JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

YOUTUBE ANALYTICS: DECIPHERING YOUTUBE COMMENTS

[1]Shreedhar G Yadawad, [2] Sudhanva R M, [3] Naveen S H, [4]Prateek D, [5]Manojkumar J

[1]Assistant Professor, [2]B.E Student, [3]B.E Student, [4] B.E Student, [5]B.E Student
[1]Department of CSE,
[1]Sri Dharmasthala Manjunatheshwara College of Engineering and Technology, Dharwad, India

Abstract: In today's digital landscape, YouTube has become a vital hub for diverse content, from educational tutorials to entertainment. However, understanding viewer sentiment and engagement is a challenge for creators and marketers. To tackle this, we introduce a YouTube sentiment analysis project to extract insightsfrom user comments. With YouTube's audience expanding rapidly, decoding emotional nuances and interaction patterns becomes complex. Our automated system, powered by AI and ML, assesses sentiment, categorizes topics, identifies key entities, and gauges anticipation for sequels. By analyzing comments, it empowers creators and marketers with data-driven decisions, tailoring content to audience preferences. Through sentiment analysis, NLP, and visualization, our project aims to enhance content experiences on YouTube, addressing challenges and fostering strategic content creation in the digital realm.

Index Terms - Sentiment Analysis, Name Entity Recognition, Natural Language Processing (NLP), Podcast.

I. INTRODUCTION

In today's digital era, YouTube has become a powerhouse for content creation and distribution, offering a vast array of videos spanning educational tutorials to entertainment. However, as YouTube's userbase grows exponentially, understanding viewer sentiment and engagement becomes increasingly challenging. Viewer comments provide valuable insights, but manual analysis is impractical due to theirvolume and complexity. To address this, our proposed YouTube sentiment analysis project aims to harness AI and ML technologies to interpret viewer comments comprehensively. Beyond sentiment analysis, the project categorizes video content, identifieskey entities, and gauges viewer anticipation for sequels.By empowering content creators and marketers with data-driven insights, the project enhances engagement and fosters loyalty. This initiative not only addresses immediate challenges but also propels content creation towards personalization and impact, enriching the YouTube ecosystem for creators and viewers.

1.1 What is Sentiment Analysis?

Sentiment analysis, also known as opinion mining, is theprocess of computationally identifying and categorizingopinions expressed in text data to determine the sentiment or emotional tone conveyed by the text. The primary goal of sentiment analysis is to understand the attitude, opinion, or emotion expressed by individuals orgroups toward a particular topic, product, service, or event.

1.2 What is Named Entity Recognition?

Entity recognition, also known as named entity recognition (NER), is a natural language processing(NLP) task that involves identifying and categorizing named entities in text data into predefined categories such as names of persons, organizations, locations, dates, quantities, and more. The goal of entity recognition is to extract relevant pieces of information from text and understand the semantic meaning associated with specific entities.

II. LITERATURE

This paper introduces the VADER (Valence AwareDictionary and sEntiment Reasoner) sentimentanalysis tool, which is specifically designed for analyzing sentiment in social media text. The authors demonstrate the effectiveness of VADER in capturingsentiment nuances and handling context-specific sentiment expressions [1].

This paper introduces spaCy 2, a powerful library for natural language processing tasks including entity recognition. It discusses the architecture and features of spaCy, including its support for various entity typesand its efficient processing capabilities [2].

This paper introduces SentiStrength, a sentiment analysis tool that combines lexicon-based and machinelearning approaches. While not directly related to VADER, it discusses sentiment analysis methods and provides insights into the broader field [3].

This paper compares various sentiment analysismethods, including VADER, using different evaluationmetrics and datasets. It provides insights into the performance of VADER compared to other approaches [4].

III. IMPLEMENTATION

Our implementation focuses on two distinct functionalities: sentiment analysis of YouTube comments using the VADER sentiment analysis libraryand entity recognition of guest names mentioned in thecomments utilizing the SpaCy natural language processing framework.

3.1 Sentiment Analysis of YouTube Comments

i. *YouTube API Integration*: We utilize the YouTube Data API to fetch comments from a specified YouTube video. The API is accessed securely using OAuth 2.0 authentication.

ii. Comment Scraping: Comments are scraped from the YouTube video using the get_comment_threads function. Each comment is extracted in plain text format.

iii. Sentiment Analysis: We employ the VADER sentiment intensity analyzer to perform sentiment analysis on each comment. The sentiment analyzer provides scores for the positivity, negativity, and neutrality of the text.

iv. Classification and Visualization: Comments are classified as positive, negative, or neutral based on their sentiment scores. The distribution of sentiment categories is visualized using a pie chart, providing insights into the overall sentiment of the comment section.

