

Determining the Polarity and Statistics of Chat based on Sentiment Analysis

Piyush Anand

Galgotias University
Greater Noida
Uttar Pradesh, India
piyushanand2580@gmail.com

Raj Gupta

Galgotias University
Greater Noida
Uttar Pradesh, India
rajg33182@gmail.com

Kunal Chaudhary

Galgotias University
Greater Noida
Uttar Pradesh, India
mindhunter5613@gmail.com

Abdul Aleem

Galgotias University
Greater Noida
Uttar Pradesh, India
er.aleem@gmail.com

Abstract-Social media networks (SMNs) generate huge amounts of data every day, which can be structured, unstructured or semi-structured. Data can range from plain text to a graphic image or sound or video. Analyzing this diverse and ever-growing data is a major challenge. This paper focuses on extracting and analyzing data from the frequently used online text application WhatsApp through the process of sentimental and statistical analysis. Opinion mining, also known as sentiment analysis is the contextual mining of data to identify, extract and analyze the underlying sentiment in news and classify it as positive, negative or neutral. Data can provide a wealth of information for cutting-edge technologies such as machine learning. Providing a machine learning model with the appropriate learning environment, which is indirectly influenced by the data it receives, is crucial. No matter what subject the chat is centred upon, the developed code can be used to provide a deeper comprehension of the facts. In order to comprehend the many analyses and feelings, this study uses the Python programming language.

Keywords – Social media networks, Sentiment Analysis, Structured Emotions, Unstructured Emotions, Polarity Analysis.

I. INTRODUCTION

Chat applications generate a large amount of data every day. Each user sends a lot of text data over the network. The proposed sentiment and statistical analysis system uses this data to learn and predict the user's sentiment or mood at a given moment. One can correctly predict a user's mood by reading their chat messages. To enable the system to perform a similar task, it needs to convert this text data into functions that the machine understands. Once the features are selected, the system classifiers are trained using the training set [1]. There are two common approaches to text mining, use a bag-of-words model, or consider word structure and grammar. The main aspect of text mining is data cleaning. Working with loosely structured data requires a large amount of pre-processing to prepare the data for analysis. Mining social media conversations using natural

language processing techniques to extract useful insights from subjective information is quickly becoming popular [2].

On a larger scale, businesses use them to understand customer sentiment toward their products or services. At the periphery, this type of analysis could provide an understanding of the social behaviour and engagement of a small group of people, such as a department or classroom. The growth of social networking sites has created many areas dedicated to analysing these networks and their content in order to obtain the necessary information [3]. Sentiment analysis deals with deriving the feelings communicated by a text excerpt from its content. Curiosity is one of the primary matters which driven many human beings to do many mind-blowing matters also identical interest kills man or woman internally, absolutely everyone have the interest approximately their picture inside the different person's mind based at the conversation that look location between them, also the person's mood and emotion may be based totally best at the past conversations as properly [4]. So the solution this paper affords is making use of sentiment evaluation to the messages which might be exported from the chat.

All of the messages are exported from WhatsApp using the app's internal export tool. All of the messages in this exported chat must be split for higher calculation, and each one is given a sentiment evaluation. Sentiment analysis is a technique, or sub-technique, of natural language processing and data mining that is frequently used to determine a person's attitude or emotion [5]. Also, this notion is effective at the fundamental principles of system mastery, where a corpus of material is used to train and technique that information, and a version is developed and used to assess the emotional content of test chat. This method has developed into a discipline for text analysis that is utterly effective.

II. LITERATURE SURVEY

People are thought to be subjective by nature and frequently prejudiced by previous views. Nearly all realms and disciplines of human civilization have opinions that have a significant impact on modern human cultures. Therefore, sentiment analysis may be thought of as a crucial technique for examining the sentiment and general mood of any sample group of people. When such human sentiments are articulated in terms of any entity, sentiment analysis ranks significantly as a study of human attitudes, opinions, and emotions. Either supervised learning or unsupervised learning is typically used to perform sentiment analysis. Sentiment analysis is based on the idea of computational linguistics and natural language processing text analysis. It can identify, extract, quantify, and analyse different states in subjective data [6]. The capability of processing natural language is a computer programme that can recognise spoken language in people.

