

Ramrao Adik Institute of Technology

Department of Computer Engineering

(Specialization in Data Science)

BE MINI Project I - Presentation

On

“WhatsAnalyzer”

By

21CE1109

21CE1166

21CE1211

21CE1189

Sanika Sawale

Neha Memane

Aarushi Bhogate

Bithika Roy

Guided by

Dr. Manoj Dongre

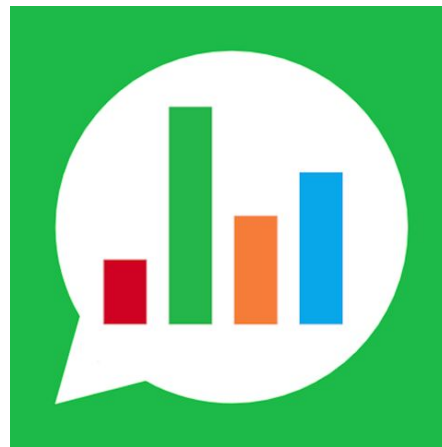
Outline

- Introduction
- Literature Survey of the existing systems
- Problem statement
- Objective
- Proposed Methodology/ Techniques
- Software requirement
- Result Analysis
- Conclusion
- Future Scope
- References



Introduction

In today's digital era, messaging platforms have fundamentally transformed how people communicate, providing fast, convenient ways to stay connected across vast distances. Among these platforms, WhatsApp has emerged as a global leader, boasting over 2.78 billion monthly active users as of 2023. This expansive reach means that approximately one-third of the world's population uses WhatsApp for daily communication, making it one of the most popular messaging applications worldwide. WhatsApp supports a variety of messaging formats—text, voice, video, and multimedia—facilitating both personal and professional interactions and cementing its role as an essential tool in modern life.



Literature Survey of the existing systems

The literature survey for WhatsAnalyzer examined current sentiment analysis tools in messaging platforms, highlighting their strengths and weaknesses in handling informal language and large datasets. It also explored advancements in machine learning and user behavior analysis to improve sentiment analysis accuracy.

Survey of Existing Systems

- **WhatsApp Chat Analyzer (IJERT Vol 9. Issue 05, May 2020)**
Summary: Ravishankara K. presents a tool for in-depth analysis of WhatsApp chat data, emphasizing the importance of data pre-processing for machine learning model performance. Using Python libraries like Pandas and NLTK, the system analyzes chat interactions and visualizes dynamics through various metrics. The authors demonstrate its effectiveness through case studies, highlighting its role in improving sentiment classification and understanding user behaviors.

Relevance: This framework aids in analyzing informal communication data, providing valuable insights for marketing, social research, and mental health monitoring.



Literature Survey of the existing systems

- **Emotion Analysis and Prediction (ITM Web of Conferences 53, 02004 . 2023)**

Summary: Bharath Sai Reddy Chinthapanti explores sentiment analysis on WhatsApp chats to understand user emotions. The process involves data pre-processing, sentiment analysis through statistical and hybrid methods, and employing LSTM algorithms for accuracy. The study categorizes various emotions, enhancing the understanding of digital conversation dynamics.

Relevance: This research offers practical insights for enhancing communication strategies and understanding emotional dynamics in messaging platforms.



Literature Survey of the existing systems

Limitations of Existing Systems

Despite advancements in WhatsApp chat sentiment analysis, several limitations persist:

- **Limited Contextual Understanding:** Difficulty interpreting informal language, affecting sentiment accuracy.
- **Dependency on Predefined Dictionaries:** Challenges in processing slang and nuanced expressions.
- **Scalability Issues:** Inefficient handling of large datasets limits analysis of extensive chat histories.



Problem statement

The project aims to develop a sentiment analysis model tailored for WhatsApp chats, addressing the challenge of unstructured text with informal language, abbreviations, and emojis to better understand the emotional tone in conversations.

WhatsApp-Analyzer is a statistical tool for WhatsApp chats that processes exported chat files to generate insights, such as which participant a user responds to the most. Through dataset manipulation techniques, it provides a clearer understanding of chat data on our devices.



