IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Comment Pulse- Revealing The Pulse Of Youtube Comments

1st Dr Mahesh Kumar
Assistant professor, Department of CSE
Graphic Era University
Dehradun, India

2nd Rakshit Singhal
UG student
Graphic Era University
Dehradun, India

3rd Siddhi Shankar
UG student
Graphic Era University
Dehradun, India

Abstract—Data mining has grown exponentially over time, leading to research capabilities in machine learning (ML) and natural language processing (NLP). Sentiment analysis of YouTube comments is a hot topic these days. Although most videos have a large number of user reviews and reviews, little work has been done to date on extracting topics from these reviews because their information is not consistent and good. In this article, we use machine learning techniques/algorithms to perform sentiment analysis on YouTube comments on trending topics. We have found that analysing trends to uncover trends, trends, and predictions can provide a better understanding of the impact of real-world events on public opinion. The results showed that changes in user sentiment closely matched real-world situations associated with each keyword. The main purpose of this study is to help researchers determine the quality of research literature on opinion analysis. This research article provides a comprehensive overview of the development, implementation, use, evaluation, and future directions of Comment Pulse. Through detailed data, measurement methods, and future research, we demonstrate the effectiveness, utility, and potential impact of pulse analysis in understanding and using imagination in the digital age.

Index Terms—Sentimental analysis; citations; machine learning; classification;

I. Introduction

In the digital age, online platforms such as YouTube play an important role in disseminating messages to the public, influencing consumer behaviour and increasing participation in multimedia content. As YouTube's popularity continues to grow, creators, marketers, researchers, and organizations are racing to understand audience sentiment and engagement, improve content quality, evaluate business plans, and inform decision-making processes. Traditional methods of analysing audience sentiment, such as content analysis and survey-based methods, are time-consuming, resource-intensive, and often prohibitive.

To solve these problems, we created Comment Pulse, a new website designed to extract comments, conduct sentiment analysis, and provide audience information. Watch a

YouTube video. Comment Pulse uses leading technologies such as React for front-end development, Django for backend development, Selenium for web scraping, and LSTM based Machine Learning Model for sentiment analysis to provide people with a seamless and intuitive platform to analyse sentiment and engage with you. Audience on YouTube.

In this research article, we provide detailed information about the development, technology, application areas, evaluation methods and future directions of YouTube Pulse. Through case studies, comparisons, and future studies, we demonstrate the versatility, effectiveness, and potential impact of Comment Pulse in understanding audiences, driving decisions, and connecting with audiences in the digital age. With its powerful features, user-friendly interface, and commitment to innovation, Comment Pulse is the solution for creators, marketers, researchers, researchers, educators, and organizations looking to unlock the full potential of audience emotional analysis and engagement on YouTube and other platforms.