Political elections can be predicted using sentiment analysis, which demonstrates that data from Twitter is more trustworthy as a platform. Twitter data has been found to have a 94% correlation with survey data, and it has the potential to become a platform that can compete with sophisticated polling methods.

The study of sentiment analysis has a solid foundation in the 20th century. In 2005, there were 101 papers on sentiment analysis, whereas in 2015, there were 5699. This indicates a 50-fold growth in sentiment analysis studies over a ten-year period [7]. As more people started using social media to communicate any form of worry, sentiment analysis has become increasingly crucial. Using social media Suicidal attempts, online fraud, and cyberbullying all rise. 3.02 billion people worldwide are using social media, according to statistics [8].

In order to investigate the significance of social media applications among young people, a study was carried out in the southern region of India on the age range of 18 to 23 years. Despite the fact that this survey discovered that students used these apps for 8 hours each day and were online for over 16 hours each, WhatsApp was the most popular social networking app overall [9]. Everyone who responded affirmed that they use WhatsApp to talk to their friends. They use WhatsApp to send each other pictures, audio, and video and document files. Furthermore, it has been shown that children use only one app while they're hanging out. The methodologies utilised in this study are intended to analyse the volume of usage of WhatsApp and its well-known services and to determine the extent to which using WhatsApp has either good or negative consequences [10]. Launched in 2009, WhatsApp was one of the top 20 apps in the US at the beginning of 2011. It takes the position of SMS by giving users some more information, such as a blue double tick, and notifying them when their message has been seen by the recipient. The primary objective of WhatsApp was to offer a simple messaging service. Interesting features were gradually added to the software.

Voice memos, location information, and now a payment gateway are some of its primary features [11].

Sentiment analysis is incredibly helpful for social media monitoring since it gives us a better understanding of how the general public feels about certain issues, goods, and services. Organizations all over the world are utilising this technology to enable the practise of extracting insights from social data [12]. In addition to political elections and social media monitoring, sentiment analysis has also found applications in customer service. Many businesses are using sentiment analysis tools to analyse customer feedback and reviews, enabling them to identify areas where they are excelling and areas where they need improvement. Even manuscripts could be rectified using sentiment analysis [13].

Sentiment analysis has also find its utility in recommendation of entertainment systems [14] [15]. Businesses may better satisfy the wants and expectations of their consumers by analysing customer sentiment before making changes to their products or services. This not only enhances customer satisfaction but also helps businesses to retain their existing customers and attract new ones.

III. PROPOSED METHODS

The section describes how the data is collected for the research and the system architecture to be followed.

A. Data Collection

Understanding the initial stage of any data analysis project - data collection, is crucial before going into the technical specifics of the data pre-processing phase. In this instance, data was gathered by exporting a data file from a chat group. Pre-processing the data was the next step after obtaining it in order to draw valuable conclusions from it. Python's multiple built-in modules and its programming language come into play in this situation.

An effective technique to get understanding of group dynamics, communication patterns, and other pertinent information is by gathering data from WhatsApp chats. In order to gather information from a WhatsApp conversation, the chat history must first be exported as a text file so that it may be used for further investigation. This stage involves the collection of data. This was done by visiting the chat group to export the data file that was used.

These straightforward steps can be used to export the chat history:

1. To get data from a WhatsApp conversation group, open the group.
2. For access to the chat settings, click on the three dots in the top right corner of the screen.
3. Then click "Export chat" under "More" in the drop-down menu.
4. The inclusion of media assets (such as pictures and movies) is then an option that will be presented to you.

B. System Architecture

The project's initial phase is data pre-processing, which involves learning how to use several built-in Python modules and implement them. These different components improve user comprehension and code representation. NumPy, pandas, matplotlib, re, emoji, seaborn, URL extract, collections, etc. are among the libraries utilised. The finest statistics are provided after data analysis, including total messages, total media, links, shared photographs, and graphs displaying a weekly activity map and emojis are used, along with polarity and subjectivity, as well as monthly and daily timelines, primarily busy users, most commonly used terms in graph.

