Whatsapp Chat Analysis Based On NLP Using ML

[1]Dr.Sneha Jondhle, [2]Mr.Harsh Jondhle,[3]Miss.Insha Mulla , [4]Mr.Ibad Nachan [5]Mr.Prashik Tayde

[1]Associate Prof, [2]Asst.Prof,[3,4,5]UG Student,[1,2]EXTC Engg, [3,4,5]Computer Engg.

[1]Shivajirao S. Jondhale College of Engineering, Dombivli, [2,3,4,5]Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

[1]snehajondhle@gmail.com,[2]harshjondhle@gmail.com,[3]inshasmulla@gmail.com,

[4]ibadnachan@gmail.com, [5]tprashik505@gmail.com

Abstract - WhatsApp has an extremely user-friendly user interface, which is one of its main features. WhatsApp is the first messaging app that springs to mind when you want to connect or interact with someone. Everyone finds it convenient to communicate via WhatsApp. To send a message, just open the app, choose Contact, and start typing. You're done there. The user can utilize WhatsApp chats to examine discussions with peers, customers, or big groups of people, as the messaging app receives over 65 billion messages daily. To discover fascinating details about the most active people, the ghosts (those who do conversation research on a college student WhatsApp group conversation is used to uncover fascinating information like who is most active, who are ghosts (those who don't reply often), the most popular emoji, the busiest times of the day, or the most common phrases. These would surely be some enlightening findings from the chat data. [1]

Keywords- WhatsApp Chat, Python, Streamlit, Analysis, Nature Language Processing, Emoji, Pandas, Matplotlib

I. INTRODUCTION

Web applications can be developed on a server using Python. It can be used in conjunction with programs to create workflows.. Python allows for execution of intricate mathematical operations. Numerous Python libraries are utilized, such as: An open-source Python framework called Stream-lit that enables users to create and distribute visually appealing, complex data apps may be developed and implemented.

The main purpose of the open-source pandas library is to make dealing with relational data efficient. It offers a range of methods and data structures for handling numerical and time series data. Based on NumPy arrays,A multiplatform data visualization tool is called Matplotlib ,toolkit intended to be used with the extensive SciPy stack. A Python class called URL Extract gathers (extracts) URLs from input. Emoji analysis makes use of emoji. WhatsApp chat review is also great important application as it helps a lot as most of the time, a person buys or cancel a product only based on reviews. Thus, it is clear to show the importance of reviews. Although, it will be quite difficult to go through thousands of reviews whenever a person thinks of buying a product. Thus, it will be good to scratch out useful info from these reviews [1][10]

II.AIMS AND OBJECTIVE

a) Aim

The project makes sure to offer a thorough exploratory data analysis on a range of WhatsApp chats

b) Objective

Recognizing consumer behavior: Businesses can learn about the requirements, preferences, and behavior of their customers by examining chat interactions. This data

can be utilized to enhance product offers, create more focused marketing campaigns, and provide better customer service.

Enhancing team collaboration: Teams can better understand communication patterns, pinpoint bottlenecks, and enhance communication by working on a chat analysis project.

Better results and increased productivity may result from this and help the one to reach the correct taught which increases productivity.

III.LITERATURE SURVEY

Paper 1: "Content Analysis of WhatsApp Conversations: An Analytical Study to Evaluate the Effectiveness of WhatsApp Application in Karachi".

As to the August 4, 2016, publication of the UK Communications Market Report (CMR), instant messaging

 DOI : 10.35291/2454-9150.2024.0256

has emerged as a highly favored mode of communication. In 2016, the percentage of individuals utilizing instant messaging services increased from 28% to 43%. In the UK, there is a fall in texting and emailing for the first time. According to 82% of respondents, using smartphone apps like SnapChat and WhatsApp facilitates happier and simpler communication with family and friends. As a result, the number of users of instant messaging apps has significantly increased. According to the study's findings, Facebook Messenger has the most user base (22.5 million), followed by WhatsApp (16.7 million).

