video Graphics:	Intel Iris Xe Graphics
Internet:	A stable internet connection for web scraping and API usage

Software Requirements:

Operating System:	Windows 10/11, MacOS, or any Linux distribution
Programming Language:	Python 3.x
Backend Libraries:	Selenium (for web scraping YouTube comments)
	BeautifulSoup (for parsing the HTML structure of page)
	Flask/Django (for backend development)
	TextBlob/VADER (for sentiment analysis)
Frontend Framework:	React.js/Next.js (for building the user interface)
Data Visualization:	Chart.js/D3.js (for visualizing the sentiment results)
Dataset Source:	YouTube (for extracting real-time comment data)

Additional Features for the Project That could be added :

1.Trend Analysis:

- Include a feature to analyze sentiment trends over time for specific videos or channels. This would enable users to see how sentiment changes as a video gains more views or as content evolves.

2. Advanced Sentiment Metrics:

- Add deeper sentiment analysis with subcategories such as emotions (joy, anger, sadness, etc.), sarcasm detection, or context-based sentiment using models like BERT or GPT.

3. Comment Filtering & Sorting:

- Provide options to filter comments by sentiment (positive, negative, neutral) or by specific keywords. This will help users focus on the most relevant feedback.

4. Comment Toxicity Detection:

- Integrate a feature to detect toxic, spammy, or harmful comments, helping content creators manage the quality of their comment sections.

5. Recommendation System:

- Based on sentiment analysis, suggest improvements or content ideas that align with positive feedback or help address common complaints.

6. User Analytics Dashboard:

- Provide users with a comprehensive dashboard that visualizes key metrics such as the number of comments, overall sentiment distribution, top keywords, and emotion categories.

References:

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10026245/
2. https://www.irjmets.com/uploadedfiles/paper/volume3/issue_4_april_2021/9272/1628083377.pdf
3. https://developers.google.com/youtube/v3/docs/
4. https://textblob.readthedocs.io/en/dev/
5. https://www.iosrjournals.org/iosr-jce/papers/Vol23-issue3/Series-2/A2303020105.pdf
6. https://flask.palletsprojects.com/en/2.0.x/