DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Comparative Study on Sentimental Analysis in Youtube Comments

Under the Guidance of

MUNI BABU SIR

SIDDIQ TAHASEEM
R190054.

ABSTRACT

This project focuses on developing a model to analyze YouTube comments, with the goal of determining the sentiment expressed by users. The process begins with collecting data using the YouTube API, which allows for real-time scraping of comments from a specified video. By extracting various attributes such as the author's name, comment timestamp, like count, and the comment text itself, we form a comprehensive dataset.

Once the data collection phase is complete, the dataset is cleaned and preprocessed. This involves removing stopwords, punctuation, and applying techniques like lemmatization to normalize the text.

These steps are crucial in preparing the data for analysis, as they ensure that the text is in a consistent format and free of noise that could interfere with accurate sentiment analysis.

With the cleaned data, we proceed to the sentiment analysis phase. Here, various natural language processing (NLP) techniques and machine learning models are employed to classify the comments into different sentiment categories, such as positive, negative, or neutral. Visualization tools like matplotlib and seaborn are used to illustrate the distribution of sentiments, providing a clear understanding of the general mood and opinions expressed by users in the comments.

Overall, this project not only demonstrates the technical process of scraping and analyzing YouTube comments but also highlights the importance of sentiment analysis in understanding audience feedback. By leveraging machine learning and NLP, we can gain valuable insights into user sentiments, which can be beneficial for content creators, marketers, and researchers.

INTRODUCTION

What is Sentiment Analysis?

Sentiment analysis, also known as opinion mining, is the process of using natural language processing (NLP) to determine the emotional tone behind a body of text. It identifies whether the expressed opinion in a piece of text is positive, negative, or neutral.

The main goal is to gain an understanding of the attitudes, opinions, and emotions expressed by users. This is widely used in social media monitoring, customer service, and market research.

Example:

Positive Sentiment: "I love this video! It's so informative and well-made.**"**
Negative Sentiment: "This video is terrible. I didn't learn anything useful.**"**
Neutral Sentiment: "This video is about Python programming**."**
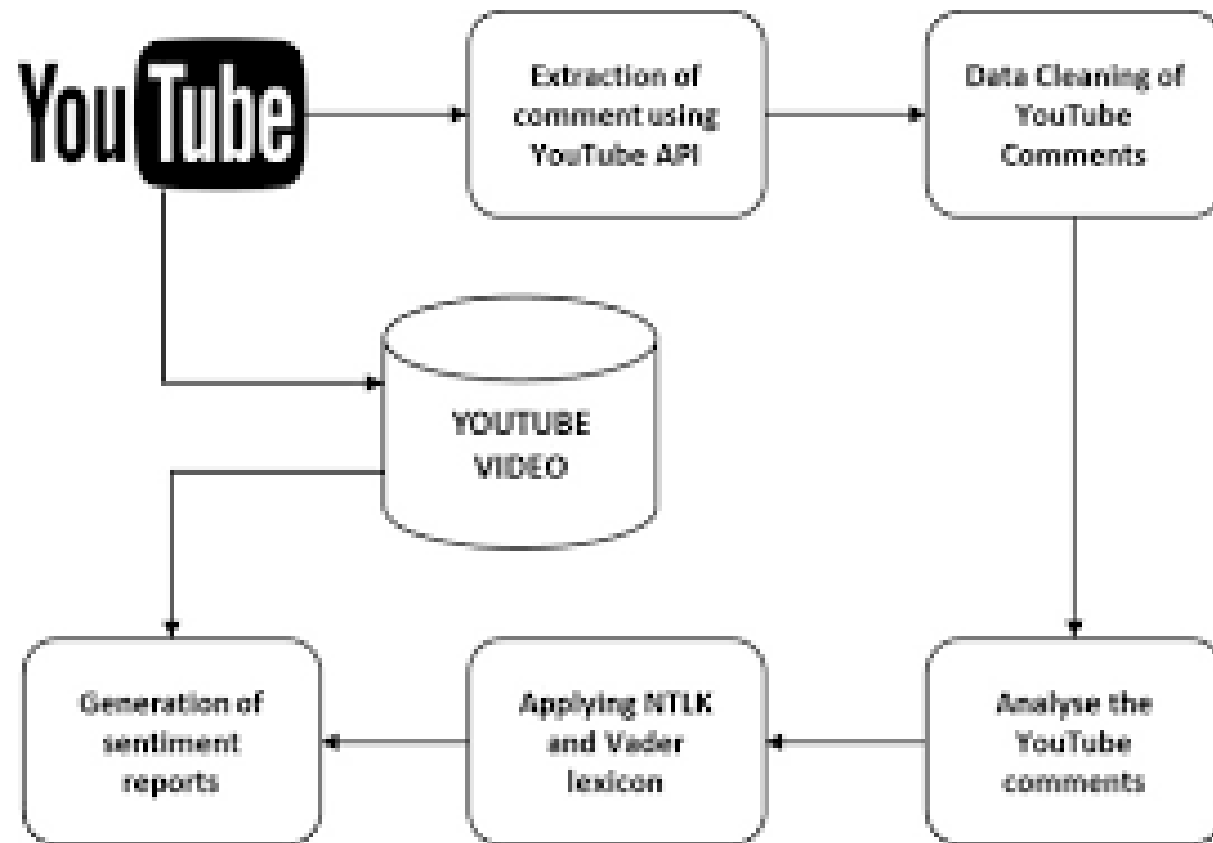
Figure : Working flow of Sentimental analysis on youtube comments