An investigation into the potential impact of instant messaging on interpersonal relationships was conducted qualitatively at the Chinese University of Hong Kong. In this context, ten informants of various ages and backgrounds participated in a structured interview. The interviewees listed a number of advantages and disadvantages of instant messaging. Several informants highlighted the benefits of affordaability, effectiveness, portability, and simplicity of use.[3]

Paper 2: "Survey Analysis on the usage and Impact of WhatsApp Messenger"

These studies cover WhatsApp's effects on students and young people. According to the survey, people in the range of 18 and 23 in the southern region of India use WhatsApp for roughly 8 hours per day and are occasionally online for up to 12 to 16 hours per day. The majority of them acknowledged that WhatsApp was their preferred website. They share videos, audio, and photos. This study also demonstrated that WhatsApp is the program that is used on smartphones the most, compared to all other apps. Since WhatsApp is the most popular app among young people and other generations, this project can give them insights into their discussions and reveal unknown truths to them.This study was done to determine WhatsApp usage's benefits and drawbacks. As this survey has revealed, of course. [4]

Paper 3: "WhatsApp Group Data Analysis with R."

The WhatsApp group chat dataset used for the analysis spans a year, from May 2015 to May 2016, with a total of 5,5563 records. It includes various attributes that indicate the extent to which a specific user uses the group, including the number of years of usage, the amount of time spent using it each day, response levels, the types of messages posted by group members (Smiley, Text, and Count) which age-groups have the most active members, and so forth.The primary attributes used in this analysis are th age sent, how long it has been used for in a year, month, week, day, or hour, the timestamp (AM/PM), the senders' age group, and their gender (male or female).Because it is open-source, RStudio is the most widely used integrated development environment for R and is used for exploratory data analysis and visualization of the gathered data. [5]

IV.EXISTING SYSTEM

The existing system has undergone significant development. No features for sharing documents, sharing locations, or displaying status were available in the previous version. All of these features are present in the current version.

The user were unable to exchange photographs via the Doc format on previous versions. This approach allows users to use WhatsApp on Windows via the web application, which can be connected via QR code. Another function is export chat, which allows users to communicate, share, or get chat information for analysis of data. A The latest version also introduces a robust export chat feature, enabling to analyze chat data across various platformsPeople rely on online products from food to cloth and from home to electronics, rather than going outside. Several products are available on these platforms by different brands. Thus, it will be quite difficult to choose a product that is useful and reliable. To get a useful product, a user goes through the reviews of the product, to understand the product and to decide whether to purchase it or not. [2][10]

V.COMPARATIVE STUDY

Sr. No.	System Title	Authors	Objective	Methodology /Technique	Result/Performance
1.	Analyzing WhatsApp Chat Using Python Libraries	Vanshika Kooper, Bhoomi Gupta, Deepika Bansal.	Analyse WhatsApp chats using Python libraries to derive insights into user behaviour and communication patterns.	Semantic Analysis	For Analysing WhatsApp chat data
2.	Content Analysis of WhatsApp Conversations: An Analytical Study to Evaluate the Effectiveness of WhatsApp Application in Karachi	Sana Shahid	Analyze WhatsApp usage: message frequency, media sharing, day-night activity, conversation differences.	Matplotlib and Seaborn, Word Cloud, Word Cloud Generation, Sentiment Analysis	WhatsApp is integral for students and professionals, facilitating communication
3.	Survey Analysis on the usage and Impact of WhatsApp Messenger	Naveen Kumar and Sudhansh Sharma	Investigate the usage patterns and impact of WhatsApp Messenger	Pandas, NumPy, Communication tools, Data analysis tools.	. Limited Indian literature on WhatsApp usage, but studies indicate increased connectivity and

 DOI : 10.35291/2454-9150.2024.0256

			through survey analysis.		intimacy through the app, particularly among youth
4.	WhatsApp Group Data Analysis with R	Sanchita Patil	Project aims to understand user behavior and engagement patterns.	Methodology involved exploratory data analysis and visualization in RStudio	Insights included user activity trends, response levels, and demographics.

VI.PROBLEM STATEMENT

WhatsApp-Analyzer provides statistical analysis for WhatsApp talks. WhatsApp conversation files can be exported to make charts indicating which participants a user most frequently interacts with. This project uses dataset modification approaches to improve understanding Whatsapp Communication on our phones.[1].

