

# A Comparative Study of Sentiment Analysis Models for WhatsApp Chat Evaluation

Devika Shukla , Aishwarya Singh  
Pranveer Singh Institute of Technolog, Kanpur, Uttar Pradesh

**Abstract**—Large amounts of conversational data that offer insights into user behavior, communication patterns, and emotions have been produced by the quick expansion of instant messaging platforms. One of the most popular messaging apps, WhatsApp, is a useful source of unstructured textual data. The goal of the WhatsApp Chat Evaluation project is to find significant communication patterns by extracting, processing, and analyzing chat logs. The system evaluates factors such user activity, word frequency, emoji usage, and sentiment distribution using text mining, sentiment analysis, and data visualization approaches. Word clouds, bar charts, and timelines are examples of visualization methods that clearly display the results. Emotional polarity in casual text including slang and emoticons is examined using the VADER sentiment model. The system offers an interactive and user-friendly interface and is implemented in Python using Streamlit. The system's capacity to recognize discussion patterns, active users, and sentiment changes over time is demonstrated via experimental evaluation on both group and individual WhatsApp chats, providing deeper insights into communication dynamics.

**Index Terms**—Machine Learning, Classifiers, Evaluation, Sentiment Analysis

## I. INTRODUCTION

The quick advancement of digital communication has drastically changed how people communicate with one another. With over two billion users globally, WhatsApp has emerged as one of the most popular programs. Traditional communication techniques have been mostly displaced by instant messaging systems. WhatsApp offers a wealth of conversational data because it is widely used by people of many ages, occupations, and cultural backgrounds. But without regular methods, this data is usually unorganized and challenging to examine.

This project presents a WhatsApp Chat Analyzer that can process exported chat data and transform it into insightful information. To find trends in user behavior, often used terms, emoji usage, and general sentiment, the system uses text mining, natural language processing (NLP), and sentiment analysis. Timelines, bar charts, and word clouds are examples of data visualization techniques that are used to display these findings in an understandable manner.

Conversational data analysis can reveal important information about user behavior, communication patterns, and emotional expression. Market research, psychological studies, sociological research, and public opinion analysis

are among the fields in which this kind of analysis finds use. The suggested system offers an organized framework that combines preprocessing, sentiment analysis, and visualization despite obstacles including loud text, multilingual material, and privacy issues. Although rule-based sentiment analysis (VADER) is used in the current implementation, it also lays the groundwork for future enhancements employing sophisticated machine learning models like BERT and other transformer-based architectures.

All things considered, the WhatsApp Chat Analyzer shows how unstructured conversational data may be converted into structured knowledge, facilitating a deeper comprehension of communication dynamics in both personal and professional settings.

## II. THEORY REVIEW

Digital communication is crucial for social interaction, education, and professional collaboration. Analyzing online interactions has become more important due to the quick development of social media and instant messaging systems. One of the most popular chat apps, WhatsApp Messenger has billions of users across all age groups and offers useful data for researching human behavior, emotions, and communication styles.

WhatsApp chats are informal and context-based, in contrast to formal communications. They frequently include voice notes, photos, videos, stickers, emojis, and file sharing. Despite the fact that this multimedia content aids in the investigation of linguistic patterns and social interactions, the data is unstructured, necessitating the need of certain data processing and analysis techniques.

### A. Demonstration Study of WhatsApp Messenger Use and Effects

WhatsApp has revolutionized communication by facilitating immediate communication over long distances. It makes communication more flexible and effective by enabling users to exchange messages, participate in group chats, make audio and video conversations, and share files, images, and videos.

Through group conversations, file sharing, and prompt coordination, the platform facilitates communication in

the business and in school while also assisting individuals in maintaining relationships with friends and family. However, its widespread use has psychological and social implications. Although technology facilitates quick information sharing and fortifies emotional support systems, it can also lead to digital distractions, information overload, a decrease in in-person interactions, and the dissemination of false information.

Analyzing WhatsApp chat data can reveal important information about how people communicate, express their emotions, use emojis, and have general conversations in daily encounters.