Why Analyze YouTube Comments?

- **Audience Feedback:** YouTube comments provide direct feedback from viewers about the content they consume. Analyzing these comments helps understand viewer satisfaction and engagement.

- **Content Improvement:** Creators can identify areas of improvement or topics that resonate well with their audience.

- **Brand Monitoring:** Companies can gauge public perception of their products or services featured in videos.

  Example:

  A tech reviewer can see if viewers appreciate their reviews or if there are recurring complaints about the depth of analysis.

How Sentiment Analysis Works?

- **Data Collection:** Gather comments from YouTube videos using the YouTube API.

- **Data Cleaning:** Remove irrelevant data, punctuation, and stopwords (common words like "and", "the") to focus on meaningful content.

- **Text Processing:** Apply techniques like tokenization (breaking down text into words) and lemmatization (reducing words to their base form).

- **Sentiment Classification:** Use NLP algorithms to categorize the sentiment of each comment.
  Example:

  Raw Comment: "I really liked the camera quality, but the battery life is poor."
  Cleaned Comment: "really like camera quality battery life poor"
  Sentiment Classification: Mixed (positive about camera quality, negative about battery life).

How Sentiment Analysis Works?

COLLECT DATA

CLEAN DATA

ANALYZE SENTIMENTS

GIVE A SCORE

USE THE RESULT

Real Applications of Sentimental Analysis :

The project on sentiment analysis of YouTube comments can have several real-world applications across various domains. Here are some practical applications:

1.Content Moderation and Quality Control:
   Application: Platforms like YouTube can use sentiment analysis to automatically flag and filter out offensive or inappropriate comments.
   Benefit: Enhances user experience by promoting constructive engagement and maintaining a safe online environment.

 2.Audience Engagement and Feedback Analysis:
      Application: Content creators and marketers can analyze sentiment to understand audience reactions and preferences towards their videos.
      Benefit: Helps creators tailor content to better resonate with their audience, leading to increased viewer engagement and retention.

3.Market Research and Brand Perception:

   Application: Brands can analyze sentiment to gauge public perception and sentiment towards their products or services based on YouTube comments.

   Benefit: Provides valuable insights for brand management, marketing strategies, and product development.

4.Trend Analysis and Virality Prediction:

   Application: Sentiment analysis can help identify emerging trends and predict the virality of videos based on viewer sentiments.

   Benefit: Enables content creators and marketers to capitalize on trending topics and optimize content strategies.