VII. PROPOSED SYSTEM

The suggested solution would use automatic data extraction techniques to compile information from WhatsApp talks. This would mean using specialized software or algorithms to extract the data straight from WhatsApp servers, rather than requiring users to copy and paste chat logs by hand. This approach would reduce the likelihood of biases or errors and ensure that the data is collected consistently and reliably. After the data of WhatsApp chat has been retrieved, the proposed system will apply advanced data cleaning and pre-processing techniques to get it ready for analysis.In order to prepare the data for analysis, it would be necessary to eliminate duplicate data, deal with missing values, and arrange the data in a systematic manner. The suggested system would use state-of-the-art data cleaning and pre-processing techniques to confirm that the data is of the highest quality and ready for more in-depth analysis. It would use advanced analysis techniques to mine the data of WhatsApp chat to identify patterns and insights. This would mean using methods for natural language processing (NLP) to identify keywords and the conversation's emotional undertone in addition to machine learning algoritnms to identify trends and patterns in the data.By utilizing these techniques,the suggested system would be able to provide more accurate and comprehensive insights regarding user behavior and preferences. sophisticated analytic methods on pages. Thanks to interactive visualization and reporting capabilities, users may review the data and see the outcomes of the research. This would require the creation of interactive dashboards and reports that enables users to investigate more deeply and examine the data in more depth. By providing more dynamic and user-friendly visualization and reporting features, the suggested system would aid users in understanding and interpreting the study's findings. Several essential elements of the proposed WhatsApp Chat Analysis system contribute to the process's increased scalability, accuracy, and speed.By leveraging automated data extraction, advanced analysis methods, interactive visualization, and reporting, the proposed system would be able to provide more accurate and thorough knowledge about user behavior and preferences. putting data security and privacy first , the proposed approach will also boost user confidence and ensure that the analysis is conducted in an ethical and responsible manner. The proposed methodology holds significant potential to enhance the utility and worth of WhatsApp Chat Analysis for researchers and enterprises.

VIII. ALGORITHM

1. User Interface Setup:

Display the Streamlit sidebar with the title "WhatsApp Chat Analyzer".

Provide a file uploader for users to upload a WhatsApp chat data file.

```
import streamlit as st
st.sidebar.title("WhatsApp Chat Analyzer")
uploaded_file= st.sidebar.file_uploader("Choose a file")
```

2. Data Processing:

Read the uploaded chat data.

Preprocess the data to extract useful information like user, message content, date, etc.

```
import preprocessor
if uploaded_file is not None:
bytes_data = uploaded_file.getvalue()
data = bytes_data.decode("utf-8")
 df=
preprocessor.preprocess(data)cnn_model.add(Dense(128,
activation='relu'))
cnn_model.add(Dense(2, activation='softmax'))
```

3. User Selection and Analysis Trigger:

Allow the user to select a specific user for analysis.

Trigger the analysis upon clicking the "Show Analysis" button.

```
if st.sidebar.button("Show Analysis"):
 # Perform analysis...
```

4. Analysis and Visualization

Calculate statistics like total messages, words, shared media, and links shared.

Display the statistics using Streamlit components.

```
import helper
num_messages, words, num_media_messages, num_links
= helper.fetch_stats(selected_user, df)
st.title("Top Statistics")
st.header("Total Messages")
st.title(num_messages)
```

 DOI : 10.35291/2454-9150.2024.0256

Display other statistics...

5. Data Processing and Analysis:

Import necessary libraries and modules.
Define functions for statistical analysis, semantic analysis, etc.

```
from urlextract import URLExtract
from wordcloud import WordCloud
from collections import Counter
import pandas as pd
import re
import emoji
import matplotlib.pyplot as plt
import seaborn as sns
from model import load_sematic_analysis_pipeline
from model import load_sentiment_analysis_pipeline
import numpy as np
from sklearn.cluster import KMeans
import streamlit as st
```

6. Semantic Analysis:

Define functions for semantic analysis using pre-trained models.
Perform clustering and sentiment analysis on the chat data.

```
def plot_sematic_analysis(selected_user, df):
# Semantic analysis implementation...
```

7. Streamlit Integration:

Set up Streamlit interface with sidebar and file uploader.
Perform data processing and analysis upon file upload.
Integrate semantic analysis into the Streamlit app.

```
if __name__ == "__main__":
st.sidebar.title("WhatsAppChat Analyzer")
uploaded_file=  st.sidebar.file_uploader("Choose a file")
# Display statistics...
# Perform semantic analysis
plot_sematic_analysis(selected_user, df)
```

IX.MATHEMATICAL MODEL

Let's denote the following variables:
(N): Total messages exchanged in the chat.
(W): Total word count across all messages.
(M): Number of media files (images, videos, etc.) shared.
(T_d): Timeline of chat activity (daily basis).
(T_m): Timeline of chat activity (monthly basis).
(U_i): User (i) (where (i = 1, 2, \ldots, n)) in the chat.
(A_i): Count of messages sent by user (i).
(C_{ij}): Count of messages from user (i) to user (j).
(W_f): Word frequency distribution.
(E): Emoji analysis.
Now, let's express some key insights:
Total Chat Activity:
(N = \sum_{i=1}^{n} A_i)
Word Count:
(W = \sum_{i=1}^{n} \text{word_count}(U_i))
Media Sharing:
(M = \text{count_media_files})
Busiest Days and Months:
(T_d): Plot daily chat activity.
(T_m): Plot monthly chat activity.
User Analysis:Most active user: (max(A_i))Interactionmatrix:C{ij}

X.SYSTEM ARCHITECTURE

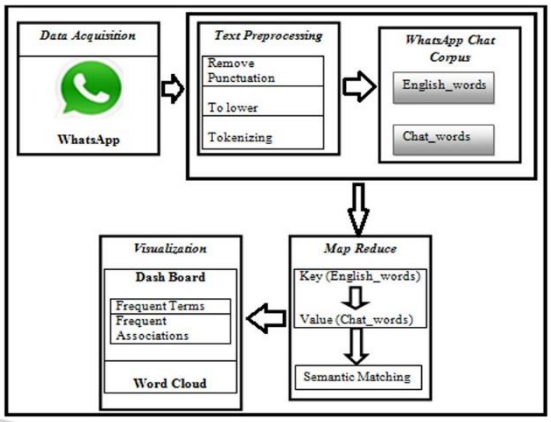

Fig.1: System Architecture

Here's a step-by-step guide for the architecture flow in the WhatsApp Chat Analysis Project:

1. Data Extraction:
Obtain chat data directly from WhatsApp servers through human or automated means.
Store the extracted data securely in a database for further processing.

2. Data Pre-processing and Cleaning: Organize and transform the data for analysis.
Clean the data by resolving missing values and eliminating noise, and structuring the data appropriately.

3.Natural Language Processing (NLP):
Analyze the text data from WhatsApp chats using NLP techniques.
Identify keywords, gauge sentiment, and recognize specific entities mentioned in the chats.

4. Machine Learning:
Apply machine learning methods to discover patterns and trends within the data.
For labeling, use classification algorithms and clustering algorithms to group related discussions.
Create predictive models using past data to forecast future trends.

5. Visualization and Reporting:
Create interactive dashboards and reports to present analysis results.
Use visualization equipments are used to generate charts and graphs.

6. Data Security and Privacy:
Implement robust Limits on access and encryption to safeguard information from unwanted access.
Adhere to industry best practices for data privacy and security.

 DOI : 10.35291/2454-9150.2024.0256

This structured approach enables businesses and researchers to gain applicable insights from WhatsApp chat data while upholding high standards of data security and privacy.

XI.ADVANTAGES

- Sentiment analysis drives proactive feedback response and boosts customer and employee satisfaction.
- Automated responses and issue prioritization streamline customer support, enhancing response times and resolution efficiency.
- Chat data analysis informs trend identification and market research, facilitating informed decision-making and adaptive strategies.
- Feedback extraction and anomaly detection aid in product improvement and fraud detection, respectively, fostering trust and product integrity.
- Multilingual support through NLP-driven translation expands market reach and fosters global communication, positioning businesses for sustained growth and inclusivity.