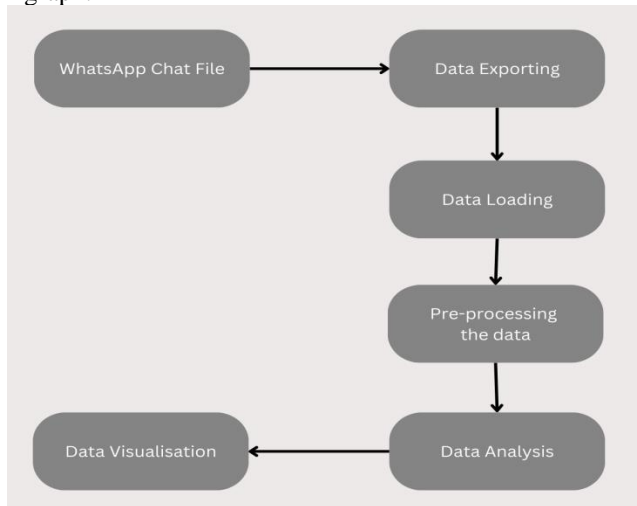


Fig. 1. Proposed System's Flowchart

In addition to the libraries described above, other libraries are also used in the data pre-processing stage to carry out various natural language processing tasks like tokenization, stemming, lemmatization, and sentiment analysis. These libraries include NLTK (Natural Language Toolkit) and TextBlob. These libraries assist with data cleansing, stop word removal, and text sentiment and polarity identification. After the data has been pre-processed, several visualisation techniques are used using Python libraries like Matplotlib and Seaborn to produce graphs and charts that aid in comprehending the chat group's communication patterns. It is possible to spot patterns and trends in the activity of the group using the created graphs and charts. The project's final product comprises a variety of statistics, charts, and graphs that offer a thorough analysis of the communication patterns in the WhatsApp chat group. The most popular topics, popular users, sentiment analysis, and other pertinent insights can all be found using this data. The project's output is useful for better comprehending the behaviour of the chat group and can be used to inform data-driven decisions.

IV. FEATURES

A. Fetch Unique User

Identifying unique users is one of the main components of the data analysis process in a WhatsApp chat. This entails obtaining the names or phone numbers of each chatter's contributors. This allows for the measurement of user activity and the identification of the chat's most active participants.

B. Statistics area

It includes number of words, messages, number of media messages and number of links shared in chat. The statistics section is a crucial component of the data analysis procedure because it offers crucial details about the chat group. The total amount of messages, media messages, shared links, and chat terms are all shown in this section. It is possible to learn more about the chat group's communication preferences and trends by analysing this data.

C. Monthly and daily timeline

The creation of daily and monthly timelines of the most active users is a crucial component of the data analysis process. This enables a deeper comprehension of the chat group's activity patterns and can be used to track down the contributors who are most engaged over time.

D. Activity map

The activity map shows the chat group's activity patterns visually, showing things like the busiest days of the week and the month. This feature offers insightful information regarding the group's communication styles and can be applied to spot trends and patterns in the group's behaviour.

E. Busiest user in the group

The busiest user in the group can be identified using this functionality at the group level. It is possible to discover the most active group members and the user who has contributed the most to the discussion by analysing the data.

F. Word-Cloud

The most frequently used words in the chat are represented visually in the word cloud. This function enables quick and simple comprehension of the themes and subjects covered in the chat group.

G. Most common words

The most frequently used words in the chat are listed in this feature. It is possible to learn more about the subjects and ideas that are most significant to the chat group by analyzing this data.

H. Emoji Analysis

The most frequently used emoticons in the discussion can be found using the emoji analysis function. It is possible to comprehend the feelings and sentiments conveyed by the chat group by analysing this data.

I. Polarity and subjectivity

The ability to identify the emotions and sentiments stated in the chat is provided by the polarity and subjectivity analysis feature. It is possible to comprehend the tone of the discussion and the emotions expressed by the chat group by looking at the text's polarity and subjectivity.