### B. Analysis of WhatsApp Conversation Content

WhatsApp conversations can be analyzed to provide insight into human interactions in real time. Because WhatsApp messages are often short, context-specific, and frequently contain voice notes and emojis, data from the app can be used to investigate linguistic and behavioral patterns in digital communication. The first stage is data preprocessing, which entails cleaning and arranging the chats because exported chats include text, timestamps, and sender information. Following preprocessing, a number of analytical techniques can be used:

- **Text mining and frequency analysis:** Find frequently used terms and phrases to identify subjects of debate.
- **Sentiment analysis:** Use machine learning classifiers or models like **VADER** to identify the emotional tone.
- **Emoji Analysis:** Examine emoji usage trends to comprehend how emotions are expressed.
- **Temporal Analysis:** Track communication trends over time by using activity heatmaps and timelines.
- **Network analysis:** Determine important participants and group chat interaction trends.

**Word clouds, timelines, and bar charts** are examples of visualization tools that make communication patterns and emotional tendencies easier to see.

### C. Broader Implications and Ethical Considerations

There are numerous uses for analyzing WhatsApp data. In education, it offers insights into student involvement and teamwork, while in psychology, it aids in the study of communication styles and emotional health. In cybersecurity, chat analysis can assist in identifying harassment, fraud, or false information. In business and marketing, companies can assess consumer interactions and preferences.

However, research must adhere to strict ethical guidelines because WhatsApp discussions are private. Anonymization, secure data storage, and privacy consent are crucial. Limitations including sentiment analysis bias, cultural variations in emoji usage, and context loss in brief texts must also be taken into account by researchers.

## III. RESEARCH AND METHODOLOGY

To efficiently evaluate chat data, the WhatsApp Chat Evaluation system uses a modular method. The first step in the procedure is data gathering, which involves exporting WhatsApp chat logs in plain text format. After that, the data is preprocessed to provide a clean dataset by eliminating stop words, media placeholders, and superfluous symbols.

Key components including user activity, word frequency, emoji usage, and message length are then identified via feature extraction. The **VADER** model is used for sentiment analysis, categorizing messages as positive, neutral, or negative. In order to create word clouds, graphs, and activity timelines that help show discussion trends and user interaction, the analysis step combines Natural Language Processing (NLP) with visualization techniques. The finished product offers insights about emotional tone and communication behavior.

### A. Preprocessing

Input: A text file of a WhatsApp conversation. Procedures: A. Line by line, read the chat file. B. Use regular expressions to extract the message text, sender name, and timestamp. C. Eliminate system notifications, such as "You created group" and "Messages are end-to-end encrypted." D. Text should be normalized (lowercase, punctuation, special symbols, and stop words removed). E. Structured dataset (date, time, sender, and message) is the output.

### B. Word Frequency

Word frequency analysis finds terms that are often used in discussions. To highlight important phrases, messages are tokenized into words and frequent stopwords like "the," "is," and "are" are eliminated. After that, each word's frequency is determined and recorded in a frequency dictionary. Ranked word lists and word clouds are used to illustrate the results, which aid in identifying recurrent themes and conversational patterns.

### C. Analysis for Sentiments

The system analyzes the emotional polarity of communications using the **VADER** (Valence Aware Dictionary