5.Customer Support and Sentiment-based Recommendations:

Application: Companies can use sentiment analysis on YouTube comments to identify customer issues, sentiment towards products, and provide targeted support.

Benefit: Improves customer service by addressing concerns proactively and enhancing customer satisfaction.

6.Content Curation and Personalization:

Application: Platforms can use sentiment analysis to personalize content recommendations based on user preferences and sentiments expressed in comments.

Benefit: Enhances user experience by delivering more relevant and engaging content recommendations.

7.Political and Social Sentiment Analysis:

Application: Governments, NGOs, and political analysts can analyze sentiments in YouTube comments to understand public opinion on social and political issues.
Benefit: Facilitates informed decision-making and policy formulation based on public sentiment.

Educational Insights and Feedback:

Application: Educational institutions and online learning platforms can analyze sentiments in comments to gauge student feedback and improve course offerings.

Benefit: Helps in optimizing curriculum design, improving teaching methodologies, and enhancing student engagement.

Fragrance-1 (Lavender)

REVIEWS

1. Smells amazing! A perfect purchase :)
2. Must buy! Super amazing.
3. Quite satisfactory

Fragrance-1 (Rose)

REVIEWS

1. A decent purchase
2. Quite okayish! Smells average
3. Could have been better in lot terms

Fragrance-1 (Lemon)

REVIEWS

1. An absolute waste of money.
2. Total waste of money
3. Terrible smell, not worth buytng

SENTIMENT ANALYZER

POSITIVE (81%)

NEUTRAL (88%)

negative (91%))

IMPORTANCE OF SENTIMENTAL ANALYSIS IN SOCIAL MEDIA?

" A sentiment analysis system helps businesses improve their product offerings by learning what works and what doesn't ."

" The content creaters are able to gain better insights into thier audience's preferences , reactions and emotional engagement. "

LITERATURE

Reference	Authors	Year	Objective	Contribution	Outcome
1	Alhujaili, R. F., & Yafooz, W. M.	2021	To analyze sentiments in YouTube video comments.	Explored various techniques and approaches specifically for analyzing YouTube comments.	Provided a detailed study on how sentiment analysis can be applied to YouTube comments using different methods.

Reference	Authors	Year	Objective	Contribution	Outcome
2	de Arriba, A., Oriol, M., & Franch, X.	2020	To apply transfer learning for sentiment analysis in social media.	Demonstrated the effectiveness of transfer learning in handling diverse user-generated content on social media platforms.	Showed how transfer learning can improve sentiment analysis performance, which can be applied to YouTube comments.

Reference	Authors	Year	Objective	Contribution	Outcome
3	Raza, M. R., Hussain, W., Tanyıldızı, E., & Varol, A.	2019	To apply deep learning techniques for sentiment analysis in cloud environments.	Discussed the use of deep learning models for sentiment analysis and their adaptability to cloud-based systems.	Highlighted the power of deep learning for effective sentiment analysis, applicable to analyzing YouTube comments.

Reference	Authors	Year	Objective	Contribution	Outcome
4	Alec Go, Richa Bhayani, Lei Huang	2018	To perform sentiment analysis on Twitter data.	Examined various sentiment analysis techniques for Twitter, a platform with user-generated content similar to YouTube.	Provided insights and methods that can be adapted for sentiment analysis of YouTube comments.

MOTIVATION

Understanding and Improving Viewer Experience

1 . Audience Insights: By analyzing comments, content creators can gain valuable insights into what their audience likes or dislikes. This helps in creating more engaging and relevant content.

Example: Positive comments on a tutorial video might highlight clarity and helpful tips, guiding the creator to maintain these elements in future videos.

2 . Content Enhancement: Feedback from comments can pinpoint areas needing improvement, such as video quality, pacing, or content depth.

Example: If viewers frequently mention that a video is too long, the creator can consider making shorter, more focused videos.