J. Positive, neutral and negative words

The ability to distinguish between positive, neutral, and negative words makes it possible to determine the mood of the chat. It is possible to comprehend the feelings and sentiments spoken by the chat group and spot any patterns or trends in their conversation by analyzing this data.

V. TESTING

To determine the quality of the tested product, testing is similar to research. Software testing offers an unbiased perspective on the product, allowing engineers to comprehend implementation hazards. In order to detect bugs or weaknesses, testing approaches entail the practise of sequentially executing a programme or application.

Table1. Test Case Of Work.

Test case type	Description	Expected result	Status
Top-Statistics	Includes total message and words, shared media and shared links	Statistics overall for the entire messages, words, media and links.	Pass
Timeline	Involves monthly timeline and Daily timeline	Monthly and Daily timeline graph.	Pass
Activity Heatmaps	Involves busiest and least busy day in week and busiest and least busy month in year and weekly activity heat map	Heatmap showing the week's activity and Bar Graph showing busiest day and month of the year	Pass
User map	Involves most busy user in group	Bar graph of most busy user in group	Pass
Word-Cloud	The words which are used most frequent in group and by individual user	Word cloud of most used words	Pass
Emoji Analysis	Most commonly used emoji in group and by user	Pie chart of most used emoji	Pass
Polarity and subjectivity of chat	Involves category of sentiment like neutral, positive and negative	Bar chat of polarity and subjectivity and classify word in positive, negative and neutral	Pass

VI. RESULTS AND DISCUSSION

The outcomes of this investigation demonstrated a number of actions on particular dates as determined by system at the time as depicted in fig 2. The data that can be

viewed includes information on the number of messages, words, and shared image links. To locate the links, an URL extract was used after converting the complete conversation file into a dataframe and separating the words and messages. The analysis process can provide valuable insights into the communication patterns of the chat group, in addition to the quantity of messages, phrases, and shared picture links. By transforming the entire conversation file into a dataframe and separating the words and messages, it is possible to determine the most active people in the chat and how often they contribute. To understand the communication trends within a chat group, it is important to analyze the frequency of messages on a daily basis (Fig 3). A graph was created using Matplotlib, where the number of messages was calculated and plotted against the corresponding days. These graphs represent the timeline of users' monthly and daily activity within the chat group. The hectic days and months are evident. To plot the graph, Matplotlib is utilised. A month's or day's worth of messages are assigned to that specific day or month

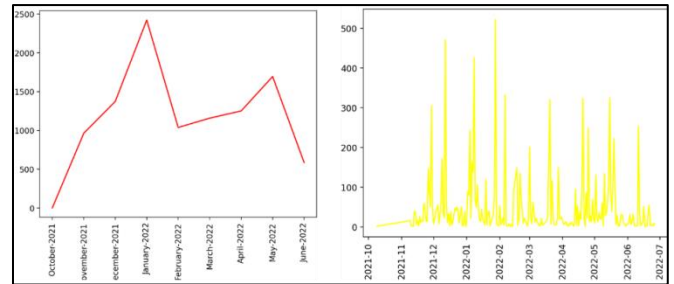


Fig. 2. Timeline Based on Monthly & Daily Analysis

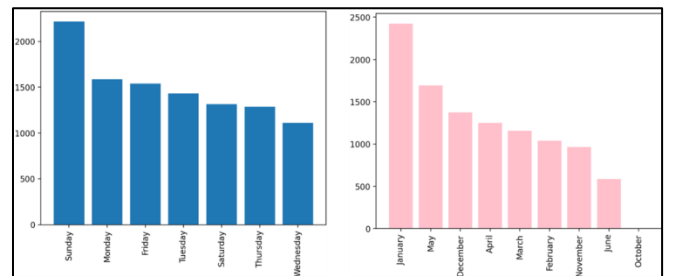


Fig. 3. Most Busy Day & Most Busy Month

The graphical representation of data uses color-coding to display different values. This graph as shown in fig 4 provides a visual representation of the activity patterns within the chat group, allowing for the identification of the busiest times and the preferred communication methods of the members. Group administrators can leverage this information to enhance group effectiveness and ensure that all members receive important communications.