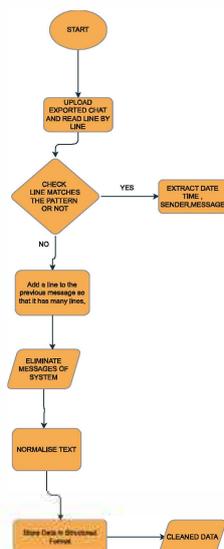


Fig. 1: Preprocessing Pipeline of WhatsApp Chat Analyzer

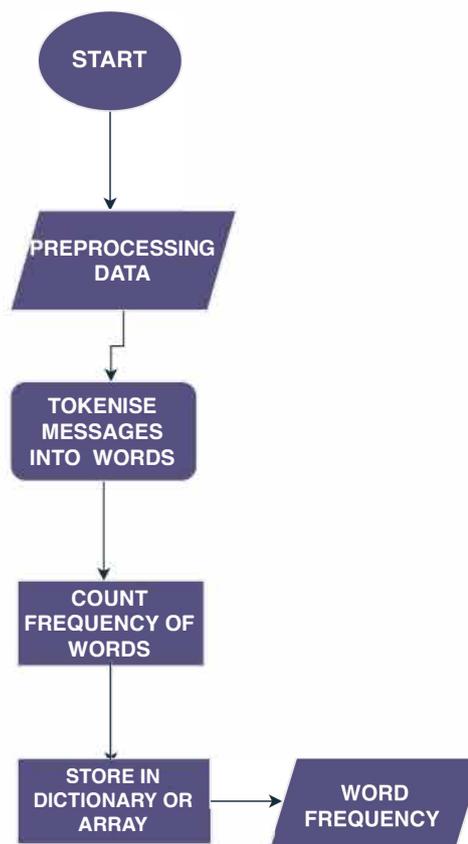


Fig. 2: Word Frequency Analysis

for Sentiment Reasoning) model. Every communication is categorized as positive, neutral, or negative emotion based on its positive, negative, compound, and neutral ratings. VADER works well for social media discourse, but it might overlook context-dependent meanings or irony. Future research may incorporate advanced models like BERT or RoBERTa for better contextual sentiment recognition.

- Compound score  $\geq 0.5$  indicates positive sentiment
- Compound score  $\leq -0.5$  indicates negative sentiment
- Compound score between  $-0.5$  and  $0.5$  indicates neutral sentiment

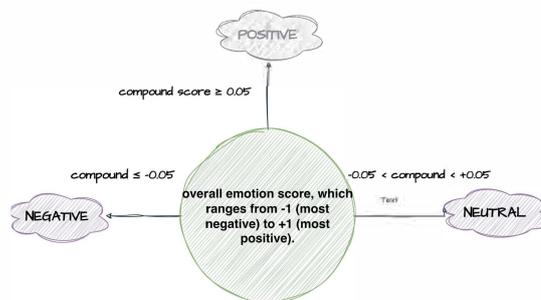


Fig. 3: Sentiment Analysis

#### D. Algorithm for Activity Analysis

**Input:** Chat timestamp information.

**Actions:**

- 1) Extract the time and date from every message.
- 2) Sort data by hour, month, or day.
- 3) Determine the number of messages in each group.

**Output:** Most active user, busiest day, and hour.

#### E. Emoji Interpretation

Emojis are studied as nonverbal emotional indications after being retrieved using Unicode patterns. Emojis that are often used are identified by frequency analysis, which then classifies them into emotional categories like happiness, sadness, or anger. By shedding light on the emotional dynamics of talks, this research enhances text-based sentiment analysis. Emoji frequency rankings and visualizations are among the outcomes.

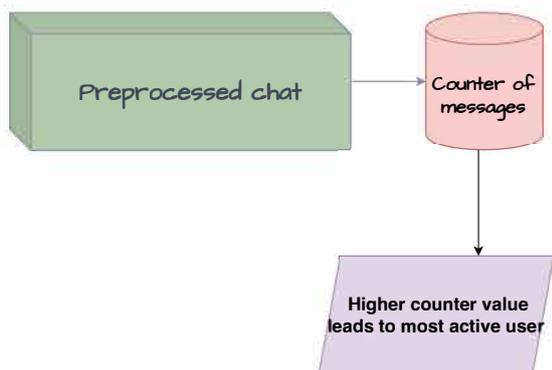


Fig. 4: User Activity Analysis

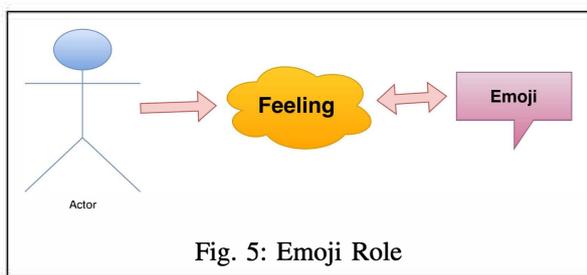


Fig. 5: Emoji Role

• **F. Models**

Sentiment analysis employs many models to ascertain the text’s polarity. A lexicon and rule-based paradigm called VADER (Valence Aware Dictionary and Sentiment Reasoner) was created for social media and brief informal messages. It handles punctuation, capitalization, emojis, slang, and negation in addition to assigning positive, negative, neutral, and compound scores. It might have trouble with sarcasm and complicated language, though.