Business and Brand Benefits

1 . Market Research: Companies can use sentiment analysis to gauge public opinion on their products or services featured in videos. This real-time feedback is invaluable for product development and marketing strategies.

   Example: A beauty brand can analyze comments on makeup tutorials to see if viewers like their products and identify any common issues.

2 . Brand Reputation Management: Monitoring sentiment helps in identifying and addressing negative comments quickly, maintaining a positive brand image.

   Example: A company can respond to complaints about a new

 product feature, providing solutions and showing commitment to

 customer satisfaction.

CONTRIBUTION

Automated Sentiment Analysis System for YouTube Comments

Comprehensive Approach:

- Utilized a combination of Natural Language Processing (NLP) techniques, web scraping, the YouTube API, and machine learning models.

- Integrated three key tools for sentiment analysis: the VADER Lexicon Sentiment Analyzer and the Roberto model from Hugging Face Transformers.

Data Collection:

- Employed the YouTube Data API to scrape comments from targeted videos.

- Implemented efficient methods to retrieve user-generated content, ensuring a robust dataset for analysis.

METHODOLOGY

Data Collection

- Using the YouTube Data API to scrape comments.

- Example of API usage and data retrieval process.

PACKAGES USED FOR ANALYSIS :

In this module, we import all the necessary libraries and dependencies

needed for our project. Each library serves a specific purpose:
NumPy: This library is essential for numerical computations. It supports large multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.

Pandas: This library is crucial for data manipulation and analysis. It provides data structures like Series (1-dimensional) and DataFrame (2-dimensional) which are highly flexible and easy to use for handling and analyzing structured data.

Matplotlib and Seaborn: These libraries are used for creating visualizations. Matplotlib is a plotting library that provides a variety of plotting functions. Seaborn is built on top of Matplotlib and provides a high-level interface for drawing attractive statistical graphics.

NLTK (Natural Language Toolkit): This library provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and more.

  stopwords: A collection of common words that are usually filtered out before processing the text.

  WordNetLemmatizer: A tool for lemmatizing words (reducing them to their base or root form).

Google API Client: This library allows us to interact with Google's APIs, such as the YouTube Data API, which is used to fetch comments from   YouTube videos.

TextBlob: This library provides simple APIs for common natural language processing (NLP) tasks, including part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

TQDM: This library helps in adding progress bars to loops, making it easier to monitor the progress of long-running tasks

ipywidgets and Jupyter Extensions: These tools are used to create interactive widgets in Jupyter notebooks, enhancing the interactivity and usability of the notebooks.

Understanding VADER Model :

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool specifically tuned to sentiments expressed in social media. Here's a detailed look at its components and functionality:

Lexicon-Based Approach: VADER uses a dictionary of words that are mapped to sentiment scores. Each word in the lexicon has a pre-defined sentiment intensity score.

For example, words like "happy" have positive scores, while words like "sad" have negative scores.

Intensity Modification: VADER takes into account the intensity of the sentiment expressed by handling capitalizations, punctuation, and degree modifiers. For instance, "VERY good" has a stronger positive sentiment than just "good".

Context Handling: VADER is designed to understand the context and nuances in the text, such as negations ("not good"), booster words ("very good"), and contractions.

