International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 04 | April - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Whatsapp Chat Analyzer Using Machine Learning

Bollepalli Sri Sai Hemanth[1], Guntupalli Lakshmi Tejaswini[2],

Bobbala Sai Niharika[3], Ratnala Mehar Baba[4], Yalavarthy Swathi [5]

[1,2,3,4] Student, Department of Computer Science and Engineering, Tirumala Engineering College
[5] Associate Professor, Department of Computer Science and Engineering, Tirumala Engineering College

---

Abstract - WhatsApp has emerged as the go-to method for communication. Conversations on WhatsApp cover a wide range of topics among individuals or groups. This data can be valuable for advancing technologies like machine learning, which rely on quality data for effective learning experiences. Our tool is designed to offer comprehensive analysis of WhatsApp data, regardless of the subject of the conversation. By using our developed code, a deeper insight into the data can be achieved. One great benefit of this tool is that it utilizes common Python libraries like Pandas, Matplotlib, Seaborn, Streamlit, Numpy, Re, Emojis, and sentiment analysis to generate data frames and visualizations. These are then showcased in a streamlit web app that is efficient and requires fewer resources. This makes it ideal for analyzing large datasets.

***Key Words*:** Inspecting, Examining, Research, Data Analysis, Matplotlib, Pandas, Streamlit.

1.Introduction

WhatsApp Chat Analyzer is a tool that uses data analysis and processing to improve machine-learning algorithms. The key to implementing a successful algorithm is understanding the learning process and starting with the right experience. Data preprocessing is crucial for machine learning success. To enhance model efficiency, a large amount of data is necessary. Our focus is on WhatsApp, a major data producer owned by Facebook, which claims to send nearly 140 billion messages daily. In 2023, it was found that the average WhatsApp user spent around 195 minutes per week on the platform and was part of several groups. Realizing the wealth of information available, our goal was simple: to uncover insights from the large number of messages shared daily. By carefully examining chat backups, our tool sifts through the data to reveal interesting patterns. Through visual representation such as pie charts showing message distribution and diagrams highlighting key contributors, our findings offer a clear view of WhatsApp behavior. Going deeper, we analyze message timing, interpret emoji usage trends, and assess the emotions behind conversations. In addition, our tool goes beyond just plain text by creating eye-catching word clouds that visually display common terms. We also analyse group interactions, providing a detailed look at communication patterns within different chat groups. By offering a thorough analysis, users can gain valuable insights into their WhatsApp conversations, turning unprocessed data into useful knowledge.

2. Literature Survey

2.1. Significance of WhatsApp in Modern Messaging:

WhatsApp is a game-changer in messaging today, completely changing how we communicate globally. Its easy-to-use interface, ability to work on any device, and affordable messaging options have made WhatsApp more popular than traditional SMS. With over two billion users around the world, WhatsApp connects people of all backgrounds and locations, making communication easy and seamless. From text messaging to video calls, group chats, and sharing media, WhatsApp has everything you need to stay connected easily. Furthermore, the strong emphasis on privacy and security with end-to-end encryption has increased user trust in WhatsApp. This has made it a top choice for private discussions. In addition to its practical functions, WhatsApp has become a significant part of modern society, impacting social norms and behaviors related to communication. Whether it's personal chats or crucial updates during emergencies, WhatsApp is crucial in shaping how people communicate, exchange information, and nurture relationships in the digital era.

2.2. WhatsApp Chat Content Analysis:

Analyzing WhatsApp conversations includes various important metrics. First, evaluating how often messages are sent gives a glimpse into communication activity. Looking at the distribution of word counts shows common message lengths. The rate of sharing media indicates engagement with multimedia content, and tracking emojis and sticker usage shows emotional expressions. Sentiment analysis reveals the overall tone of conversations, while noting how often keywords appear identifies recurring topics. Analyzing response times measures communication efficiency, and evaluating group chat engagement assesses participation levels. By analyzing how people talk and using topic modeling, we can better understand communication patterns and what topics interest them. This helps us gain a deep insight into the content of WhatsApp conversations, revealing valuable information about user behavior and preferences online.

