

---

## Research Paper

# WhatsApp Analyzer: A Tool to Measure the User Performance in Social Platform

Amrut Ranjan Jena<sup>1\*</sup> , Pratyush Kumar<sup>2</sup> , Rafiqul Islam<sup>3</sup> 

<sup>1,2,3</sup>CSE, GNIT, Kolkata, India

\*Corresponding Author: amrutranjan.jena@gnit.ac.in

---

**Abstract:** An application called WhatsApp has emerged as the most popular and effective means of communication in recent years. The heroku web application called WhatsApp Chat analyzer provides analysis of WhatsApp groups. In this paper authors applied matplotlib, streamlit, seaborn, re, pandas, and certain NLP concepts for analyzing WhatsApp chat. Here authors combine machine learning with NLP. This WhatsApp conversation analyzer imports a user's WhatsApp chat file, analyses it, and produces various visualizations as a consequence.

**Keywords:** WhatsApp chat analyzer, NumPy, Pandas, NLP, Matplotlib

---

## 1. Introduction

The authors of this study suggested a WhatsApp chat analyzer. Different forms of communication between groups and individuals are included in WhatsApp chats. Different subjects are covered in this talk. This could give machine learning technology more data to work with. The correct learning experience is offered by machine learning models, which is a crucial factor that is indirectly impacted by the data supplied to that model [1]. This programme offers analysis of the information that WhatsApp provides. This application has the benefit of being implemented by straightforward Python modules, such as seaborn, pandas, NumPy, streamlit, and matplotlib, which are frequently used for building data frames and other graphs [2]. This is shown on the web via a Heroku link, which is accessible from any device with a browser [3].

## 2. Related Work

### Existing System:

There was no WhatsApp chat analysis in the past. There isn't a CSV file available for analysis if someone wants to do it [4]. The raw formatted text file can be exported from the WhatsApp application. It is really difficult to analyse. Therefore, we must disregard that system and instead use the WhatsApp Chat Analyzer.

### Disadvantages of Existing System:

- Raw data.
- Time consuming.
- Difficult to Analyze.
- Analysis is not accurate.

- Proposed System

The "WhatsApp Chat Analyzer" gives users a platform to analyze WhatsApp talks online via a Heroku connection [5]. With this programme, users can read exported WhatsApp (.txt) files, import them into the WhatsApp chat analyzer, and receive analysis based on those txt files. By selecting the Show Analysis option, the user can also analyze [6] [7].

### Advantages of WhatsApp Chat Analyzer:

- Runs on all devices.
- Programs based on the WhatsApp chat log.
- Sum of all words.
- Sharing of connectivity.
- Monthly timeline.
- Most busy day.
- Most busy month.
- Weekly activity.
- Most busy users.
- Most used words.
- Emoji analysis.
- Displays several visuals. Sum of all messages.
- Sharing of media.

## 3. Design Methodology

This section explores the tools used to develop the working model [8] [9] [10].

### Python

It is a programming language with several uses. It offers various library types that give projects various functionalities. In order to make predictions and identify patterns, Python is

employed. Python has a large number of libraries that offer mathematical and statistical functions that aid in deriving analysis of data.

### Pandas

The data science and machine learning industries frequently use the open-source Python library. This package provides data analysis software, and these tools are used to analyse time series analysis and numerical data utilising its data structures.

### NumPy

NumPy, which derives its name from "Numerical Python," is a Python data analysis library that includes a number of numerical methods and functions for numerical analysis as well as multi-dimensional array objects and a collection of routines to process them.

### Matplotlib

Matplotlib, a visualization toolkit for Python, is an outstanding and user-friendly tool. It uses the larger SciPy stack and is based on NumPy arrays. It has a variety of plots, including pie, line, bar, graph, scatter, histogram, etc. For the numerous visualizations utilised in this project's analysis of WhatsApp messages, Matplotlib is used. The use of visualizations like pie charts, bar charts, and line charts.

### Seaborn

Python's Seaborn package is mostly used for statistical graphing. It offers lovely colour palettes and default styles to make statistics charts more aesthetically pleasing. Seaborn is utilised in this project to create a heatmap that displays 24 hours, 7 days, and various colour gradations for messages ranging from the most important to the least important.

### Streamlit

Using numerous charts and visualizations on streamlit, this library is used in this project to give elegant web items and objects for expressing WhatsApp chat analysis.

### NLP

Here, NLP features like text parsing, stop word removal and text analysis are used [11]. Text is parsed to separate messages into words for analysis, such as word counts and frequently used words. The Python program is instructed to present only significant words by removing all stop words from a file that contains all stop words. To determine how many links and/or media are shared, text analysis is employed.