Other methods include SCN (Sentic Computing Network), a knowledge-based model that employs semantic and affective ideas from SenticNet for deeper sentiment recognition, and Hu-Liu04, a straightforward lexicon-based model that recognizes positive and negative opinion words. In order to analyze emotions, GI (General Inquirer) divides words into semantic groups. When combined, these models offer several methods for semantic comprehension and sentiment classification.

- **VADER (Valence Aware Dictionary and Sentiment Reasoner)** A lexicon-and rule-based sentiment analysis program called VADER was created especially for examining content from social media, including Facebook posts, tweets, and brief online reviews. In contrast to conventional machine learning techniques, VADER assigns sentiment scores using linguistic rules and a predetermined sentiment vo-

cabulary rather than training data.

a) **Score of Polarity**:: The VADER lexicon’s  $k(x_i)$  sentiment intensity score is allocated to each word in a text  $P$  that contains the terms  $x_1, x_2, \dots, x_n$ . It is as follows:

$$K_{raw} = \sum_{i=1}^n k(x_i)$$

b) **Rule Adjustments**:: Linguistic heuristics are used to alter the raw score:

- **Negation**: Decreases the polarity or inverts it when negative phrases (e.g., “no”, “not”) present. Intensifiers: Intensify the polarity words (such as “very” and “extremely”) and increase their magnitude.
- **Punctuation/Capitalization**: Exclamation marks or ALL-CAPS increase emphasis.
- **Emojis/Slang**: Symbols such as “:)” or “lol” contribute to polarity.

After applying these rules, the adjusted sentiment score is:

$$K_{adj} = f(K_{raw}, \text{negations}, \text{intensifiers}, \text{punctuation}, \text{emojis})$$

c) **Normalization**:: Normalization of the adjusted score to the range  $[-1, +1]$ :

$$K_{norm} = \frac{K_{adj}}{\sqrt{K_{adj}^2 + \gamma}}$$

where  $\gamma$  is a normalization constant (default  $\gamma = 15$ ).

d) **Classification**:: Lastly, sentiment polarity is calculated as follows:

$$\text{Sentiment}(P) = \begin{cases} \text{Positive,} & \text{if } K_{norm} > 0.05 \\ \text{Negative,} & \text{if } K_{norm} < -0.05 \\ \text{Neutral,} & \text{otherwise} \end{cases}$$

e) **Main Concept**:: VADER is especially useful for brief, informal texts like tweets, reviews, and chat messages since it blends rule-based changes with lexicon-based scoring.

• **Hu-Liu04**

The Hu Liu (2004) Opinion Lexicon (Hu-Liu04) is a lexicon-based technique that counts the number of positive and negative terms from a predefined dictionary to identify the polarity of a text. There are roughly 4,783 negative terms and 2,006 positive words in the vocabulary.

f) **Word Counts**:: The counts of positive and negative words in a text  $P$  with the following words:

$$x_1, x_2, \dots, x_n$$

$$Pos = \sum_{g=1}^n \mathbf{1}(x_i \in L_{pos}), \quad Neg = \sum_{g=1}^n \mathbf{1}(x_i \in L_{neg})$$

in which:

The collection of positive terms in the lexicon is item  $L_{pos}$ . The collection of negative terms in the lexicon is item  $L_{neg}$ . The indicator function is item  $1(\cdot)$ , which equals 1 if the condition is true and 0 otherwise.

g) *Rule of Decision*:: By comparing the counts, the text's sentiment polarity is ascertained:

$$\text{Sentiment}(P) = \begin{cases} \text{Positive,} & \text{if } Pos > Neg \\ \text{Negative,} & \text{if } Neg > Pos \\ \text{Neutral,} & \text{if } Pos = Neg \end{cases}$$

h) *Main Concept*:: Compared to models like VADER or SCN, this rule-based approach is less accurate for informal social media writing because it ignores context, negations, and intensifiers. However, it is straightforward and computationally efficient. Sentences like "I'm not feeling bad at all," for instance, can be incorrectly categorized. It also has trouble with irony and sarcasm, where the intended meaning is not reflected in the literal words. Despite these drawbacks, its ease of use makes it a valuable baseline model and a suitable place to start when learning more complex sentiment analysis methods.

