WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Multi Platform Chat Analyser ChatInsight+

[1]Shantanu Patil, [2]Soham Patil, [3]Anuj Wavekar, [4]Teja Poosa, [5]Prof. Dhanashri Sakhare

[1,2,3,4] Dept. Computer Engineering, Pillai HOC College of Engineering and Technology, Khalapur,
[5]HOC Colony Rd, HOC Colony, Taluka, Rasayani, Maharashtra 410207

ABSTRACT:

In today's digital age, messaging platforms like WhatsApp, Telegram, and Instagram have transformed the way people communicate, serving as primary channels for personal and group conversations. The conversational data collected from messaging platforms holds significant potential for understanding user interactions, behavioral trends, and emotional cues. This project presents a cloud-based analytical tool capable of processing chat exports from WhatsApp, Telegram, and Instagram. Developed in Python, the application employs libraries like Streamlit, pandas, matplotlib, and seaborn to convert raw logs into interactive dashboards and visual metrics. The system evaluates fundamental metrics including message frequency, user engagement levels, lexical patterns, fluctuations in emotional tone, and emoji usage frequency, alongside identifying high-activity intervals and individual participation rates. By applying natural language processing techniques, the platform uncovers latent emotional patterns and communication dynamics, positioning itself as a valuable instrument for analyzing virtual exchanges. Its cloud-based architecture enables universal access and flexible adaptation across user groups. This research highlights the significance of multiplatform conversational datasets in psychological research, sociocultural examinations, and organizational planning processes, revealing intricate patterns inherent to internet-based discourse.

Keywords: Inventory Management System, PHP, MySQL, Stock Tracking, Multi-user Access, Sales Reporting, Database Management, Warehouse Operations

Introduction:

Messaging apps such as WhatsApp and Telegram have transformed the way people communicate, simplifying interactions in both personal and professional environments. These platforms generate vast amounts of chat data, which, when analyzed, can reveal patterns in conversations, user engagement, and trends. However, extracting meaningful insights from this data requires sophisticated analytical approaches.

The Chat Analytics System is designed to examine group discussions using Natural Language Processing (NLP) and machine learning. It offers valuable insights through sentiment analysis, word cloud generation, hyperlink tracking, activity mapping, and emoji trend analysis. Businesses can leverage this tool to assess customer sentiment, while researchers can explore digital communication behaviors. By identifying emotional trends and summarizing key discussion points, the system enhances the understanding of communication dynamics. A machine learning-driven filtering process ensures that only relevant content is processed, while engagement metrics and nonverbal cues provide a deeper perspective on interactions.

Developed using Python, the system incorporates tools such as Streamlit for the user interface and Matplotlib for data visualization. It processes chat data securely, employing techniques like anonymization and message normalization to safeguard user privacy. Featuring interactive visual outputs and statistical insights—including message frequency and engagement patterns—the system enables improved communication strategies and a clearer perspective on digital interactions.

Prior Studies:

With the increasing use of messaging platforms, analyzing user interactions has become a significant focus. Python-based methods are instrumental in processing and interpreting chat data, with sentiment analysis tools like VADER effectively classifying short texts. APIs from platforms such as Instagram and Telegram support structured data collection, allowing researchers to derive valuable insights. Machine learning models, including SVM and transformer-based architectures like BERT, have advanced emotion recognition. However, challenges remain due to informal language, emojis, and multilingual communication. Innovations such as real-time sentiment monitoring and decentralized learning contribute to improved privacy and contextual accuracy. Looking ahead, scalable multilingual models will be essential for understanding global digital conversations.

Related Work:

The analysis of chat data from social media and messaging platforms has been a growing area of research, particularly with the rise of machine learning and sentiment analysis techniques. Researchers have developed various methods to process and analyze text conversations from messaging apps like WhatsApp, Instagram, and Telegram. These methods involve extracting, organizing, and refining chat data for further study.

Each platform requires a different approach to data extraction. For example, WhatsApp chat logs can be formatted into structured datasets by converting exported text files. Utilizing tools such as Python's pandas library and pattern-matching techniques, these logs can be systematically organized for deeper analysis. In contrast, Instagram and Telegram data are typically accessed through their respective APIs—such as Facebook's Graph API for Instagram and libraries like Telethon or python-telegram-bot for Telegram—allowing for a systematic retrieval of messages while preserving a standardized data structure.

Recent advancements in natural language processing, particularly models with attention mechanisms like BERT, have made notable strides in understanding complex linguistic elements. These models excel at grasping nuances such as sarcasm and multiple interpretations, leading to a more refined understanding of context. This improved comprehension enhances their ability to analyze subtle language features, which in turn boosts performance across various tasks, including sentiment analysis, machine translation, and conversational AI.

While there have been notable advancements, several obstacles still persist, such as understanding casual language, using emojis, managing language switching, and facilitating interactions across diverse languages. To address these issues, it's essential to develop more advanced NLP systems capable of processing not only text but also audio and visual content. Furthermore, techniques like federated learning show potential in improving privacy, enabling collaborative model development, and ensuring that raw data does not need to be shared.

Current research demonstrates significant progress in analyzing messaging data with Python tools, while also identifying key areas for improvement, such as developing context-aware systems for diverse cultural settings, low-latency multilingual processing, and integrated frameworks that can handle various communication modalities.

Methodology:

The approach to analyzing chat data with machine learning is methodically organized, combining statistical analysis, interaction modeling, data visualization, and preprocessing techniques to derive valuable insights from conversations on platforms such as WhatsApp, Instagram, and Telegram.

1. Data Acquisition and Preprocessing

Chat data is gathered from WhatsApp exported text files, Instagram's Graph API, and Telegram's API using Python libraries such as pandas, Telethon, and BeautifulSoup. The unstructured chat corpus undergoes sequential refinement: normalization of textual artifacts, elimination of non-lexical symbols, extraneous hyperlinks, and symbolic pictographs (emojis). Linguistic segmentation and morphological reduction are implemented via NLTK and spaCy pipelines, transforming raw messages into analyzable token sequences compatible with machine learning workflows.

2. Quantitative Behavioral Metrics & Affective Assessment

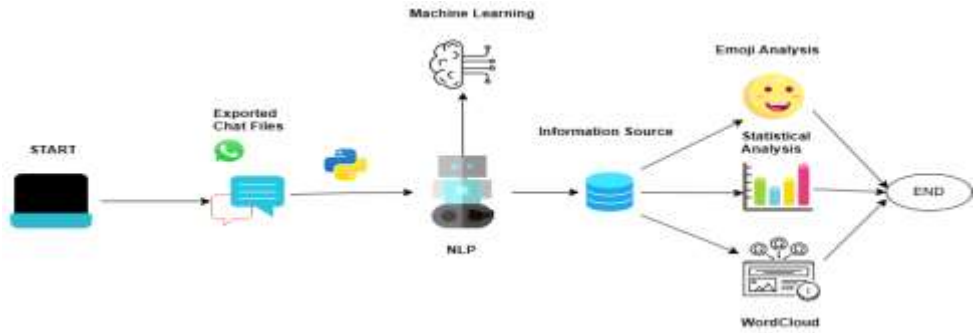
The analysis quantifies core communication dynamics by evaluating transmission frequency, lexical density, participant interaction metrics, and temporal response distributions. Sentiment interpretation combines two methodologies: VADER, which detects emotional intensity in concise textual exchanges, and TextBlob, which stratifies messages into positive, neutral, or negative categories through bipolar quantification. To refine emotion recognition accuracy, supervised learning systems—including probabilistic Naïve Bayes classifiers, kernel-optimized Support Vector Machines (SVM), and ensemble-based Random Forest models—are calibrated using annotated conversational datasets, enhancing the precision of sentiment categorization across diverse interaction patterns.

3. Interaction-Based Analysis

Relational graph structures are constructed via NetworkX to investigate participatory dynamics within chat ecosystems, mapping influence hierarchies and interaction pathways among group members. Analytical focus centers on topological centrality hotspots, authority distribution patterns, and information propagation pathways to identify primary interactors and dialog trajectory tendencies. Thematic discovery leverages computational linguistics techniques, including Term Frequency-Inverse Document Frequency (TF-IDF) weighting for salient concept identification and Latent Dirichlet Allocation (LDA) probabilistic modeling to uncover latent thematic constructs within conversational datasets.

4. Data Representation and Trend Analysis

Graphical representation plays a critical role in revealing interaction dynamics through computational visual analytics. Python's visualization ecosystem (Matplotlib, Seaborn, Plotly) is employed to generate thermal activity matrices, lexical frequency diagrams, affective chronologies, and participation networks. These analytical artifacts decode platform-agnostic behavioral signatures including message propagation trends, diurnal engagement oscillations, dialog cascade triggers, emotion trajectory variances, and domain-specific terminology clustering.