3.2 Entity Recognition of guest names.

In addition to sentiment analysis, our implementation incorporates entity recognition to identify guest namesmentioned in the YouTube comments. This functionality is achieved using the SpaCy natural language processing framework. The implementation entails:

i. *YouTube API Integration*: Similar to sentiment analysis, we utilize the YouTube Data API to fetch comments from the target video.

ii. *Comment Processing*: Each comment is processed using SpaCy's pre-trained transformer-based model (en_core_web_trf). The model extracts entities, particularly focusing on identifying persons mentionedin the comments.

iii. *Entity Extraction:* We extract person entities recognized by SpaCy, indicating guest names mentioned in the comments.

iv. *Visualization*: The top ten mentioned guest names are visualized in a horizontal bar chart. This visualization provides insights.

3.3 Deployment and Scalability:

Our implementation is designed to be easily deployable and scalable to accommodate varying network environments and workload demands. The modular architecture allows for seamless integration with existing systems and easy adaptation to different use cases. Additionally, the utilization of cloud-based services for API access ensures scalability and reliability in handling large volumes of data and user requests.

Overall, our implementation of sentiment analysis and entity recognition offers a comprehensive solution for analyzing YouTube comments, providing valuable insights into sentiment trends and identifying key entities mentioned in the comments.'

IV. APPROACH

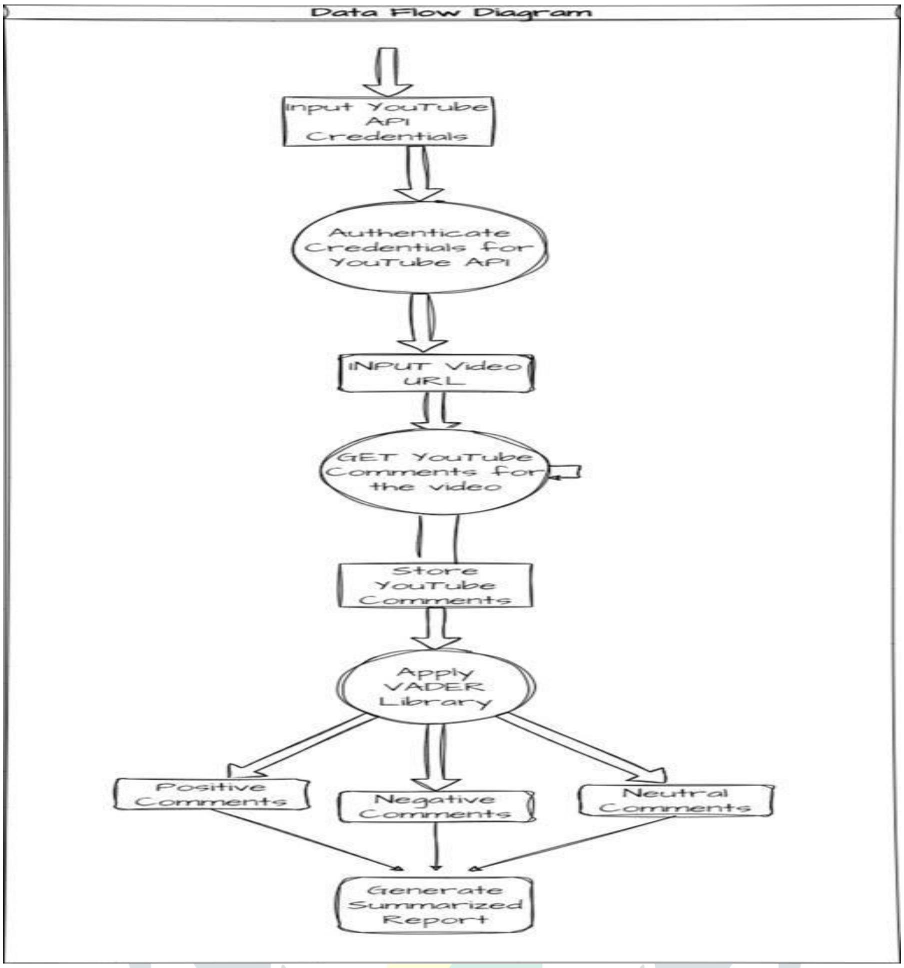

Figure 1

i. *Importing Necessary Libraries*: The initial step in thescript involves importing essential libraries to facilitate various functionalities. These libraries, including os, sys,httplib2, csv, argparser from oauth2client.tools, SentimentIntensityAnalyzer from vaderSentiment, andmatplotlib.pyplot, serve distinct purposes such as filehandling, HTTP requests, command-line argument parsing, sentiment analysis, and data visualization.