2.3. Extracting Insights from WhatsApp Messenger:

Gaining valuable information from WhatsApp Messenger requires a comprehensive strategy that covers different aspects of online communication. Utilizing forensic analysis methods, investigators can carefully inspect message content, metadata, and attachments to reveal important evidence. By analyzing timestamps, they can reconstruct conversation timelines, and by decrypting encrypted messages, they can access information that was previously locked. Sentiment analysis techniques also help understand the emotional context of conversations, giving insights into the tone and dynamics of interactions. Analyzing keyword frequency helps identify common topics and themes, which in turn helps understand user behavior and interests. Additionally, studying call logs and contact lists can uncover communication patterns and relationships between people. By using these techniques, forensic specialists can uncover valuable insights from WhatsApp Messenger data that can assist in investigations and legal matters.

3. Methodology

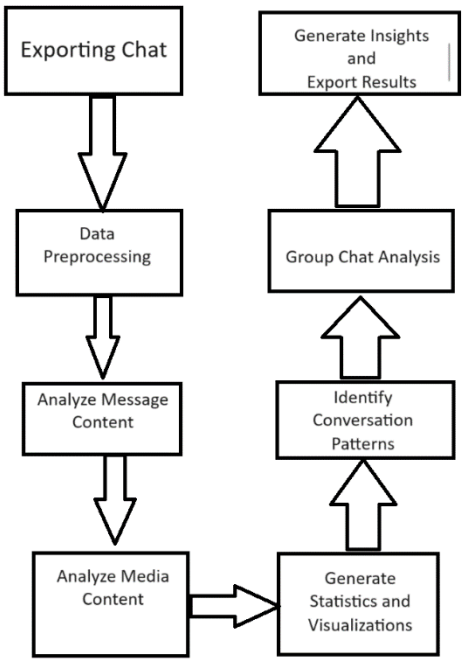
3.1. Proposed System:

The "WhatsApp Chat Analyzer" allows users to analyze WhatsApp conversations online through a Heroku connection. Users can import WhatsApp-exported (.txt) files into the analyzer and view an analysis based on the text files. Users can also click the Show Analysis button to see the results. The first step in this process is applying a sentiment analysis algorithm to the data. This method categorizes the conversations as positive, negative, or neutral and creates a pie chart based on these categories. Creating visualizations such as bar charts to show the number of messages and authors for each date, the message count per author, a chart showing the relationship between date and message count, and the media sent by authors. Also, display messages without authors and a graph representing the message frequency hourly. These modules, including Numpy, Scipy, Pandas, CSV, Sklearn, Matplotlib, Sys, Re, Emoji, and NLTK Seaborn, improve code readability and user comprehension.

The website has the following features:

- It displays chats from WhatsApp in a user-friendly format.
- It offers different visual representations of the data.
- It provides the total number of messages.
- It shows the word count of the messages.
- Users can easily share links and media.
- There is a monthly schedule available.
- Users can see which day is the busiest.
- They can also view which month is the busiest.
- The website tracks the most active users.
- It shows weekly activity trends.
- It highlights common words used in the chats.
- It also identifies the most frequently used emojis.

The working of the system is given below:



3.2. Functioning

To export a chat from WhatsApp, follow these steps:

- Open the WhatsApp chat for a group and click on the menu.
- Then, click on "more" and select "export chat."
- Choose to export the chat without media.

Here's how WhatsApp chat analysis works:

- Start by opening the WhatsApp chat analyzer web page.
- Choose the date format for the analysis.
- Upload the exported chat file.
- A trained model will analyze the data.
- The data will be preprocessed by the trained model.
- Decide whether you want an overall analysis or analysis for a single person.
- The trained model will provide analysis, including top statistics, a word cloud, an activity map, a monthly timeline, a daily timeline, and emoji analysis.

3.3. System Modules

This project consists of five main modules, each with their own sub modules:

3.3.1 Data extraction:

To extract data from WhatsApp, simply click on the export button. This will create a text file containing the raw data. The raw data is then handed over to the WhatsApp chat analyzer for preprocessing.

3.3.2. Data Collection:

Utilize the “Export Chat” feature to send the entire conversation in text format to your

email ID. Retrieve the exported chat from your email inbox.

3.3.3. Data Preprocessing:

During the data preprocessing stage, unnecessary content is eliminated from the raw data. Only the essential data needed for analysis is retained.

3.3.4. Importing Data to the Website:

Users are given the option to select either a general group analysis or a more focused analysis on a specific user. Once the user makes their selection, they can view the analysis results after importing the file. The analysis includes imported WhatsApp text files.

3.3.5. Statistical Representation:

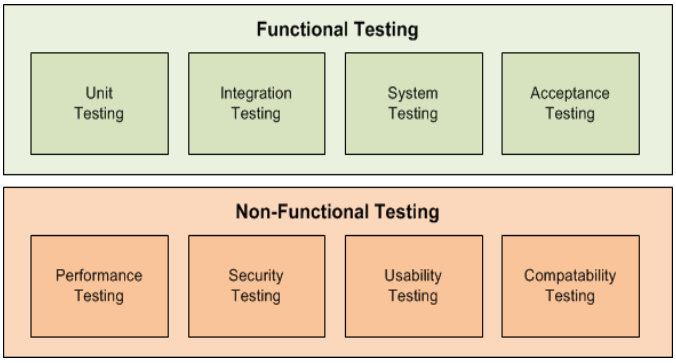
Different types of graphical representations are utilized to present the preprocessed data visually on the website.

4. Testing

Software testing is similar to conducting an investigation to assess the quality of the product being tested. It offers an unbiased perspective on the software, helping developers grasp potential risks in its implementation. Through various testing techniques, such as executing programs or applications, software bugs and defects can be uncovered.

To put it simply, Software Testing is the process of confirming the functionality of the Application under Test (AUT).

1. Functional Testing
2. Non-Functional Testing



1. Functional Testing

Testing the functionality of software is crucial in ensuring that each part of an application works correctly according to the requirements set out. This includes evaluating aspects like the interface, APIs, database operations, security features, client/server setups, and overall performance. Tests can be carried out manually or automatically using specific tools.

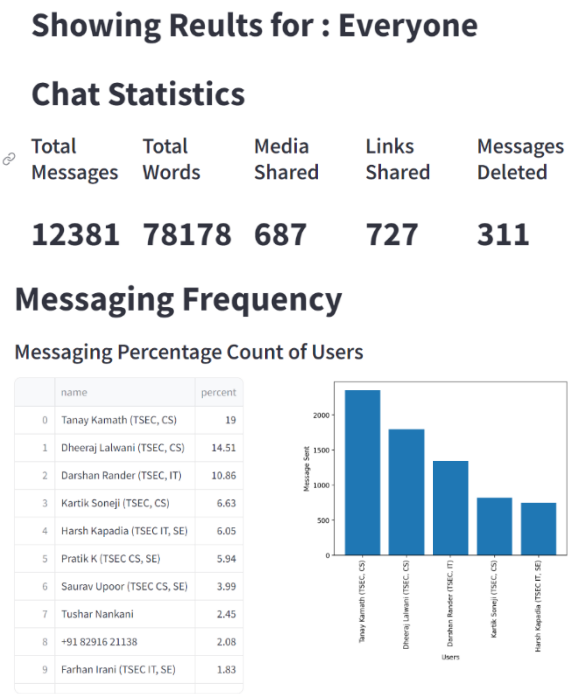
S. No	Test Objective	Pre condition	Steps	Test Data	Expected result
1	To check the extension of the file	chat file exported from whatsapp	Upload file from a panel in the left side	Valid file with extension txt	Show analysis button will appear in the panel
2	To check time format	Text file uploaded	Select time format Click on Show Analysis button	Valid time format of the file	Analysis result will be displayed
3	To select specific user	Overall analysis will displayed	From the left side in the panel select specific user from dropdown And click show analysis	Name of the user from the text file	Analysis result with respect to the selected user
4	To remove text file	File already uploaded and show analysis button will be visible	Click on the cross mark beside the file	Valid file	Browse file button will re appear

2. Non-Functional Testing

Non-functional testing looks at things like how well a system performs, how easy it is to use, and how reliable it is - things that functional testing doesn't consider. Its goal is to evaluate a system's preparedness using criteria beyond just how it works. For instance, a non-functional test could examine the ability of a software to handle multiple users logging in simultaneously. This kind of testing is essential for guaranteeing client happiness and should enhance user friendliness, effectiveness, and ease of maintenance.

S. No	Test case	Domain
1	Application load time is about 10 to 15 sec.	Performance Testing
2	Application should run on any browser	Compatibility Testing
3	Minimum storage on browser 30MB	Scalability Testing