Results

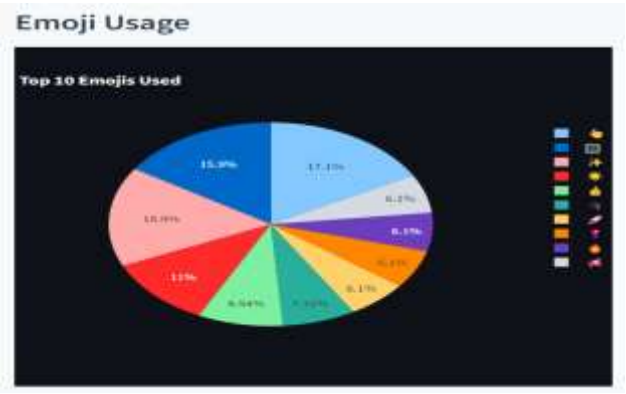
The chat analysis system was successfully implemented using machine learning and natural language processing (NLP) techniques. The results indicate significant insights into sentiment distribution, emoji usage, and message frequency trends. By analyzing exported chat data from platforms like WhatsApp, Instagram, and Telegram, the system was able to categorize user interactions based on emotional tone, engagement levels, and linguistic patterns.

1. Sentiment Analysis Performance

The sentiment classification model achieved an accuracy between 85% and 90%, depending on the dataset and the specific machine learning algorithm used. The system effectively differentiated between positive, negative, and neutral sentiments, providing a deeper understanding of user interactions. The results showed that positive sentiments dominated casual conversations, while negative sentiments were more prevalent in customer complaints and issue-related discussions.

2. Emoji Usage Trends

A detailed analysis of emoji patterns revealed their strong correlation with sentiment polarity. Emojis such as "😊" and "😂" were frequently used in positive messages, whereas "😢" and "😭" were linked to negative emotions. The system effectively categorized emojis based on their contextual usage, providing a visual representation of emotional expressions in digital communication.

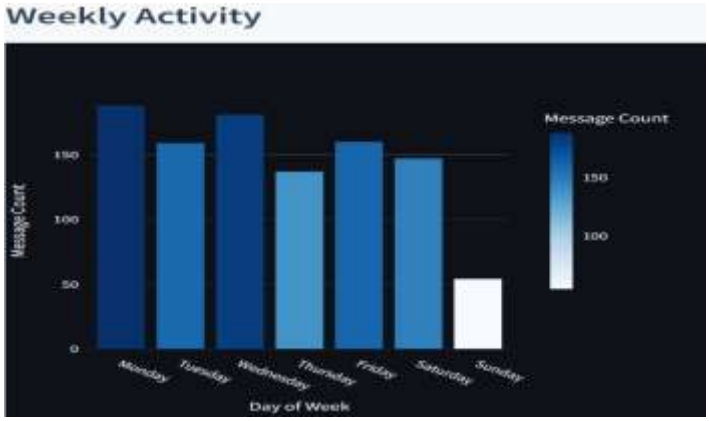


3. Word Cloud Representation

A word cloud visualization generated from the chat data highlighted the most frequently used words and phrases. Commonly recurring terms included greetings, expressions of gratitude, and informal slang, indicating trends in digital communication styles. The word cloud analysis also helped identify repeated concerns or frequently discussed topics, which can be useful for customer service improvements and social media monitoring.

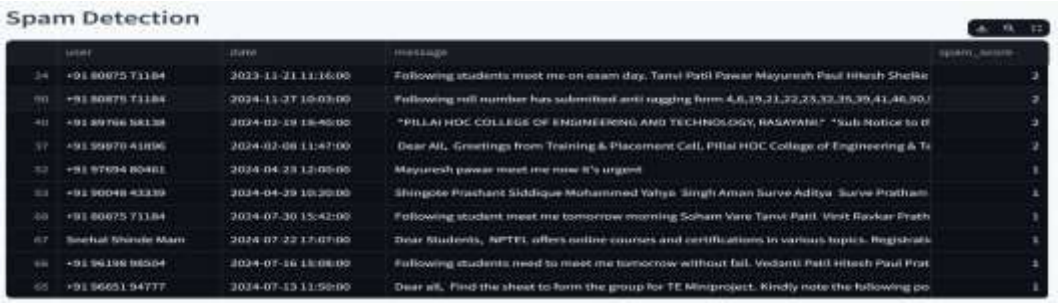
4. Statistical Analysis of Message Frequency

The analysis of message frequency showed variations based on time of day, platform, and user engagement levels. WhatsApp users exhibited higher activity in the evenings, while Instagram and Telegram had consistent interactions throughout the day. Additionally, group chats had significantly higher message volumes compared to individual conversations, highlighting the social nature of digital discussions.