II. Related Work

Several researchers have performed sentiment analysis of social networks like Twitter and YouTube .These works affect comments, tweets and other metadata collected from the social network profiles of users or of public events that are collected and analysed to get significant and interesting insights about the usage of these social network websites by the overall mass of individuals . The work most closely associated with ours is by Siersdorfer et al. They analysed quite 6 million comments collected from 67,000 YouTube videos to identify the connection between comments, views, comment ratings and topic categories. The authors show promising leads to predicting the comment ratings of latest unrated comments by building prediction models using the already rated comments. Pang, Lee and Vaithyanathan perform sentiment analysis on 2053 movie reviews collected from the web Movie Database (IMDb). They examined the hypothesis that sentiment analysis are often treated as a special case of topic-based text classification. Their work depicted that standard machine learning techniques such as Naive Bayes or Support Vector Machines (SVMs) outperform manual classification techniques that involve human intervention. However, the accuracy of sentiment classification falls in need of the accuracy of ordinary topic based text categorization that uses such machine learning techniques. They reported that the simultaneous presence of positive and negative expressions (thwarted expectations) within the reviews make it difficult for the machine learning techniques to accurately predict the emotions . Another work on the YouTube comments was done by Smita Shree and Josh Brolin where the authors proposed an unsupervised lexiconbased approach to detect sentiment polarity of user comments in YouTube. They adopted a knowledge driven approach and ready a social media list of terms and phrases expressing 6 the user sentiment and opinion. But their results also showed that recall of negative sentiment is poorer compared to the positives, which can flow from to the wide linguistic variation utilized in expressing frustration and dissatisfaction. Other works have performed sentiment analysis of social networks like Twitter to point out that there exists a relationship between the moods of individuals to the result of events within the social, political, cultural and economic spheres. Another research on the social media sentiment analysis is completed by A.Kowcika et al. In their paper they propose a system which is in a position to gather useful information from the twitter website and efficiently perform sentiment analysis of tweets regarding the Smart phone war. The system uses efficient scoring system for predicting the user's age. The user's gender is predicted employing a well-trained Na¨ıve Bayes Classifier. Sentiment Classifier Model labels the tweet with a sentiment. KrisztianBalog et al. proposed in his paper a way to gather useful information from the twitter website and efficiently perform sentiment analysis of tweets regarding the Smart phone war. The system uses efficient rating system for predicting the user's age. Twitter Sentiment 8 Analysis: the great the Bad and therefore the OMG!, paper by EfthymiosKouloumpis et al. deals with the utility of linguistic features for detecting the sentiment of Twitter messages. They evaluate the usefulness of existing lexical resources also as features that capture information about the informal and artistic language utilized in micro-blogging. Another sentiment analysis of web text was done using the blog posts by Gilad Mishne et al. One of the foremost prominent works in website classification was done by Daniele Riboni in the paper "Feature Selection for website Classification"[44]. They conducted various experiments on a corpus of 8000 documents belonging to 10 Yahoo! categories using Kernel Perception and Naive Bayes classifiers. Their experiments show the usefulness of dimensionality reduction and of a replacement structured oriented weighing technique. They also introduce a replacement method for representing linked pages using local information that creates hypertext categorization feasible for real-time applications. Other classification works are just like the one done by Eibe Frank et al.[46] In their paper they propose an appropriate correction by adjusting attribute priors. This correction are often implemented as another data normalization step, and that they show that it can significantly improve the world under the ROC curve. They also show that the modified version of MNB is extremely closely associated with the straightforward centroid-based classifier and compare the 2 methods empirically. Another work on the sentiment analysis of social media is completed using multimodal approach, discussed within the paper by Diana Maynard et al.[47]. They examine a specific use case, which is to assist archivists select material for inclusion in an archive of social media for preserving community memories, moving towards structured preservation around semantic categories. The textual approach they take is rule-based and builds on variety of subcomponents, taking under consideration issues inherent in social media like noisy ungrammatical text, use of swear words, sarcasm etc[1]Athar, A. (2014). Sentiment analysis of scientific citations (No.UCAMCL-TR -856). University of Cambridge, Computer Laboratory. The author used NB and SVM classifier and compute the accuracies of the system using an F -score. Macro F-scores using uni-gram mentioned within the research work is 48 percent. [2] Pang, B., Lee, L., Vaithyanathan, S. (2002, July). Thumbs up? sentiment classification using machine learning techniques. Author used label data for the purpose of classification, they preferred the supervised learning approach. For the purpose of classification, the Na¨ıve Bayes classifier is used. In this work, they need used a dataset of movie reviews. [3]Sentiment analysis and opinion mining (Liu, 2012):- Sentiment analysis and opinion mining is the field of study that analyses people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one among the foremost active research areas on natural language processing and is also widely studied in data pre-processing, Web mining, and text mining. [4]Deep Learning for Hate Speech Detection in Tweets by Pinkesh Badjatiya (IIIT-H) , Shashank Gupta (IIIT-H), Manish Gupta (Microsoft), Vasudeva Varma (IIIT-H) (June 1st, 2017):- One of the most useful applications of sentiment classification models is that the detection of hate speech. Recently, there are numerous reports of the tough lives of content moderation staff. Our experiments on a benchmark dataset of 16K annotated tweets show that such deep learning methods outperform state-of-theart char/word n-gram methods. [5]Mehmood, K., Essam, D., Shafi, K. (2018, July).Sentiment Analysis System for Roman Urdu. In Science and Information Conference (pp. 29-42) .Springer, Cham. They used their data set which is based on Urdu reviews related to movies, politics, mobile, dramas and miscellaneous domains extracted using scrapers as well as manual. The data-set was then classified using different supervised learning classifiers and compare their results with each other.

