

WHATSAPP CHAT ANALYZER

Shaikh Mohd Saqib^{*1}, Prof. Sujata Bhosle^{*2}

^{*1}Student, Department of MCA, JNEC MGM University, Aurangabad, MH, India

^{*2}Assistant Professor, Department of MCA, JNEC MGM University, Aurangabad, MH, India

ABSTRACT

The most used and efficient method of communication in recent times is an application called WhatsApp. WhatsApp chat analyzer is the application deployed on heroku web which provide analysis of WhatsApp group. There are various methodologies available for analysis but here matplotlib, streamlit, seaborn, re, pandas libraries of python and some concept of NLP is used. This is the combination of machine learning and NLP. This whatsapp chat analyzer take import whatsapp chat file from user and analyze it and give different visualizations as a result.

I. INTRODUCTION

In this report I have proposed a WhatsApp Chat Analyzer. WhatsApp chats contains of different types of communications held among groups and personal chats. This chat contains of different topics. This can provide more data for technologies like machine learning. Machine learning models provides right learning experience which is important thing and indirectly affected by the data which provided to that model. This application provides analysis of this data which is WhatsApp provides. The advantage of this application is that it is implemented by simple python libraries like seaborn, pandas, numpy, streamlit and matplotlib which are commonly use for creating data frames and different graphs. This is displayed in web using heroku link which can run on all devices which supports browser.

II. LITERATURE REVIEW

2.1 Existing System

In olden days' there is no analysis for whatsapp chat. If someone wants to analyze there is no CSV file to analyze. WhatsApp Application provide export txt file which is in raw format. It is very complicated for analysis. So we have to forget that system and switch over to the WhatsApp Chat Analyzer.

Disadvantages of Existing System

1. Raw data.
2. Time consuming.
3. Difficult to Analyze.
4. Analysis are not accurate.

2.2 Proposed System

The "WhatsApp Chat Analyzer" provides a platform to the user which enables user to analyze whatsapp chats online on heroku link. This application allows user to browse whatsapp exported (.txt) file and import it to WhatsApp chat analyzer and get analysis according to that txt file. And user can Analyze by clicking Show Analysis button.

Advantages of WhatsApp Chat Analyzer.

- Runs on all devices.
- Shows based on whatsapp chat file.
- Shows different visualizations.
- Total Messages.
- Total words.
- Media shared.
- Link shared.
- Monthly timeline.
- Most busy day.

- Most busy month.
- Weekly activity.
- Most busy users.
- Most used words
- Emoji analysis.

III. METHODOLOGIES USING TECHNICAL THINKING

Python

It is a general-purpose programming language. It provide different types of libraries which provides different functionality to project. Python is used for predictions and pattern using test and data. Python consist many libraries which provide mathematical, statistical functions and help to find insights from data.

Pandas

This is an open-source Python libraries which is mainly used in Data Science and machine learning subjects. This library provides analysis tool for data manipulation, using its data structures this are used for analyzing data for manipulating time series analysis and numerical data.

Numpy

Numpy can be name come from Numeric Python, it is a data analysis library for Python that contains various numerical functions and methods for numerical analysis and also having multi-dimensional array objects and to process these arrays contains collection of routines.

IV. ANALYZE METHODOLOGIES BASED UPON MEASURES OR PERFORMANCE

Matplotlib

Matplotlib is easy to use and an amazing visualizing library in Python. It is built on NumPy arrays and it work with the broader SciPy stack and consists of several plots like pie, line, bar, graph, scatter, histogram, etc. In this project, Matplotlib is used for various visualizations for analysis of whatsapp chats. Visualizations like bar charts, line charts, pie charts are used.

Seaborn

Seaborn is a library mostly used for statistical plotting in Python. To make statistical plots more attractive it provides beautiful color palettes and default styles. In this project, Seaborn is used for heatmap visualization for showing 24 hours with 7 day with different scale of color for getting hour with max to min messages.

Streamlit

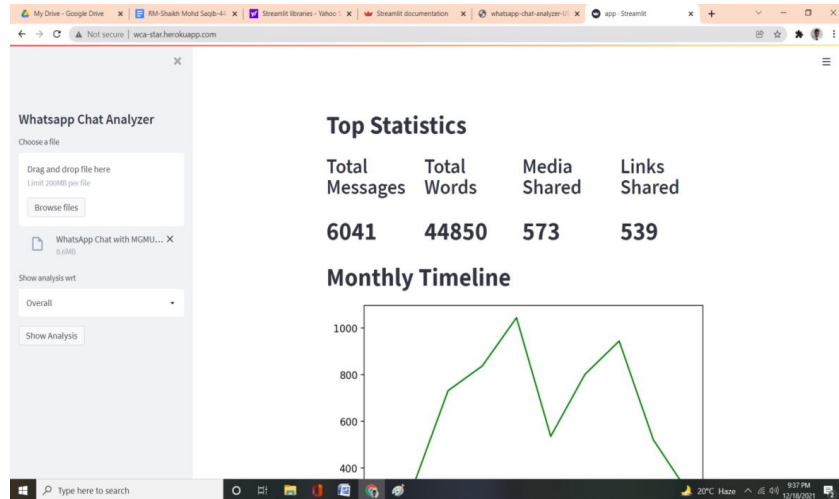
In this project, this library is used for creating beautiful web items and objects for representing Whatsapp chat analysis with different types of charts and visualizations on Streamlit.