ii. *Loading Language Model and Sentiment Analyzer*: Following the library imports, the script proceeds to loadthe English transformer-based model from spaCy (en_core_web_trf) and the sentiment analyzer from VADER (SentimentIntensityAnalyzer). These models play a crucial role in processing natural language text anddetermining sentiment scores.

iii. *Setting Up Authentication with YouTube API*: Authentication with the YouTube API is established using OAuth 2.0 credentials stored in the client_secrets.json file. The script defines a function (get_authenticated_service) to facilitate this authentication process, enabling access to YouTube data.

iv. *Fetching Comments from YouTube Video*: A function(get_comment_threads) is defined to fetch commentsfrom a specified YouTube video using the YouTube API.Comments are retrieved in plain text format and stored ina list for further analysis.

v. *Validating YouTube Video ID*: To ensure the providedYouTube video URL is valid, the script defines a function (validate_video_id) to extract the video ID. This step is crucial for subsequent processing of the video comments.

vi. *Main Script Execution*: The main part of the script prompts the user to input a YouTube video URL, validates the URL, fetches comments from the video,and performs sentiment analysis on the retrieved comments.

vii. *Sentiment Analysis and Data Processing*: Using theVADER sentiment analyzer (commentbot), each comment is analyzed for sentiment. Comments arecategorized as positive, negative, or neutral based on sentiment scores. Additionally, comments containing links are identified for further processing.

viii. *Generating Reports*: After sentiment analysis, thescript generates a comprehensive report displaying the number and percentage of positive, negative, neutral comments, and comments containing links. Theoverall feedback is classified as positive, negative, or neutral based on sentiment percentages.

ix. *Writing Segregated Comments to Files*: Segregatedcomments, including positive, negative, neutral, and those containing links, are written to separate CSV files. This facilitates further analysis or storage of the comments.

x. *Visualizing Results*: Finally, the script visualizes thedistribution of sentiment categories and comments containing links using a pie chart generated with Matplotlib. This visualization provides a clear overview of the sentiment distribution within the YouTube comments dataset.

Named Entity Recognition (NER).

This script employs the YouTube API, spaCy, andMatplotlib to analyze comments on YouTube videos. Itidentifies mentions of specific guests, conducts sentiment analysis, and visualizes data. With OAuth 2.0 authentication, it securely accesses YouTube data, providing insights into viewer engagement and sentiment towards featured guests.

Figure 2

1. *Importing Necessary Libraries*: The script begins by importing essential libraries required for various functionalities. These include os, sys, httplib2,apiclient.discovery, apiclient.errors, oauth2client.client,oauth2client.file, oauth2client.tools, spacy, and matplotlib.pyplot. These libraries provide capabilities for file handling, HTTP requests, command-line argument parsing, accessing the YouTube API, natural language processing, and data visualization.

2. *Loading Language Model and Setting Up* Authentication: The English transformer-based model(en_core_web_trf) is loaded from the spaCy library.Additionally, OAuth 2.0 authentication with theYouTube API is established.

3. *Fetching YouTube Comments*: A function is definedto retrieve comments from a specified YouTube video using the YouTube API. Comments are fetched in plaintext format and stored in a list for further analysis.

4. *Main Script Execution and Guest Mentions Analysis*In the main part of the script, the user is prompted to input a YouTube video URL. The script then fetches comments from the video and analyzes them to identifyguest mentions using the loaded language model. The top ten mentioned guest names are extracted and visualized in a bar chart.

V. CONCULSION

In conclusion, the YouTube sentiment analysis project stands as a pivotal advancement in deciphering the invaluable insights embedded within viewer comments. Through the integration of state-of-the-art AI & ML technologies, this initiative offers content creators and marketers a powerful toolset for informed decision-making. By unraveling the complexities of viewer sentiment, engagement, and content preferences, the project facilitates a deeper understanding of audience dynamics and fosters the cultivation of vibrant, loyal communities. As we stride forward into the ever-evolving landscape of digital content creation, this endeavor not only addresses immediate challenges but also heralds a future of more personalized, impactful, and enriching content experiences for creators and viewers alike.

