IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

Whatsapp Text Analyzer

Harshitha K V[1], Hemadri A N [2], Keerthi S[3], Madhushree K[4]
B.E, CSE[1-4]
Kalpataru Institute of Technology, Tiptur, India

Abstract: *The WHATSAPP TEXT ANALYZER is a data analysis project designed to extract and interpret information from WhatsApp chat data. With the increasing use of WhatsApp as a primary mode of communication, large volumes of textual data are generated daily. This project focuses on transforming that unstructured data into meaningful insights. The system accepts exported WhatsApp chat files as input and performs data cleaning and preprocessing to remove unwanted characters, timestamps, and media indicators. Various analytical techniques are applied to evaluate message frequency, user activity levels, commonly used words, and conversation trends over time. Additionally, natural language processing (NLP) methods are used to conduct sentiment analysis and visualize emotional patterns within the chat. The results are represented through charts, graphs, and word clouds for better interpretation. This project demonstrates how data analytics and text processing techniques can be applied to everyday communication data to understand user behavior, communication patterns, and social interaction dynamics.*

The analyzer employs regular expressions for accurate extraction of timestamps, sender names, and messages, followed by comprehensive preprocessing to remove noise, system messages, and formatting inconsistencies. Statistical computations are performed to determine key metrics such as total messages, word count, emoji usage, media shared, and most active participants. In addition, the system generates time-based analytics including daily, monthly, and hourly activity patterns, as well as a weekly activity heatmap. The integration of Natural Language Processing (NLP) enables the creation of a word cloud for frequently used terms, while optional sentiment analysis provides insights into the emotional tone of the conversation.

Visualization modules using Matplotlib and Seaborn produce intuitive graphical representations of chat trends, making the analysis more insightful and accessible. The system is designed to be user-friendly, efficient, and capable of handling large group chats with thousands of messages. The WhatsApp Chat Analyzer thus offers a powerful and automated approach to understanding communication behavior, supporting academic research, social analysis, and personal reflection on messaging habits.

Keywords: WhatsApp Text Analyzer, Chat Data Analysis, Text Mining, Natural Language Processing, Data Analytics, Message Statistics, Emoji Analysis, Word Frequency, Social Media Analysis, Unstructured Data

I. INTRODUCTION

In the modern digital era, communication has undergone a rapid transformation with the widespread adoption of instant messaging applications. Among these, **WhatsApp** has emerged as one of the most popular and influential platforms, enabling individuals and groups to exchange messages, multimedia, documents, and voice notes seamlessly. With billions of active users globally, WhatsApp has become an essential tool for personal communication, education, business interactions, and community collaborations. As a result, a massive volume of conversational data is generated daily across diverse contexts.

Every interaction on WhatsApp—whether a personal conversation, family group chat, academic discussion, or professional communication—produces valuable textual patterns. These patterns can reveal communication frequency, user engagement levels, conversational topics, sentiment, emotional behavior, information flow, and social dynamics

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

within groups. However, despite the abundance of such rich data, very few users possess the time or technical ability to manually analyze their chat histories.

As WhatsApp chats grow over months or years, manually scrolling through long message threads becomes inefficient, overwhelming, and practically impossible. A typical chat may contain thousands or even millions of messages, including text, emojis, reactions, multimedia, and deleted content. Extracting insights such as "Who is the most active participant?", "What time do conversations peak?", or "What emotions dominate the chat?" becomes extremely challenging. Furthermore, important patterns—such as changes in communication behavior, social trends, or sentiment fluctuations—remain hidden unless analyzed using systematic and computational approaches.

To address these challenges, the field of text analytics and natural language processing (NLP) has introduced innovative techniques to extract meaningful information from unstructured textual data. Techniques like tokenization, frequency analysis, sentiment detection, keyword extraction, and data visualization help transform raw text into structured knowledge. By applying such methods to WhatsApp chat exports, it is possible to build tools that convert everyday conversations into comprehensible analytical reports.

II. PROBLEM STATEMENT

With the exponential growth of digital communication, WhatsApp has become one of the most widely used platforms for exchanging personal, professional, and social messages. Every day, millions of conversations generate vast amounts of unstructured textual data that reflect users behavior and interaction patterns. Despite this abundance of data, there are limited tools available to analyze and interpret chat histories in a meaningful and automated manner.

To design and implement an intelligent system capable of analyzing WhatsApp chat data using natural language processing (NLP) and data visualization techniques to derive meaningful insights into user behavior and communication patterns.

III. METHODOLOGY

The methodology of the WhatsApp Text Analyzer describes the systematic process followed to collect, preprocess, analyze, and visualize WhatsApp chat data. The entire workflow is designed to transform raw, unstructured text data into meaningful insights using text analytics techniques.

Data Collection

The first step involves exporting WhatsApp chat data from the mobile application in text (.txt) format. WhatsApp allows users to export individual or group chats without media, which helps reduce data size and complexity. These exported files serve as the primary input for the analyzer.

Data Preprocessing

Since WhatsApp chat data is unstructured, preprocessing is essential. This step includes:

Removing timestamps, phone numbers, and system-generated messages.

Handling missing or incomplete messages.

Converting text into a standardized format.

Tokenization of text into words.

Data Parsing

The cleaned chat data is parsed to extract key elements such as sender name, message content, date, and time. Regular expressions are used to identify message patterns and separate messages from metadata. The parsed data is then stored in a structured format such as a table or dataframe.

Text Analysis

Various text analysis techniques are applied to extract insights, including:

Message Count Analysis: Calculates total messages and identifies the most active users.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

Word Frequency Analysis: Identifies commonly used words and phrases.
Emoji Analysis: Extracts and counts emojis to understand emotional expression.
Time-Based Analysis: Examines chat activity by hour, day, week, and month.

Data Visualization

The analyzed results are presented using charts and graphs such as bar charts, line graphs, and word clouds. Visualization helps users easily understand communication patterns and trends within the chat data.

Result Interpretation

The final step involves interpreting the analytical results to draw meaningful conclusions about user behavior, communication frequency, and interaction trends. The insights can be used for social behavior studies, group interaction analysis, or personal chat evaluation.

IV. RESULTS AND DISCUSSION

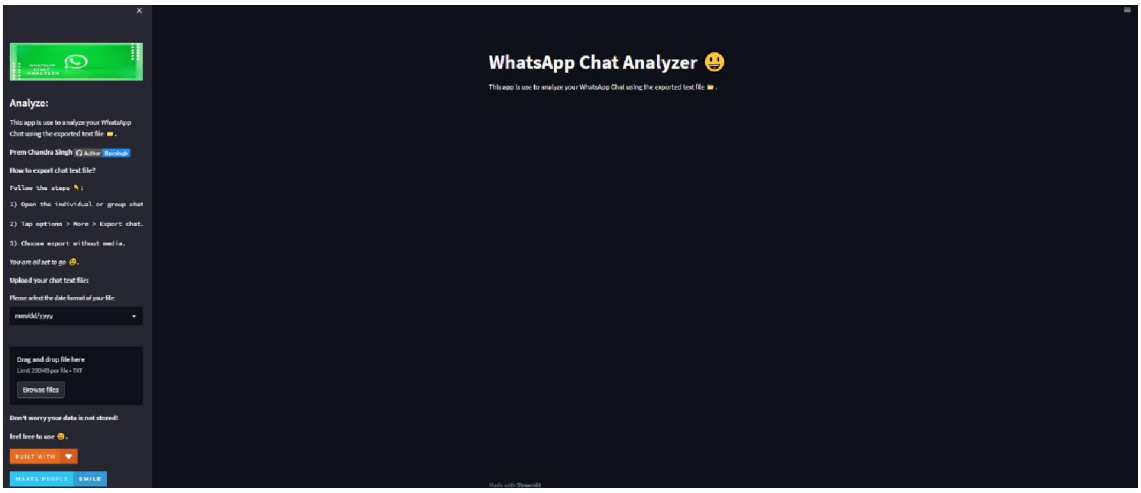

FIG 8.1 User interface setup

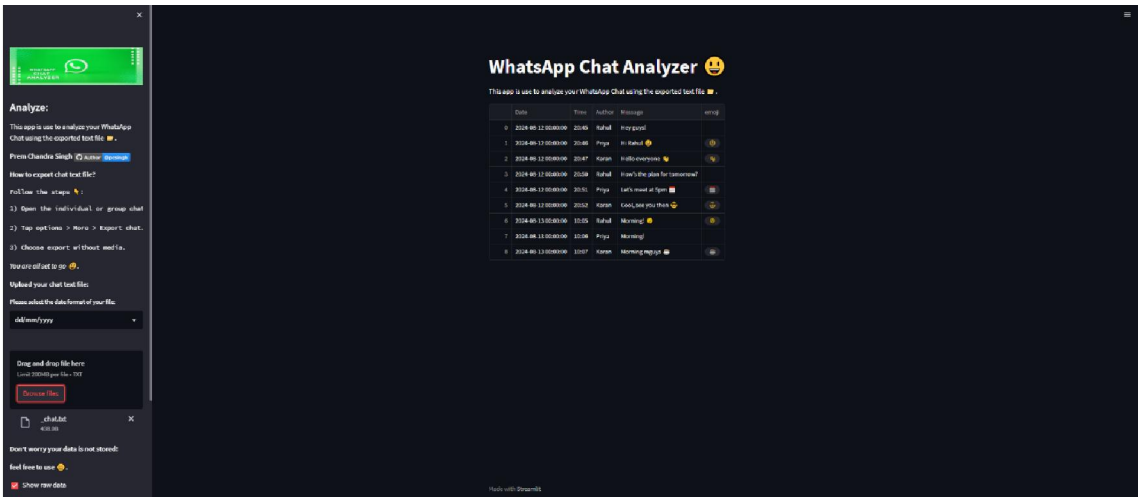

FIG 8.2 Data Loading

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

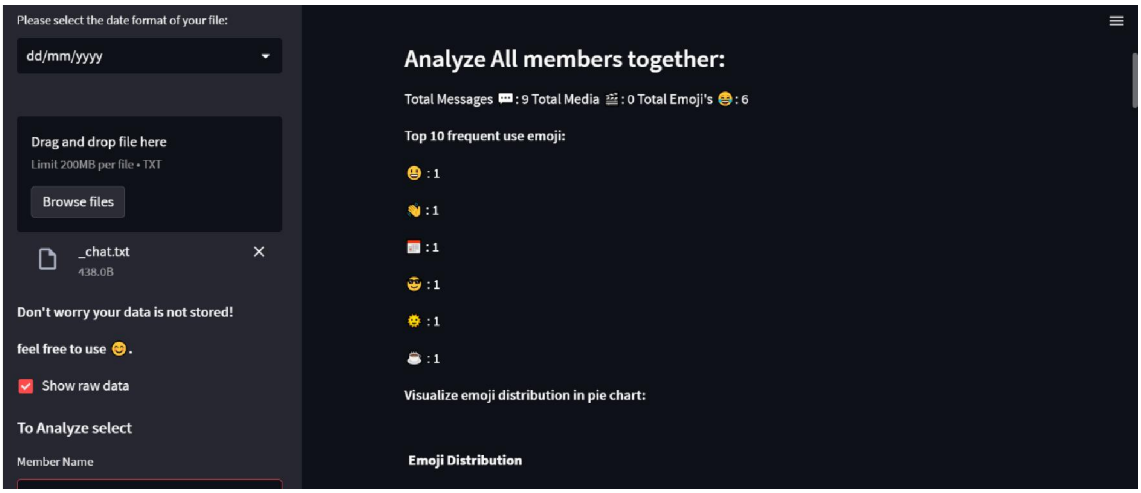

FIG 8.3 Analyze all members together

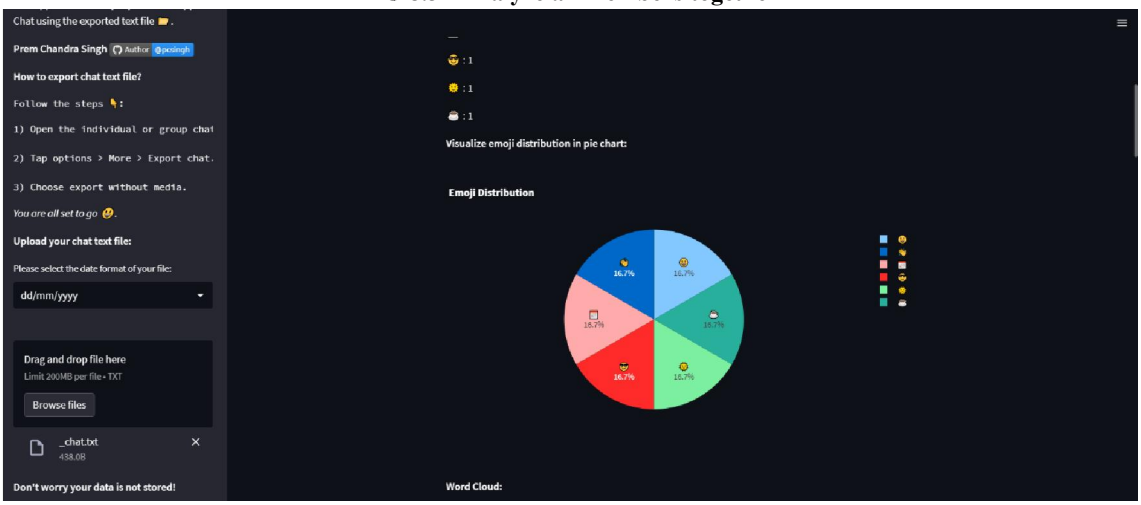

FIG 8.4 Emoji distribution in pie chart

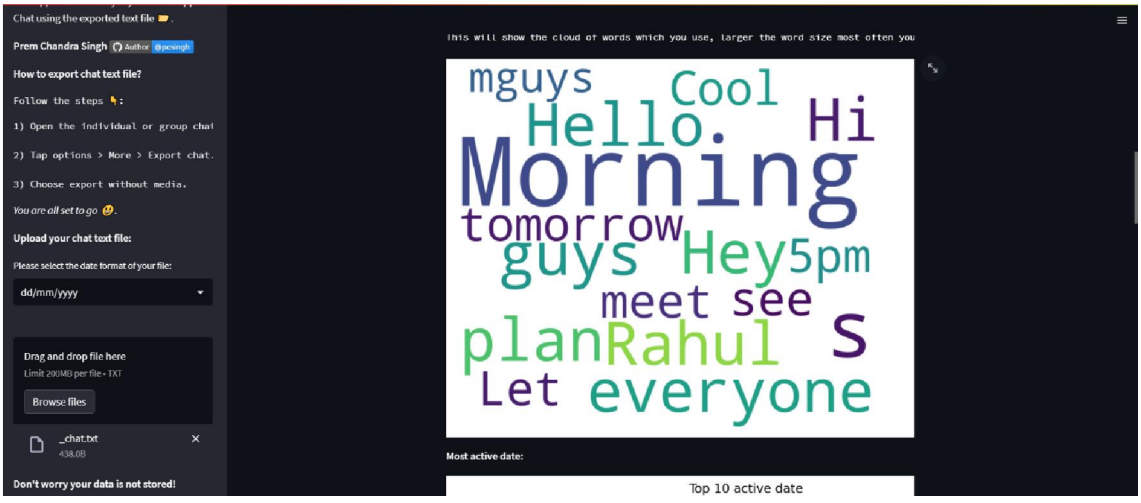

FIG 8.5 Word cloud

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

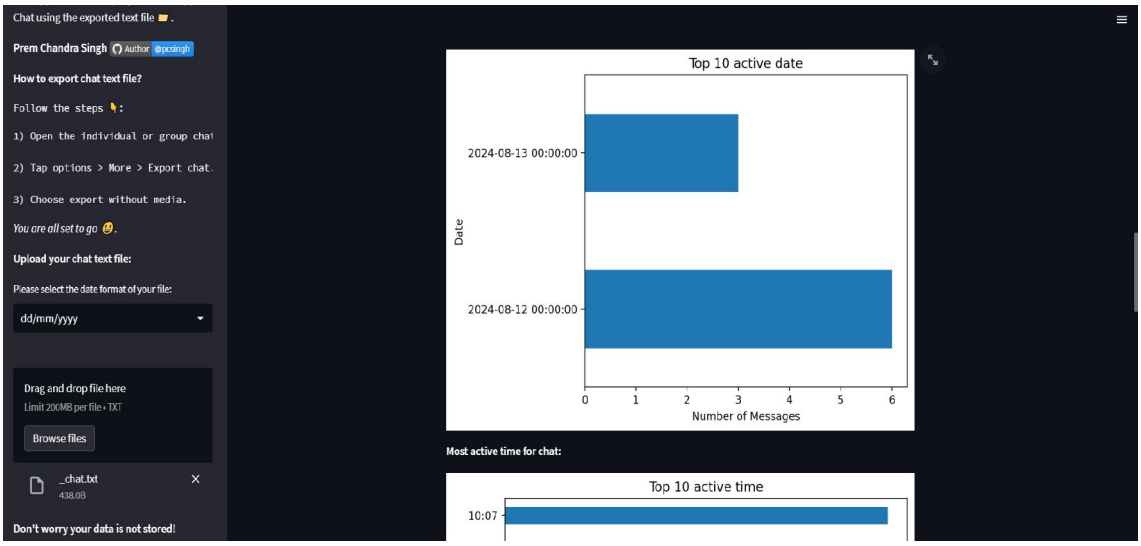

FIG 8.6 Most Active Date

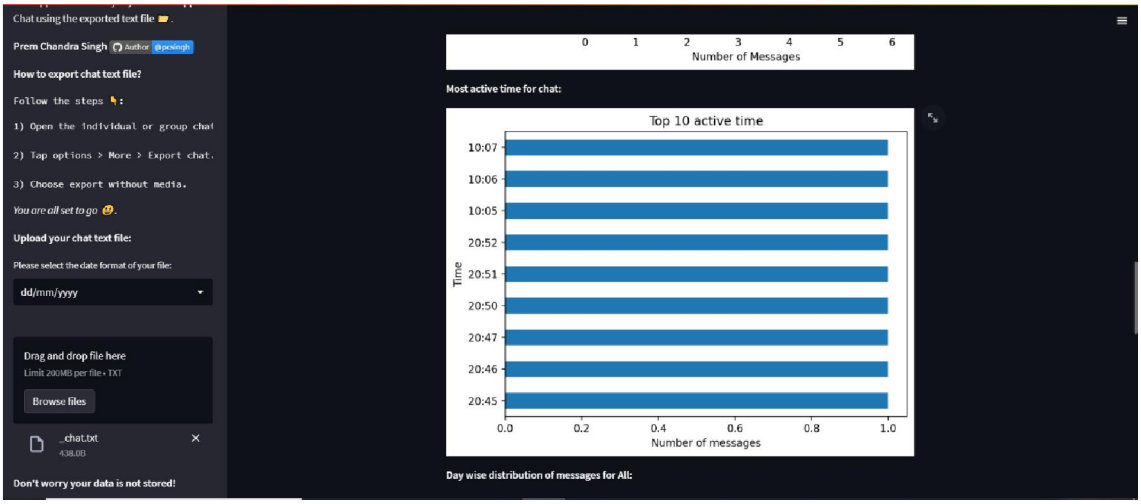

FIG 8.7 Most Active Time

The WhatsApp Text Analyzer successfully demonstrates how unstructured chat data can be transformed into meaningful and actionable insights using text processing and data analytics techniques. Through the analysis of exported WhatsApp chat files, the system provides a clear understanding of communication behavior and interaction patterns among users.

One of the major strengths of the system is its ability to handle large volumes of chat data efficiently. By performing preprocessing tasks such as data cleaning, removal of irrelevant symbols, and stop-word elimination, the analyzer ensures that the extracted information is accurate and reliable. This preprocessing step plays a crucial role in improving the quality of results, especially in word frequency and emoji analysis.

