International Journal of Computer Sciences and Engineering
Vol.**11**, Issue.**10**, pp.**64-70**, October **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Paper

Analyzing a WhatsApp Conversation for both Textual Contents and Emotional Sentiments

Vineethalakshmi Boyapati[1*], Jyothireddy Dronadula[2], Smilysrinidhi Gollapati[3], Udayasri Bodapati[4], Shakeelahmed Md[5]

[1,2,3,4,5]Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India

*****Corresponding Author: *bvineethalakshmi25@gmail.com*

Received: 07/Sept/2023; **Accepted**: 09/Oct/2023; **Published**: 31/Oct/2023. **DOI**: https://doi.org/10.26438/ijcse/v11i10.6470

Abstract: WhatsApp, one of the most widely used instant messaging applications, has transformed the way people communicate globally. It presents an overview of WhatsApp chat analysis, a burgeoning field that leverages the vast amount of textual data generated through WhatsApp with the help of this platform users now have a convenient way to connect with their social networks, professional networks, and commercial partners. This give an analysis of the WhatsApp group data in order to determine the degree of participation and involvement among the group's members. Additionally, it requires analyzing the most active day in the group, the quantity of messages sent on that date, the most active user overall, the list of active admins in the group, the overall user count, the quantity of posts made by each user in the group, and the most frequently used term on the platform. The analysis was able to demonstrate the level of participation of the various people on the specified WhatsApp group.

Keywords: WhatsApp Chat, Sentiment Analysis, Stream lit, Nature Language Processing, NMF, Emotion Analysis, Vader.

1. Introduction

WhatsApp, a ubiquitous instant messaging application, has revolutionized the way people communicate in the digital age. With over two billion users worldwide, this platform generates a vast amount of textual data through individual and group chats on a daily basis. WhatsApp chat analysis is an emerging field that focuses on extracting valuable insights from this wealth of conversational data, offering a myriad of opportunities for research, sentiment analysis, and business intelligence. The essence of WhatsApp chat analysis lies in the examination and interpretation of the conversations that unfold within the app. Researchers and organizations harness a spectrum of tools and techniques, including Natural Language Processing (NLP) algorithms and machine learning, to make sense of this data. By doing so, they can uncover hidden patterns, sentiments, and information that are integral to understanding human interaction in the digital realm. The applications of WhatsApp chat analysis are diverse and multifaceted. Social researchers use it to explore communication patterns, delve into the dynamics of social groups, and observe the diffusion of information within these contexts. Sentiment analysis plays a crucial role in gauging public opinion on various topics, products, or events, thereby assisting decision-makers in shaping strategies and policies. In the business sphere, WhatsApp chat analysis helps organizations gain valuable insights into customer feedback, enabling them to enhance their products and services, thus fostering customer satisfaction and loyalty.

Application	WhatsApp
Cost	Free of cost
Number of users	2.7 billion users
Maximum message length	4096 characters includes text, emojis, and media attachments like images and videos.
Protection mechanism	end-to-end encryption, two-factor authentication.

As the volume of WhatsApp messages continues to escalate, WhatsApp chat analysis is poised to play an increasingly significant role in academic research, marketing strategies, and decision-making processes. This field's growth is contingent upon the responsible and ethical handling of data, ensuring that user privacy and consent are always respected. In this exploration of WhatsApp chat analysis, we will delve into its methodologies, applications, and the ethical considerations that shape its future. WhatsApp chat analysis can be useful in various scenarios, including social media, monitoring, customer service analysis, and market research.

2. Problem Statement

In the era of digital communication, WhatsApp has become a ubiquitous platform for interpersonal and group conversations. However, there is a pressing need to gain deeper insights into WhatsApp chat data. Challenges include understanding communication patterns, sentiment analysis,

social network dynamics, the spread of misinformation, privacy and security implications, linguistic and cultural variations, and the psychological and sociological effects of WhatsApp usage. Additionally, businesses require better tools for customer service, consumer behavior analysis, and marketing insights. This research aims to develop robust methods and tools for WhatsApp chat analysis to address these multifaceted challenges, enhancing our understanding of digital communication's impact on society, individuals, and businesses.

3. Literature survey

Various researches have done work in the field of emotion and text analysis of WhatsApp chat data. This section discussed some of their work.

