



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.379**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# WhatsApp Chat Analyzer

Suraj Gangarde, Rutvik Patil, Smith Gholap, Vivek Vishwakarma

UG Student, Dept. of Computer Engineering, ISB&M College of Engineering, Pune, India

**ABSTRACT:** Nowadays, almost everyone uses the WhatsApp application for messaging and chatting. We can communicate with each other and transfer photos, videos, and messages through this platform. However, some people have become addicted to using their mobile phones and WhatsApp, causing them to waste valuable time chatting. Therefore, we have developed a model that can analyze individual or group chats on WhatsApp using machine learning and data science.

In this model, we utilized various Python libraries such as pandas, matplotlib, seaborn, word cloud, and Emoji. The model provides valuable insights into the time users spend on chatting and their level of involvement in the conversation. WhatsApp Analyzer is a tool that allows us to analyze our WhatsApp group activities by tracking our conversations and determining how much time we spend, or rather "waste," on the app. This article aims to provide a step-by-step guide to building our own WhatsApp analyzer using Python. To extract useful information from raw data, we utilized various Python libraries. We selected our college's official WhatsApp group to analyze the patterns followed by students. As a result, we had to blur the contact information of my college faculty and classmates in some of the snapshots. We apologize for any inconvenience caused by this.

**KEYWORDS :** Machine Learning, WhatsApp, Streamlit, Python, Seaborn.

## I. INTRODUCTION

The initial step in applying a machine learning algorithm is to determine the appropriate learning experience to start with, which will help the model improve. Data pre-processing is crucial in machine learning. To enhance the efficiency of the model, a large amount of data is required. Therefore, we have directed our attention to WhatsApp, one of Facebook's largest data producers. According to WhatsApp, approximately 55 billion messages are transmitted every day. The average user spends around 195 minutes each week on WhatsApp and is a member of various groups. This abundant source of data provides us with an opportunity to gather insights into the messages we receive on our phones. Hence, it is essential to explore this valuable resource.

WhatsApp is a messaging app that permits users to send messages, chat and share media files, documents and applications. It is available across all platforms and can be used on various operating systems like Android, Apple, and Windows iOS. The app enables communication between multiple users in a group, and broadcast messages can be sent to up to 256 users at once. It facilitates instant messaging, voice calls and sharing of images and videos via an internet connection.

This model involves an analysis of WhatsApp group chats and users to determine their level of participation and involvement in the conversation. The goal is to evaluate the level of engagement and participation in a particular WhatsApp group. The analysis includes identifying the most active date in the group, the number of messages sent on the most active date, the most active user overall, the list of active group admins, the total number of users, the number of posts made by each individual in the group, and the most frequently used word on the platform.

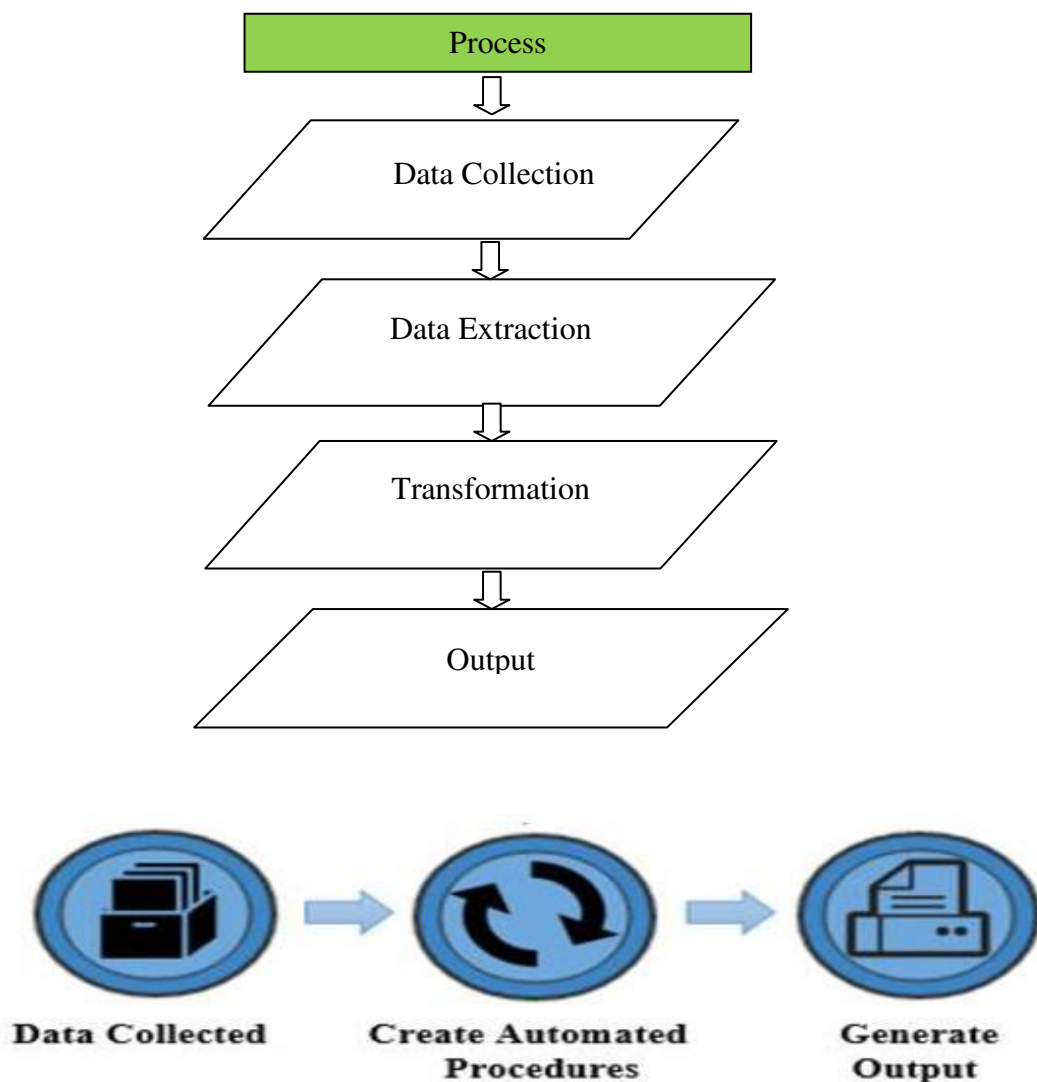
## II. LITERATURE REVIEW

An examination was conducted on the use of WhatsApp groups in various settings by different researchers. In one study, the aim was to predict an individual's addiction level to WhatsApp groups based on their gender and age group using R statistics software. The analysis was performed on data from a specific WhatsApp group to determine the most preferred communication medium, the most active day of the week, and the age group that was most active. The study also aimed to determine if males were more addicted to the WhatsApp group than females.

In another study, the focus was on the use of WhatsApp groups in high school classrooms. The goal was to understand how the application affected the communication between teaching staff, faculty staff, and students, as well as the impact it had on educational and academic output.

In a third study, WhatsApp was analyzed as a communication tool among emergency surgery teams in a London hospital. The study examined the response times and communication types of team members who used WhatsApp for 19 weeks. Safety events were also reported. Finally, a review of the literature was conducted to explore the various applications of WhatsApp and Instagram in health and healthcare. The review described the main issues surrounding the uses of these applications in health and medicine, using evidence from published literature and case reports indexed in PubMed and other sources. One issue identified was that WhatsApp can consume students' study time and lead to a lack of concentration during lectures.

### III. FLOWCHART



### IV. STEPS TO EXPORT CHAT

- Step 1: Go to WhatsApp.
- Step 2: Select the group that you want to analyze.
- Step 3: Click on the three dots in the top right corner.
- Step 4: Click on the "More" option.
- Step 5: Select the "Export Chat" option.

Step 6: Choose either "Without Media" or "With Media."

Step 7: Your chat is now ready for analysis.