The results obtained from message frequency analysis clearly highlight the most active users and peak communication periods. Time-based analysis helps in identifying daily, weekly, and monthly activity trends, which can be useful for understanding user engagement levels in both personal and group conversations. Emoji and word usage analysis further reveals the emotional tone and common discussion topics within chats.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, January 2026
Impact Factor: 7.67

V. CONCLUSION

The WhatsApp Chat Analyzer project is a powerful tool designed to extract and visualize meaningful insights from WhatsApp chat data. By allowing users to upload their exported chat files, the application parses the data and provides comprehensive analyses, including statistical summaries, emoji usage, word clouds, and patterns of activity over time. Implemented using Streamlit for the web interface and Pandas for data manipulation, the project ensures a user-friendly and interactive experience. The visualizations, created with Plotly, present the results in an easily interpretable format, making it simple for users to understand their communication habits and chat dynamics. Looking ahead, the project has significant potential for further enhancements, such as integrating sentiment analysis to classify messages by emotional tone and topic modeling to identify common themes in conversations. Advanced visualizations like network graphs could provide deeper insights into interactions between chat participants. Additionally, enabling real-time analysis could allow users to gain continuous insights into their ongoing chats. Overall, the WhatsApp Chat Analyzer project effectively demonstrates the practical application of data science and machine learning to everyday data, offering valuable tools for understanding digital communication patterns and laying the groundwork for more sophisticated analysis tools in the future.

VI. ACKNOWLEDGMENT

We would like to express our sincere gratitude to our project guide and faculty members for their continuous support, valuable suggestions, and encouragement throughout the development of this work. Their guidance helped us understand the concepts and complete the project successfully.

We also extend our appreciation to our institution for providing the necessary resources and a supportive learning environment. Finally, we thank all our friends and family members for their motivation and constant support during this project.

REFERENCES

[1]. Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processsing (3rd Edition). Stanford University.

[2]. Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of ICWSM.

[3]. Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates.

[4]. Cambria, E., Poria, S., Hazarika, D., & Kwok, K.s (2018). SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis.

[5]. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.