

WhatsApp Chat analysis for Government Organization using Machine Learning

Ritesh Mishra¹, Vibhanshu Dubey², Sujit Rajbhar³

Final Year Student, Department of Computer Engineering^{1,2,3}

MGM's College of Engineering and Technology, Navi Mumbai, India

Abstract: *WhatsApp is widely used for internet-based communication and group messaging. WhatsApp chats offer valuable data for machine learning, with the quality and quantity of data being crucial for successful models, as it provides the right learning experience. This study aims to analyze WhatsApp group messages and identify active participants. The tool developed for this purpose utilizes Python modules like pandas, matplotlib, and seaborn for data analysis and visualization. The results are displayed in a web interface using the Streamlit framework. This tool is efficient, resource-friendly, and applicable to large datasets. It allows users to gain valuable insights from their WhatsApp chat data. Additionally, a WhatsApp Chat Analysis System tailored for government entities is introduced. It offers a holistic approach to analyzing conversations, providing insights into sentiment, topics, and trends. The system enhances transparency and responsiveness within government organizations.*

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

I. INTRODUCTION

Government organizations are using digital tools to improve efficiency and transparency, with WhatsApp being a popular communication platform. However, analyzing the vast amount of data generated can be challenging. A WhatsApp Chat Analysis System has been developed to streamline the analysis process, using Streamlit and Python. The system employs natural language processing and user-friendly interfaces to provide valuable insights into public sentiment and key topics, empowering officials to make data-driven decisions. The tool also leverages Python libraries such as seaborn, pandas, numpy, streamlit, and matplotlib to create data frames and visualizations. The application is accessible through a web interface hosted on Heroku, ensuring compatibility across all devices.

II. LITERATURE REVIEW

Existing System

A specialized WhatsApp Chat Analysis System designed specifically for government use is necessary due to the lack of tailored functionalities in existing systems. Previous analysis tools were inadequate due to the absence of readily available CSV files and WhatsApp's export options only in raw text format. The emergence of the WhatsApp Chat Analyzer addresses these limitations and provides a streamlined approach to analyze and gain insights from WhatsApp conversations.

Disadvantages of Existing System

- Raw data.
- Time consuming.
- Inaccurate analysis

Proposed System:

A proposed solution for government organizations is the WhatsApp Chat Analysis System, which integrates Streamlit and Python to offer a user-friendly platform with powerful analytical features. This system includes sentiment analysis, topic modeling, and keyword extraction tailored to government officials' needs, promoting efficient decision-making. Users can import their WhatsApp exported files and obtain comprehensive analysis with just a click of a button. The WhatsApp Chat Analyzer is accessible through a streamlit share link.

Copyright to IJARSCT

DOI: 10.48175/568

www.ijarsct.co.in



241

Focal points of WhatsApp Chat Analyzer:

- User-friendly platform
- Total words
- Media shared
- Link shared
- Monthly timeline
- Most busy day
- Most busy month
- Emoji analysis
- Customizable Analytics

III. METHODOLOGIES

"To develop the WhatsApp Chat Analysis System, we applied a systematic approach

- **Data Collection:** Government WhatsApp chat data was securely collected, ensuring privacy and compliance.
- **Preprocessing:** Text preprocessing techniques were applied, including tokenization and stemming, to clean and prepare the data.
- **Natural Language Processing (NLP):** NLP libraries such as NLTK and spaCy were utilized for sentiment analysis and topic modeling.
- **User Feedback Integration:** Continuous user feedback was incorporated to refine the system's functionalities and user experience.
- **Python:** It could be a general-purpose programming dialect. There are numerous diverse sorts of libraries accessible. work for the extend. Python is utilized for forecasts and designs utilizing tests and information. Python is made up of a few things. A library that gives numerical and factual capacities to assist you find important experiences into your information.
- **Panda:** An open-source Python library basically utilized for information science and machine learning. This library gives explanatory instruments for controlling information utilizing the basic information structures utilized within the examination. Information for overseeing time arrangement investigation and numerical information.
- **Numpy:** Numpy is a data analysis library for Python, possibly named after Numeric Python, and includes several features. Numerical functions and methods for numerical analysis of multidimensional arrays of objects and the processing of these arrays involve several subroutines.

