ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/04.2024-78841519/IJAEAST.2024.01.0003, January 2024, Volume-2, Issue-1, pp.21-30
International Journal of Advances in Engineering Architecture Science and Technology

Pragmatic Analysis of WhatsApp Chats Using NLP

Pranav Chavan[1*], Harshraj Deshmukh[2], Aakash Dhotre[3], Aditya Gharat[4], Sanjay Waykar[5]

[1,2,3,4] Student of Department of Information Technology, MGM College of Engineering and Technology, Navi Mumbai, Maharashtra, India.

[5]Faculty of Department of Information Technology, MGM College of Engineering and Technology, Navi Mumbai, Maharashtra, India.

[1]pranavchavan815@gmail.com, [2]harshrajdeshmukh777@gmail.com, [3]aakashdhotre12@gmail.com, [4]gaditya7506@gmail.com

Received: 10-01-2024 **Accepted:** 24-01-2024 **Published:** 28-01-2024

Abstract

Background: WhatsApp, an extensively espoused medium for communication, has come a fat source of different exchanges in the digital age. This design, named" realistic Analysis of WhatsApp Chats Using NLP," delves into data analysis and Natural Language Processing (NLP) to prize precious perceptivity.
Objectives: It harnesses Python libraries and NLP ways to unleash idle knowledge in WhatsApp exchanges, transcending conversational boundaries. WhatsApp exchanges contain a wealth of data, gauging, professional, and social conversations, essential for Machine literacy models. This tool employs popular Python modules like pandas, matplotlib, seaborn, NumPy, and sentiment analysis, enabling the creation of data frames and graphical representations. These visualizations, paired with sentiment analysis, unveil sentiments, trends, and patterns.
Methods: This study presents a pragmatic analysis of WhatsApp chat data using Natural Language Processing (NLP) techniques. The research encompasses three primary modules: preprocessing, statistical analysis, and visualization.
Statistical Analysis: In the preprocessing phase, the raw chat data is parsed and organized into a structured format, extracting crucial information such as user, message content, and timestamp. Subsequently, statistical metrics including message count, word frequency, and user activity patterns are computed, offering insights into communication dynamics.
Findings: Finally, various visualizations such as word clouds, emoji usage charts, and activity heatmaps are generated to provide intuitive representations of the chat data.
Applications and Improvements: Through this interdisciplinary approach, the study aims to elucidate nuanced aspects of communication behaviour within WhatsApp chats, facilitating a deeper understanding of linguistic pragmatics in digital discourse.

Keywords: Natural Language Processing (NLP), Timestamp, Word clouds, Visualization, Preprocessing and Statistical analysis.

1. Introduction

This tool is grounded on data analysis and processing. The first step in enforcing a machine learning algorithm is to understand the right literacy experience from which the model starts

perfecting. Data preprocessing plays a crucial part when it comes to machine literacy. To make the model more effective we need lots of data, so we turned our focus primarily on one of the large- scale data directors possessed by Facebook which is nothing but WhatsApp. WhatsApp claims that 55 billion dispatches are transferred each day. The average stoner spends 195 twinkles per week on WhatsApp and is a member of plenitude of groups. With this treasure house of data right under our very tips, it is imperative that we embark on a charge to gain perceptivity on the dispatches which our phones are forced to bear substantiation to. A list that uses pie maps and plates to represent the intriguing data that it collects after analysing your WhatsApp exchanges. You know the drill by now. You will take a backup of your converse and shoot it to a dispatch id listed on the point. WhatsApp claims that 55 billion dispatches are transferred each day. The average stoner spends 195 twinkles per week on WhatsApp and is a member of plenitude of groups with this treasure house of data right under our very tips, it is imperative that we embark on a charge to gain perceptivity on the dispatches which our phones are forced to bear substantiation to. A list that uses pie maps and plates to represent the intriguing data that it collects after analysing your WhatsApp exchanges. You know the drill by now You'll take a backup of your converse and shoot it to a dispatch id listed on the point.