## 4. Outcome and Discussions

Results should be given in the text in a logical order. This model was built using streamlit in Python and published to Heroku Web. The work flow of the model is as follows.

- The user clicks on browse file in the sidebar.
- Select the text file of a WhatsApp chat and import it for examination.
- Users can choose between a group-wide analysis and a detailed examination of a single person.

- The user then clicks the show analysis button to evaluate the imported file after selecting the user.
- Analysis of the imported WhatsApp text file is displayed.
- The user can view the total number of messages, words, files, and links shared inside the group.
- Following that, line charts are used to display the message's daily and monthly timelines.
- Activity Map, where the bar charts depict the busiest month and day.
- A weekly activity map that utilizes a heat map to display user activity on an hourly basis for each corresponding day.
- A list of users with percentages of use and a graph showing the top five busiest users in a group.
- Word clouds provide an intriguing visual representation of the most popular words.
- A bar graph is used to depict the top twenty most popular words.
- A list of emojis together with their usage statistics.
- A pie chart displaying the usage rates of the top five emojis. This is the outcome of the project and how it is operating.

The following figures from Fig.1 to Fig. 9 shows the results coming up from the chat analysis in a period of time. From this analysis we can count the total number of message, words, media shared, and link shared in day wise, week wise, and month wise. As a result the user performance over WhatsApp is easily measured. Let us visualize the outputs through graphically.

Top Statistics:

Top Statistics			
Total Messages	Total Words	Media Shared	Links Shared
1733	27322	181	252

Figure 1. Statistic report of the web application

Monthly Timeline:

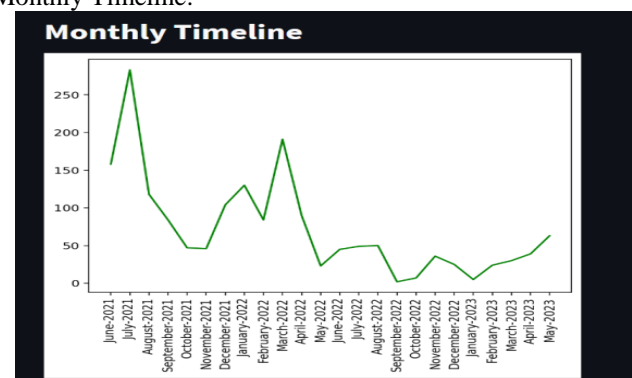


Figure 2. Monthly timeline of the web application

Daily Timeline:

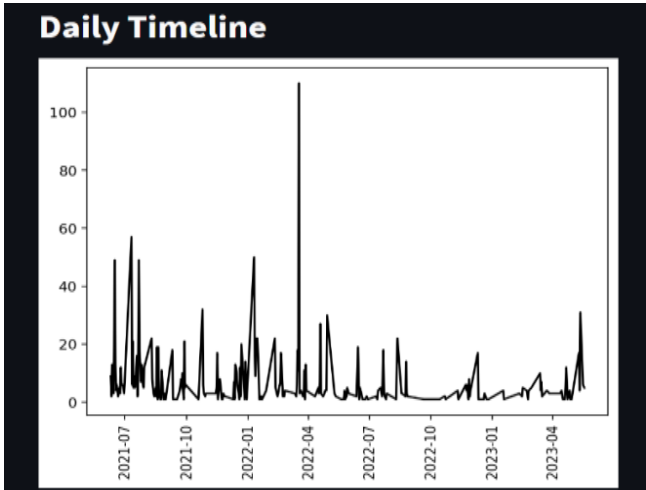


Figure 3. Daily timeline of the web application

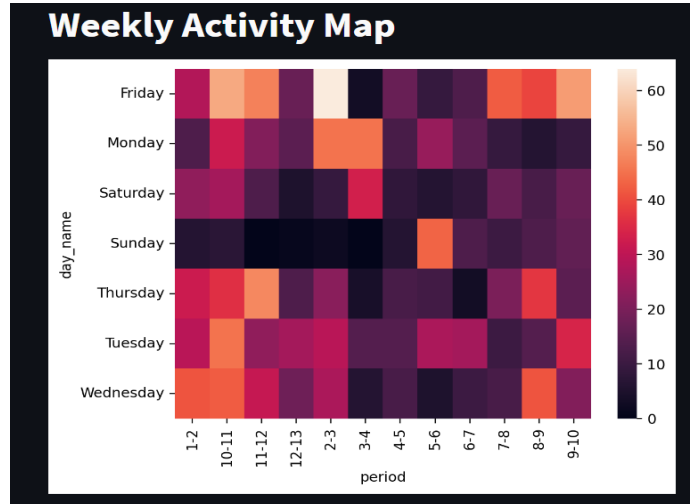


Figure 6. Weekly activity map of the web application

Activity Map Day Wise:

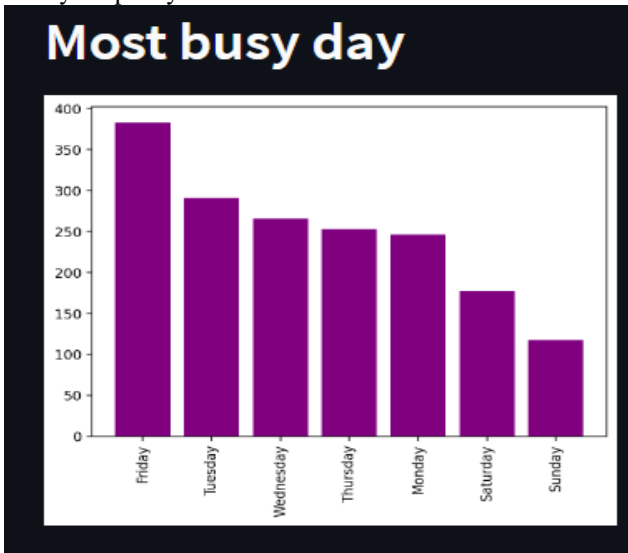


Figure 4. Day wise activity map of the web application

Busy User:

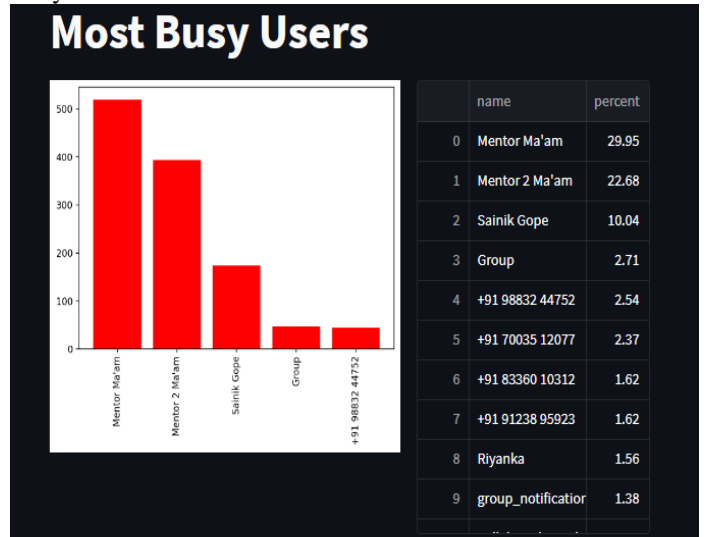


Figure 7. Most busy user selection

Activity Map Month Wise:

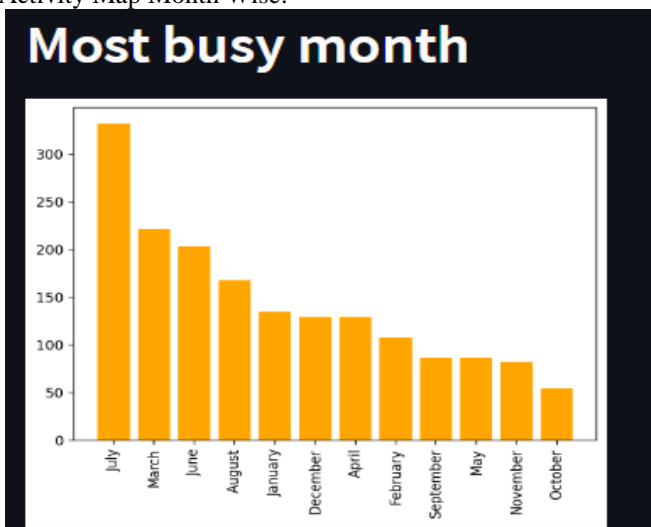


Figure 5. Month wise activity map of the web application

Common Words:

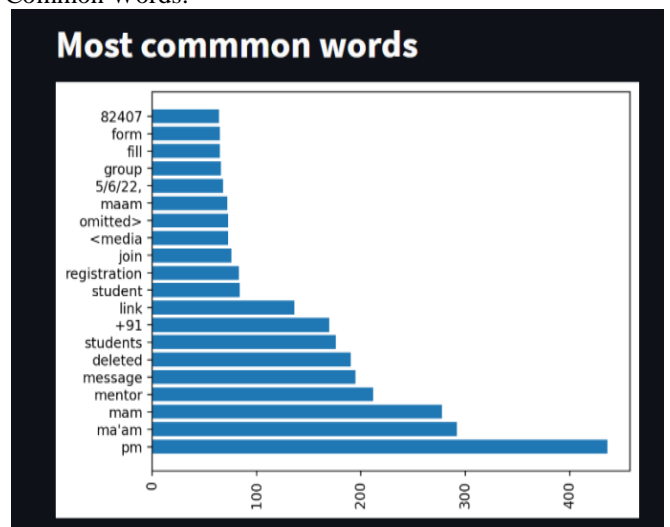


Figure 8. Most common word selection

Weekly Activity:

Emoji Analysis:

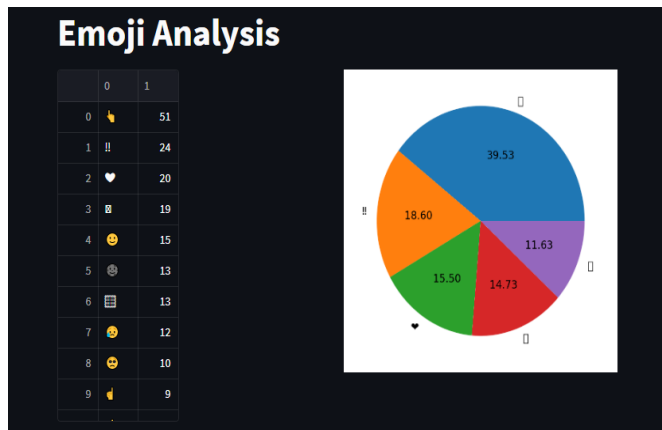


Figure 9. Analysis of emojis used

## 6. Conclusion

Successful completion of the key target set during a requirement analysis's earliest stage. Following installation, the model delivers accurate outcomes. The suggested model is entirely menu-based and intuitive, simplifying this for people to use even if they have no experience with computer environments. The system completely removes the possibility of incorrect data entry thanks to its validation feature, which also avoids the shortcomings of the previous manual system. The model has the following attributes:

- Convenient.
- Saving time.
- Works with all devices.
- Examines any imported WhatsApp file.
- Dependability.
- Accurate and user-friendliness.

## References

- [1]. Li, Diya, Harshita Chaudhary, and Zhe Zhang. "Modeling spatiotemporal pattern of depressive symptoms caused by COVID-19 using social media data mining." *International Journal of Environmental Research and Public Health* **17.14**, 4988, 2020.
- [2]. Python for Everybody: Exploring Data in Python 3 by Dr. Charles Russell Severance.
- [3]. Neupane, S. Developing a static website and deploying it to Heroku. 2020.
- [4]. Watson, Hugh J. "Preparing for the cognitive generation of decision support." *MIS Quarterly Executive* **16.3**, 2017.
- [5]. Kanala, S., Cheng, C. F., Gherasoiu, I., & Tekeoglu, A. NYS Fair Events Mobile Application With Client-Side Caching (Doctoral dissertation). 2017.
- [6]. Karabatak, S., & Alanoğlu, M. The Relationship between Teacher Candidates' Technology Addictions and Their Social Connectedness: A Data-Mining Approach. *Malaysian Online Journal of Educational Technology*, **10(4)**, 265-275, 2022.
- [7]. Bhati, V. S., Bansal, J., & Villa, S. Social media and Indian youth. *International Journal of Computer Sciences and Engineering*, **7(1)**, 818-821, 2019.
- [8]. Chen, H., Harinen, T., Lee, J. Y., Yung, M., & Zhao, Z. Causalm: Python package for causal machine learning. *arXiv preprint arXiv:2002.11631*, 2020.
- [9]. Bloice, M. D., & Holzinger, A. A tutorial on machine learning and data science tools with python. *Machine Learning for Health Informatics: State-of-the-Art and Future Challenges*, 435-480, 2016.
- [10]. Bach, P., Chernozhukov, V., Kurz, M. S., & Spindler, M.

DoubleML-An Object-Oriented Implementation of Double Machine Learning in Python. *J. Mach. Learn. Res.*, **23**, 53-1, 2022.

- [11]. Ahuja, Ravinder, et al. "The impact of features extraction on the sentiment analysis." *Procedia Computer Science* **152**, 341-348, 2019.

## AUTHORS PROFILE

I am Dr. Amrut Ranjan Jena, Associate Professor, CSE, GNIT, Kolkata and project guide of co-authors.

I am Pratyush kumarna , B.Tech final year student, CSE from GNIT, Kolkata.

I am Rafiqul Islam, Assistant Professor, CSE, GNIT, Kolkata and co-authors of this paper.