ECHO INSIGHTS

YouTube Sentiment Analysis App

Sharan Adhitya C D[1], Senthilnaathan K[2]

Computer Science and Engineering Department,

Rajalakshmi Engineering College, Chennai

[1]*210701241@rajalakshmi.edu.in*

[2]*210701238@rajalakshmi.edu.in*

ABSTRACT

In an era where social media is omnipresent, ECHO INSIGHTS emerges as a streamlined solution for analyzing YouTube video sentiment. This project focuses on developing a robust social media analytical tool designed specifically for YouTube sentiment analysis. By leveraging YouTube API V3, ECHO INSIGHTS extracts comments from any given video link and organizes them into a data frame using Pandas. To ensure accessibility for a global audience, these comments are translated into English with the Google Translate module in Python. Subsequently, sentiment analysis is performed using pre-trained Vader models, known for their effectiveness in interpreting social media text. The analysis results are showcased in an interactive dashboard created with Streamlit, enabling users to visualize and interpret sentiment patterns effortlessly. The dashboard features a range of visualizations, including the best and worst comments, providing a clear view of the most positive and negative feedback. Additionally, an overall sentiment score is calculated and displayed, summarizing the general tone of the video's reception. One of the standout features of ECHO INSIGHTS is the word cloud, which highlights the most frequently used words in the comments, offering a quick overview of prevalent topics and trends within the audience's feedback. ECHO INSIGHTS aims to empower content creators, marketers, and analysts by providing a comprehensive understanding of audience sentiment and engagement on YouTube videos. By offering detailed insights into viewer opinions and reactions, users can make more informed decisions regarding content creation and strategy development. This tool not only helps in understanding the current sentiment but also in tracking changes over time, thereby enabling continuous improvement and optimization of content. Overall, ECHO INSIGHTS is designed to be an invaluable asset for anyone looking to delve deeper into the sentiment behind YouTube videos. By transforming raw comment data into meaningful insights, it facilitates a deeper connection with the audience and aids in crafting content that resonates more effectively.

1. INTRODUCTION

In today's fast-paced world, Social media sites like YouTube are effective means for content producers to interact with their viewers in the current digital era. To properly customize their content strategies, marketers and content

creators must have a thorough understanding of audience sentiment and engagement dynamics. The proposed project, ECHO INSIGHTS, provides a thorough social media analytical tool targeted at YouTube videos in order to meet this demand.

To give consumers practical insights into audience interactions, ECHO INSIGHTS makes use of cutting-edge technology including the YouTube API, sentiment analysis algorithms, and interactive data visualization tools. Through the process of extracting, translating, and analyzing comments from YouTube videos, ECHO INSIGHTS allows users to identify important topics and sentiment patterns that emerge from audience interactions. Streamlit was used to create an intuitive dashboard interface that allows users to view word clouds representing popular themes, analyze sentiment trends, and access comprehensive sentiment analysis results. To help with additional analysis and documentation, ECHO INSIGHTS also provides easy ways to print or save analysis findings as PDFs.

Content producers, marketers, and researchers can obtain more profound understanding of the attitudes and behaviors of their YouTube audience using ECHO INSIGHTS. This will enable them to make well-informed decisions and maximize their content strategies for maximum effect and success

2. LITERATURE SURVEY

The realm of social media analytics offers a dynamic environment where audience engagement is crucial for shaping content strategies. Understanding audience sentiment on platforms like YouTube is increasingly important. To address this need, recent research explores innovative techniques for sentiment analysis of YouTube comments, leveraging technology to provide valuable insights for content creators.

Sentiment Analysis for YouTube Comments:

• The work from R. Pradhan, "Extracting Sentiments from YouTube Comments," from 2021 gave us idea of fetching comments and assigning sentiments [3]

• For tackling sentiment analysis for YouTube comments, This paper addressed challenges like informal language and propose machine learning methods for sentiment classification, potentially improving the accuracy of your project's sentiment analysis.[1]

Data Analysis and Visualization:

• For EDA and data formation we referred [4] it uses Python libraries like Pandas. These techniques can be helpful for analyzing the sentiment data obtained from YouTube comments within your project.

• For the showcase of sentiment analysis, using word clouds, a potential visualization option for our project that attracts and visually appeals the user. A. I. Kabir, R. Karim, S. Newaz, and M. I. Hossain, "The power of social media analytics: Text analytics based on sentiment analysis and word clouds in R" [9]

Translation and Multilingual Analysis:

• For exploration of approaches to enhance bilingual machine translation using Python. While not directly related to sentiment analysis, this highlights techniques for handling potentially multilingual comments you might encounter in your project.[2]

• An AI-based language translation platform. If you encounter multilingual comments, this research suggests potential

solutions for incorporating translation into the project.[12]

• For a comparison of techniques for multilingual sentiment analysis. While you might not be focusing on multiple languages initially, this work offers a foundation for future expansion of your project.[13]

Additional Considerations:

• M. Khorasani, M. Abdou, and J. H. Fernández, Web Application Development with Streamlit. Apress. This book covers web application development with Streamlit, a potential tool for building the dashboard to visualize your sentiment analysis results.[5]

• For guidance on Python libraries like Pandas for data manipulation, crucial for preparing your sentiment data for visualization.[8]

These [10][15] papers explore sentiment analysis using machine learning techniques. If VADER sentiment analysis proves insufficient, we can explore these resources for alternative approaches.

3. RESOURCES

This project has been developed through widespread secondary research of accredited manuscripts, standard papers, business journals, white papers, analysts' information, and conference reviews. Significant resources are required to achieve an efficacious completion of this project.

The following prospectus details a list of resources that will play a primary role in the successful execution of our project:

• A properly functioning workstation (PC, laptop, net-books etc.) to carry out desired research and collect relevant content.

• Stable internet connection throughout the process is required for utilizing all the features without any interruption or delays, since this project utilizes certain modules like YouTube API, which requires a stable internet connection in order to retrieve all the comments from the given YouTube Video link provided as the input.

4. METHODOLOGY

With the help of your YouTube sentiment analysis app, you may gain a deeper understanding of your audience. Through a user-friendly Streamlit interface, you may examine comments related to videos you provide by utilizing the YouTube Data API. The sentiment analysis program VADER then analyzes each comment's underlying emotions and assigns a score to classify it as positive, negative, or neutral. Even comments written in foreign languages can be easily translated for thorough examination with Google Translate.

The software does more than just provide emotion scores. It offers a basic sentiment that encapsulates the atmosphere of the comments area. Additionally, it highlights the most encouraging and unfavorable remarks, which can help you figure out what appeals to your audience and where you might make improvements. An eye-catching word cloud illustrates the most frequently used words and insights.

With an interactive dashboard that lets you easily study trends, filter comments, and dig deeper into the data, Streamlit excels once more. This user-centered strategy guarantees a seamless and enlightening encounter. After deployment and testing, the app is prepared to provide insightful information about the mood of your audience with a few clicks.
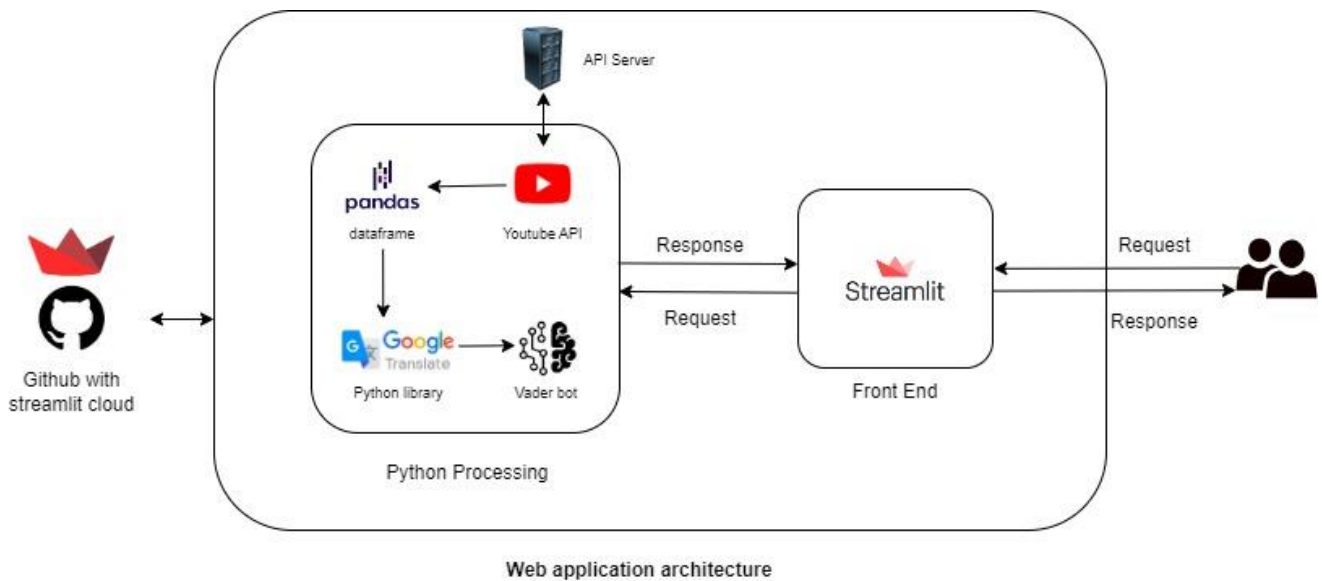
Figure 1 System Architecture

4.1 DASHBOARD

The main dashboard, uses streamlit, a python-based web framework that ensures easy integration with py models and other components much easier. Here, the Input, i.e., the YouTube video which is to be analyzed, is taken as URL from user and manipulates it with the processes and returns the result for user in intuitive dashboard.

4.2 FETCH COMMENTS

Then initial module is "fetch_comments". This module utilizes googleclient's YouTube API V3 to fetch the comments the given youtube video link. Here, we set the max counts to 100. This API will take the given URL of the YouTube video, get the video ID from it, and returns the Author, published time, Updated time, Like count and text of each comment. This data will be stored in a dataframe (a Data table).

4.3 TRANSLATE

The "Translate_cmt" translates YouTube comments. It uses Google Translate's Python library (googletrans) to turn comments into a chosen language. This module gets the dataframe from the fetch_comments module as an input, translates it using the googletrans library and finally outputs the translated comments as another dataframe for further analysis. Beware, translation speed can vary depending on your device's processing power.

4.4 SENTIMENT

Next comes "vader_sentiment". This module uses the vader bot, which is a pre trained model used to alter the sentiments for the given texts and updates the sentiment value of each comment as a new column to dataframe alongside the previous data of the comments. Vader bot provides a better result, in terms of finding the sentiment of the given text, when compared to other pre trained models that can be used for sentimental analysis.

4.5 INSIGHTS

The "Insights" module is used to provide the analysis of the comments after all the processing stages like translating and sentiment assigning. This module displays the top 10 comments based on the sentiment

as a table, worst10 comments based on the sentiment as a table. Also, this module shows the overall sentiment of the given YouTube video's comments, thereby giving an overview of the video's overall sentiment, by aggregating the sentiments of each comment. Then, this module shows the graph stating the distribution of the sentiment in the given YouTube video. This graph is typically like a bell-curve. If it is leaned towards right side, then the video is positive and if it is leaned towards left side, then the video is negative. Then, another graph is also displayed showing the comments over time, by plotting the published date against the comment count

using plotly library of python.

4.6 WORDCLOUD

"Word cloud", a simple python library that uses matplotlib to plot most frequently used words based on sizes for a better visual glance and return to key page. This module utilizes this Wordcloud python library, to visualize the most frequently used words in the given YouTube video's comments. The size of the words displayed is correlated to the frequency of the word. The bigger the word is, the frequent it is present in the comments of the given YouTube video
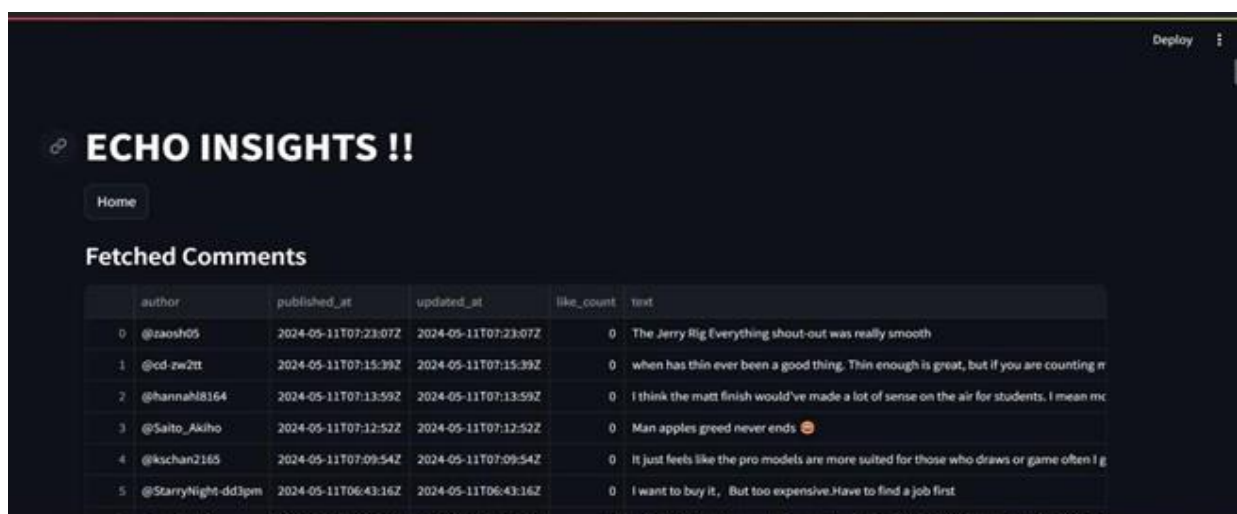
5. OUTPUT

Figure 2 Welcome page

Figure 3 Display Fetched Comments

Translated Comments

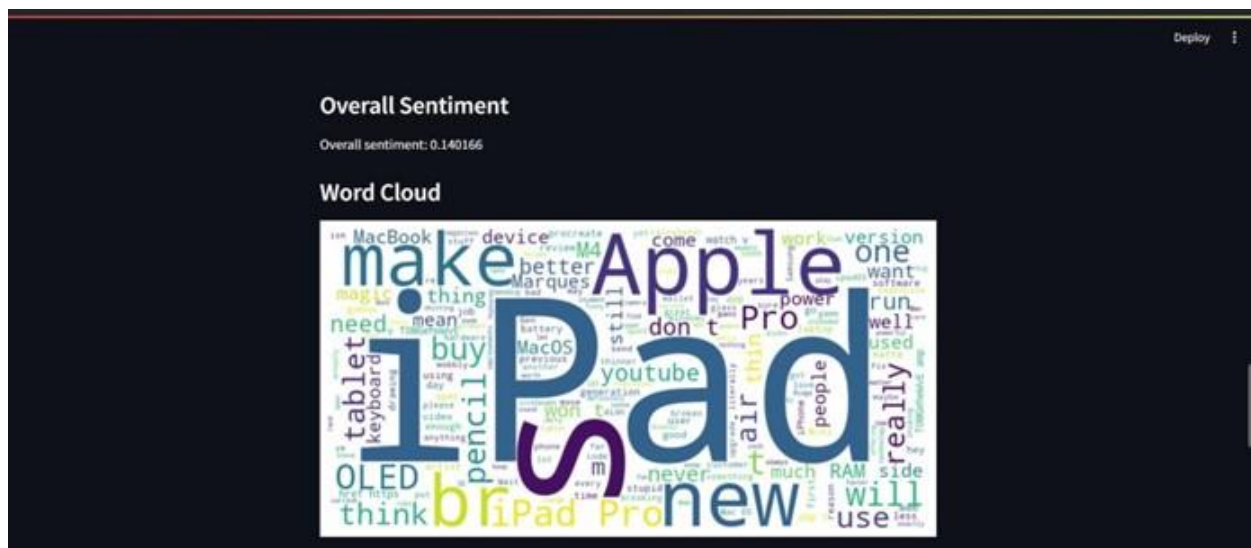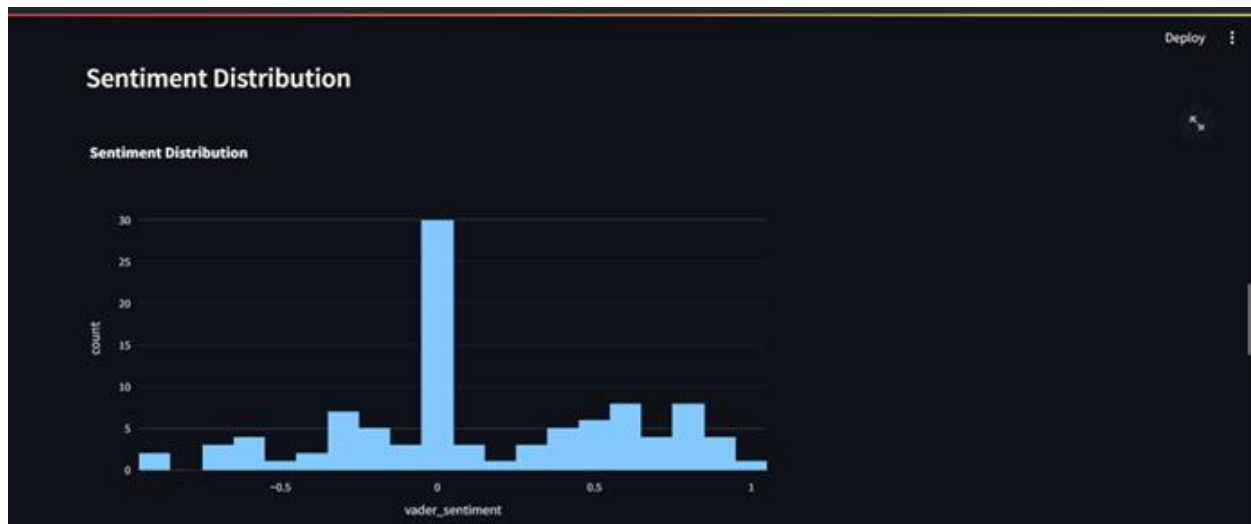|	text	translated_text
0	The Jerry Rig Everything shout-out was really smooth	The Jerry Rig Everything shout-out was really smooth
1	when has thin ever been a good thing. Thin enough is great, but if you are counting m	when has thin ever been a good thing. Thin enough is great, but if you are counting m
2	I think the matt finish would've made a lot of sense on the air for students. I mean mc	I think the matt finish would've made a lot of sense on the air for students. I mean mc
3	Man apples greed never ends 😩	Man apples greed never ends 😩
4	It just feels like the pro models are more suited for those who draws or game often I g	It just feels like the pro models are more suited for those who draws or game often I g
5	I want to buy it, But too expensive.Have to find a job first	I want to buy it, But too expensive.Have to find a job first
6	2:48(	2:48(
7	Anyway, did we get a calculator?	Anyway, did we get a calculator?
8	@MkHBD, it may not be relevant for you and it's ridiculous in terms of power. But, You	@MkHBD, it may not be relevant for you and it's ridiculous in terms of power. But, You
9	Until Mac OS comes to IPad they can't do anything to make this thing worth it to me	Until Mac OS comes to IPad they can't do anything to make this thing worth it to me

Figure 4 Display Translated Comments

Top 10 Comments based on Sentiment

	author	text	vader_sentiment
27	@fitdude	With your sarcasm and flippant style of video you have lost my brother as a subscribe	0.9844
35	@BrockBarr	Haha! This is MKB at peak sarcasm. Hilarious! 😂 😂 😂	0.9199
18	@nickranly6522	It's impressive. I love my Pixel Tablet. No complaints.	0.8676
23	@TheWildFatKid	Keep it up Marques. I love the videos, and the content. The MKBHD Ridge Wallet look	0.8622
66	@rcfamily4life74	I've replaced my Windows laptop with a 4th Gen iPad Pro and love it!!! Well I rec	0.8565
14	@mr-jq2ny	the spec bumped ipads are hella useful for artists and animators. they're not made fc	0.8425
45	@brandonlee7382	Wow, it's so thin!! I honestly didn't think that was possible	0.8247
36	@The3pleSix	I think the ultimate question for thin devices like that is "how sturdy it is? I mean isn't	0.8156
47	@DustinDAmour	I will be upgrading to the new iPad Pro 13 from a 2nd generation iPad Pro 12.9. The 1:	0.8074
55	@njnjco	Great hardware that you will never be able to take full advantage of. Yep, that's	0.8074

Figure 5 Sorted Comments - Top 10

Worst 10 Comments based on Sentiment

	author	text	vader_sentiment
93	@titania7768	Apple's new plan with the ipad is to instill its users will a constant fear of literally anyt	-0.9109
39	@truetech4158	Yay, lets make yet another stupid wobbly otherwise flat sitting device that has that ca	-0.886
77	@V1rusOne	Zack can make that iPad's bend test a live paid event and YouTube would need to ass	-0.7351
40	@bobd429	damn $999 for the 11 inch? it cost $1,260 here in Malaysia 😩	-0.7003
85	@MrRapmaster19	Apple desperately needs to simplify its iPad lineup. It's a product that is far less requi	-0.6652
83	@vedantdesai1	Matte OLED = No OLED. There is a reason why the OLED gaming community is shitting	-0.6124
15	@StanleyKC	MKBHD in his villain stage. Here for it!	-0.5983
67	@user-ti7hp8ld4y	Apple has run out of ideas since Steve Jobs died	-0.5574
48	@24kiryusuf	I'm guessing Apple's gonna add some crazy stuff to iPadOS next month, sin	-0.5574
80	@liban4695	Speaking of overpriced, why are ridge wallets so damn expensive.	-0.541

Figure 6 Sorted Comments - Worst 10

Figure 7 Sentiment Distribution Graph

Figure 8 Wordcloud of Comments

Figure 9 Comments over time Graph

6. CONCLUSION

To sum up, ECHO INSIGHTS is an essential resource for researchers, marketers, and content creators that want to understand the nuances of audience emotion and interaction on YouTube. Through the utilization of cutting-edge technology such as the YouTube API and sentiment analysis algorithms, customers are enabled to obtain practical insights regarding audience interactions. Comprehensive sentiment analysis results and interactive data visualization tools are easily accessible through the dashboard's user-friendly design, which is powered by Streamlit. Users may maximize effect and refine content strategy by using ECHO INSIGHTS to uncover sentiment trends, emerging topics, and patterns within audience interactions

7. FUTURE ENHANCEMENTS

To give consumers a complete picture of their digital footprint, extend ECHO INSIGHTS to evaluate sentiment and interaction patterns across a variety of social media sites, including Facebook, Instagram, and Twitter. Introducing configurable reporting options will enable customers to adapt analysis results to their own requirements. These options include the capacity to produce comprehensive reports in a variety of formats, including interactive dashboards, CSV files, and PDFs. Allow users to proactively modify their content strategies for optimal effect and success by integrating predictive analytics tools to foresee future audience behaviour and trends based on historical data.

REFERENCES

[1] R. Singh and A. Tiwari, "YOUTUBE COMMENTS SENTIMENT ANALYSIS," Int. J. Sci. Res. Eng. Manag.(IJSREM, no. May, 2021, [Online]. Available: https://www.researchgate.net/profile/Ritika-Singh-38/publication/351351202_YOUTUBE_COMMENTS_SENTIMENT_ANALYSIS/links/6092e8b2299bf1ad8d7d79a7/YOUTUBE-COMMENTS-SENTIMENT-ANALYSIS.pdf

[2] R. Manish, A. Mohammad, Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati, India, and Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati, India, "Language translation: Enhancing bi-lingual machine translation approach using python," I-manag. S J. Comput. Sci., vol. 7, no. 2, p. 36, 2019.

[3] R. Pradhan, "Extracting Sentiments from YouTube Comments," in 2021 Sixth International Conference on Image Information Processing (ICIIP), IEEE, Nov. 2021, pp. 1–4.

[4] K. Sahoo, A. K. Samal, and J. Pramanik, "Exploratory data analysis using Python," Technology, 2019, [Online]. Available: https://www.researchgate.net/profile/Dr-Subhendu-Pani/publication/337146539_IJITEE/links/5dc70b124585151435fb427e/IJITEE.pdf

[5] M. Khorasani, M. Abdou, and J. H. Fernández, Web Application Development with Streamlit. Apress.

[6] S. Elbagir and J. Yang, "Sentiment Analysis on Twitter with Python?s Natural Language Toolkit and VADER Sentiment Analyzer," in IAENG Transactions on Engineering Sciences, WORLD SCIENTIFIC, 2019, pp. 63–80.

[7] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. "O'Reilly Media, Inc.," 2012.

[8]   J. Bernard, "Python Data Analysis with pandas," in Python Recipes Handbook: A Problem-Solution Approach, J. Bernard, Ed., Berkeley, CA: Apress, 2016, pp. 37–48.

[9] A. I. Kabir, R. Karim, S. Newaz, and M. I. Hossain, "The power of social media analytics: Text analytics based on sentiment analysis and word clouds on R," Inform. Econ., vol. 22, no. 1/2018, pp. 25–38, Mar. 2018.

[10] N. Rathee, N. Joshi, and J. Kaur, "Sentiment Analysis Using Machine Learning Techniques on Python," in 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE, Jun. 2018, pp. 779–785.

[11] R. F. Alhujaili and W. M. S. Yafooz, "Sentiment Analysis for Youtube Videos with User Comments: Review," in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), IEEE, Mar. 2021, pp. 814–820.

[12] M. Kolhar and A. Alameen, "Artificial intelligence based language translation platform," Intell. Autom. Soft Comput., vol. 28, no. 1, pp. 1–9, 2021.

[13] K. Dashtipour et al., "Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques," Cognit. Comput., vol. 8, pp. 757–771, Jun. 2016.

[14] H. Saif, Y. He, and H. Alani, "Semantic Sentiment Analysis of Twitter," in The Semantic Web – ISWC 2012, Springer Berlin Heidelberg, 2012, pp. 508–524.

[15] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," the Journal of machine Learning research, vol. 12, pp. 2825–2830, 2011.