Problem Statement

Despite the prevalence of WhatsApp as a primary mode of communication, there exists a gap in understanding the nuanced aspects of digital discourse and communication behaviour within WhatsApp chats. The lack of comprehensive tools and methodologies for pragmatic analysis inhibits insights into linguistic pragmatics and social interaction dynamics embedded within these conversations. Current approaches often rely on manual inspection or lack the scalability to handle large volumes of chat data effectively. Additionally, there is a need for interdisciplinary research integrating Natural Language Processing (NLP) techniques and visualization methods to unravel the complexities of digital discourse and communication patterns. Thus, the problem statement revolves around the development of an automated system for pragmatic analysis of WhatsApp chats using NLP, aiming to address the gaps and provide valuable insights into communication dynamics, linguistic nuances, and social interactions within WhatsApp communities.

Objective

The objective of this study is to develop an automated system for pragmatic analysis of WhatsApp chats using Natural Language Processing (NLP) techniques. The primary aim is to extract meaningful insights from raw chat data by preprocessing it to obtain crucial information such as user identities, message content, and timestamps. Subsequently, NLP techniques will be applied to conduct statistical analysis, including metrics such as message counts, word frequencies, and user activity patterns. Moreover, visualization methods such as word clouds, emoji usage charts, and activity heatmaps will be implemented to provide intuitive representations of the analyzed data, thereby facilitating a deeper understanding of linguistic pragmatics and communication dynamics within WhatsApp communities. By integrating NLP methodologies and visualization techniques, the study seeks to unravel the complexities of digital discourse and communication behaviour, aiming to offer valuable insights applicable across various domains, from sociolinguistics to digital marketing strategies.

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/04.2024-78841519/IJAEAST.2024.01.0003, January 2024, Volume-2, Issue-1, pp.21-30
International Journal of Advances in Engineering Architecture Science and Technology

2. Literature Review

Existing System

There is a lot of development in the current system. In the aged interpretation there was no point to display status, there was no point to partake documents and there was no point to partake position. In the current interpretation, all these features are available. In aged interpretation we could not partake images through croaker ’s format. In this system stoner can pierce WhatsApp in windows through WhatsApp web operation, which can be connected through QR law. There is another point called import converse where stoner can shoot or partake or get the converse detail for data analysis through dispatch, Facebook, or some runner operation. The being system provides WhatsApp druggies with an introductory set of features for communication. In the aged interpretation, it demanded several functionalities that have ago been addressed and bettered in the current interpretation. Some notable aspects of the being system include

- Status Feature: The aged WhatsApp interpretation demanded a stoner status display point, limiting druggies' capability to express themselves. The current interpretation introduced a status point, enabling druggies to partake textbook, prints, vids, and GIFs as their status.
- Document participating: In the aged interpretation, participating documents within exchanges was inconvenient. The current interpretation allows druggies to partake colourful document formats, making it easier to change lines.
- Position participating the aged interpretation did not support position sharing. The current system enables real- time position sharing for meetups and environment.
- Image participating in Doc Format participating images in croaker format was insolvable in the aged interpretation. The current system now allows image sharing in croaker format.
- WhatsApp Web operation druggies can pierce WhatsApp on Windows via a web operation connected through a QR law.
- Export Chat for Data Analysis The system introduced" Export Chat," enabling druggies to shoot or gain converse details for analysis via dispatch or other messaging apps.

These advancements have enhanced the stoner experience in the current system, making WhatsApp a more protean and stoner-friendly platform for communication and data analysis.

Feasibility Study

The main objective of the feasibility study is to treat the technical operational and economic feasibility of developing the application. Feasibility is the determination of whether project is worth doing. The process followed in making this determination is called feasibility study. All systems are doable, given unlimited coffers and horizonless time. The feasibility study to be conducted for this design involves:

- Technical Feasibility
- Operational Feasibility
- Economic Feasibility

Technical Feasibility

The specialized feasibility study reports whether there exists correct needed coffers and technologies which will be used for design development. It is the measure of the specific specialized result and the vacuity of the specialized coffers and moxie. In our design we will be

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/04.2024-78841519/IJAEAST.2024.01.0003, January 2024, Volume-2, Issue-1, pp.21-30
International Journal of Advances in Engineering Architecture Science and Technology

using Jupyter tablet (web- grounded operation) and VS law (textbook editor), both are open-source software's. Along with these colourful python libraries and will be used.