Emoji and Slang Interpretation: VADER can interpret common internet slang and emojis, which are frequently used in social media comments and can significantly influence the sentiment.
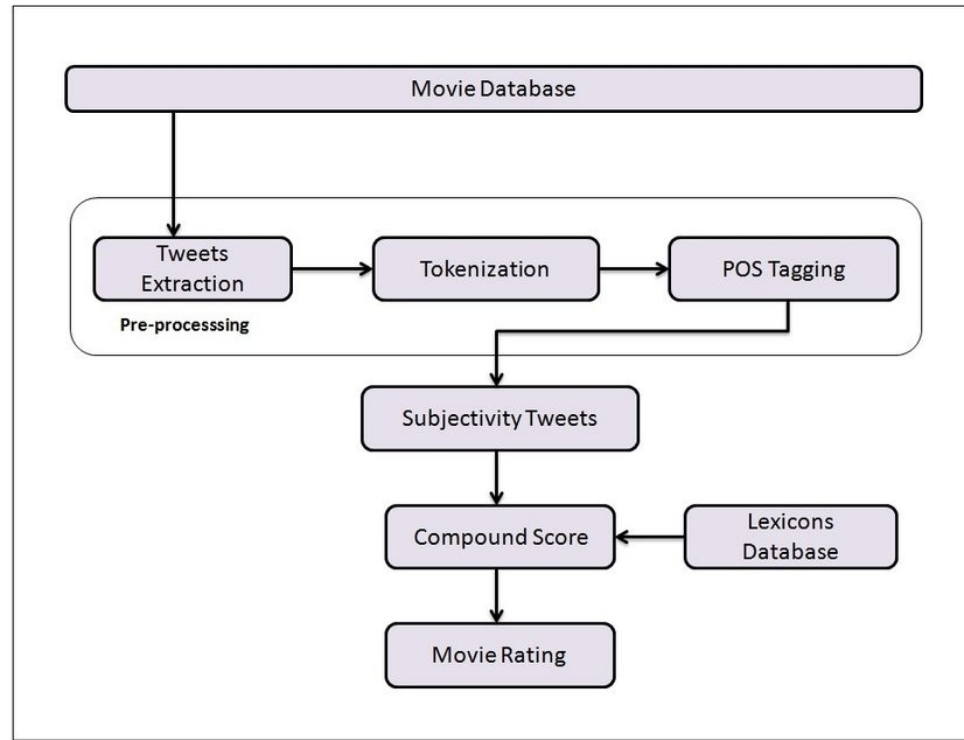
Figure : Framework of the VADER Sentimental Analysis System.

Loading and Preprocessing Steps :

This module focuses on loading YouTube comments and preparing the text data for analysis:

1.Loading Comments:

We use the YouTube Data API to fetch comments from a specific video. The API provides endpoints to access YouTube data, including video details, comments, and user information.

We specify the video ID and the maximum number of comments to retrieve. The API returns a JSON response containing the comments and their metadata.

We parse the JSON response and extract relevant fields such as the author's name, comment text, published date, updated date, and like count. This data is stored in a pandas DataFrame for easy manipulation.

2.Preprocessing Steps:

Convert Text to Lowercase: To ensure uniformity, we convert all characters in the comment text to lowercase. This step helps in treating words like "Happy" and "happy" as the same word.

Remove Symbols and Special Characters: Using regular expressions, we strip out any non-alphanumeric characters from the text. This includes punctuation marks, emojis, and special symbols, which can introduce noise in the data.

Remove Stopwords: Stopwords are common words like "the", "is", and "and", which do not contribute to the sentiment of the text. We use NLTK's list of English stopwords to filter them out.

Lemmatization: Lemmatization is the process of reducing words to their base or root form. For example, "running" is reduced to "run". This step helps in normalizing different forms of the same word, making the analysis more accurate.

Applying the Sentiment Analysis Model

In this module, we apply the VADER sentiment analysis model to the preprocessed comments to determine their sentiment:

Initialize VADER:
We create an instance of the `SentimentIntensityAnalyzer` class from NLTK's VADER module. This instance is used to analyze the sentiment of the text.

Analyze Sentiments:
For each comment in the DataFrame, we use VADER to compute sentiment scores. VADER provides four scores:
Positive: The proportion of the text that is positive.
Negative: The proportion of the text that is negative.
Neutral: The proportion of the text that is neutral.
Compound: A normalized score ranging from -1 (most extreme negative) to +1 (most extreme positive), which provides an overall sentiment score.

We store these scores in a new DataFrame, with each row representing a comment and its corresponding sentiment scores.

Store Results:

The computed sentiment scores are stored in a DataFrame, along with the original comment and its metadata. This structured format allows us to easily analyze and visualize the results.