VI. RESULTS

The script successfully extracts and analyzes commentsfrom YouTube videos, identifying top mentioned guests and categorizing sentiment as positive, negative,or neutral. It generates visual reports illustrating sentiment distribution and top mentioned guests, offering valuable insights into viewer engagement andfeedback.

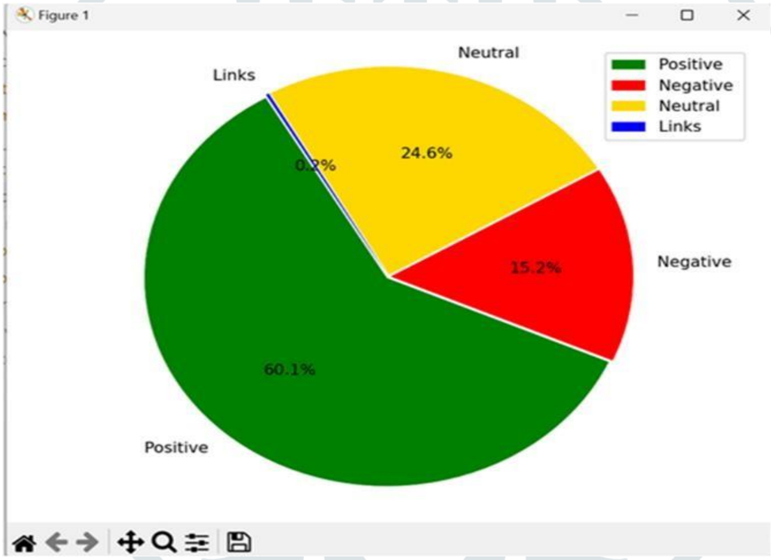

Figure 3

Figure 3 presents the distribution on positive, negative and neutral comments. Sentiment analysis is conducted on each YouTube comment using the VADER sentiment analyzer, categorizing them as positive, negative, or neutral based on their sentiment scores. This analysis provides valuable insights into viewer opinions and attitudes towards the video content, helping creators gauge audience reception and identify areas for improvement. The results facilitate data-driven decision-making and content optimization strategies.

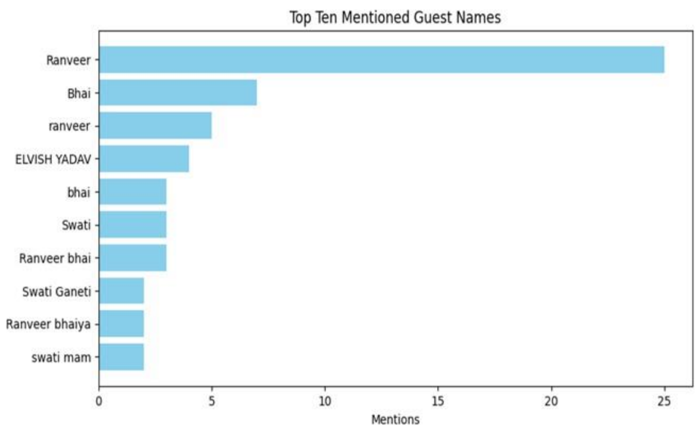

Figure-4

Figure 4 visualizes the entity recognized. Named EntityRecognition (NER) identifies and extracts entities suchas names, organizations, and locations from YouTube comments using the spaCy language model. This process enables the detection of top mentioned guest names, providing creators with valuable insights into viewer engagement and preferences. By visualizing thefrequency of mentioned entities, content creators can tailor their content to better resonate with their audience.

VII. REFERENCES

[1] Hutto, C.J. & Gilbert, Eric. (2015). VADER: A Parsimonious Rule-based Model for SentimentAnalysis of Social Media Text. Proceedings of the 8thInternational Conference on Weblogs and Social Media, ICWSM 2014.

[2]. https://sentometrics-research.com/publication/72/

[3].Paltoglou, Georgios. (2014). Sentiment Analysis in Social Media. 10.1007/978-3-7091-1340-0_

[4]. Devika, M.D. & C, Sunitha & Ganesh, Amal. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. Procedia Computer Science. 87. 44-49. 10.1016/j.procs.2016.05.124.