Economic Feasibility

Cost and benefit of the project is analysed in economic feasibility, which means what will be the cost of final development of the product. This project has not cost in development since all the software and technologies used are open source. This project is not economical as it depends on the analysis of data between two more devices (phones).

Operational Feasibility

It is to determine whether the system will be used after the development and perpetration. In functional Feasibility degree of furnishing service to conditions is anatomized. This involves the study of application and performance of the product. Our design shows the whole analysis of the exchanges among people. It can be two people or a group of people and provides colourful information using maps in fluently readable format.

3. Proposed System

The proposed system aims to conduct a pragmatic analysis of WhatsApp chats using Natural Language Processing (NLP) techniques. The system comprises several key components: preprocessing, statistical analysis, and visualization. In the preprocessing stage, raw chat data is processed to extract essential information such as user identities, message content, and timestamps. Subsequently, statistical analysis techniques are applied to derive insights into communication dynamics, including metrics such as message counts, word frequencies, and user activity patterns. Finally, the system employs various visualization methods such as word clouds, emoji usage charts, and activity heatmaps to provide intuitive representations of the analyzed data. Through the integration of NLP methodologies and visualization techniques, the proposed system aims to offer a comprehensive understanding of linguistic pragmatics within WhatsApp conversations, facilitating insights into digital discourse and communication behaviour.

4. System Design

Algorithm

The algorithm for pragmatic analysis of WhatsApp chats using Natural Language Processing (NLP) involves several sequential steps. Firstly, the raw chat data is pre-processed to parse and organize it into a structured format, extracting pertinent information such as user identities, message content, and timestamps. This preprocessing step also includes tokenization and removal of stop words to prepare the text for analysis. Following preprocessing, statistical analysis is conducted to derive insights into communication dynamics. This includes calculating metrics such as message counts, word frequencies, and user activity patterns. Additionally, specialized analyses may be performed to identify familiar words, top emojis, busiest days or months, and other relevant statistics. Finally, visualization techniques are applied to represent the analyzed data in an intuitive manner. This involves generating word clouds, emoji usage charts, activity heatmaps, and other visualizations to provide a comprehensive understanding of linguistic pragmatics within the WhatsApp conversations. Through the integration of NLP

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/04.2024-78841519/IJAEAST.2024.01.0003, January 2024, Volume-2, Issue-1, pp.21-30
International Journal of Advances in Engineering Architecture Science and Technology

techniques and visualization methods, the algorithm facilitates a pragmatic analysis that aids in elucidating digital discourse and communication behaviour within WhatsApp chats.