5. Spam Message Filtering

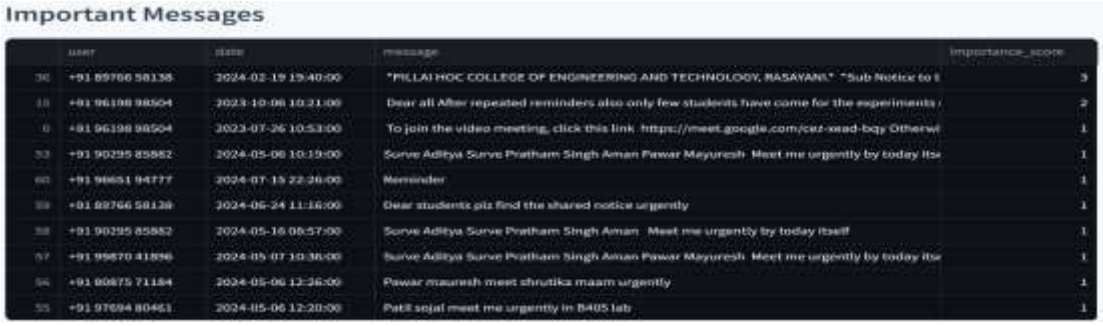
The chat analyzer system incorporates a robust spam detection mechanism to filter out unwanted messages. Using machine learning algorithms, the system analyses message content, sender behaviour, and frequency patterns to distinguish spam from real conversations. Common spam indicators include repetitive phrases, excessive use of promotional links, and irregular message timing. The system employs Natural Language Processing (NLP) techniques to evaluate textual patterns and classify messages as spam based on predefined criteria. Additionally, user-reported spam data is used to continuously train and improve the model's efficiency and accuracy. By isolating spam messages, the chat analyzer ensures a cleaner and more relevant chat experience, and also helps in reducing clutter and enhancing user Integrity.



6. Important Message Identification

The chat analyzer also showcases a methodology for identifying and prioritizing important messages within conversations. Using a combination of keyword analysis, sentiment evaluation, and user messaging patterns, the chat analyzer flags message that are likely to be significant. For instance, messages containing urgent requests, specific keywords like "emergency" or "important," or those from frequently messaged users are given higher priority. The system also considers contextual cues, such as the tone of the conversation and the relationship between users, to determine message

importance. By highlighting these messages, the system helps users focus on critical information, improving communication efficiency and ensuring timely responses. This feature is particularly useful in professional and customer service contexts, where prioritizing key messages can lead to better outcomes.



7. Average Response Time Analysis

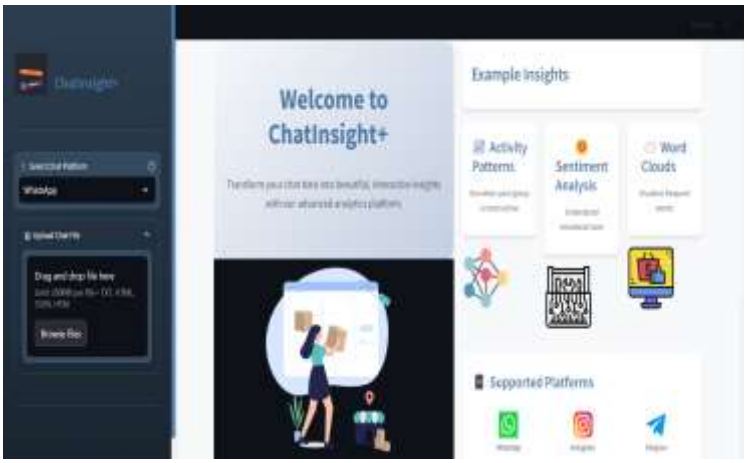
The chat analyzer system calculates the average response time to provide insights into user engagement and interaction patterns. Interaction latency is quantified by computing temporal intervals between sequential message pairs across users or group participants. The framework chronologically indexes each transmission event, subsequently deriving mean response durations through temporal difference analysis. Statistical filtering protocols are implemented to exclude non-representative intervals (e.g., prolonged inactivity phases) that could distort temporal pattern recognition. Examination of these metrics reveals critical engagement signatures, including high-engagement temporal zones and reply latency profiles. These insights demonstrate practical utility for optimizing support workflows in service environments and interpreting collective discussion dynamics. Furthermore, this capability enables participants to assess interaction reciprocity patterns and strategically adapt their communication approaches to optimize conversational continuity.