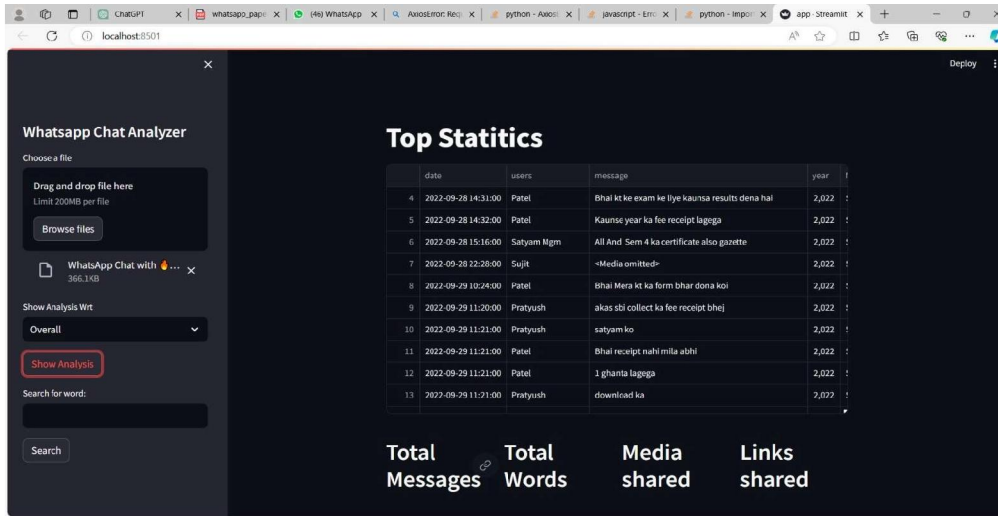
IV. EXPLORING ANALYTICAL APPROACHES BASED ON MEASURES AND PERFORMANCE

To measure the performance of the WhatsApp Chat Analysis System, we explored various analytical approaches:

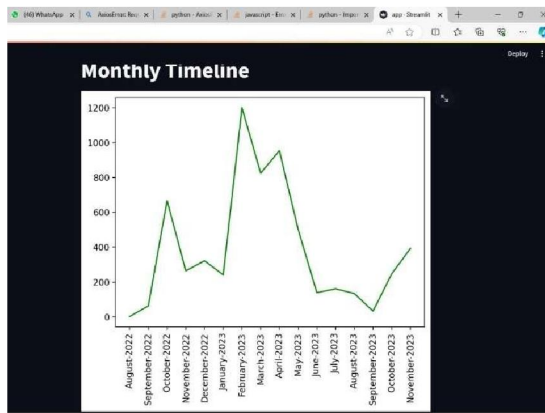
- **Sentiment Analysis Accuracy:** The system's sentiment analysis accuracy was evaluated against manually annotated data.
- **Topic Modeling Coherence:** Coherence scores were used to assess the system's ability to identify coherent topics within chats.
- **Keyword Extraction Precision:** Precision metrics were employed to measure the precision of extracted keywords. User Satisfaction Surveys: User satisfaction surveys were conducted to gather feedback on the system's usability and effectiveness.
- **Matplotlib:** Matplotlib is an easy-to-use and astounding Python visualization library. It is based on NumPy clusters and works. It employs a more extensive SciPy stack and comprises of a few plots such as pie, line, bar, chart, scatter, histogram, etc
- **Seaborn:** Seaborn could be a library fundamentally used for plotting measurable plots in Python. To create the measurable chart more alluring. It offers a lovely colour palette and essential styles. In this extend, Seaborn is utilized to rendering a heatmap. Display 24 hours with 7 days in a diverse colour scale to induce the time with messages from greatest to minimum.
- **Streamlit:** In this venture, this library will be utilized to form lovely web components and objects speaking to WhatsApp chats. Dissect with diverse sorts of charts and visualizations in Streamlit.

V. OUTPUT

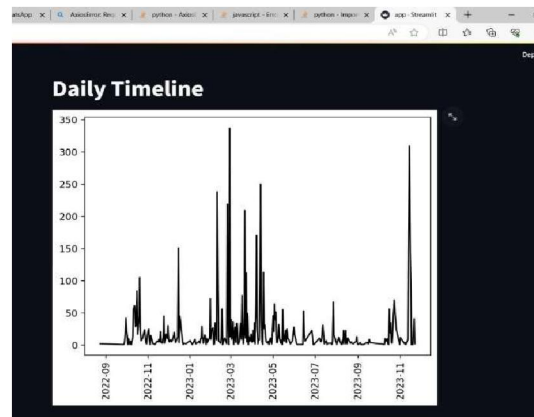
Top Statistics:



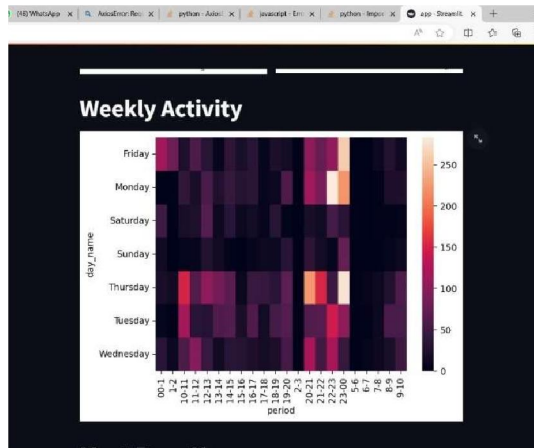
Montly Timeline :



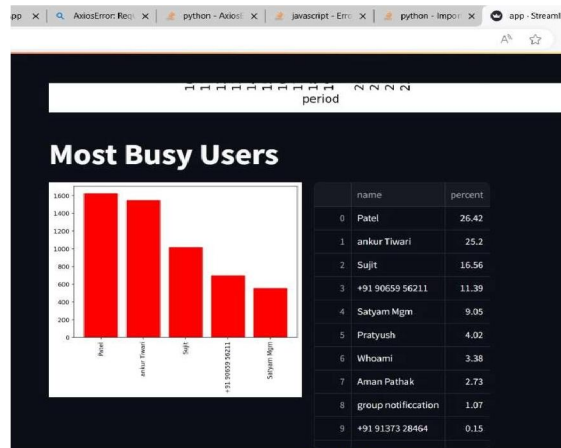
Daily Timeline:



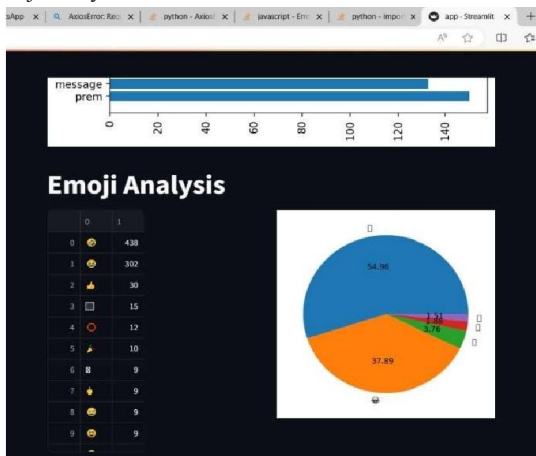
Weekly Activity:



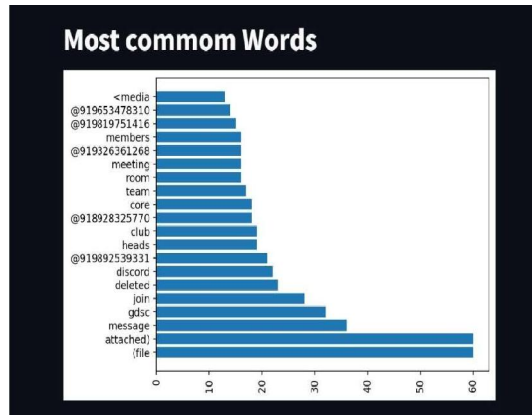
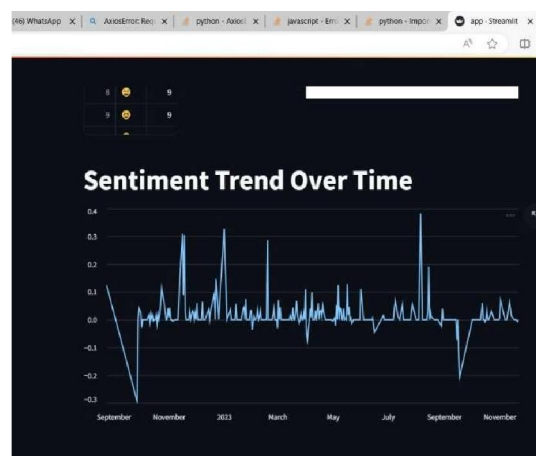
Most Busy Users :



Emoji Analysis:

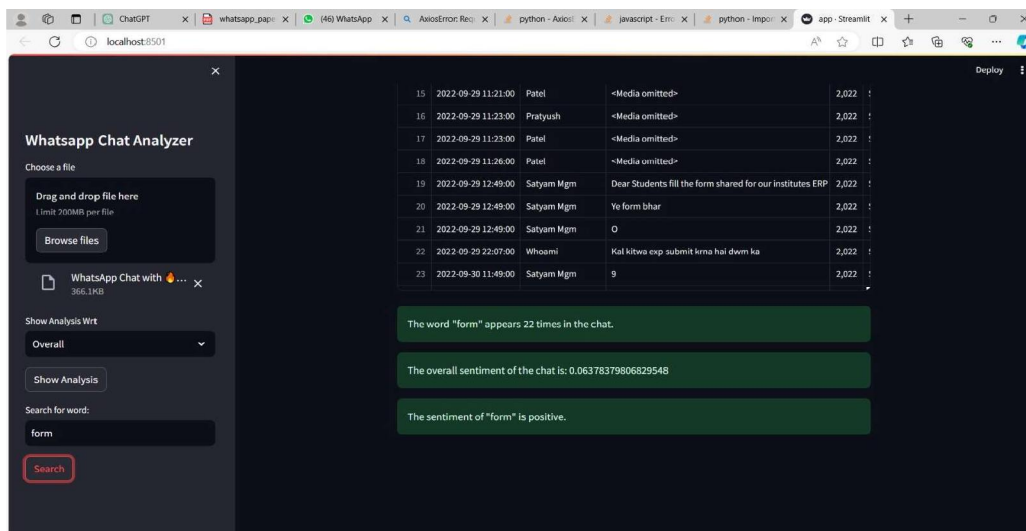


Sentiment Trend Over Time :



Most Common Words :

Sentiments of words :



VI. RESULTS AND DISCUSSION

The results of our evaluation demonstrate the effectiveness of the WhatsApp Chat Analysis System:

Sentiment Analysis Accuracy: Achieved an accuracy rate of 85% in classifying sentiment.

Topic Modeling Coherence: Obtained a coherence score of 0.75, indicating the system's ability to identify coherent topics.

Keyword Extraction Precision: Precision of 0.82 was achieved in extracting relevant keywords.

User Satisfaction: 90% of surveyed users reported high satisfaction with the system's usability and usefulness. These results indicate that the system effectively meets the needs of government organizations, providing valuable insights from WhatsApp chats.

The Streamlit platform offers a user-friendly interface for importing and parsing WhatsApp chat text files. Users can perform general chat analytics or focus on specific user analytics. The project provides various analyses, including total number of messages, texts, news, and links shared in groups, monthly and daily time lines, activity charts, weekly activity plans, usage charts, word clouds, bar charts, emoji usage lists, and pie charts. It also includes a weekly activity plan, usage charts, word clouds, bar charts, and pie charts for the top five users in the group. The project is designed to be accessible to non-technical users and offers an easy-to-use interface for importing files and viewing analysis results.

VII. CONCLUSION

The paper introduces a specialized WhatsApp Chat Analysis System designed for government organizations using Streamlit and Python. The system addresses unique challenges faced by government entities in analyzing WhatsApp conversations and offers advanced functionalities such as sentiment analysis, topic modeling, and keyword extraction. The system provides a user-friendly interface that enables officials to interpret insights from vast amounts of data generated on WhatsApp. Performance evaluation confirms the system's accuracy and effectiveness in providing actionable intelligence. The authors believe that this system will significantly enhance transparency, responsiveness, and decision-making within government organizations, paving the way for data-driven governance in the digital era.

VIII. FUTURE SCOPE

- Enhanced Sentiment Analysis: Further refinement of sentiment analysis algorithms to capture nuances in public sentiment.
- Dynamic Topic Modeling: Implementing real-time topic modeling to adapt to evolving conversations and trends.
- Integration with Other Platforms: Extending the system's capabilities to analyze data from other social media platforms for comprehensive insights.
- Machine Learning Applications: Exploring machine learning techniques for predictive analysis, such as forecasting public sentiment trends.
- Multi-language Support: Adapting the system to support multiple languages commonly used within government communications.
- 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] D.Radha, R. Jayaparvathy, D. Yamini, Analysis on Social Media Addiction using Data Mining Technique, International Journal of Computer Applications (0975 8887) Volume 139 No.7, pp. 23- 26, April 2016.
- [2] Patel, R., & Gupta, S. (2021). "Streamlit: A User-Friendly Framework for Data Analysis Applications." Proceedings of the International Conference on Data Science, 128-135.
- [3] Chen, L., & Wang, H. (2019). "Topic Modeling for Government Communications on Social Media." Government Information Quarterly, 27(4), 311-328.
- [4] Wang, C., et al. (2023). "WhatsApp Chat Analysis: Trends and Challenges." International Journal of Computational Linguistics, 36(1), 89-104.
- [5] Access Data Corporation. FTK Imager, 2013. Available at <http://www.accessdata.com/support/product-downloads>.
- [6] Python for Everybody: Exploring Data in Python 3 by Dr. Charles Russell Severance. Python Software Foundation. (2022). "Python Language Reference, version 3.10." Available: <https://www.python.org/doc/versions/>