Proposed System Architecture

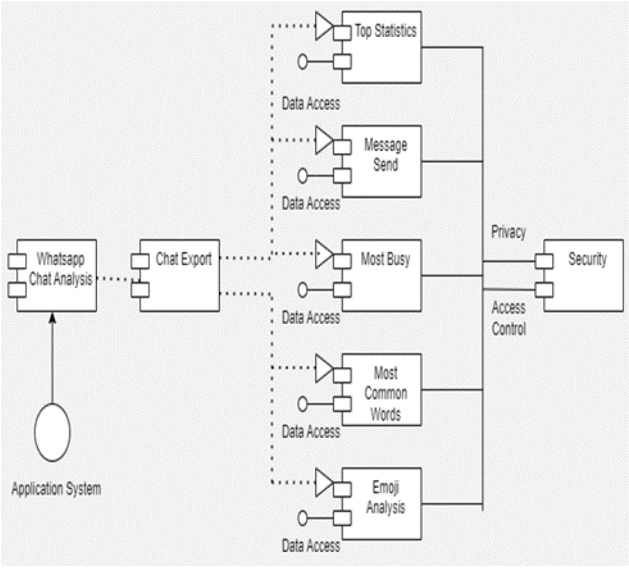

Figure 1. Architecture

Data Flow Diagram

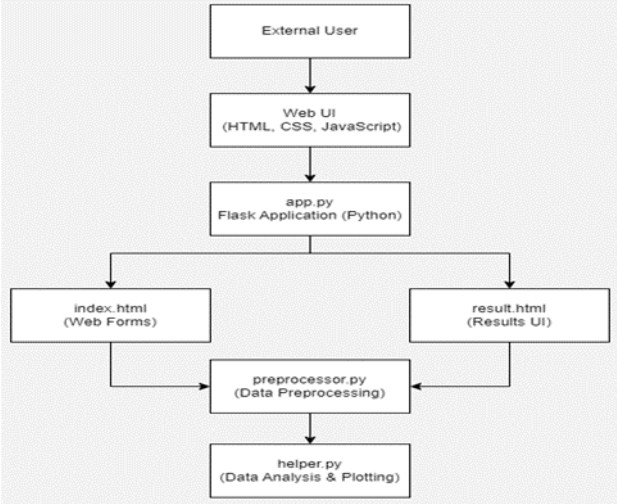

Figure 2. Data Flow Diagram

System Implementation

The system's implementation includes several key components, each aimed at providing users with valuable insights into their WhatsApp chat data. Below are the sections with their corresponding figure numbers:

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/04.2024-78841519/IJAEAST.2024.01.0003, January 2024, Volume-2, Issue-1, pp.21-30
International Journal of Advances in Engineering Architecture Science and Technology

Home Page (Figure 3): The website Pragmatic allows you to analyze your WhatsApp conversations. The homepage features a prominent button to "START NOW" analyzing your chats, alongside a "HOW TO" guide and FAQs. We can upload a chat file and the website will deliver insights into your WhatsApp communication, revealing details about chat content or communication patterns.

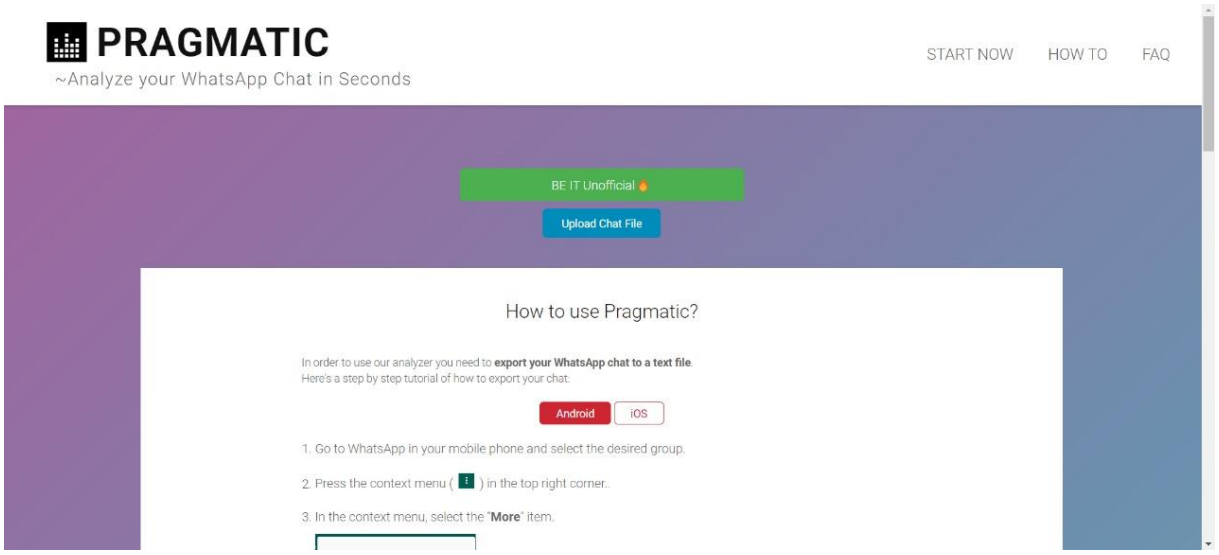

Figure 3. Home Page

Specific Chat (Figure 4): The image shows a specific chat named "WhatsApp Chat with BE IT Unofficial". The chat includes 79 users, 15,914 messages, and 90,984 words. The analysis includes details like the number of links and media shared, messages per day, and the most talkative user in the group.