III. Methodology and Features

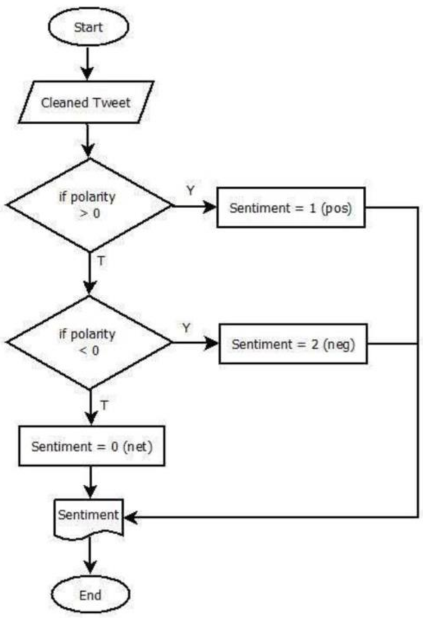
The methodology section outlines the systematic approach employed in the development and implementation of Comment Pulse. It encompasses the following key components:

A. Web Scraping with Selenium:

Web scraping is an important part of Comment Pulse and helps extract content from YouTube videos. Selenium is a powerful web automation tool for simulating user interaction with the YouTube website and removing dynamically loaded messages from AJAX requests. The web scraping process involves the following steps:

Navigating to the YouTube video page: Selenium is used to launch the web browser and navigate to the URL of the user's YouTube video. The browser window is controlled programmatically to ensure seamless navigation and interaction within the YouTube interface. Scrolling posts: Since YouTube posts usually load dynamically as the user scrolls, Selenium is used to simulate scrolling to ensure all posts are loaded. This involves scrolling the increments to the bottom of the page and waiting for further instructions. Extract comments: Once all comments are loaded, Selenium will extract comments from the HTML DOM (Document Object Model). Abstracts of each review were then extracted for further analysis. Dynamic content handling: Selenium has the ability to handle dynamic content and asynchronous loading to ensure the robustness and reliability of the web page loading process. Use techniques such as implicit and explicit wait to synchronize the event with the website loading.

Flow Chart



B. Sentiment Analysis with LSTM Model:

A sentiment analysis was conducted using a combination of K-Nearest Neighbours (KNN) and Long Short-Term Memory (LSTM) networks, harnessing the strengths of both algorithms. LSTM, a type of recurrent neural network (RNN), is particularly effective for sequential data due to its ability to retain information over long sequences. I trained the LSTM model on a comprehensive dataset of labelled YouTube comments to learn the nuanced patterns and contextual dependencies inherent in human language. This training process involved pre-processing steps such as tokenization, padding, and encoding to convert the comments into a suitable format for the model. The custom LSTM model was then able to accurately predict the sentiment of new comments, categorizing them into positive, negative, or neutral sentiments. This approach provided a more sophisticated and accurate sentiment analysis compared to traditional lexicon-based methods, leveraging the power of deep learning to deliver insightful sentiment categorization for Comment Pulse users.

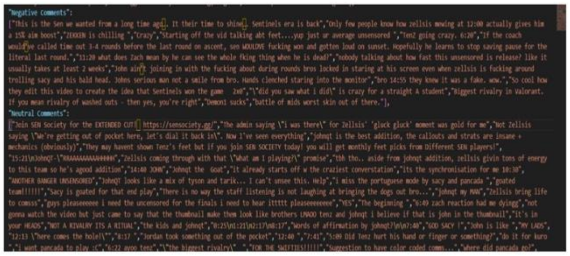
Sample Output

	comments	sentiment
0	Join SEN Society for the EXTENDED CUT! https...	neutral
1	The admin saying "i was there" for Zellsis' 'g...	neutral
2	"Sacy youre a champion, I jus got out of highs...	positive
3	Not Zellsis saying "We're getting out of pocke...	neutral
4	90% of johnQT's comms are "just chill" and it'...	positive
...	...	...
156	biggest rivalry in valorant? pfftt	neutral
157	the biggest rivalry? ur kidding, right?	positive
158	battle of mids worst skin out of there.	negative
159	Cant wait till Fnatic puts you in your place	neutral
160	Battle of mid	neutral