Displaying the Results

The final module focuses on visualizing and interpreting the results of the sentiment analysis:

Sentiment Distribution:
We calculate the number of positive, negative, and neutral comments. This involves counting the number of comments that fall into each category based on their compound scores.

We also compute the percentage of comments in each category relative to the total number of comments. This provides a clear understanding of the overall sentiment distribution.

Bar Chart:
	Using Seaborn, we create a bar chart to display the number of comments in each sentiment category (positive, negative, and neutral). The bar chart provides a visual representation of the sentiment distribution, making it easy to compare the different categories.

Pie Chart:
	We create a pie chart to show the proportion of positive, negative, and neutral comments. The pie chart helps in understanding the relative share of each sentiment category and provides a visual summary of the results.

Detailed Report:
	We print a detailed report that includes the percentage of comments in each sentiment category. This report provides a concise summary of the sentiment analysis results and helps in drawing meaningful conclusions from the data.

RESULTS :

No. of reviews with respect to categories
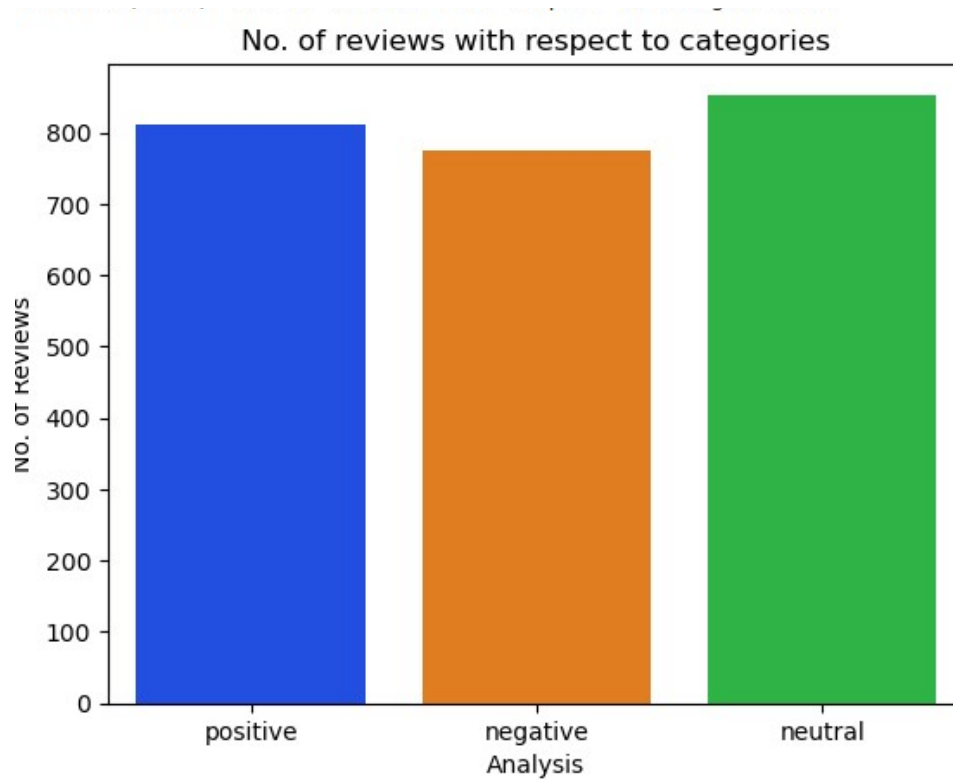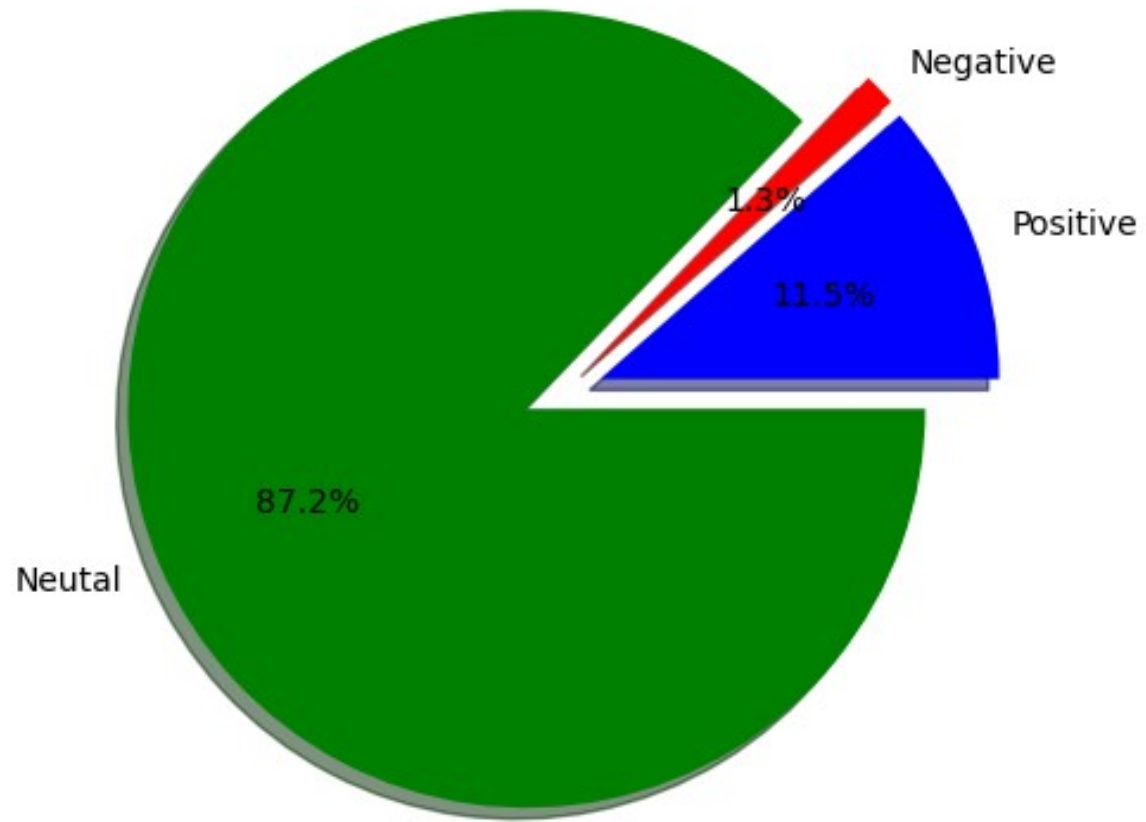
REFERENCES