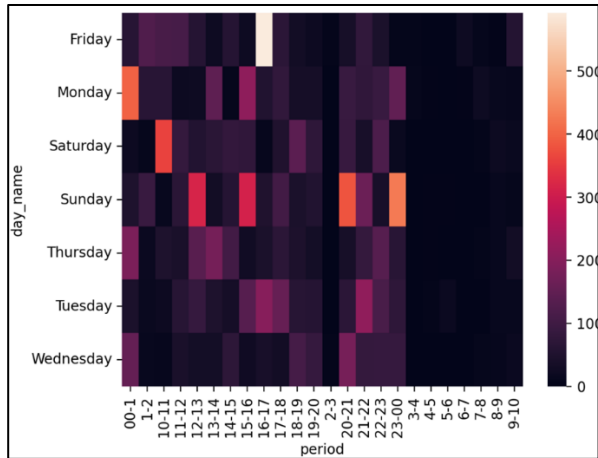


Fig. 4. Weekly Activity HeatMap

Matplotlib is used to calculate the frequency and plot the chart for the graph, which displays the active users and their contributions to the conversation as illustrated in fig 5. Which helps in identifying the most active users and those who contribute the most to the conversation is an essential aspect of the data analysis process.

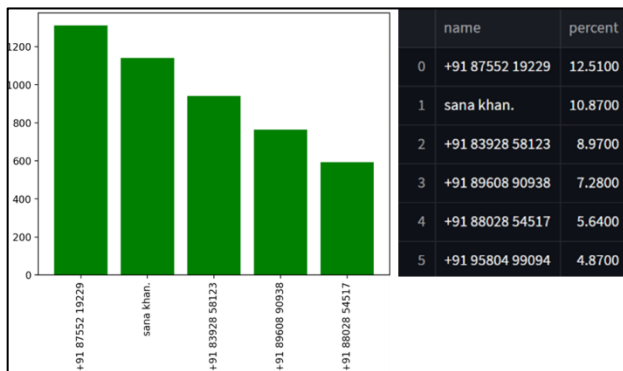


Fig. 5. Most Busy user

A graph is plotted to display the group's most frequently used word (fig 6). It is possible to determine the language patterns and communication preferences of the group by looking at the most frequently used words. In order to bring emphasis to the most frequently used words in the discussion, the words with the highest frequency are often shown in larger font sizes or bolded. This function can also be utilized to spot any odd or unexpected communication patterns inside the group, including the sudden appearance of a new subject or trend.

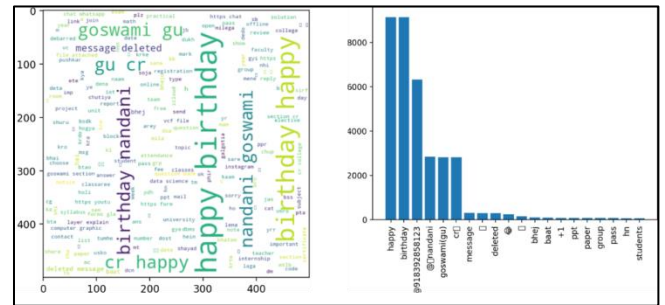


Fig. 6. Word Cloud & Most Common Word

Text sentiment is examined using "Text-Blob" and "Vader-Sentiment." It will determine whether the text falls under subjectivity (refer to one's own opinion, emotion, or judgement while objective refers to factual facts) or polarity (float that falls between [-1,1], with 1 denoting a positive statement and -1 denoting a negative statement) in metric form. Under token emotion, it will also display words that are positive, negative, and neutral.