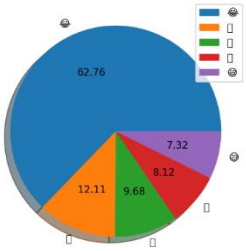
5. Output Screens



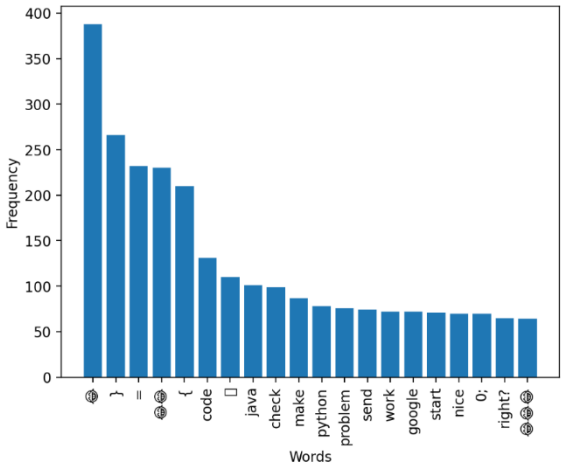
The data displayed includes the number of messages, words, and shared images links. We transformed the entire chat file into a data frame, isolated the words and messages, and utilized URLextract to identify links.

Emoji Analysis

	0	1
0	😂	1,886
1	👍	364
2	□	291
3	🔥	244
4	🥰	220
5	💯	180
6	♂	165
7	🤦	136
8	🤣	128
9	🍋	101

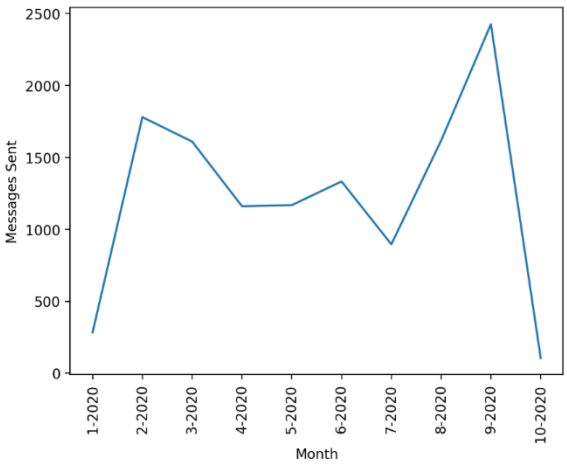


Most Frequent Words Used In Chat

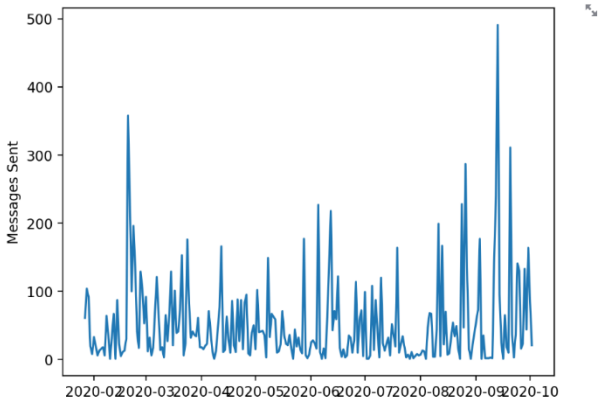


We utilized the Emoji library to identify and display the most frequently used emojis in messages. These emojis were then represented in a pie chart created with matplotlib.

Monthly Timeline

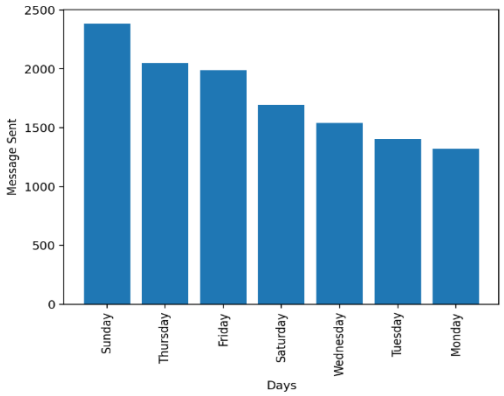


Daily Timeline

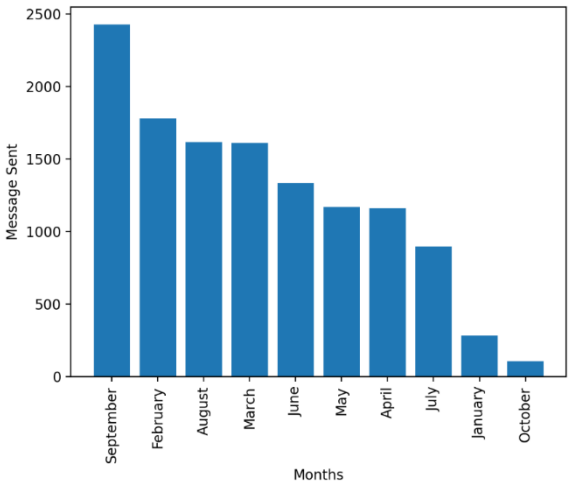


The graph shows how often messages are sent each day and each month. We used matplotlib to create the graph based on the number of messages sent each day and each month.

Most Busy Days

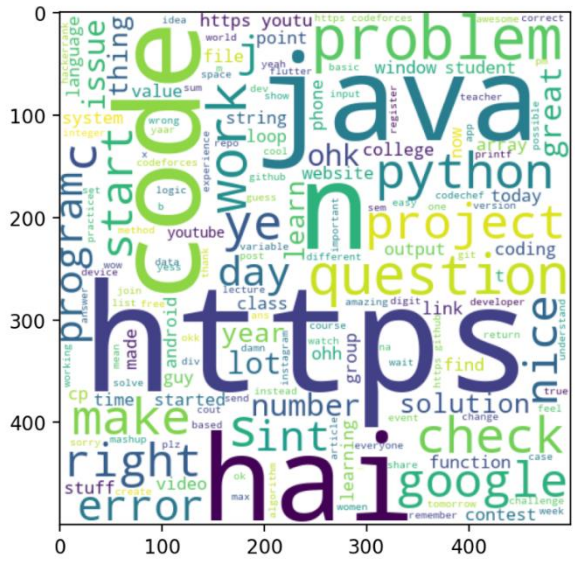


Most Busy Months

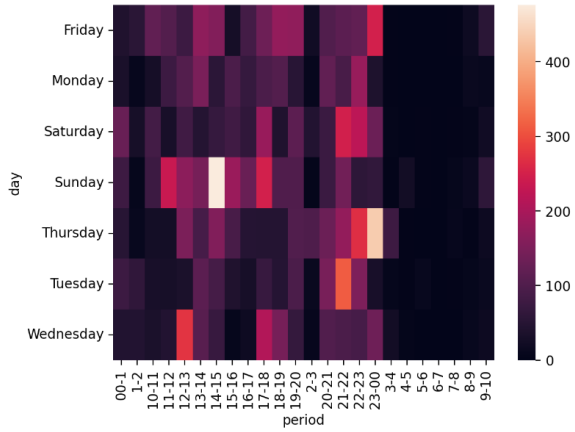


The graph depicts the hectic schedules we experience throughout the days and months. Utilizing the matplotlib library, we plotted a graph that correlates the number of messages with each specific day or month.

Wordcloud



Weekly Activity Map



The weekly activity map shows how messages change every week in a WhatsApp chat. The word cloud gives a quick look at the most common words or phrases in the chat.

6. Conclusion

To summarize, the WhatsApp application and the Python programming language are highly essential for conducting thorough data analysis. This project focused on analyzing a WhatsApp chat by utilizing dataset manipulation techniques to uncover meaningful insights. The analysis revealed commonly used emojis and frequently repeated words, providing insights into our conversations and time spent chatting. The project was carried out using Python, along with libraries such as NumPy, Pandas, Matplotlib, and Seaborn. The work concluded with successful outcomes, revealing the extent of individuals' participation in the specified WhatsApp group chat. This system showcases the capability to analyze any WhatsApp conversation inputted into it.

7. References

1. Ravishankara K, Dhanush, Vaisakh, Srajan I S, “International Journal of Engineering Research & Technology (IJERT)”, ISSN: 2278-0181, Vol. 9 Issue 05, May-2020

2. Marada Pallavi, Meesala Nirmala, Modugaparapu Sravani, Mohammad Shameem, Dr. K. Soumya, “International Research Journal of Modernization in Engineering Technology and Science (IRJMETS)”, eISSN: 2582-5208, Volume:04/Issue:05/May-2022

3. https://www.studocu.com/in/document/shri-vaishnav-vidyapeeth-vishwavidyalaya/bachelor-oftechnology/whatsapp-chat-analyzer/39681617

4. Dr.T.Siva Ratna Sai, T.Naga Nandini, M.Harsha Vardhan, B.Sivaji, R.Sai Kalyan, “Journal of Engineering Sciences”, Vol 14 Issue 04,2023.

5. https://medium.com/@avikumart_/a-complete-guide-to-building-streamlit-data-web-apps-withcode-examples-6cc1e42b2397

6. https://www.analyticsvidhya.com/blog/2023/02/whatsapp-chat-analysis-end-to-end-data-analysisproject-with-deployment/