NLP

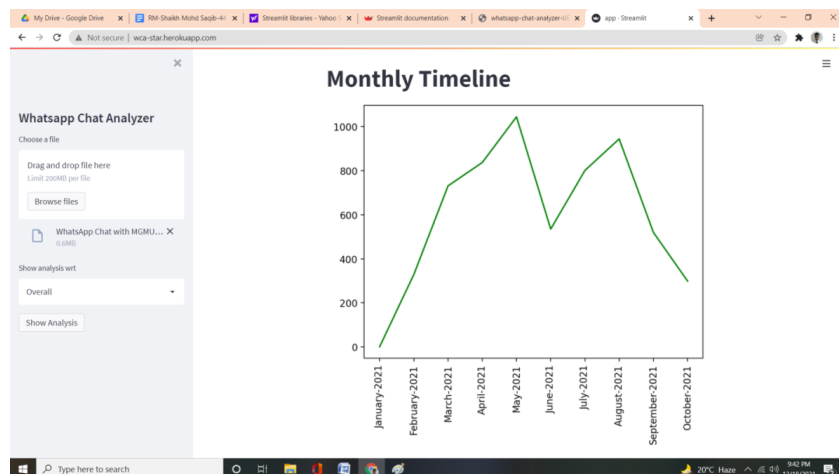
In this project, Features of NLP are used like Parsing Text, Eliminating stop words and Analyzing Text. Parsing text is used for splitting messages into words for analysis like total words and mostly used words. A file is used that contains all stop words which is given to the python program to show meaningful words only by eliminating all stop words. Text analysis is used to identify how many media are shared, how many links are shared.

OUTPUT:

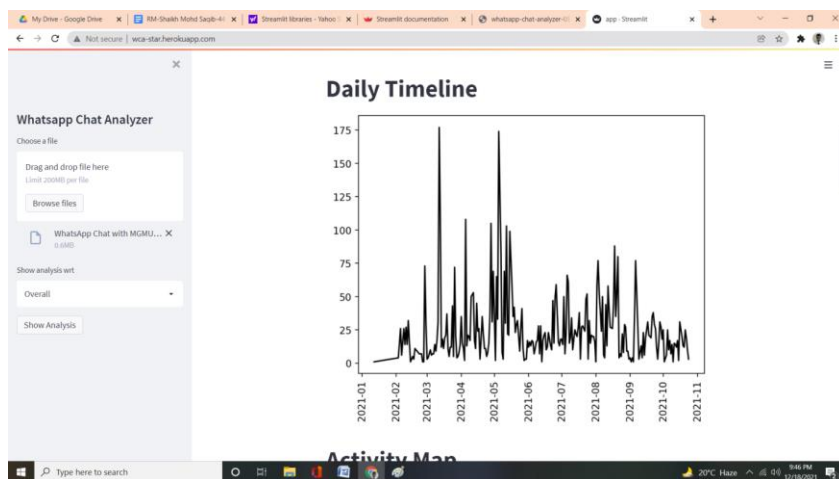
Top Statistics:



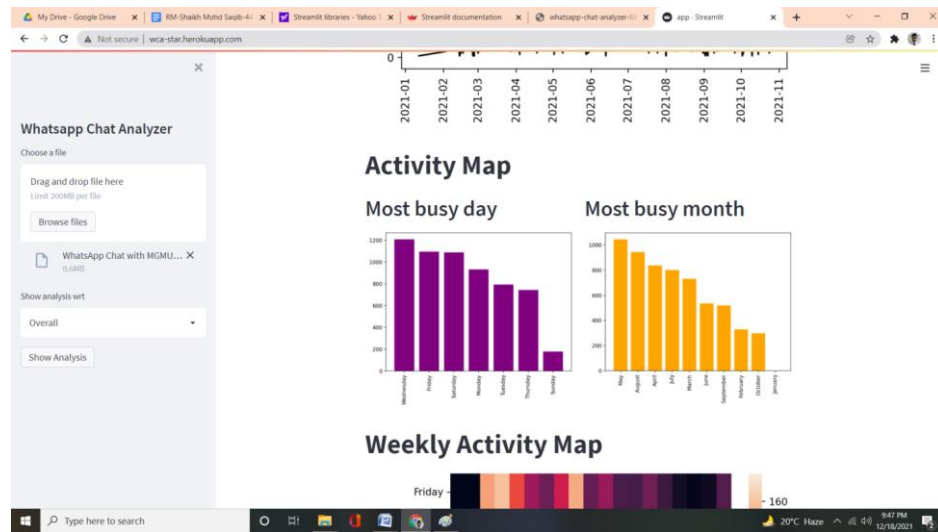
Monthly Timeline:



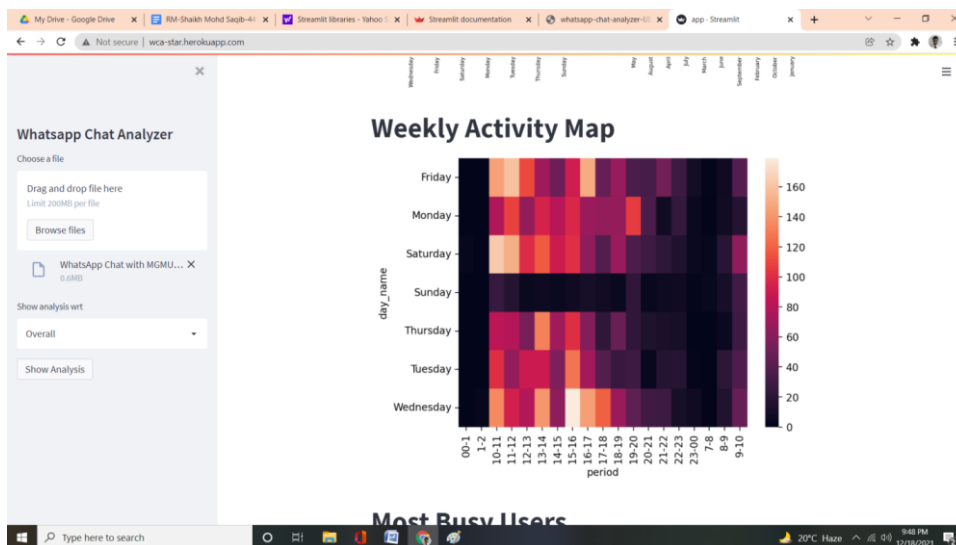
Daily Timeline:



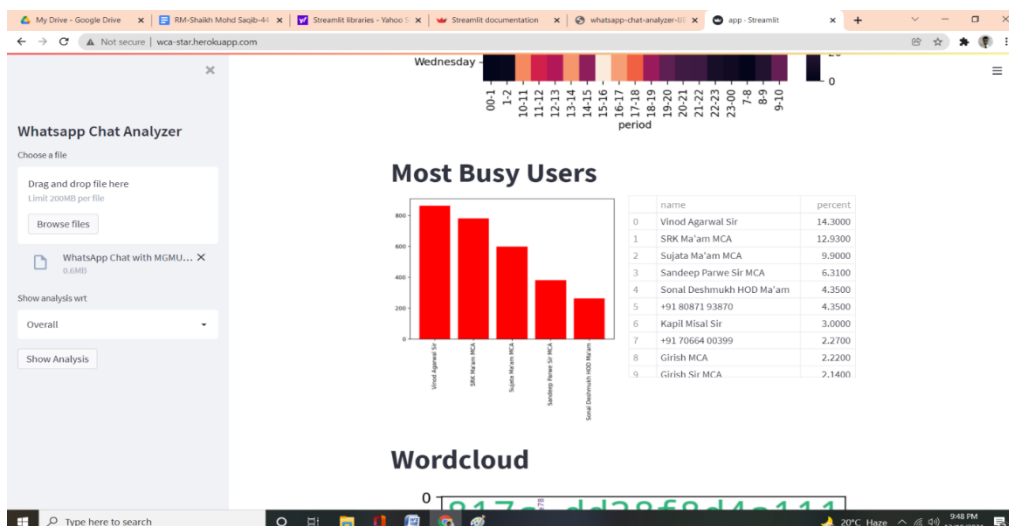
Activity Map:



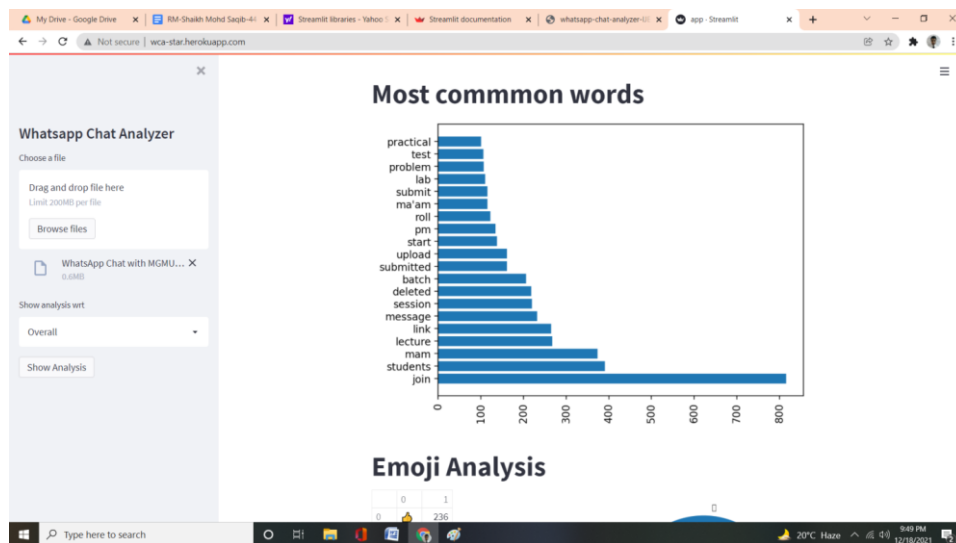
Weekly Activity:



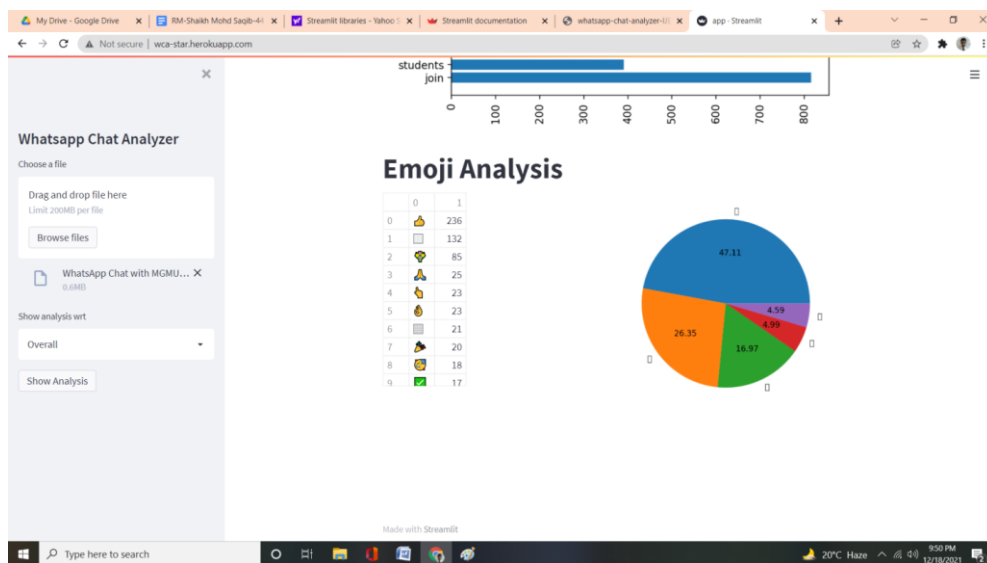
Busy User:



Common Words:



Emoji Analysis:



V. RESULTS AND DISCUSSION

This project is created in python using streamlit and deployed on heroku web.

Working of project:

1. User go to sidebar and click on browse file .
2. Select whatsapp chat text file and import it for analysis .
3. User have choice for overall analysis or specific user analysis from whole group.
4. After selecting user, User click on show analysis button to analyze imported file.
5. It shows analysis of imported whatsapp text file.
6. User can see Total messages, words, media and link shared in the group.
7. then monthly and daily timeline for the message is shown using line charts.
9. Activity Map in which most busy month and day is shown by the bar charts.
10. then weekly activity map which shows hourly activity of users with corresponding day using heat map.
11. Top five busy users in group using graph and list of users with percentage of use.

12. WordCloud shows an interesting visualization of most common words.
13. Top twenty most common word represented by using bar chart.
14. List of Emojies with number of times it is used.
15. Pie chart which shows top five emojis percentage of use.

This is the result of project and how project is working.

VI. CONCLUSION

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. The system avoids the drawbacks of the existing manual system and the validation facility of the system totally eliminates the chances of wrong data entry.

It has following features:

- User friendly.
- Time saving.
- Runs on any devices.
- Analyzes any WhatsApp imported file.
- Accuracy.
- Reliability.
- Easy to use.

VII. REFERENCES

- [1] D.Radha, R. Jayaparvathy, D. Yamini, "Analysis on Social Media Addiction using Data Mining Technique", International Journal of Computer Applications (0975 – 8887).
- [2] Python for Everybody: Exploring Data in Python 3 by Dr. Charles Russell Severance.
- [3] Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic.