Whatsapp Chat Analyzer

Amarnath Nishad[1], Shiva Singh Patel[2], Sachin Kumar[3], Rohit Saini[4], Rakhi[5]

Student, CSE, ITS Engineering College Greater Noida, India[1,2,3,4]

Professor, CSE, ITS Engineering College Greater Noida, India[5]

Abstract: In recent times, the widely used and efficient communication method is the WhatsApp application, which facilitates conversations among groups of individuals covering a variety of topics. These WhatsApp chats contain a wealth of data that can be harnessed for cutting-edge technologies like machine learning. The key factor for successful machine learning models is providing the right learning experience, a factor directly influenced by the quality and quantity of the data provided to the model. This tool aims to offer comprehensive analysis of the data derived from WhatsApp conversations, regardless of the topic being discussed. The code developed for this purpose allows for a deeper understanding of the data. The tool leverages the simplicity and effectiveness of Python modules such as pandas, matplotlib, and seaborn, which facilitate the creation of data frames and the visualization of diverse graphs. The results are then displayed in a web interface using the Stream lit framework, enabling compatibility with all devices that support web browsing. The advantage of this tool lies in its efficient implementation, as it consumes fewer resources and can be easily applied to large datasets. By utilizing this tool, users can gain valuable insights and unlock the potential hidden within their WhatsApp chat data.

Keywords: WhatsApp Chat Data, Pandas, Seaborn, Matplotlib, Streamlit.

I. INTRODUCTION

This tool utilizes data analysis and processing techniques to improve machine learning algorithms. The initial step involves identifying the optimal learning experience to enhance the model's performance. Data pre-processing plays a pivotal role in machine learning, ensuring data quality and efficiency. To acquire ample data for model training, we focus on leveraging WhatsApp, a prominent data producer owned by Facebook. With nearly 55 billion messages sent daily and users spending an average of 195 minutes per week on the platform, WhatsApp presents a vast repository of data to extract valuable insights from. The advantage of this application lies in its implementation using user-friendly Python libraries such as seaborn, pandas, numpy, streamlit, and matplotlib. These libraries facilitate the creation of data frames and various types of graphs for visualization. The application is accessed through a web interface hosted on Heroku, ensuring compatibility across all devices supporting web browsers.

II. LITERATURE REVIEW

2.1 Existing System

In the past, there was a lack of dedicated analysis tools for WhatsApp chats. Analyzing these chats was a challenge as there were no readily available CSV files for analysis.

WhatsApp only provided export options in raw text format, making the analysis process complex and cumbersome. As a result, the existing systems for chat analysis were inadequate, and a shift to a more efficient solution was necessary. Hence, the emergence of the WhatsApp Chat Analyzer, which addresses these limitations and provides a streamlined approach to analyze and gain insights from WhatsApp conversations.

Disadvantages of Existing System

1. Raw data.
2. Time consuming.
3. Difficult to Analyze.
4. Analysis is not accurate.

Copyright to IJARSCT
www.ijarsct.co.in

DOI: 10.48175/IJARSCT-10179

55

ISSN
2581-9429
IJARSCT

2.2 Proposed System

The "WhatsApp Chat Analyzer" offers users a convenient online platform, accessible through a streamlit share link, for analyzingWhatsApp chats. With this application, users can effortlessly import their WhatsApp exported (.txt) files and obtain comprehensive analysis based on the content of the text file. By simply clicking the "Show Analysis" button, users can delve into the insights derived from their WhatsApp chats.

Advantages of WhatsApp Chat Analyzer

- Runs on all devices
- Shows based on Whatsapp chat file
- Shows different visualizations
- Total Messages
- Total words
- Media shared
- Link shared
- Monthly timeline
- Most busy day
- Most busy month
- Weekly activity
- Most busy users
- Most used words
- Emoji analysis

III.APPLYING TECHNICAL THINKING IN METHODOLOGIES

Python

Python is a versatile programming language that offers an extensive range of libraries catering to various functionalities required for different projects. It excels in tasks like predictive analysis and pattern recognition by leveraging data and conducting tests. Python's strength lies in its rich library ecosystem, encompassing numerous modules that delivermathematical and statistical functions essential for deriving insights from data.

Pandas

This open-source Python library is widely utilized in the fields of Data Science and machine learning. It serves as a valuable resource for data analysis, offering powerful tools for data manipulation. With its robust data structures, this library is particularly well- suited for analyzing and manipulating numerical data and conducting time series analysis.

Numpy

Derived from the term "Numeric Python," Numpy is a Python library renowned for its data analysis capabilities. It provides a wide array of numerical functions and methods that are invaluable for performing numerical analysis tasks. Moreover, Numpy is equipped with powerful multi-dimensional array objects that enable efficient processing of arrays.

IV.EXPLORING ANALYTICAL APPROACHES BASED ON MEASURESAND PERFORMANCE

Matplotlib

Matplotlib, a user-friendly and powerful visualization library for Python, proves to be an excellent choice for creating stunning visual representations. Leveraging the capabilities of NumPy arrays and integrating seamlessly with the broader SciPy stack, Matplotlib offers a diverse range of plot types, including pie charts, line charts, bar charts, scatter plots, histograms, and more.

Seaborn

Seaborn, a popular Python library, specializes in statistical plotting, enhancing the visual appeal of plots with its exquisite colorpalettes and predefined styles. In this project, Seaborn is utilized to create visually appealing heatmap visualizations. These heatmaps represent 24 hours across 7 days, utilizing varying colors to denote the intensity of messages from highest to lowest.
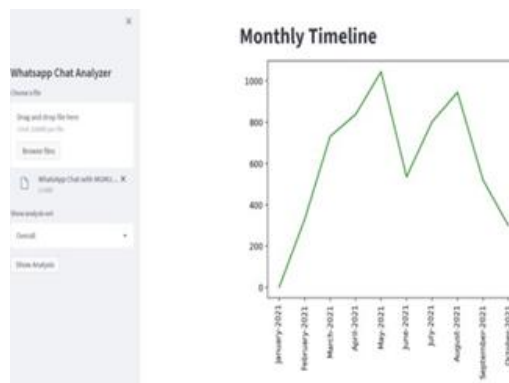
Streamlit

In this project, the mentioned library is employed to generate visually appealing web elements and objects. These components are utilized to represent the analysis of WhatsApp chats using a variety of charts and visualizations on the Streamlit platform.
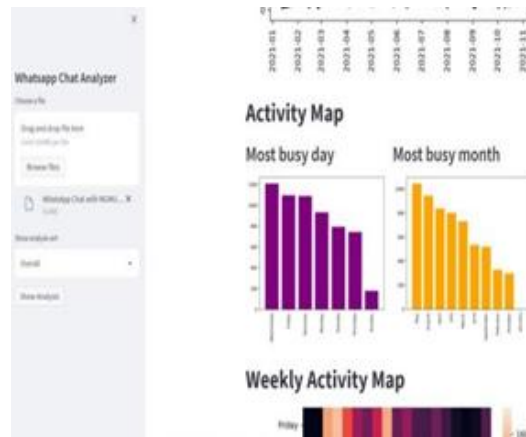
OUTPUT:
Top Statistics:

Monthly Timeline:

Daily Timeline:

Activity Map:

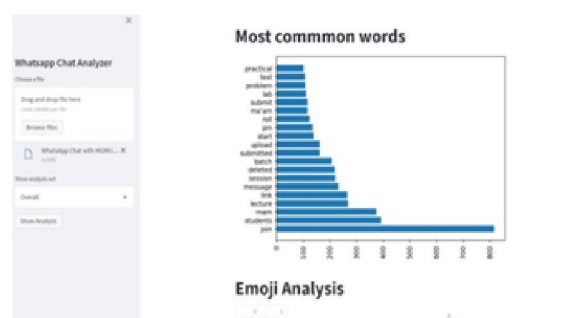**Weekly Activity:**

Busy User:
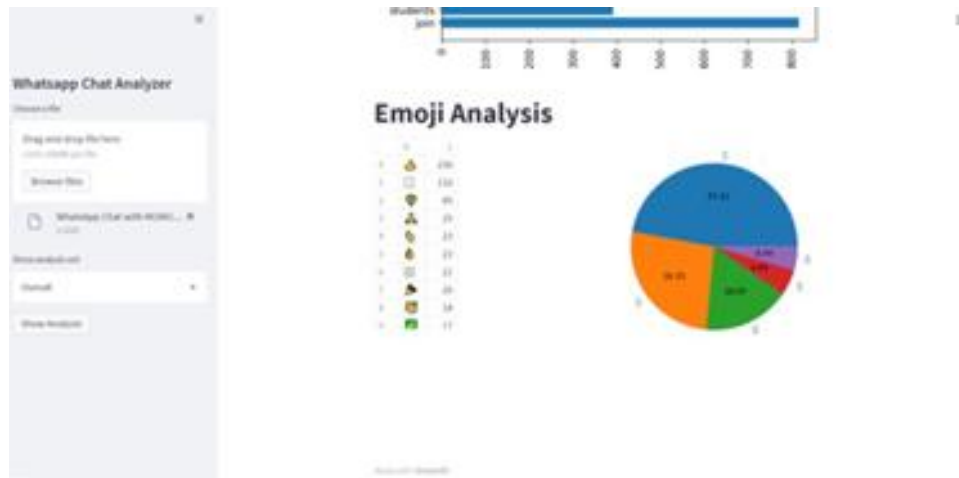

Common Words:

Emoji Analysis:

V.RESULTS AND DISCUSSION

This project was developed using Python and Streamlit and is deployed on the Streamlit Share web platform. Here's an overview of how the project works:

1. The user navigates to the sidebar and clicks on the "Browse File" option.
2. They select the WhatsApp chat text file they want to analyze and import it into the application.
3. The user can choose between conducting an overall analysis or analyzing specific users within the group.
4. After selecting the desired user, the user clicks on the "Show Analysis" button to initiate the analysis process.
5. The application then displays the analysis results based on the imported WhatsApp text file.
6. Users can view statistics such as the total number of messages, words, media files, and shared links within the group.
7. The application presents monthly and daily timelines, visualized through line charts, showcasing message activity over time.
8. An activity map is provided, presenting the busiest month and day using bar charts.
9. A weekly activity map illustrates the hourly activity of users throughout the week, with corresponding days represented using a heat map.
10. The top five most active users in the group are displayed using a graph and a list, indicating their respective usage percentages.
11. The application generates a WordCloud, offering a visually appealing representation of the most common words used in the chat.
12. A bar chart showcases the top twenty most frequently used words.
13. The application presents a list of emojis along with the number of times each emoji was used.
14. A pie chart illustrates the percentage of usage for the top five most commonly used emojis.
15. These features collectively form the results of the project, showcasing the analysis capabilities and functionality of the application.

VI.CONCLUSION

The primary objective identified during the initial phase of requirement analysis has been effectively accomplished. Following the implementation, the system consistently delivers dependable results. It boasts a fullyuser-friendly and intuitive interface, ensuring ease of use for individuals with limited computer proficiency. By addressing the limitations of the previous manual system, the developed system mitigates potential data entry errors through comprehensive validation mechanisms.

It has following features:

- User friendly.
- Time saving.
- Runs on any device.
- Analyze any WhatsApp imported chat file.
- Accuracy.
- Reliability.
- Easy to use.

VII.FUTURE SCOPE

1. Emotion recognition: Future research can focus on developing techniques to recognize emotions in WhatsApp conversations accurately. This can be useful for understanding sentiment analysis better.
2. Multilingual analysis: WhatsApp chat analyzer can be developed to analyze conversations in multiple languages. This can be useful for analyzing conversations in diverse groups and communities.
3. Automated responses: Future development can focus on developing chatbots that can provide automated responses to frequently asked questions or requests.
4. Integration with other platforms: WhatsApp chat analyzer can be integrated with other social media platforms to provide a more comprehensive analysis of social interactions.

REFERENCES

[1] In the publication titled "Analysis on Social Media Addiction using Data Mining Technique," authors D. Radha, R. Jayaparvathy, and D. Yamini delve into the examination of social media addiction through the utilization of data mining techniques. This research is published in the International Journal of Computer Applications (0975 - 8887).

[2] "Python for Everybody: Exploring Data in Python 3" is a comprehensive book written by Dr. Charles Russell Severance. This book serves as a resource for individuals seeking to explore and analyze data using the Python programming language.

[3] "Storytelling with Data: A Data Visualization Guide for Business Professionals" authored by Cole NussbaumerKnaflic offers valuable insights and guidance for professionals in the business domain who wish to effectively communicate data-driven narratives through compelling data visualizations.Res Lett. 2021 Dec 5;16(1):173. DOI: 10.1186/s11671-021-03628-6. PMID: 34866166; PMCID: PMC8645667.