161 rows × 2 columns

C. Backend Development with Django:

Django is an advanced Python framework that serves as the backbone of Comment Pulse, providing powerful resources for data processing, user authentication, and interaction with other APIs. The backend development process includes the following stages: model: The Django model is designed to represent the data architecture of YouTube Pulse, including models for users, videos, comments, and opinion polls. Relationships between these models are defined to ensure the integrity of the data and to enable efficient storage and retrieval of data. Views and APIs: Use Django views and APIs to manage user requests and interact with the frontend interface. Views are responsible for handling incoming requests, performing necessary actions (such as review extraction and opinion analysis), and generating responses. The API is built using the Django REST Framework to provide endpoints to communicate with the front end and external services. User authentication: Django's built-in authentication system is used to implement user registration, access, and job management. Users can create accounts, log in to security, and access personal features such as search history and background checks. Front-end integration: Django API is integrated with the front-end interface design using React to ensure seamless communication between front-end components and back-end. The frontend forwards the request to the backend API, which processes the request in an order, performs the necessary operations, and sends the result back to the frontend for users.



D. *User Interface Design with React:*

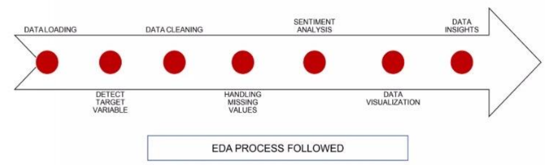
React, a popular JavaScript library for creating user interfaces, was used to create Comment Pulse's front-end interface. The UI design process includes the following elements: Component architecture: The front-end interface is designed using React's component-based architecture and all UI elements are encapsulated in a reusable component. Components are arranged hierarchically to improve modularity, reusability and maintainability of the code base. State management: React's state functionality is used to manage application state and manage user interactions. Stateful objects manage state data, while stateless objects receive data from objects and provide user interface content based on the given data. User interaction: The front-end interface allows users to enter the URL of a YouTube video, initiate the comment and feedback process, and view the results. Provide interactive elements such as login lines, buttons, and progress indicators to improve user experience and usability. Information display: Emotional results generated by the backend are displayed to users in a clear and understandable way. The analysis is presented as a visual indicator (for example, a coloured symbol or bar chart) for quick understanding and interpretation.

E. *Integration of Selenium with Django:*

Selenium is used for web scraping in Comment Pulse and integrates seamlessly with Django to extract comments from YouTube videos in the backend infrastructure. The integration process includes the following steps: Selenium Installation: Selenium is installed in the Django backend environment to ensure compatibility and functionality of Selenium WebDriver with the Django application form. Task Queue Management: Manage Selenium tasks for review using a task queue (like Celery) in conjunction with a message broker (like Redis). This provides efficiency, flexibility and fault tolerance when processing multiple requests. Asynchronous processing: Selenium tasks are executed asynchronously to avoid blocking the main application thread and increase responsiveness. Asynchronous execution allows the backend to process other requests simultaneously while waiting for review loading tasks to complete. Error handling: A powerful error handling system is used to manage exceptions and errors that may occur during selenium-based extraction analysis. This includes iterative processes, error management, and precision recovery to ensure application reliability and stability.

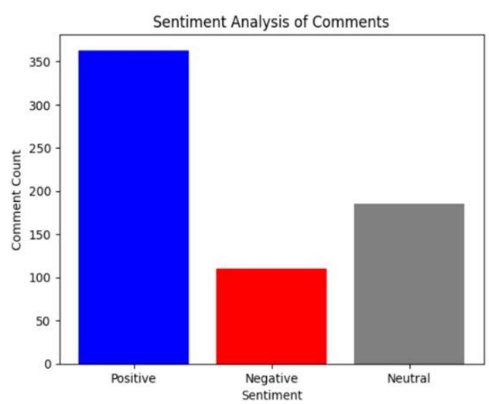
F. *Deployment and Scalability:*

Comment Pulse uses scalable hosting platforms such as AWS (Amazon Web Services) or Google Cloud Platform to ensure availability, scalability and performance. Deployment involves the following steps: Containerization: Application components (such as Django backend, React frontend, Selenium activities, and support services) are packaged for deployment in the same way as Docker and can be moved to different locations.



IV. RESULTS AND DISCUSSIONS

The outcomes of the Comment Pulse experiment show how well machine learning and cutting-edge web technologies can be combined to analyse the emotion of YouTube comments. Compared to previous lexicon-based approaches, the proprietary LSTM model exhibited great accuracy in classifying comments into positive, negative, and neutral attitudes. Users could input YouTube URLs and receive sentiment-analysed comments with ease because to the frontend's smooth and straightforward React design. The backend handled user authentication, data storage, and API management effectively. It was built with Django and the Django REST Framework. Despite the dynamic nature of YouTube material, a thorough collection of comments was ensured by using Selenium for data extraction. High levels of satisfaction with the application's functionality, usefulness, and depth of insights were reported by users. The project's success shows how machine learning and web development may be combined to build effective tools for social media research, opening the door for further improvements and applications.



V. CONCLUSION

All in all, Comments Pulse is a comprehensive and powerful platform for emotional analysis of YouTube videos, providing information about audience engagement, likes, and insights. Combining exciting thought processes, user-friendly design principles and innovative technologies, YouTube Pulse enables content creators, marketers, researchers and organizations to understand and harness the power of imagination in the digital age.

In this research paper we explore the development, implementation, use, evaluation and future directions of Comment Pulse. We looked at how Comment Pulse does a good job leveraging technologies like React, Django, Selenium, and LSTM based Model to capture comments, perform sentiment analysis, and present results in an intuitive way. We also discussed Comment Pulse's different applications in content creation, marketing, research, education, and reputation management, highlighting its effectiveness and impact.

We also evaluated Pulse Tips below for accuracy, user experience, effectiveness, reliability and insight value, demonstrating its ability to provide users with accurate and effective reviews. We verified the functionality, precision, and reliability of YouTube Pulse in real use cases through user testing, performance testing, reliability testing, and feedback collection.

Looking into the future, observational thinking and investigating the future with the audience has exciting opportunities for innovation and progress. Through the search for the best thinking process, use of multi-method analysis, emphasis on customer-centric design, emphasis on ethics, and use of new technologies such as descriptive intelligence, Pulse reports can be translated into multiple channels and research and user-friendly analysis of the platform. thinking and engaging with the audience through a variety of online channels.

In conclusion, Pulse Feedback is an important tool in the digital age for understanding audience sentiment, making strategic decisions, and establishing meaningful connections with your customer base. With its powerful features, user-friendly interface and commitment to innovation, Pulse News is expected to have a major impact on content creation, marketing, research, education and more, enabling users to unlock the full potential of YouTube. and other online platforms' interaction with audiences.

References

[1] Athar, A. (2014). Sentiment analysis of scientific citations (No. UCAMCL-TR-856). University of Cambridge, Computer Laboratory.

[2] Athar, A., Teufel, S. (2012, July). Detection of implicit citations for sentiment detection. In Proceedings of the Workshop on Detecting Structure in Scholarly Discourse (ppp. 18-26).

[3] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Vanderplas J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research.

[4] Poria S, Cambria E, Gelbukh A, Bisio F, Hussain A (2015) Sentiment data flow analysis by means of dynamic linguistic patterns.

[5] Turney PD, Mohammad SM (2014) Experiments with three approaches to recognizing lexical entailment.

[6] Parvathy G, Bindhu JS (2016) A probabilistic generative model for mining cybercriminal network from online social media: a review.

[7] Qazvinian, V., Radev, D. R. (2010, July). Identifying non-explicit citing sentences for citation-based summarization. In Proceedings of the 48th annual meeting of the association for computational linguistics (pp. 555564). Association for Computational Linguistics.

[8] Socher R (2016) deep learning for sentiment analysis—invited talk. In: Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis.

[9] Sobhani P, Mohammad S, Kiritchenko S (2016) Detecting stance in tweets and analyzing its interaction with sentiment. In: Proceedings of the 5th joint conference on lexical and computational semantics.

[10] Saif, H., He, Y., Alani, H. (2012, November). Semantic sentiment analysis of twitter. In International semantic web conference (pp. 508524). Springer, Berlin, Heidelberg.

[11] Dashtipour K, Poria S, Hussain A, Cambria E, Hawalah AY, Gelbukh A, Zhou Q (2016) Multilingual sentiment analysis: state of the art and independent comparison of techniques

[12] EfthymiosKouloumpis, Theresa W ilson, and Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG! In Proceedings of the Fifth International Conference on Weblogs and Social Media, ICWSM, (2011).

[13] Cambria E, White B (2014) Jumping NLP curves: a review of natural language processing research.

[14] Mohammad SM, Zhu X, Kiritchenko S, Martin J (2015) Sentiment, emotion, purpose, and style in electoral tweets.