A survey analysis on the usage and impact of WhatsApp Messenger [6] has been conducted and various studies and analysis have been found. These studies include the impact of WhatsApp on the students (youth). In the survey, it was found that in the southern part of India, ages 18 to 23 spend around 8 hours using WhatsApp and sometimes be online almost 12-16 hours a day. Most of them agreed to be using WhatsApp than any other site [5] they exchange images, audios and videos. This survey also proved that the WhatsApp has been the most widely used app on the smart phones than any other app. This survey was conducted to know the positive and negative impacts of using WhatsApp. As we can know that from this survey, WhatsApp is most used app by the youth and other generations so, our project can give them the insights of their chats and provide them unknown facts,[1]Researchers have studied the patterns of communication in WhatsApp, including message frequency, response times, and the use of multimedia elements (e.g., images, videos, and emoji's). They have explored how these patterns differ among various user groups.[4] WhatsApp chat analysis has been used to explore linguistic aspects, such as language variation, slang usage, and code-switching. Researchers investigate how language evolves within the platform and how it may differ between groups or regions. Sentiment analysis is a common theme in WhatsApp chat analysis.[8] In this research paper, we propose a model to provide analysis for WhatsApp chat like text, emoji using Natural Language Processing. We make use of NMF in python to do tasks such as topic modelling this model will work accurate for smaller texts and to determine the sentiment of messages—whether they are positive, negative, or neutral. This analysis can be applied to study user emotions and attitudes within conversations.

4. Methodologies

Data Collection: Gather the WhatsApp chat data you intend to analyze. You can export chat data from WhatsApp by going to the chat/group you want to analyze, selecting "More," and then choosing "Export chat." You can export the chat with or without media files, depending on your needs.

Data Preparation: Before analysis, you may need to clean and preprocess the data. This can include removing duplicates, handling missing messages, and extracting relevant information, such as timestamps, sender names, and messages. Data pre-processing serves as the foundational phase of any project, facilitating a comprehensive grasp of how diverse built-in Python modules are implemented and employed. These modules not only enhance user comprehensibility but also contribute to a more organized code representation. Notable libraries utilized include NumPy, pandas, matplotlib, sys, re, emoji, sea born, among others.

Text Analysis: This step involves analyzing the text content of the chat. Common text analysis techniques include Sentiment analysis. Word frequency analysis to identify common words or phrases.. Topic modeling to identify themes or topics within the chat.

Emotion Analysis: This step involves analyzing the text content of the chat. Classify each message or segment of the chat into specific emotions (e.g., happy, sad, angry) based on the analysis and shows the sentiment analysis of positive, negative and neutral.

Visualizations: Create data visualizations to help interpret the results. You can use tools like Python's Matplotlib or sea born, and used s Stream lit is an open-source Python library that makes it easy to create web applications for data science and machine learning. It is designed to be simple and user-friendly, allowing developers to quickly turn data scripts into shareable web apps.

Tools and Libraries: Explore libraries like NLTK, Apply NMF to uncover underlying topics in the chat, Features of NLP are used like Parsing Text, Eliminating stop words and Analyzing Text. Parsing text is used for splitting messages into words for analysis like total words and mostly used words and determine the optimal number of topics using techniques like silhouette scores and VADER (Valence Aware Dictionary and sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool designed to analyze text data and determine the sentiment polarity (positive, negative, or neutral) of the text. It is particularly useful for analyzing and understanding sentiment in social media posts, online reviews, and other short, informal texts. to aggregate these scores for individual messages or the entire chat to assess the overall sentiment. Emotion analysis and sentiment analysis.

The Methodologies used for Analyzing and Visualization are:

NLP: In this project, Natural Language Processing (NLP) features are applied, including text parsing, stop word removal, and text analysis. Text parsing is employed to break down messages into individual words, facilitating analysis such as word count and identifying frequently used words. A predefined file containing stop words is utilized to filter out these common, non-significant words, ensuring that the analysis focuses on meaningful content. Text analysis is conducted to determine the number of media files shared and the quantity of links shared within the text data.

Stream lit: This library is leveraged to design visually appealing web elements and objects that depict WhatsApp

chat analysis. These elements include a variety of charts and visualizations, and they are integrated into the Stream lit framework for presenting the analysis results.

Sea born: Sea born is a Python library primarily employed for creating visually appealing statistical plots. It enhances the aesthetics of statistical visualizations by offering a wide range of color palettes and predefined styles. In this project, sea born is utilized to generate a heat map visualization, which displays a 24-hour and 7-day grid with varying color intensity to represent the distribution of messages throughout different hours, with the colors indicating the range from the highest to the lowest message counts.