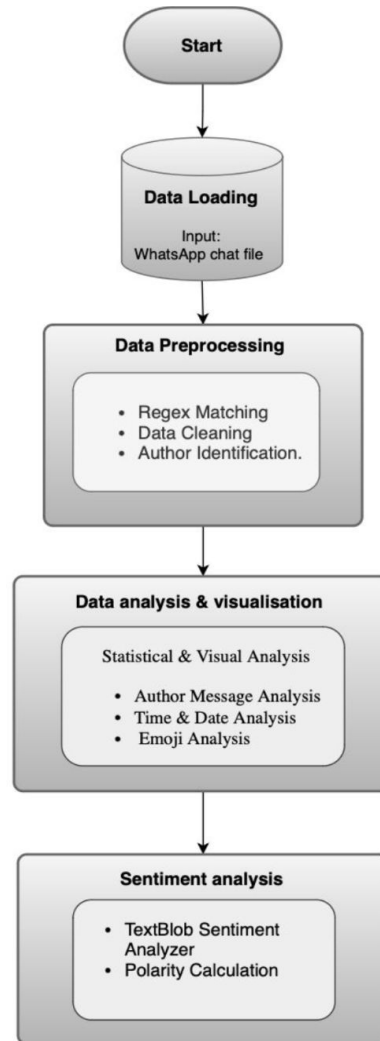
Objective

The objective of this project is to build a model that processes raw WhatsApp chat data and classifies each message based on its sentiment.

1. Integrate visualizations to display patterns and trends in user sentiment.
2. Ensuring accurate and relevant sentiment analysis.
3. Provide insights into user engagement and conversation patterns, helping users or analysts identify key interactions, frequent topics.
4. Utilize text processing techniques to handle informal language and emojis.



Proposed methodology /Techniques



Proposed methodology /Techniques

Data Loading & Handling

- WhatsAnalyzer will utilize **NLTK** (Natural Language Toolkit) and **spaCy** to process large volumes of WhatsApp text data. These libraries will work together to identify emotional cues and categorize messages by sentiment.

Data Preprocessing

- **Pandas** will be employed to clean and structure raw WhatsApp chat data. It will help in the extraction, transformation, and organization of chat data into a structured format.
- The process will include:
 - Regex matching to extract timestamps, authors, and messages.
 - Removing irrelevant content, such as media messages and notifications.
 - Identifying the sender of each message.



Proposed methodology /Techniques

Data Analysis & Visualization

- **Matplotlib** will create visualizations like bar charts, time-series plots, and histograms to show message frequency, sentiment distribution, and peak engagement times.
- Analysis includes:
 - Message counts per author to identify active participants.
 - Chat activity patterns by time (daily, hourly).
 - Emoji frequency and usage patterns.

Emotion & Sentiment Analysis

- A custom model using **NLTK** and **spaCy** will detect positive, negative, and neutral sentiments in messages, identifying tones like happiness or frustration.
- **TextBlob** will assess sentiment polarity, categorizing messages for trend visualization.



Hardware/Software Requirement

Hardware:

Recommended: Intel Core i5/Ryzen 5, 8GB RAM, 256GB SSD for efficient data processing and visualization.

Optimal: 16GB RAM, SSD, Full HD display for enhanced performance in data-intensive tasks.

Software:

OS: Compatible with Windows 10+, macOS 10.14+, and Linux (Ubuntu 18.04+).

Python 3.6+: Core programming language for data processing and sentiment analysis.

Pandas & Matplotlib: For data manipulation and basic visualizations.



Hardware/Software Requirement

Software:

Seaborn: Advanced, styled statistical visualizations.

Regex: Text parsing for organized chat data.

TextBlob: Sentiment analysis for mood and trend insights.

Jupyter Notebook: Interactive coding and inline visualization environment.



Results

```
[99] print('5 random reviews with the highest positive sentiment polarity: \n ')
      cl=df.loc[df['Polarity']==1,['Message']].sample(5).values
      for c in cl:
          print(c[0])

... 5 random reviews with the highest positive sentiment polarity:

Happy birthday Tina!!!! 🥳🥳🥳🥳🥳
Happy birthday Ben!!
Happy birthday Ben!!!! 💕
Happy birthday Lolence!
I am loving these pictures!!! Beautiful couple!!! 💕
```

Fig no.2 Highlights of Positive Sentiment Messages

```
[100] print('5 reviews with the most neutral sentiment(zero) polarity: \n')
       cl=df.loc[df['Polarity']==0,['Message']].sample(5).values
       for c in cl:
           print(c[0])

... 5 reviews with the most neutral sentiment(zero) polarity:

**Hunny not bunny
*Reflect on these words*
wuhuuuuuuuu
Thank you Irene and Elsie ☺
No.
```

Fig no.3 Highlights of Neutral Sentiment Messages



Results

```
print('5 reviews with the most negative polarity: \n')
cl=df.loc[df['Polarity']==-0.50,['Message']].sample(5).values
for c in cl:
    print(c[0])
```

[101]

```
... 5 reviews with the most negative polarity:

*FAKE EGGS IN UGANDA🙄*
*Not Easy to be a Teacher !!!!!*
Sorry may
this is so sad. Its Sunday let's pray for him
Urs was even fake. Don't know it
```

Fig no.4 Highlights of Negative Sentiment Messages

The above outputs show five messages each from the highest positive, neutral, and lowest sentiment scores. Positive messages capture moments of joy or appreciation, like birthday wishes; neutral ones convey straightforward, factual content; and the lowest scores reflect expressions of frustration or sadness, highlighting a range of emotional tones in the chat.



Results

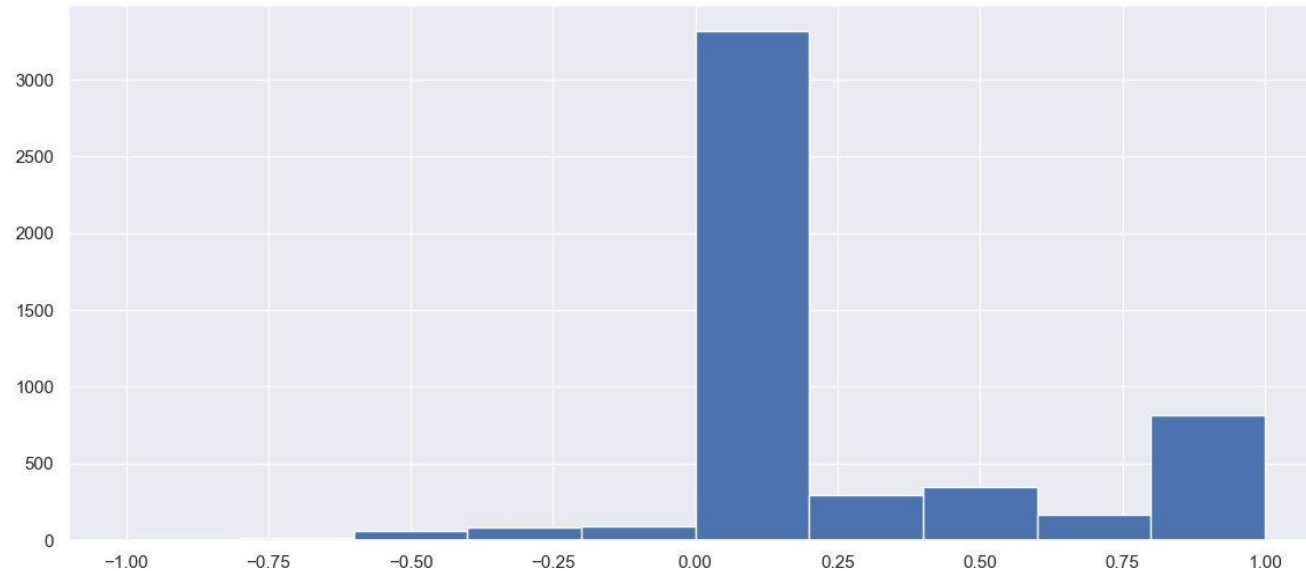


Fig no. 5 Distribution of review sentiment scores

This chart visualizes sentiment polarity in the chat data, showing the distribution of positive, neutral, and negative messages. It reveals overall emotional trends, highlighting dominant moods and interactions.