1 . Alhujaili, R. F., & Yafooz, W. M. (2021). Sentiment analysis for YouTube videos with user comments. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) (pp. 814-820). IEEE. Retrieved from

Link : https://ieeexplore.ieee.org/iel7/6287639/9312710/09612169.pdf

2 . de Arriba, A., Oriol, M., & Franch, X. (2021). Applying transfer learning to sentiment analysis in social media. In 2021 IEEE 29th International Requirements Engineering Conference Workshops (REW) (pp. 342-348). IEEE. Retrieved from

Link :
https://upcommons.upc.edu/bitstream/handle/2117/361570/2021152892.pdf?sequence=
3

3 . Raza, M. R., Hussain, W., Tanyıldızı, E., & Varol, A. (2021). Sentiment analysis using deep learning in cloud. In 2021 9th International Symposium on Digital Forensics and Security (ISDFS) (pp. 1-5). IEEE. Retrieved from

Link :
https://asafvarol.com/makaleler/Sentiment_Analysis_using_Deep_Learning_in_Cloud.pdf

4 . Go, A., Bhayani, R., & Huang, L. (2009). Sentiment analysis on Twitter. In 2009 8th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE. Retrieved from

Link :
https://www.researchgate.net/profile/Alec-Go/publication/228485464-Sentiment-Analysis-on-Twitter/links/0deec532e7f308d9c5000000/Sentiment-Analysis-on-Twitter.pdf

THANK YOU