Matplotlib: Matplotlib is a user-friendly and highly effective Python library for creating captivating visualizations. It is based on NumPy arrays and seamlessly integrates with the broader SciPy ecosystem. Matplotlib offers a rich array of plot types, including pie charts, line charts, bar charts, scatter plots, histograms, and more. In this project, Matplotlib plays a central role in crafting diverse visual representations for the analysis of WhatsApp chat data. These visualizations encompass bar charts, line charts, and pie charts, adding depth and clarity to the analysis process.

The Methodologies used for technical thinking are:
Pandas: This open-source Python library is primarily applied in the realms of Data Science and machine learning. It offers a suite of powerful tools for data analysis and manipulation. Leveraging its data structures, this library is instrumental in the examination and manipulation of data, particularly in tasks involving time series analysis and numerical data.

Numpy: The name "Numpy" is derived from "Numeric Python," and it serves as a data analysis library for Python. This library encompasses a wide range of numerical functions and methods for conducting thorough numerical analysis. Numpy is particularly renowned for its multi-dimensional array objects, and it offers a comprehensive set of routines for processing these arrays effectively.

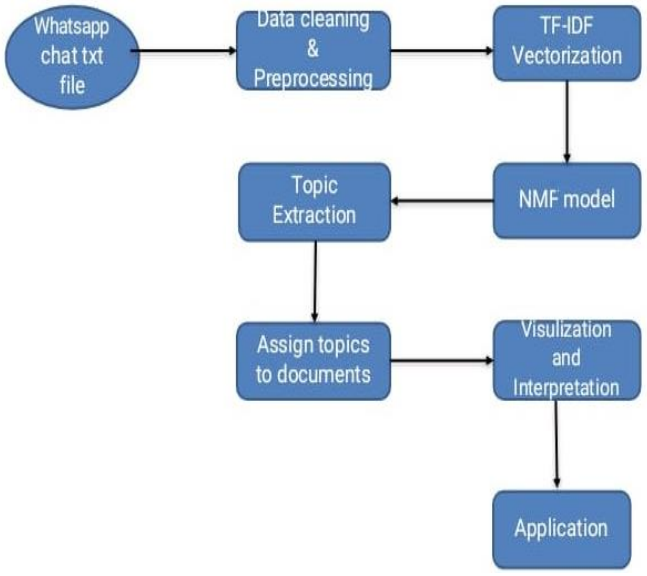

Fig.1. Flowchart of various phases of methodology

5. Data Analysis

Data analysis involves a structured approach that encompasses the cleaning, transformation, scrutiny, and modeling of data. Its primary aim is to unearth valuable information and derive meaningful conclusions. The essence of data analysis lies in breaking down a complex dataset into its individual components for in-depth examination. Tokenize the text: Break down the text into individual words or tokens. The most common words in a text or corpus of text are typically called "stop words." These common words are often filtered out during text analysis because they do not carry significant meaning on their own and are found frequently in the English language. Remove stop words, Eliminate common, uninformative words like "and," "the," "I," etc. Stemming or Lemmatization Reduce words to their base form. Conduct spell-check and text normalization to enhance the quality of the text.

The fundamental goal is to take the initial chat data and convert it into informative content that can aid decision-making or provide a comprehensive understanding of the chat content. This process typically encompasses data cleaning and preprocessing, conducting various analyses, which may include statistical and text analyses, and ultimately presenting the results in a coherent manner.

This type of analysis can be instrumental in revealing communication patterns, sentiment trends, frequently discussed topics, and other pertinent insights within WhatsApp chats. It serves various purposes, be it for monitoring team interactions, analyzing customer feedback sentiment, or simply gaining insights into one's personal chat history.