## V. MODULE

### 1. Machine learning And Data Science:

Machine learning (ML) is a field of inquiry devoted to understanding and building methods that 'learn', that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, agriculture, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks. A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers, but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning. Some implementations of machine learning use data and neural networks in a way that mimics the working of a biological brain. In its application across business problems, machine learning is also referred to as predictive analytics.

### 2. LIBRARIES:

- a) **Streamlit:** Streamlit is a free and open-source python framework. We can quickly develop web apps for Machine Learning and Data Science by using Streamlit. Streamlit can easily integrates with other popular python packages such as NumPy, Pandas, Matplotlib, Seaborn. Streamlit provides fastest way to develop and deploy web apps
- b) **Matplotlib:** Matplotlib is a popular Python packages used for data visualization. It is a cross-platform library for making plots from data in arrays. It helps in creating static, animated and interactive visualizations in python. Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications. A Python matplotlib script is structured so that a few lines of code are all that is required in most instances to generate a visual data plot. The matplotlib scripting layer overlays two APIs: • The pyplot API is a hierarchy of Python code objects topped by matplotlib.pyplot • An OO (Object-Oriented) API collection of objects that can be assembled with greater flexibility than pyplot. This API provides direct access to Matplotlib's backend layers.
- c) **Seaborn:** Seaborn is the data visualization library. It is used for making statistical graphs. Visualization is the central part of seaborn. Seaborn provides exploration and better understanding of data. Seaborn closely integrates into the data structures from python. Seaborn is an amazing visualization library for statistical graphics plotting in Python. It provides beautiful default styles and color palettes to make statistical plots more attractive. It is built on the top of matplotlib library and also closely integrated to the data structures from pandas. Seaborn aims to make visualization the central part of exploring and understanding data. It provides dataset-oriented APIs, so that we can switch between different visual representations for same variables for better understanding of dataset.
- d) **Word Cloud:** Word Cloud is a data visualization library used for representing most frequently used words within a given text. Most frequent and important words are represented in bigger and bolder size. Word Clouds (WordClouds) are quite often called Tag clouds, but I prefer the term word cloud. It think this term is more general and easier to be understood by most people. The term tag is used for annotating texts and especially websites. This means finding out the most important words or terms characterizing or classifying a text. In the early days of web development people had to tag their websites so that search engines could easier classify them. Spammer used this to manipulate the search engines by giving incorrect or even misleading tags so that their websites ranked higher. Google changed this by automatically finding out the importance of the text components. Google more or less disregarding the tags which the owners of the websites assigned to their pages. "Word clouds" as we



use them also find out automatically what are the most important words. Of course, we do it naively by just counting the number of occurrences and using stop words.

- e) **Pandas:** Pandas is an open-source python library. Pandas used to convert string data into Data frame. Data frame is the representation of data into 2-dimensional table of rows and columns. We can work with large data sets using Pandas library. Pandas library has many built-in functions for data analysis, data cleaning, data exploration and data manipulation In 2008, developer Wes McKinney started developing pandas because he needed a high performance, flexible tool for analysis of data.

### VI. MODULE WORKING AND OUTPUT

#### A. Working Steps:

- 1) Firstly, access the webpage for WhatsApp Chat Analysis.
- 2) Choose the preferred date format.
- 3) Upload the exported chat file to the webpage.
- 4) The trained model will then begin analyzing the data.
- 5) The trained model also performs pre-processing of the data.
- 6) Choose between overall or single person analysis.
- 7) The trained model provides a variety of analyses, including top statistics, word cloud, activity map, monthly timeline, daily timeline, and emoji analysis.