• **SCN**

For sentiment analysis, a deep learning model called the Semantic Compositional Network (SCN) records the compositional semantics of utterances. SCN dynamically learns how words interact based on context, in contrast to lexicon-based methods or fixed composition functions in conventional RNNs.

i) *Representation of Words*: Every word has a dense vector representation:

$$w_i \in \mathbb{H}^d$$

where  $d$  is the embedding dimension.

j) *Function of Dynamic Composition*: For two intermediate representations  $j_1$  and  $j_2$ , SCN composes them using:

$$j = f(W(j_1, j_2) \cdot [j_1; j_2] + b)$$

where:

- $W(j_1, j_2)$  is a *dynamic weight matrix*, predicted by a parameter prediction network conditioned on  $h_1$  and  $h_2$ ,
- $[j_1; j_2]$  denotes concatenation of vectors,
- $f(\cdot)$  is a non-linear activation function such as ReLU or tanh.

k) *Sentence Representation*:: By recursively applying the composition function, a full sentence representation is obtained:

$$j_{sentence} = \text{SCN}(w_1, w_2, \dots, w_n)$$

l) *Classification Layer*:: The final sentence representation is passed through a softmax classifier to predict sentiment:

$$\hat{y} = \text{softmax}(W_c \cdot j_{sentence} + b_c)$$

where  $\hat{y}$  is the probability distribution across sentiment classes (e.g., negative, neutral, positive), and  $W_c$  and  $b_c$  are classifier parameters.

m) *Main Concept*:: Because of its dynamic parameterization, SCN outperforms static lexicon-based models in capturing context-dependent sentiment expressions, such as handling negations like "not good."

• **GI(General Inquirer)**

One of the first lexicon-based sentiment analysis models was the General Inquirer (GI), created by Stone et al. at Harvard University in the 1960s. More than 11,000 English words are categorized into roughly 182 semantic categories by it, including sentiment-oriented categories such as *Positive Affect (PosAff)* and *Negative Affect (NegAff)*. The GI model counts the number of positive and negative words in a given text to calculate the sentiment polarity.

Let a text  $P = \{x_1, x_2, \dots, x_n\}$ , where each  $x_i$  is a word token.

$$Pos = \sum_{g=1}^n \mathbf{1}(x_i \in L_{pos}^{GI})$$

$$Neg = \sum_{g=1}^n \mathbf{1}(x_i \in L_{neg}^{GI})$$

where  $L_{pos}^{GI}$  and  $L_{neg}^{GI}$  represent the positive and negative lexicons of the GI dictionary.

Lastly the final sentiment classification is give by:

$$\text{Sentiment}(S) = \begin{cases} \text{Positive,} & \text{if } Pos > Neg \\ \text{Negative,} & \text{if } Neg > Pos \\ \text{Neutral,} & \text{if } Pos = Neg \end{cases}$$

The GI model's extensive semantic categories and historical significance as a groundbreaking computational lexicon are its main advantages. Its drawbacks, however, are its antiquated lexicon, its inability to comprehend context (such as sarcasm or negation), and its poorer performance on contemporary informal material, including social media data.

IV. CLASSIFIER PERFORMANCE EVALUATION METRICS

It is necessary to develop some performance metrics that may be utilized to assess the quality of any classifier under consideration in order to properly evaluate the performance of a model. The following measures are commonly used for evaluation.