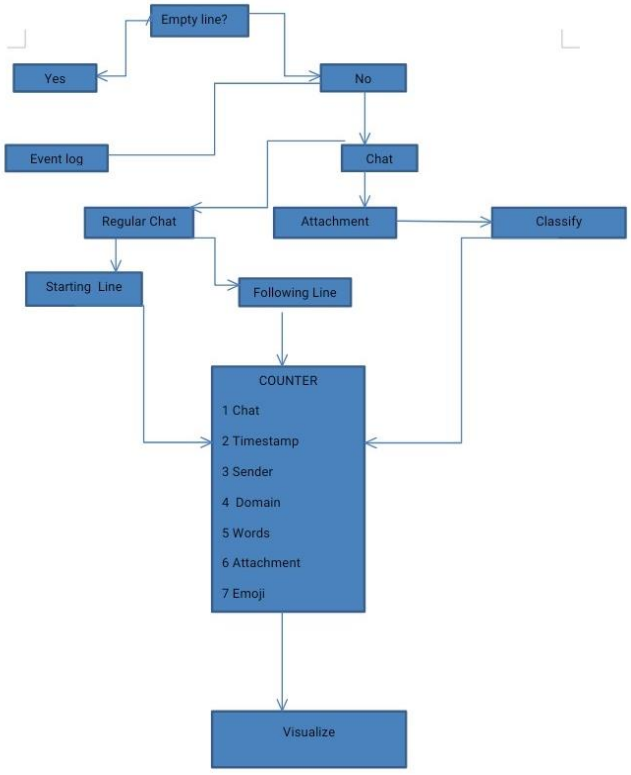

Fig.2. Analysis of Text chat data

6. Emotion Analysis

"Emotion," in simple terms, refers to the feelings or moods that people experience. An individual's emotion reflects their state of mind at a specific moment. Emotions can broadly fall into three categories: positive, negative, and neutral. Positive emotions encompass feelings like happiness, cheerfulness, contentment, and excitement. Negative emotions include emotions such as sadness, anger, anxiety, and depression. There are also statements or facts that do not carry a clear positive or negative emotion; these are considered to have a neutral emotion. Sentiment analysis, in simple terms, is the process of determining the emotional tone or attitude expressed in a piece of text. It involves assessing whether the text conveys positive, negative, or neutral sentiments. This analysis can be applied to various forms of text data, such as social media posts, reviews, or chat messages.

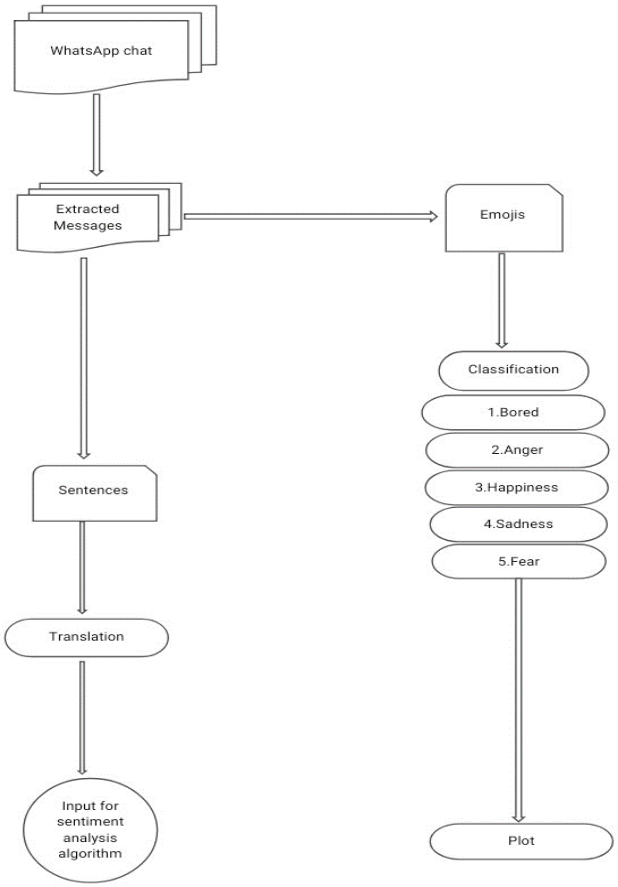

Fig.4. Analysis of emoji data

Emotion analysis involves assessing people's opinions, attitudes, views, and emotions, and classifying them primarily as positive or negative, and sometimes as neutral. Given the growth of social media and the increasing number of people using it, there is a growing demand for analyzing emotions. Emotion analysis of WhatsApp chat data, for example, entails studying individuals' views, sentiments, thoughts, attitudes, opinions, expressed in their tweets about various subjects, whether it be other individuals, products, organizations, topics, services, and more By integrating emotion analysis into your WhatsApp chat project, you can gain deeper insights into the participants' emotional states. This can be particularly valuable for organizers seeking to understand the overall mood of the group and make data-driven decisions based on emotional dynamics. Here the sentiment analysis is performed by the nltk module Vader is used for the polarity like positive, negative and neutral type of analysis and for the emoji analysis, it shows which type of emoji's are used.