#### B. Output:

Total Messages	Total Words	Media Shared	Links Shared
3956	47063	620	980

Fig.1. Shared Messages,Links,

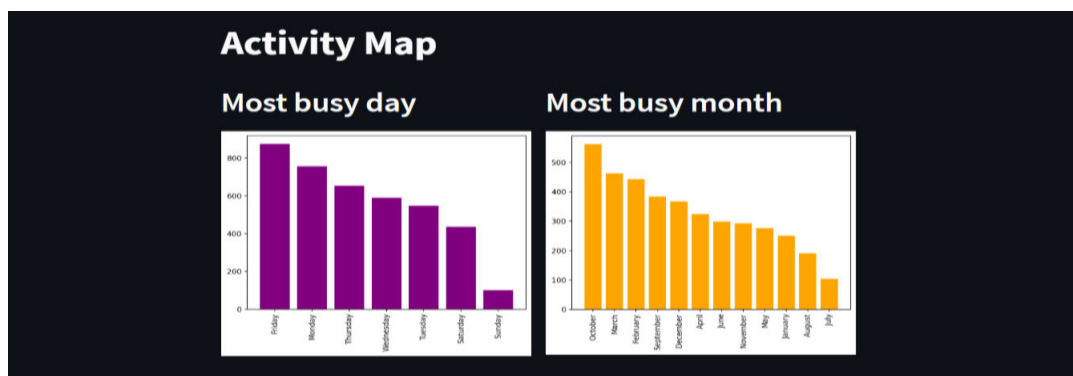


Fig. 2. Activity Map

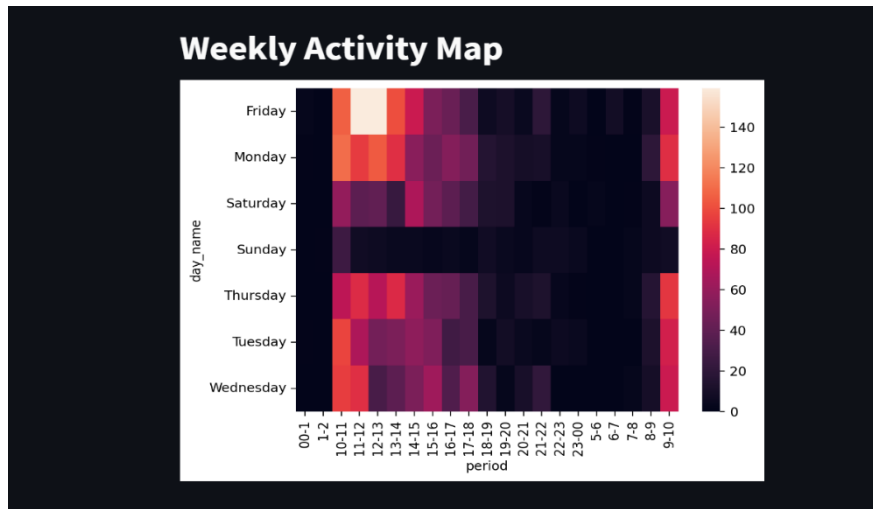


Fig. 3. Activity Heat Map

## VII. CONCLUSION

The WhatsApp application and the Python programming language are both powerful tools that can be utilized for network data analysis. In this work, we will explore the functionalities of WhatsApp and its accompanying libraries, as well as demonstrate how they can be used to analyze a group chat. Specifically, we will provide visual representations of the top 10 and top 20 users within the group chat. It is important to acknowledge the usefulness of these tools in conducting such analyses.

## REFERENCES

1. Blessing Nwamaka Idue Academic Researcher, Computer Science Department, Nnamdi Azikiwe University, Awka, Nigeria, International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS) Volume IX, Issue II, February 2020 | ISSN 2278-2540
2. Ravishankara K, Dhanush, Vaisakh, Srajan I S 1Faculty of Department of Computer Science, Srinivas Institute of Technology Mangalore, Karnataka 2,3Students of Department of Computer Science, Srinivas Institute of Technology Mangalore, Karnataka, International Journal of Engineering Research & Technology (IJERT) <http://www.ijert.org> ISSN: 2278-0181 IJERTV9IS050676 (This work is licensed under a Creative Commons Attribution 4.0 International License.) Published by : [www.ijert.org](http://www.ijert.org) Vol. 9 Issue 05, May-2020



**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.379**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details