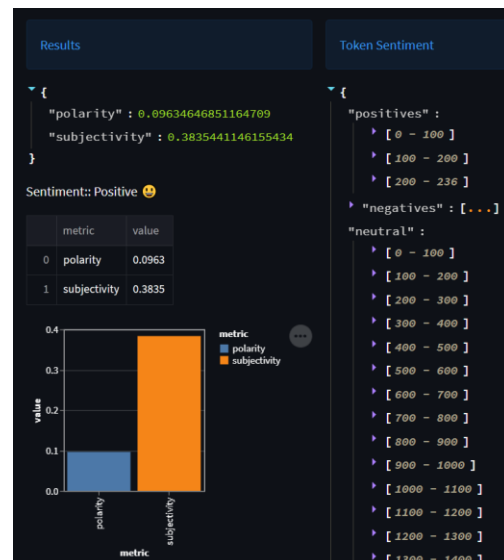


Fig. 7. Result Of Sentiment Analysis

VII. CONCLUSION AND FUTURE SCOPE

Expanding the chat dataset can have a significant impact on the accuracy and reliability of the chat analysis tool. The larger the amount of data the system can analyze, the more informative and precise its findings become. With a vast dataset, the tool can effectively detect patterns and trends in chat conversations, enabling supervisors to gain insights into their teams' communication styles. By incorporating the Streamlit library in python, it's possible to create an interactive dashboard that allows users to explore data and visualizations in real-time. This feature can be invaluable for supervisors who need to monitor chat conversations regularly. With a customized dashboard, they can quickly identify any issues or trends that require their attention.

Overall, this chat analysis tool has the potential to transform how supervisors evaluate chat interactions. By providing precise and actionable insights into team communication, it can assist them in identifying areas for improvement and enhancing the overall efficiency and productivity of their organization.

REFERENCES

- [1] M. Dickson, "Examination of Yahoo Messenger 7.0 Contact Identification," *Digital Investigation*, vol. 3, pp. 224-228, 2006.
- [2] J. Ho, P. Ji, W. Chen, and R. Hsieh, "Identifying Google Chat," in *IEEE International Conference on Intelligence and Security Informatics (ISI '09)*, 2009, pp. 157-162.
- [3] Ravishankara K, Dhanush, Vaisakh, and Srajan I S, presented paper in *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, issue 05, pp. 33-37, May 2020.
- [4] S. Prabhakaran and Dr. D. Lakshminarayanan, presented paper titled "DogoRangsang Research Journal", vol. 10, issue 07, no. 12, pp. 22-30, July 2020.
- [5] E. Winarko, A. Cherid, and co-authors, "Recognizing the Caustic Remark Made in an Indonesian-language WhatsApp Group," in *Broadband Communication, Wireless Sensors and Powering (BCWSP) conference*, IEEE, pp. 1-6, 2017.
- [6] Statista, "Number of monthly active WhatsApp users worldwide from April 2013 to February 2016 (in millions)," 2016. [Online]. Available: <http://www.statista.com/statistics/260819/number-of-monthly-active-WhatsApp-users>.
- [7] D. M. El-Din, "Enhancement of the Bag-of-Words Model to Address the Difficulties in Sentiment Analysis," *Advanced Computer Science and Applications International Journal*, vol. 7, no. 4, pp. 80-87, 2016.
- [8] Analytics Vidhya, "WhatsApp Group Chat Analysis with Python," 2021. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/04/whatsapp-group-chat-analyzer-using-python/>.
- [9] Analytics Vidhya, "Build Web App Instantly for Machine Learning - Streamlit," 2021. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/06/build-web-app-instantly-for-machine-learning-streamlit/>.
- [10] IEEE, "Natural Language Processing is Used to Analyse People's Behaviour in Chat Messages," 2021. [Online]. Available at: <http://ieeexplore.ieee.org/abstract/document/9388596>.
- [11] Medium, "Sentiment Analysis of WhatsApp Chat". 2021. [Online]. Available at: <https://medium.com/analytics-vidhya/sentiment-analysis-of-whatsapp-chat-a906105917f4>.
- [12] H. Khalid Drus, "Sentiment Analysis in Social Media and Its Applications: A Systematic Literature Review," *Z Procedia Computer Science*, vol. 158, pp. 245-252, 2019.
- [13] A. Aleem, A. Kumar, and M. M. Gore. "A Study of Manuscripts Evolution for Perfection". 2nd International Conference on Advanced Computing and Software Engineering (ICACSE-2019), pp: 278-282, 2019.
- [14] A. Aleem, P. Tewary, S. Karn, and V. Kumar. "Entertainment Advisor Using Sentiment Analysis". *Journal of Scientific Research* 66 (2), pp. 101-107, 2022