Figure 4. Specific Chat

Most Busy Day (Figure 5): The y-axis shows the number of messages, while the x-axis shows the days of the week. Saturday is the busiest day, with around 3000 messages. Friday and

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/04.2024-78841519/IJAEAST.2024.01.0003, January 2024, Volume-2, Issue-1, pp.21-30
International Journal of Advances in Engineering Architecture Science and Technology

Thursday are the next busiest days, with around 2500 messages each. Traffic drops significantly on weekdays, with Monday having the fewest messages at around 500. Sundays are also relatively quiet, with around 1000 messages.

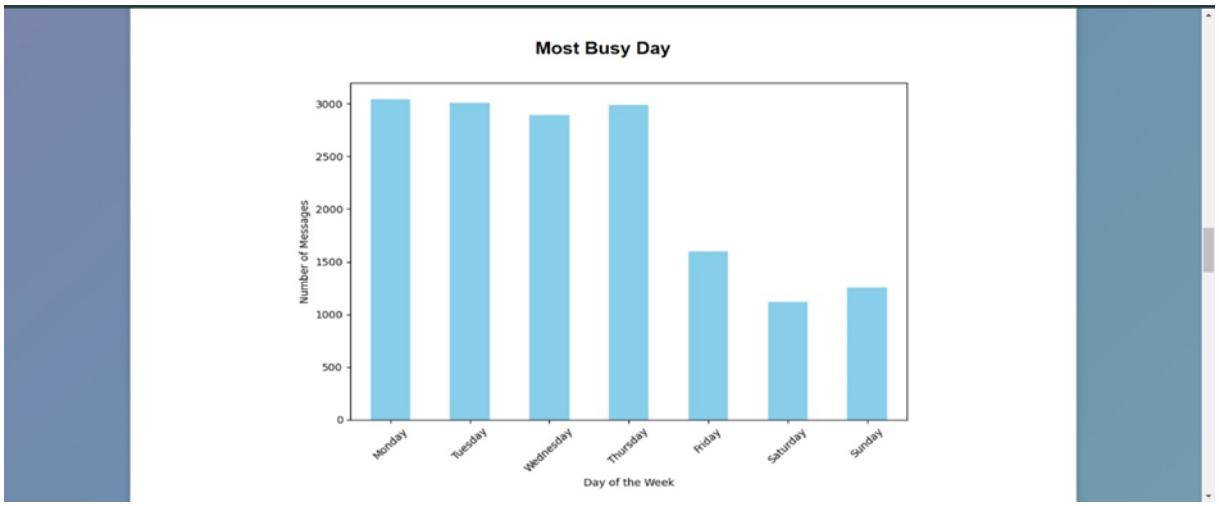

Figure 5. Most Busy Day

Common Words Bar Graph (Figure 6): This reveals the most frequently used words in a text analysis, from WhatsApp chats. "Hai" (hello), "kya" (what), and "ka" (of/the) top the list, suggesting the chats might be in Hindi due to these common greetings and question words. Other frequent terms like "nahi" (no), "CGPA" (grading system), and "percentage" hint at educational topics being discussed.