XII.DESIGN DETAILS

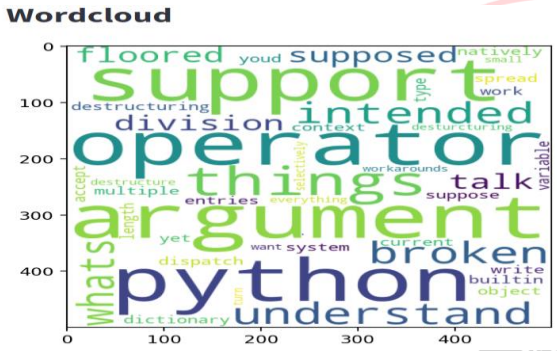

Fig 2: Wordcloud

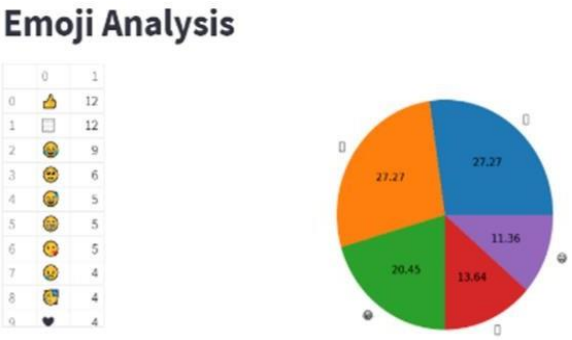

Fig 3: Emoji Analysis

Fig 4: Top Statistics

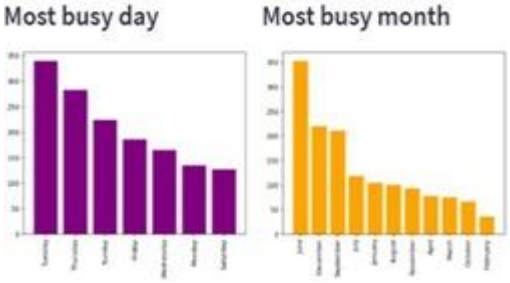

Fig 5: Activity Map

XIII.CONCLUSION

Thus this research try to impement the paper “Analyzing WhatsApp Chat Using Python Libraries” shah, Amar Jaiswal.IEEE,2023 it has concluded that data analysis and visualization of the exported WhatsApp chat may be accomplished using Python and its libraries. Whatsapp is a popular tool for communication. As a outcome, it contains a large amount of chat data that may be examined to learn more. Only members of WhatsApp groups may collect chat data, and it is presumed that the sender has not removed any messages to avoid misleading results from the research. The app displays the intended outcome, such as the monthly A range of data can be visualized with the matplotlib package, including timeliness, daily timeliness, busiest month and day, most active user, weekly activity map, etc. The user can also be able to analysis the word which he had used most of the time in the conversation as well as the user will able to analyse the emoji and analyse the behaviour of the chat. There may be additional room in the chat analysis for one to look for other pertinent information.

XIV.REFERENCE

[1] Rajan,Bhoomi Gupta,Vanshika Kapoor,Deepka Bansal Department of IT Maharaja Agrasen Institute of Technology “Analyzing WhatsApp Chat Using Python Libraries” shah, Amar Jaiswal.IEEE,2023

[2] Ravishankara K, Dhanush, Vaisakh, Srajan I S, “International Journal of Engineering Research & Technology (IJERT)”, ISSN: 2278-0181, Vol. 9 Issue 05, May-2020

[3] Sana Shahid. “Content Analysis of Whatsapp Conversations: An Analytical Study to Evaluate the Effectiveness of Whatsapp Application in

 DOI : 10.35291/2454-9150.2024.0256

Karachi". International Journal of Media, Journalism and Mass Communications (IJMJMC), vol 4, no. 1, 2018, pp. 14-26. doi:http://dx.doi.org/10. 20431/2455-0043.0401002

[4] Naveen Kumar and Sudhansh Sharma "Survey Analysis on the Usage and Impact of Whatsapp Messenger", Global Journal of Enterprise Information System. Volume-8, Issue-3,July-September,2016. (http://informaticsjournals.com/index.php/gjeis)

[5] Sanchita Patil. "Content Analysis of Whatsapp Conversations: WhatsApp Group Data Analysis with R". International Journal of Computer Applications (0975 – 8887) Volume 154 – No.4, November 2016.

[6] Ahmed, I., Fiaz, T., "Mobile phone to youngsters: Necessity or addiction", African Journal of Business Management Vol.5 (32), pp. 12512-12519, Aijaz, K. (2011).

[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] F. Meng Cai, "PubMed Central", PMCID: PMC7944036, PMID: 33732917.

[9] Radhika, Narendiran, "Kind of Crops and Small Plants Prediction using IoT with Machine Learning," International Journal of Computer & Mathematical Sciences April 2018, pp. 93-97.

[10] Prof. Satish Manje, "Sentiment Detection for Amazon Product Review" in IJREAM, ISSN : 2454-9150, Volume09,Issue 01,Special Issue, April 2023.

 DOI : 10.35291/2454-9150.2024.0256