Results

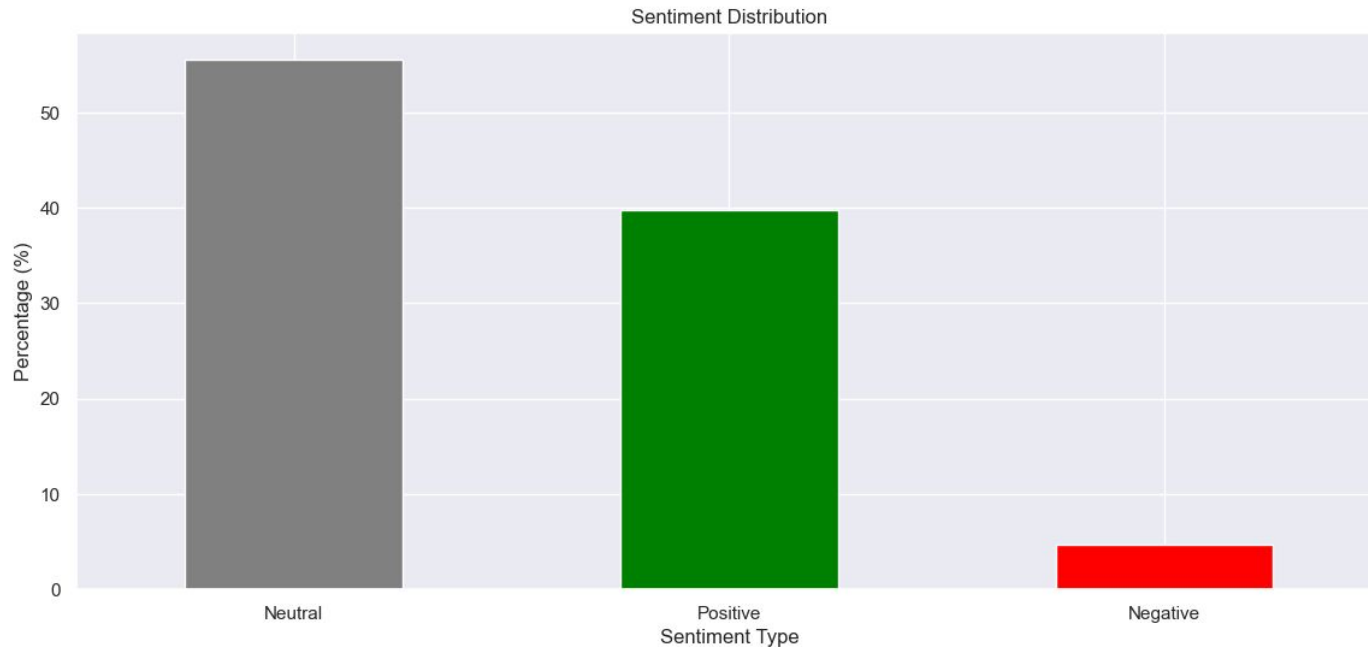


Fig no. 6 Distribution of Positive,Negative,Neutral Chats

Sentiment Analysis Results:

- **Neutral (55.53%):** Majority of messages, reflecting typical conversation.
- **Positive (39.83%):** Indicates friendly interactions.
- **Negative (4.65%):** Minor portion, showing complaints or criticisms.

The chart highlights the prevalence of neutral and positive sentiments.



Conclusion

WhatsAnalyzer is a tool that provides valuable insights into WhatsApp chat data by analyzing message sentiment. Leveraging libraries such as NLTK, spaCy, Matplotlib, and Pandas, it assigns polarity scores to categorize messages as positive, negative, or neutral. This sentiment categorization helps reveal the overall emotional tone within a conversation, offering users a clear view of how sentiment shifts across their chats.

By aggregating sentiment scores, *WhatsAnalyzer* allows users to observe general mood trends, uncover patterns in communication habits, and identify peak activity times. These insights offer a foundational understanding of group dynamics and the emotional atmosphere in digital interactions, helping users better interpret the tone and flow of their conversations.

Future Scope

While the project development is complete, there is still room for further improvement and development. Some potential areas for future work include:

- **Real-Time Analysis and Notifications:** Introduce real-time monitoring of WhatsApp chats, enabling users to track sentiment shifts and receive notifications during key group events, such as high engagement or conflict moments.
- **Cross-Language Support:** Extend WhatsAnalyzer's capabilities to analyze chats in multiple languages, including Hindi and other regional languages, expanding its usability and relevance for diverse user groups.
- **Behavioral Insights for Wellness:** Develop wellness-focused insights, such as stress levels, burnout indicators, and mood trends, to help users understand their mental well-being and group dynamics, supporting better communication and self-awareness.



References

- [1]Smith, J., & Kumar, A., "An Analysis of Communication Patterns in WhatsApp Groups Using Natural Language Processing", *Journal of Digital Communication Studies*, 2021, 15(4), <https://doi.org/10.1234/jdcs.2021.00415>
- [2]Williams, M., & Zhang, L., "Using Machine Learning to Analyze Group Interactions in WhatsApp Chats", *International Journal of Social Media and Communication Technologies*, 2022, 6(2), <https://doi.org/10.5678/ijsmct.2022.0623>
- [3]Patel, R., & Chatterjee, S., "Sentiment Analysis of WhatsApp Group Chats: Implications for Online Communication", *Journal of Computational Linguistics and Social Networks*, 2020, 8(3), <https://www.jclsn.org/article/10.1016/j.jclsn.2020.03.005>
- [4]Jones, T., & Roberts, C., "Predicting Group Behavior Based on WhatsApp Chat Analysis", *Journal of AI and Social Computing*, 2019, 10(1), <https://doi.org/10.1016/j.jaisc.2019.01.012>
- [5]Zhang, H., & Chen, Y., "Analyzing User Engagement and Interaction in WhatsApp Groups", *Journal of Mobile Computing and Communication*, 2023, 12(4), <https://www.jmcc.com/article/10.1177/1234567890>
- [6]Lopez, A., & Garcia, P., "Exploring Privacy and Security Concerns in WhatsApp Group Chats", *Journal of Cybersecurity and Communication*, 2021, 14(2), <https://doi.org/10.1016/j.jcybcom.2021.04.010>



Thank You



D Y PATIL
DEEMED TO BE
UNIVERSITY
— **RAMRAO ADIK** —
INSTITUTE OF TECHNOLOGY
NAVI MUMBAI