In the realm of WhatsApp chat analysis, emotion and sentiment analysis entails the examination of text data to decipher the underlying emotional tone and feelings conveyed in the conversation. This process employs natural language processing techniques to determine whether the expressed sentiments are positive, negative, or neutral. By assigning sentiment scores to individual messages or the overall chat, it provides valuable insights into the emotional context, helping to gauge the mood, reactions, and sentiments of the participants. This analysis is invaluable for understanding user sentiment, conducting market research, and making data-driven decisions, especially in customer service, social media monitoring, and content analysis.

7. Result and Discussion

Open WhatsApp chat for a group -> click on the menu -> click on more- -> select export chat-> choose without media. Working of WhatsApp chat analysis.

- Initially, open WhatsApp chat analyzer web page.
- Upload the exported chat file.
- Analyzing of data is done by trained model
- Preprocessing of data is done by trained model.
- Select overall or single person analysis
- Trained model shows analysis of collected data is read and processed to train our machine learning classification model on it. it includes all statistics like as the following:

• Total words, Media, Link shared.
• The daily timeline for the message is shown by line charts.
• Weekly activity map, which shows hourly activity of users with corresponding day using heat map.
• Top ten most common word represented by using bar chart.
• List of Emoji's with number of times it is used.
• Pie chart, which shows top five emoji's percentage of use.
• Bar graph with positive, negative, neutral type sentiments.

Time Analysis: Insights into message distribution over different times of the day or days of the week.

Emotional Analysis: Sentiment analysis to determine the overall tone of the conversation.

Keyword Analysis: Identification of specific keywords or topics that are frequently discussed.
It shows the busy users and their contribution to chat we have used matplotlib to plot the graph and the users and how frequently the chat is calculated and plotted. Timelines or line charts can show the frequency of messages over time. Users can see when the chat was most active or identify any notable patterns.

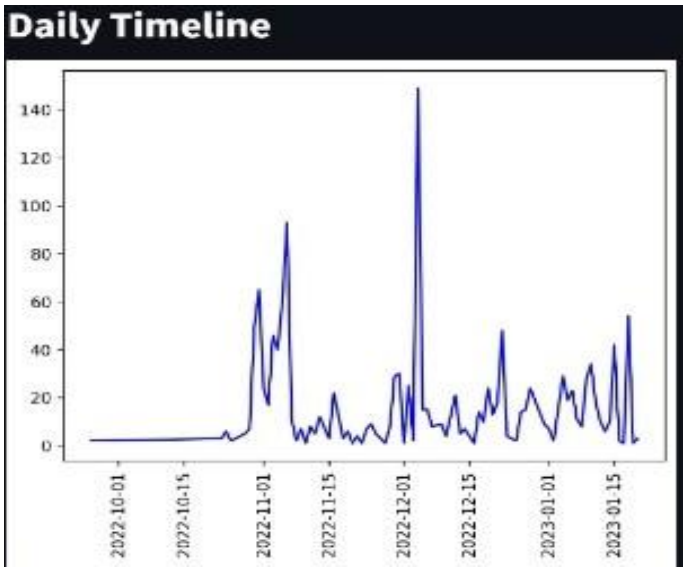

Fig.4. Daily Timeline

Analyzing weekly chat activity in this manner can be beneficial for understanding when and how often people communicate within a WhatsApp group or conversation.by using Sea born is a Python data visualization library that simplifies creating informative and attractive statistical plots for better data analysis and presentation. Heat maps can show the intensity of communication between participants, time of day, or day of the week when messages are sent. This can help identify peak activity periods.