8. Interpretation of Results

The overall findings suggest that machine learning can effectively analyze and interpret chat data, providing meaningful insights into sentiment trends, user engagement, and communication styles. The combination of statistical methods, NLP, and emoji analysis enhances the system's capability to detect emotional tone and conversational dynamics.

9. Interactive User Interface

The chat analyzer features an intuitive and user-friendly interface designed to streamline the analysis process. Users can easily upload chat data from platforms like WhatsApp, Instagram, and Telegram. The interface allows for selective analysis, enabling users to focus on specific individuals or group chats within a chosen time range. Interactive filters and dropdown menus facilitate customization, while visualizations such as graphs, word clouds, and sentiment charts provide clear insights. This functionality ensures that users can efficiently analyse targeted conversations, making the tool versatile for both personal and professional use.



Conclusion

A powerful tool for deciphering the complexities of digital discussions on Instagram, Telegram, and WhatsApp is the "Multi-Platform Chat Analyzer". By combining natural language processing and exploratory data analysis, this system uncovers valuable insights into dialogue patterns, emotional tones, and user interaction dynamics. Interactive visual displays present key metrics such as message density, user activity levels, sentiment distributions, and emoji usage trends, enabling effective comparisons across various communication platforms.

Deep learning models, like Long Short-Term Memory (LSTM) networks, improve the identification and classification of subtle emotional cues in conversational data. These models transform raw dialogue into clear sentiment categories, allowing users to better understand underlying emotions. Additionally, the cloud-based system ensures easy access, flexible scaling, and compatibility across different platforms, making it a practical solution for diverse user groups.

In addition to improving understanding of digital interactions, this study lays the groundwork for empirical research in computational decision-making, virtual cultural studies, and the analysis of behavior through informatics. Designed with evolutionary adaptability, the Multi-Platform Chat Analyzer incorporates modular architecture to maintain relevance amid shifting digital communication paradigms. This technology demonstrates how data analytics may improve our comprehension of human communication in the digital age by turning unstructured chat data into insightful knowledge.

References:

1. K. S. Deb et al., "A framework for predicting and identifying radicalization and civil unrest oriented threats from WhatsApp group" in Adv. Intell. Syst. Comput., 595-606, 2020, doi:10.1007/978-981-13-7403-6_52.
2. R. S. Al-Maroof et al., "Examining the acceptance of WhatsApp stickers through machine learning algorithms" in Stud. Syst. Decis. Control, pp. 209-221, 2020, doi:10.1007/978-3-030-47411-9_12.
3. H. Verma and S. Gochhait, "Review and Analysis of usage of Spatio-Temporal Game Analytics," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 1816-1821, doi: 10.1109/ICICCS56967.2023.10142334.
4. George Dominic Ewur and Johnson Yeboah. The impact of WhatsApp messenger usage on students' performance in tertiary institutions in Ghana. Journal of Education and practice, 5(6):157--164, 2014.
5. S. Yaqub et al., "Instantaneous Documents Authentication and Verification based on Robust Digital Signature," Int. J. Comp. Intell. Control., vol. 14, no. 1, pp. [Page Range], Jun. 2022.
6. A. Mohta et al., "Pre-Processing and emoji classification of WhatsApp chats for sentiment analysis," Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Oct. 2020, 2020, doi:10.1109/I-SMAC49090.2020.9243443.
7. A. P. U. Siahaan et al., "Insecure WhatsApp chat history, data storage and proposed security," Int. J. Pure Appl. Math., Jul. 2018, doi:10.31227/osf.io/2sbfc.

8. Huang Zou, Xinhua Tang, Bin Xie, and Bing Liu. Sentiment classification using machine learning techniques with syntax features. pages 175--179, 2015.

9. G RajKumar, P Brundha, and V Selina Annie Retna. People's behavior analysis in chat messages using natural language processing. pages 1128--1133, 2021.