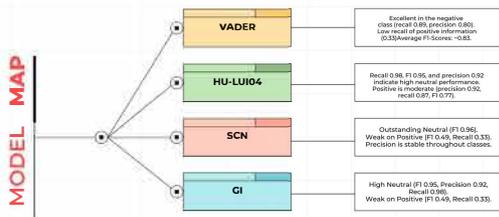


Fig. 6: Models Used for Sentimental Analysis

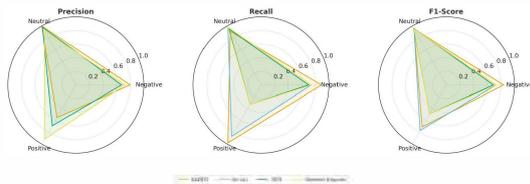


Fig. 7: Radar chart comparison of sentiment models

TABLE I: Sentiment Model Performance Comparison by Metric and Sentiment Class

Metric	Sentiment Class	VADER	Hu-Liu	SCN	General Inquirer
Precision	Negative	0.80	0.72	0.67	0.72
	Neutral	1.00	0.92	0.98	0.92
	Positive	0.55	0.92	0.69	0.92
Recall	Negative	0.89	0.74	0.70	0.74
	Neutral	0.91	0.98	0.95	0.98
	Positive	0.98	0.87	0.33	0.33
F1-Score	Negative	0.84	0.73	0.69	0.73
	Neutral	0.95	0.95	0.96	0.95
	Positive	0.71	0.77	0.49	0.49

TABLE II: Confusion Matrix Components

Component	Description
True Positive (TP)	Instances where the model correctly predicts the positive class.
False Positive (FP)	Instances where the model incorrectly predicts a positive class for actually negative data.
True Negative (TN)	Instances where the model correctly predicts the negative class.
False Negative (FN)	Instances where the model incorrectly predicts the negative class for actually positive data.

TABLE III: Classifier Performance Evaluation Metrics

Metric	Formula	Description
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Measures the overall correctness of the classifier.
Precision	$\frac{TP}{TP+FP}$	Indicates how many predicted positive cases are actually correct.
Recall (Sensitivity)	$\frac{TP}{TP+FN}$	Measures the model's ability to correctly identify positive instances.
F1-score	$\frac{2PR}{P+R}$	Harmonic mean of precision and recall.
Specificity	$\frac{TN}{TN+FP}$	Measures the model's ability to correctly identify negative instances.

## V. RESULTS

The exported group chat was successfully analyzed by the WhatsApp Chat Evaluation system, which produced information on user behavior, engagement trends, and content distribution. According to the data, the most active participant contributed 5876 messages, and the most active date was September 19, 2024, with 252 messages. In order to distinguish between active and inactive users, the system also generated a ranked list of user activity based on message counts. Common emojis that indicate the emotional tone of discussions were identified using emoji analysis. The word "happy" occurred most frequently (492 times) in the word frequency analysis, and a word cloud was used to illustrate this. Additionally, messages were classified as favorable, neutral, or negative using sentiment analysis with the VADER model. In total, 79 participants were captured by the system, which produced comprehensive data on their communication patterns and message activity.

### A. Chat Statistics

The chat dataset is quantitatively analyzed by the Chat Statistics module. Key parameters including the overall number of participants, the total quantity of messages exchanged, and the distribution of messages among users are summarized. By ranking users according to the messages they provide, it is possible to determine which members of the group are the most and least active as well as to comprehend patterns of participation. In order to identify communication patterns and times of high activity, the module also maps messages across days, weeks, and months and examines message length and activity across time. These figures show differences in participation levels and communication strategies. In general, the module transforms unstructured conversation data into structured insights that facilitate additional research, including sentiment analysis and emoji usage trends.



Fig. 8: Monthly Messages



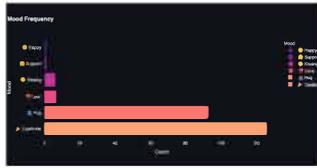


Figure 2: Mood Frequency

## VI. CONCLUSION

The WhatsApp Chat Evaluation system shows how unstructured conversational data may be turned into structured insights that emphasize communication patterns, emotional tone, and user involvement. The system finds important indicators including active users, conversation peaks, word frequency, sentiment distribution, and emoji usage using data preprocessing, feature extraction, sentiment analysis (VADER), and visualization approaches. Nevertheless, there are limitations to the study. Context, sarcasm, and multilingual expressions may not be fully captured by the VADER model, and the analysis only looks at exported text data—voice notes, pictures, and videos are not included. Future research can incorporate multimodal analysis, real-time visualization, and transformer-based models like BERT or RoBERTa to increase accuracy. All things considered, the system offers a solid basis for applications including natural language processing, social media analytics, and communication research.