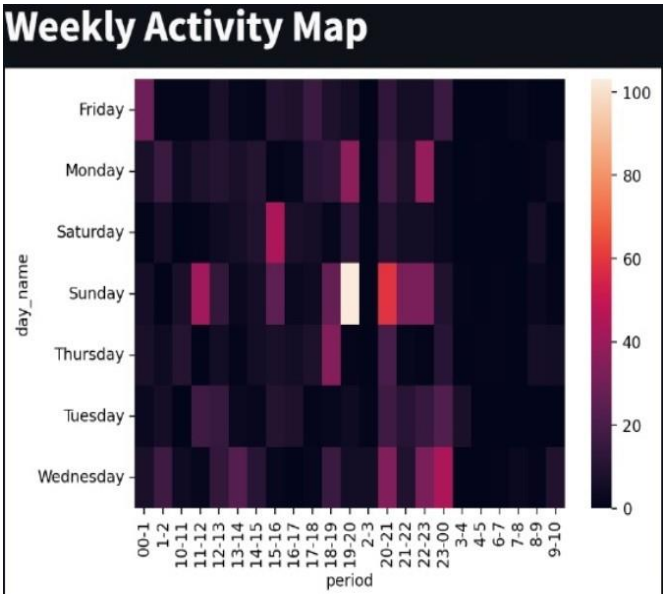

Fig.5. Weekly Activity Map

It shows the most commonly used word we have used matplotlib to plot the graph and the top frequently used words The most common words in a text or corpus of text are typically called "stop words." These common words are often filtered out during text analysis because they don't carry significant meaning on their own and are found frequently in the English language. Display a bar chart showing the frequency of specific keywords or phrases mentioned in the chat.

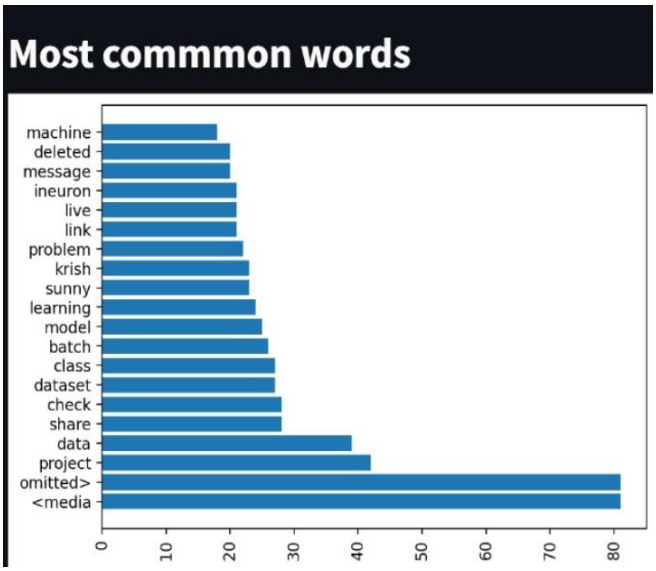

Fig.6. Most Common Words

It shows Statistics of total messages, words, links, media shared in the group.by Analyzing text messages involves counting the total number of messages, the words within those messages, and tracking the shared media and links for comprehensive data analysis and insights.

Fig.7. Total messages, words, images links shared.

It shows the most commonly used emoji's We have used the Emoji library to select or distinguish the emoji's from the messages and plotted the pie chart using matplotlib. Visualize the usage of emoticons or emoji's in the chat. This can provide insights into the emotional tone of the conversation. Emoji analysis involves studying the use of emoji's in text to gain insights into the emotional or contextual aspects of a message, enhancing communication and sentiment understanding.

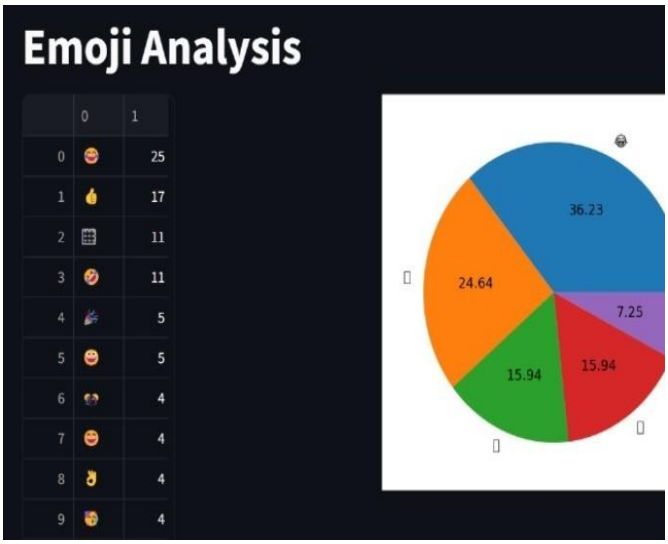

Fig.8. Emoji Analysis

It shows the sentiment analysis we have used nltk-Vader for analysis Sentiment analysis employs natural language processing to assess written or spoken text, categorizing sentiments like happiness, sadness, or neutrality, aiding in data-driven decision-making. Abled with the results positive, negative, and neutral sentiments.