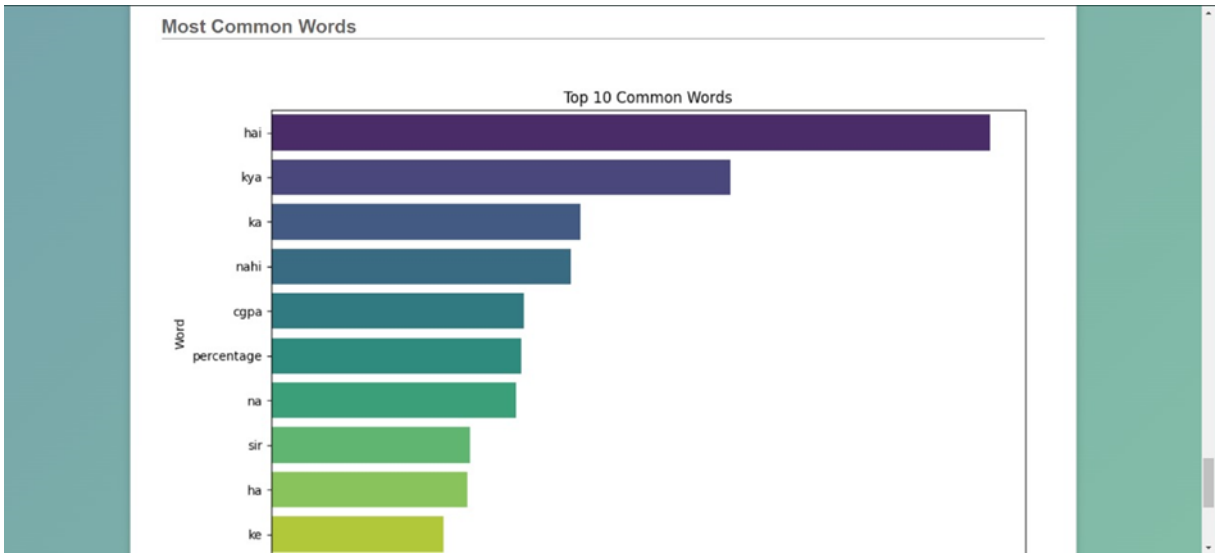

Figure 6. Common Words

Mostly Used Emoji (Figure 7): The pie chart shows the most used emoji in a WhatsApp chat analysis. The "face with tears of joy" emoji dominates, representing laughter and amusement (55.6%). The second most used emoji is similar, but stronger laughter. All other emoji combined make up a tiny fraction of total usage (far less than 44.4%). This suggests chat participants rely on laughter-based emoji to convey humour.

ISSN: 2583-7346
DOI: https://doi-ds.org/doilink/04.2024-78841519/IJAEAST.2024.01.0003, January 2024, Volume-2, Issue-1, pp.21-30
International Journal of Advances in Engineering Architecture Science and Technology

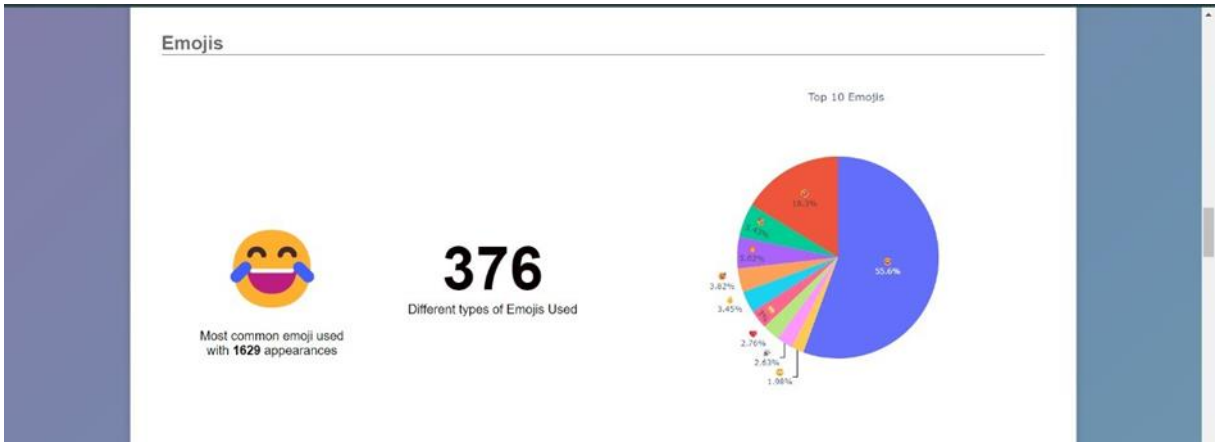

Figure 7. Emoji

Word Cloud (Figure 8): It is a visual representation of text data where word size indicates frequency. Larger words appear more often in the text analyzed. In this case, the cloud shows familiar words from a WhatsApp chat. Words like "hai" (hello), "kya" (what), and "sir" are prominent, indicating the chat is in Hindi and informal or educational.