### A. Future Scope

Advanced NLP models like BERT and RoBERTa, which offer more context-aware sentiment analysis than rule-based models like VADER, can be incorporated into the WhatsApp Chat Evaluation system to further enhance it. In order to better understand user interactions, future study may potentially incorporate multimodal analysis by looking at photographs, videos, audio communications, and emoticons. The creation of real-time monitoring technologies to quickly identify sentiment shifts and conversation trends is another extension. Cyberbullying, harassment, and poisonous language may also be detected by incorporating abuse detection algorithms (such as Detoxify or HateBERT). The system's accuracy, scalability, and use for monitoring, research, and digital communication analysis would all be improved by these changes.

## VII. RELATED WORK

Due to the abundance of publicly accessible data, research on digital communication in social media platforms has been extensively undertaken, particularly on sites like

Facebook, Instagram, and Twitter. Using methods like text mining, topic modeling, and data visualization, researchers can utilize this data to conduct sentiment analysis, trend analysis, and public opinion surveys in order to find patterns, themes, and user interactions.

However, despite being widely used, instant messaging apps like WhatsApp have gotten comparatively less academic attention than these platforms. More research is required to comprehend communication patterns and interaction trends inside WhatsApp, given the platform's significant significance in daily digital communication.

### A. Open Source Libraries and Tools

Community-developed tools for analyzing exported WhatsApp chat files include chatistics [1], whatstk [2], and chatlyzer [3]. These programs help users learn from their conversation histories by offering analytics pipelines, parsing tools, and visualizations. They extract important data, like message timestamps, sender information, message content, and media indicators (pictures, videos, or files), using text parsing algorithms. Because of these features, they might be a good place to start when developing dependable data intake pipelines and enabling more in-depth examination of communication patterns and trends in WhatsApp discussions.

### B. Tutorials And Community Dashboards

Users can upload exported WhatsApp chat files for examination on a number of community-developed dashboards, which are frequently constructed with Streamlit [4]–[7]. Word clouds, user activity rankings, emoji frequency, and activity timelines are just a few of the statistics and visuals that these systems produce after processing the data. By using local or on-device processing, many of these apps also adhere to privacy-first design principles.

### C. Studies Conducted By Scholars And Practitioners

Methodological issues in WhatsApp analysis are examined in academic research: Sentiment analysis: Research has contrasted deep learning techniques with lexicon-based strategies like TextBlob and VADER [8], [9]. Slang, code-mixed text, and excessive use of emojis are examples of domain-specific problems that are emphasized. Topic modeling: To identify latent topics in WhatsApp group discussions, methods like LDA and BERTopic have been used [9]. Analysis of networks and interactions: Studies have used social network analysis to examine sender-recipient patterns, focusing on user roles, reciprocity, and centrality [10].

### VIII. ACKNOWLEDGMENT

The authors express gratitude to their supervisors and peers for their guidance and recommendations during the design and implementation of the WhatsApp Chat Evaluation system. We would also like to express our gratitude to the open-source developer community for contributing tools and libraries like the Python visualization frameworks, VADER, and NLTK, which were essential to the analytic process. Without their ongoing donations and efforts, this work would not have been possible.