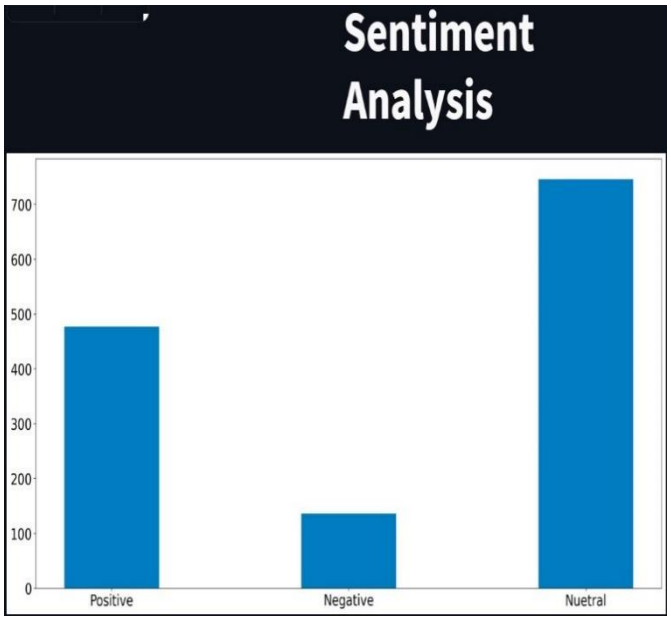

Fig.9. Sentiment Analysis

8. Conclusion

We can conclude that the capabilities of the WhatsApp application and the power of the python programming language in implementing our data analysis intended cannot be overemphasized. The system was done with python, We propose to employ dataset manipulation techniques to have a better understanding of WhatsApp chat present in our phones. It shows most used emoji and word which repeatedly most times. It tracks our conversation and analyzes how much time we are spending and the python libraries that were implemented includes, Stream lit, Emoji, NumPy, Pandas, Re, Matplotlib, URLextract, collection and Sea born. Finally, results that we intended were obtained. The future of our project is it is mainly useful for organizers. Then will get to know who is more and least active in the group. Depending on that, they can take decisions.

The major objective that has been decided in the initial phase of the requirement analysis is achieved successfully. After the implementation, the system provides reliable results. The system is totally menu and user friendly, which makes it easy for the users even with limited knowledge of computer environment to operate the developed system. These features are valuable for users who want a convenient and efficient solution for managing and analyzing WhatsApp data.

Accuracy: The tool accurately extracts and analyzes the data from WhatsApp chat files. This includes accurately identifying participants, message content, timestamps, and other relevant information.

Comprehensive Analysis: The software should offer in-depth analysis features, such as sentiment analysis, keyword search, conversation statistics, and more. Users should be able to gain meaningful insights from their WhatsApp data.

Timesaving: The software should be designed to streamline the analysis process, automating tasks where possible to save the user's time. For example, it could provide quick insights, summaries, and visualizations of the WhatsApp data.

Analyzes any WhatsApp imported file: The project or tool is capable of examining and processing data from any WhatsApp chat file that is imported into it. It suggests that the tool has the ability to handle different like as for both individual or overall analysis and versions of WhatsApp chat files for analysis.

User-Friendly Interface: The application should have an intuitive and easy-to-navigate user interface, making it accessible to a wide range of users, regardless of their technical expertise.

Conflict of interest

Financial Interests: Potential conflicts of interest may emerge when individuals involved in the project stand to gain financially based on the analysis's results. For instance, if a company is responsible for the analysis, and the outcomes could impact its financial performance, a conflict of interest may be present.

Organizational Affiliations: Conflicts of interest can arise if the organizations conducting the analysis have affiliations or connections with groups or entities with a vested interest in the analysis's outcomes. For instance, if an organization has strong ties to a particular political group and is examining chats related to that group, their analysis could be perceived as biased.

Funding source

My institution and all our team members conduct research, for WhatsApp conversation emoji and text analysis, as part of academic endeavors and as the funding source of the project.

Author's contribution

The WhatsApp chat analysis project was a collaborative effort, with each author making distinct contributions to its success. The following summarizes the roles and responsibilities of each author:

Author 1:
- Conceptualized the research project and formulated research objectives.
- Oversaw data collection and data preparation, ensuring data quality and integrity.
- Conducted data analysis, including sentiment analysis and topic modeling.
- Drafted and revised the methodology and results sections of the paper.