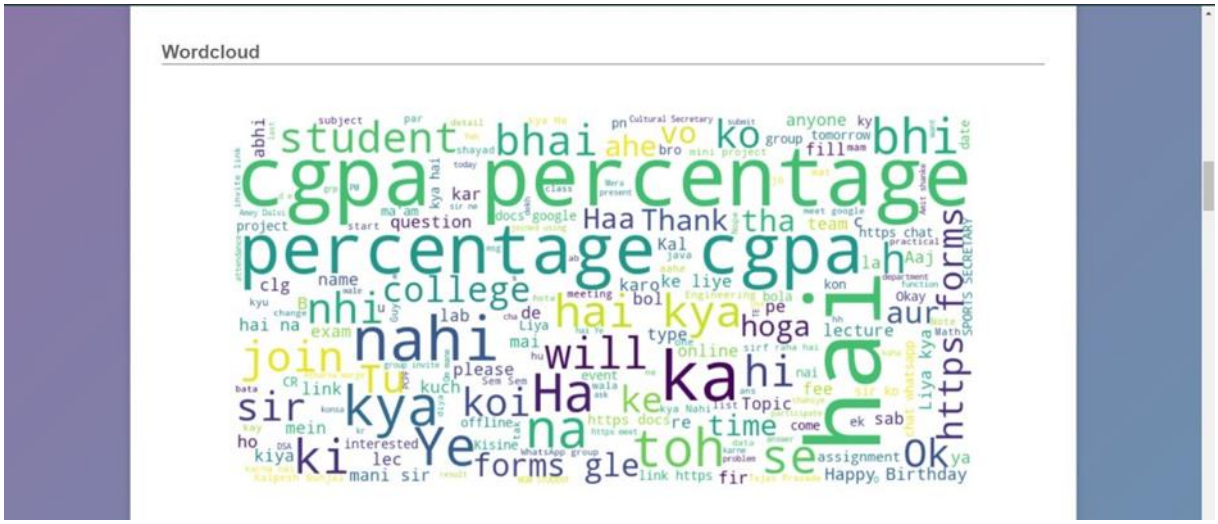

Figure 8. Word Cloud

These rudiments and data representations enhance the stoner experience by offering a comprehensive and visual understanding of WhatsApp converse data. Users can efficiently interpret these figures, gaining insights for decision-making, trend tracking, and understanding message sentiments in their chats.

5. Result Analysis

The result analysis of pragmatic analysis of WhatsApp chats using NLP involves several key findings and insights derived from the processed data. Firstly, statistical metrics such as message count, word frequency, and user activity patterns provide valuable insights into communication dynamics within the chat data. These metrics reveal the volume and frequency of interactions, the most active users, and prevalent topics of discussion. Additionally,

visualizations such as word clouds highlight familiar words and themes in the conversations, while emoji usage charts depict the emotive aspects of communication. Furthermore, activity heatmaps offer a visual representation of messaging patterns across different days and hours, elucidating peak activity periods and trends. Through this analysis, nuanced aspects of linguistic pragmatics and digital discourse within WhatsApp chats are uncovered, facilitating a deeper understanding of communication behaviour and interaction dynamics among users.