### REFERENCES

- [1] A. Esuli and F. Sebastiani, "SentiWordNet: A publicly available lexical resource for opinion mining," in *Proc. LREC*, pp. 417–422, 2006.
- [2] J. W. Pennebaker, M. E. Francis, and R. J. Booth, *Linguistic Inquiry and Word Count (LIWC): LIWC2001*, Erlbaum, 2001.
- [3] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL*, pp. 4171–4186, 2019.
- [4] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. NeurIPS*, pp. 5754–5764, 2019.
- [5] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2017.
- [6] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.
- [7] B. Liu, "Sentiment analysis and opinion mining: Recent developments," in *Handbook of Natural Language Processing*, 3rd ed., CRC Press, pp. 133–162, 2020.
- [8] "General Inquirer Categories," Harvard University. Available: <http://www.wjh.harvard.edu/~inquirer/>, Accessed: Sep. 6, 2025.
- [9] "Sentimental Analysis Using VADER," *Towards Data Science*, Available: <https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664>.
- [10] "Build Web App Instantly for Machine Learning using Streamlit," *Analytics Vidhya*, Available: <https://www.analyticsvidhya.com/blog/2021/06/build-web-app-instantly-for-machine-learning-using-streamlit/>.
- [11] M. Church and S. de Oliveira, "What's up with WhatsApp? Comparing mobile instant messaging behaviors with traditional SMS," in *Proc. 15th Int. Conf. Human-Computer Interaction with Mobile Devices*, pp. 352–361, 2013.
- [12] Marada Pallavi, Meesala Nirmala, Modugaparapu Sravani, Mohammad Shameem, "WhatsApp Chat Analysis," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 04, issue 05, May 2022.
- [13] Shaikh Mohd Saqib, "Whatsapp Chat Analyzer," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 04, issue 05, May 2022.
- [14] K. Ravishankara, Dhanush, Vaisakh, S. Srajan, "Whatsapp Chat Analyzer," *International Journal of Engineering Research and*, vol. 9, 2020, doi: 10.17577/IJERTV9IS050676.
- [15] D. Radha, R. Jayaparvathy, D. Yamini, "Analysis on Social Media Addiction using Data Mining Technique," *International Journal of Computer Applications*, 0975–8887.
- [16] S. Patil, "WhatsApp Group Data Analysis with R," *International Journal of Computer Applications*, vol. 154, no. 4, p. 0975–8887, Nov. 2016.
- [17] "Number of monthly active WhatsApp users worldwide from April 2013 to February 2016 (in millions)," Available: <http://www.statista.com/statistics/260819/numberofmonthly-active-Watsapp-users>.
- [18] I. Ahmed and T. Fiaz, "Mobile phone to youngsters: Necessity or addiction," *African Journal of Business Management*, vol. 5, no. 32, pp. 12512–12519, 2011.
- [19] J. Yeboah and G. D. Ewur, "The Impact of WhatsApp Messenger Usage on Students," *Journal of Education and Practice*, vol. 5, no. 6, pp. 157–164, 2014.
- [20] M. N. K. Boulos, D. M. Giustini, and S. Wheeler, "Instagram and WhatsApp in Health and Healthcare: An Overview," *Future Internet Creative Common Attribution MDPI*, vol. 8, no. 37, pp. 1–14, 2016.
- [21] N. Aharony and T. G., "The Importance of the WhatsApp Family Group: An Exploratory Analysis," *Aslib Journal of Information Management*, vol. 68, no. 2, pp. 1–37, 2016.
- [22] D. Radha, R. Jayaparvathy, and D. Yamini, "Analysis on Social Media Addiction using Data Mining Technique," *International Journal of Computer Applications*, vol. 139, no. 7, pp. 23–26, Apr. 2016.
- [23] C. Montag, K. Blaszkiewicz, R. Sariyska, B. Lachmann, I. Andone, B. Trendafilov, M. Eibes, and A. Markowetz, "Smartphone usage in the 21st century: who is active on WhatsApp?," 4 Aug. 2015. [Online]. Available: <https://bmcresearchnotes.biomedcentral.com/articles/10.1186/s13104-015-1280-z>. [Accessed: 12 Mar. 2019].
- [24] W. Bani, "WhatsApp," 22 Apr. 2019. [Online]. Available: <https://en.welibani.org/weli/WhatsApp>. [Accessed: 24 Apr. 2019].
- [25] M. Kolhar, R. Al-Turjman, and A. S. Alameen, "Analyzing the impact of social media on user sentiment during COVID-19 pandemic," *IEEE Access*, vol. 9, pp. 143570–143578, 2021.
- [26] Meng Cai, "PubMed Central," PMID: PMC7944036, PMID: 33732917.
- [27] E. Larson, "Automatic Checking of Regular Expressions," 2018 IEEE 18th International Working Conference on Source Code Analysis and Manipulation (SCAM), 2018, pp. 225–234, doi: 10.1109/SCAM.2018.00034.