Author 2:
- Conducted an extensive review of the literature, providing the theoretical foundation for the study.
- Assisted in the design of data collection tools and survey instruments.
- Played a significant role in data interpretation and contributed to the discussion section.
- Collaborated with the team on the project timeline and coordination.

Author 3:
- Actively participated in data collection, including chat log retrieval and organization.
- Performed statistical analyses, data visualization, and creation of graphical representations.
- Co-authored the results section, emphasizing data analysis findings.
- Contributed to the proofreading and formatting of the final manuscript.

Author 4:
- Focused on ethical considerations, including ensuring data privacy and compliance with relevant regulations.
- Assisted with data anonymization and data protection protocols.
- Reviewed and synthesized literature relevant to the introduction and ethical considerations.

Author 5:
The mentors who provided guidance, expertise, and valuable insights throughout the research process.
The collaborative efforts of all authors were instrumental in the completion of this project. This multidisciplinary approach allowed for a comprehensive examination of the data and enriched the overall quality of the research.

Acknowledgment

Data Collection Team: The data collection team played a crucial role in gathering and organizing WhatsApp chat data. Your meticulous work and commitment to data quality are highly appreciated.

Participants: We deeply appreciate the individuals who willingly shared their WhatsApp chats and made this research possible. Your participation was essential, and we thank you for your trust and cooperation.

Academic Advisors and Mentors: Our gratitude goes to our academic advisors and mentors,who provided guidance, expertise, and valuable insights throughout the research process. The mentorship was instrumental in shaping the project.

References

[1]. Ravishankara K, Dhanush, Vaisakh, Srajan I S, "International Journal of Engineering Research & Technology (IJERT)", ISSN: 2278-0181, May, Vol.**9** Issue.**5**, **2020.**
[2]. Dr. D. Lakshminarayanan, S. Prabhakaran, "Dogo Rangsang Research Journal", UGC Care Group I Journal, July, Vol.**10**, Issue.**7**, No.**12, 2020.**
[3]. Chetashri Bhadane, Hardi Dalal and Heenal Doshi, "Sentiment Analysis-Measuring Opinions", International Conference on Advanced Computing Technologies and Applications (ICACTA), Vol.**45**, pp.**808–814, 2015**.
[4]. S. Patil, "WhatsApp Group Data Analysis with R," International Journal of Computer Applications, November, Vol.**154**, No.**4**, pp.**0975 – 8887, 2016.**
[5]. Sangeeta Rani "SENTIMENT ANALYSIS AND TOPIC MODELLING ON TWITTER FOR CLEAN INDIA MISSION " Indian Journal of Computer Science and Engineering (IJCSE), Sep-Oct, Vol.**12**, No.**5, 2021.**
[6]. C. Montag, K. Błaszkiewicz, 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 August 2015. [Online]Available:https://bmcresnotes.biomedcentral.com/articles/ 10.1186/s13104- 015-1280-z. [Accessed 12 March 2019].
[7]. Aharony, N., T., G., The Importance of the WhatsApp Family Group: An Exploratory Analysis. "Aslib Journal of Information Management", Vol.**68**, Issue.**2**, pp.**1-37, 2016.**
[8]. Dr. D. Lakshminarayanan, S. Prabhakaran, "Dogo Rangsang Research Journal", UGC Care Group I Journal, July, Vol.**10**, Issue.**7**, No.**12, 2020.**
[9]. Neetu Anand ,Tapas Kumar"Text and Emotion Analysis of Twitter Data", Journal of Computer Science and Engineering (IJCSE), Jun, Vol.**5**, Issue.**6, 2017.**
[10]. D.Radha, R. Jayaparvathy, D. Yamini, "Analysis on Social Media Addiction using Data Mining Technique", International Journal of Computer Applications (0975 – 8887).
[11]. Xing Fang and Justin Zhan, "Sentiment Analysis Using Product Review Data", Journal of Big Data, Springer, pp.**1-14, 2015.**
[12]. .Hemalatha, G. P Saradhi Varma, Dr. A. Govardhan. "Preprocessing the Informal Text for efficient Sentiment Analysis", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS). July – August, Vol.**1**, Issue.**2**, pp.**58-61, 2012.**
[13]. Xia Hu, Jiliang Tang, Huiji Gao and Huan, Liu, " Unsupervised Sentiment Analysis with Emotional Signals", Proceedings of the 22nd International Conference on World Wide Web, WWW'13, ACM, pp.**607-617, 2013.**