6. Conclusion

In conclusion, the pragmatic analysis of WhatsApp chats utilizing Natural Language Processing (NLP) techniques offers a comprehensive understanding of communication dynamics and linguistic pragmatics within digital discourse. The integration of preprocessing, statistical analysis, and visualization modules enables the extraction of valuable insights from raw chat data. Through preprocessing, essential information such as user identities and message content are parsed and structured, laying the groundwork for subsequent analysis. Statistical metrics such as message counts, word frequencies, and user activity patterns provide quantitative insights into communication behaviour, highlighting engagement levels and prevalent topics of discussion. Additionally, visualizations such as word clouds, emoji usage charts, and activity heatmaps offer intuitive representations of the analysed data, enriching our understanding of linguistic nuances and social interactions within WhatsApp communities. By elucidating the intricate dynamics of digital discourse, this study contributes to the broader discourse on contemporary communication practices and facilitates insights applicable across diverse domains, from social science research to digital marketing strategies. Furthermore, it underscores the significance of leveraging NLP methodologies and visualization techniques in unravelling the complexities of online communication, paving the way for future advancements in the field of computational linguistics and sociolinguistics.

References

1. D.Radha, R. Jayaparvathy, D. Yamini, “Analysis on Social Media Addiction using Data Mining Technique,” International Journal of Computer Applications (0975 – 8887).
2. Python for Everybody: Exploring Data in Python 3 by Dr. Charles Russell Severance. Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic.
3. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" by Daniel Jurafsky and James H. Martin.
4. "Natural Language Processing with Python" by Steven Bird, Ewan Klein, and Edward Loper.
5. "Introduction to Information Retrieval" by Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze.
6. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.
7. John T Mesia Dhas, “The Functional and Storage Risks Associated to the Size Estimation of Parallel Computing Applications”, Advances in Parallel Computing, 40, 373-379, 2022, doi:10.3233/APC220052

8. [7] "The Elements of Statistical Learning" by Trevor Hastie, Robert Tibshirani, and Jerome Friedman
9. "GloVe: Global Vectors for Word Representation" by Jeffrey Pennington, Richard Socher, and Christopher D. Manning.
10. "Distributed Representations of Words and Phrases and their Compositionality" by Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeffrey Dean.
11. Tian, Ye & Galery, Thiago & Dulcinati, Giulio & Molimpakis, Emilia & Sun, Chao. (2017). Facebook sentiment: Reactions and Emojis. 11-16.10.18653/v1/W17-1102
12. Barman, Utsab & Das, Amitava & Wagner, Joachim & Foster, Jennifer. (2014). Code Mixing: A Challenge for Language Identification in the Language of Social Media. 10.13140/2.1.3385.6967.
13. Acampora, Giovanni & Loia, Vincenzo & Vitiello, Autilia. (2011). A cognitive multi-agent system for emotion-aware ambient intelligence. IEEE SSCI 2011 - Symposium Series on Computational Intelligence - IA 2011: 2011 IEEE Symposium on Intelligent Agents. 0.1109/IA.2011.5953606.
14. Posner J, Russell JA, Peterson BS. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol. 2005;17(3):715 –734. doi:10.1017/S0954579405050340.
15. Tian, Ye & Galery, Thiago & Dulcinati, Giulio & Molimpakis, Emilia & Sun, Chao. (2017). Facebook sentiment: Reactions and Emojis. 11-16. 10.18653/v1/W17-1102.
16. Krebs, Florian & Lubascher, Bruno & Moers, Tobias & Schaap, Pieter & Spanakis, Gerasimos. (2017). Social Emotion Mining Techniques for Facebook Posts Reaction Prediction. 10.5220/0006656002110220.
17. Hussien, Wegdan & Tashtoush, Yahya & Al-Ayyoub, Mahmoud & AlKabi, Mohammed. (2016). Are Emoticons Good Enough to Train Emotion Classifiers of Arabic Tweets? 10.1109/CSIT.2016.7549459.
18. WhatsApp. Retrieved from https://faq.whatsapp.com/en/android/23756533
19. Unicode (2020) Retrieved from https://unicode.org/emoji/charts/emojilist.html
20. Novak, P.K.; Smailovi´c, J.; Sluban, B.; Mozeti˘c, I. Sentiment of Emojis. PLoS ONE 2015, 10, e144296.