

CHAT ANALYSIS ON WHATSAPP USING MACHINE LEARNING

Dr.T.Siva Ratna Sai¹, T.Naga Nandini², M.Harsha Vardhan³, B.Sivaji⁴, R.Sai Kalyan⁵

¹Professor, Department of Computer Science and Engineering, Qis college of Engineering and Technology, Ongole, Andhra Pradesh, India

^{2,3,4,5}Student, Department of Computer Science and Engineering, Qis college of Engineering and Technology, Ongole, Andhra Pradesh, 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. 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. This tool aims to offer a thorough study of the information that WhatsApp provides. Regardless of the subject around which the conversation is centred, our generated code may be used to improve comprehension of the data. The benefit of this tool is that is implemented using simple python modules such as pandas, matplotlib, seaborn and sentiment Analysis that are used to produce data frames and plot various graphs are then shown in the flutter application. Because this approach is effective and resource-conserving, it can be readily applied to the largest dataset.

Keywords: Pandas, Seaborn, Matplotlib, sentiment analysis, streamlit, NLTK, and data from WhatsApp chats are some examples.

I. INTRODUCTION

We have suggested a WhatsApp Chat Analyzer in this research. Different forms of communication between groups and individuals are included in WhatsApp conversations. Different subjects are covered in this talk. This might give machine learning technology additional 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. This application has the benefit of being implemented using straightforward python libraries, such as seaborn, pandas, numpy, streamlit, and matplotlib, which are often used for building data frames and various graphs. This is presented online via a HEROKU link, which is compatible with all chat-enabled devices.

Future technologies are heavily dependent on data in this decade. Only by conducting study on the context of the tool's requirements will it be possible to gather this data. Since a lot of machine learning enthusiasts create models that assist in solving several issues, a lot of appropriate data is needed. This project seeks to provide users a better knowledge of various conversation kinds. This investigation demonstrates that it can provide machine learning models—which essentially investigate chat data—with superior input. These models need the right learning cases, which increases their accuracy. In-depth exploratory data analysis of numerous WhatsApp chat kinds is guaranteed by our project.

A .SCOPE OF THE PROJECT:

Data pre-processing, the initial part of the project is to understand implementation and usage of various python- built modules. The above process helps us to understand why different modules are helpful rather than implementing those functions from scratch by the developer. These various modules provide better code representation and user understandability. The following libraries are used such as numpy, scipy

pandas, csv, sklearn, matplotlib, sys, re, emoji, nltk
seaborn etc.

II. LITERATURE SURVEY

1. Impact of WhatsApp Messenger : The use and effect of WhatsApp have been the subject of several studies and analyses. Some of these research focus on discovering WhatsApp's effects on students, while others focus on the broader population in a specific area.

A survey was done in the southern region of India on people between the ages of 18 and 23 to find out how popular WhatsApp is with young people. Through this survey, it was shown that students use WhatsApp for 8 hours daily and are online for roughly 16 hours daily. All of the respondents said that they use WhatsApp to talk to their pals. They utilise WhatsApp to share photographs, music files, and video files with their buddies. Additionally, it was demonstrated that WhatsApp is the sole programme that young people utilise when using their smartphones. In order to determine the extent of any beneficial or bad effects of using WhatsApp, methods were utilised in this study to analyse the volume of usage of WhatsApp and its well-known services.

2. Content Analysis of WhatsApp conversation: a research project to analyse the WhatsApp application's effectiveness in Karachi. The study will be a crucial piece of information for examining if WhatsApp may overtake other mobile messaging apps in Pakistan.

The communication landscape in Pakistan has drastically transformed as a result of the development of digital technology and the fusion of mobile phones (Ali, Rizvi, & Sherdil, 2014). In Pakistan, the popularity of social networking sites and smartphones has made communication quicker and simpler than ever before. People may not have enough money to buy food, rent a place to stay, or buy clothes anymore, but they do have a mobile phone in their pockets that they may use to communicate with their loved ones, friends, and clients. The usage of quantitative and qualitative research methods has grown throughout time as a result of the changing environment. The steps were

devised for

the study of the nature and impact of communication technologies on behaviour. In the same time frame, smart phones and instant messaging services like Skype, Viber, and WhatsApp completely changed Pakistan's communication landscape.

3. WhatsApp Group Data Analysis with R : The dataset of WhatsApp group chat used for analysis spans 1 year (May 2015–May 2016), has a total of 5,5563 records, and includes various characteristics that indicate how frequently an individual uses WhatsApp chat groups, such as the years of usage, daily usage duration, response rates, the types of messages each member posts (Smiley, Text, Multitude), which age group has the most active members, and so on.

The primary variables specified for this study are the type of message sent, the length of usage during the last year, month, week, day, and hour, the timestamp (AM/PM), the age range of the senders, and the gender (Male/Female). Because it is free source, the most popular IDE for R, RStudio, is utilised to do exploratory data analysis and visualisation for the gathered data.

4. Forensic analysis of WhatsApp Messenger : User-to-user communication, broadcast messaging, and group conversations are just a few of the several types of communication that WhatsApp offers its users. Users may communicate by exchanging simple text messages, multimedia items (including voice, video, and picture files), contact cards, and geographical data. Each user has a profile, or collection of data, that includes their WhatsApp name, status update, and avatar (a visual file, usually a photo). Each user's profile is kept on a centralised system and may be accessed by other WhatsApp users who have added that person to their contacts. Other services offered by the central systems include message relay, user registration, and authentication.

III .PROPOSED SYSTEM

The "WhatsApp Chat Analyzer" gives users a platform to analyse WhatsApp talks online via a heroku connection. With this programme, users may explore whatsapp exported (.txt) files, input them into WhatsApp conversation analyzer, and receive

analysis based on those txt files. And user can Analyze by clicking Show Analysis button.

Applying a sentiment analysis algorithm to exploratory data analysis is the initial stage in this process. This approach offers the positive, negative, and neutral chatter and is used to construct a pie chart depending on these parameters. Plotting a line graph that displays the author and message count for each date, the author and message count for each author, an ordered graph of date against message count, the number of media sent by authors, etc. Show the message with no authors and a graph of the number of messages every hour.

Better code representation and user understandability are provided by these diverse modules. Numpy, Scipy Pandas, CSV, Sklearn, Matplotlib, sys, re, emoji, nltk Seaborn, and other libraries are utilised.

3.Data Preprocessing: In this module data preprocessing is done by removing all the unwanted content from the raw data. Only the data required for analysis is taken into consideration.

4.Importing Data into website: Users can choose between overall group analysis and a specific user analysis. The user then clicks the display analysis option to analyse the imported file after choosing the user. Analysis of an imported WhatsApp text file is displayed.

5.Statistical Representation: Various graphical representations are used to display the preprocessed data.

IV .DESIGN

Use Case Diagram:

The website has following features:

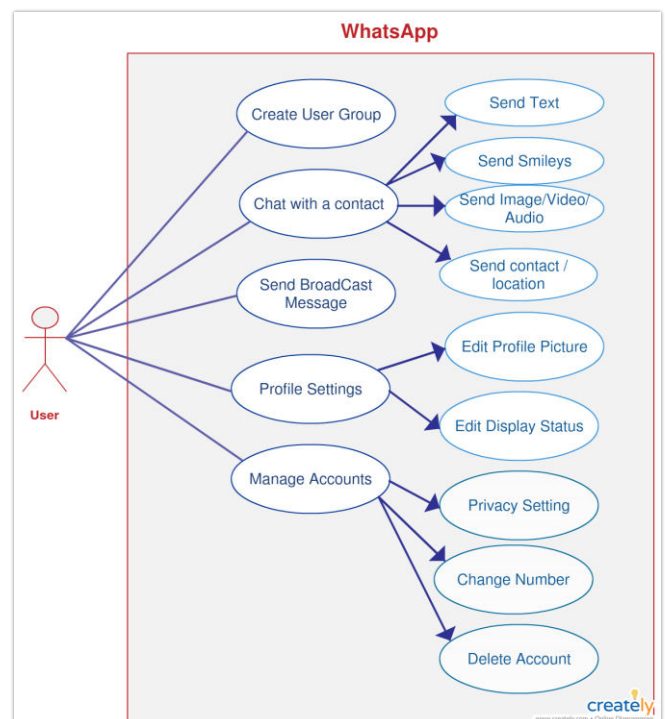
- Shows based on a chat file from WhatsApp.
- Displays various visualisations.
- Messages in total.
- Word count.
- Sharing of links and media.
- A monthly schedule.
- The busiest day.
- The busiest month.
- Most active users;
- Weekly activity.
- Common words
- Examining emojis.

A. MODULE DESCRIPTION:

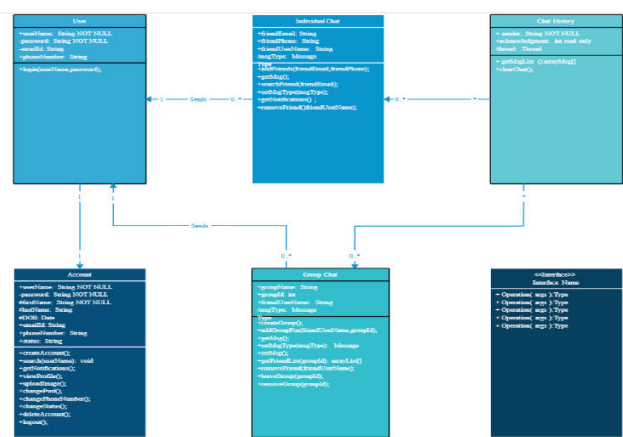
This project is composed of five main modules which also include sub modules:

1.Data extraction: Data is extracted from WhatsApp by clicking on export button. After exporting, a text file is created which includes raw data. The raw data is given to whatsapp chat analyzer to perform preprocessing

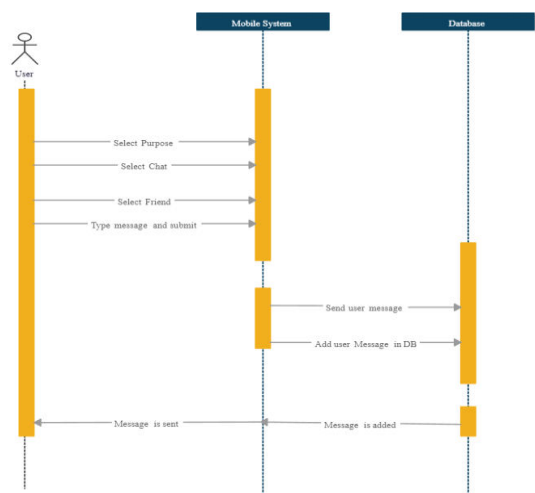
2. Data Collection: Use the “Export Chat” functionality to send the entire conversation in text format to your email ID. Download the exported chat from your email inbox.



Class diagram

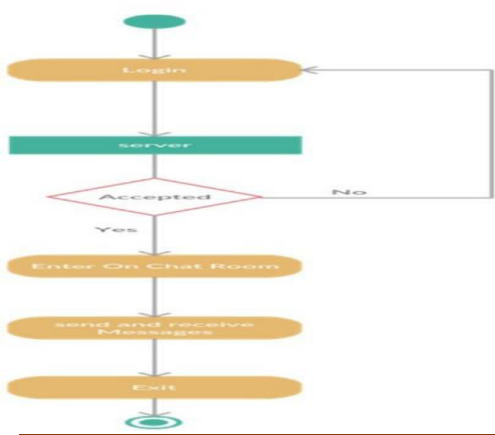


Sequence diagram



Activity diagram

Activity Diagram



V .IMPLEMENTATION

NLP

NLP In this project, NLP features including text parsing, stop word elimination, and text analysis are applied. Text is parsed to separate messages into words for analysis, such as word counts and frequently used terms. The Python programme is instructed to present only significant words by removing all stop words from a file that contains all stop words. To determine how many media files and URLs are exchanged, text analysis is employed.

Matplotlib

Python's Matplotlib is a fantastic visualising package that is simple to use. 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. In this project, numerous visualisations for the study of WhatsApp talks are done using Matplotlib. The use of visualisations like pie charts, bar charts, and line charts.

Python

Python is a robust programming language that is simple to learn. Its object-oriented programming methodology is straightforward but efficient, and it includes good high-level data structures. Python's interpreted nature, intuitive syntax, and dynamic typing make it the perfect language for scripting and quick application development across a wide range of platforms.

Installing Python Modules

As a well-liked open source development project, Python has a thriving user and contributor community that also makes its software accessible for use by other Python developers under the conditions of an open source licence.

Pycharm

PyCharm is a specialised Python Integrated Development Environment (IDE) that offers a variety of necessary tools for Python developers. These tools are tightly integrated to produce a practical environment for productive Python, web, and data science

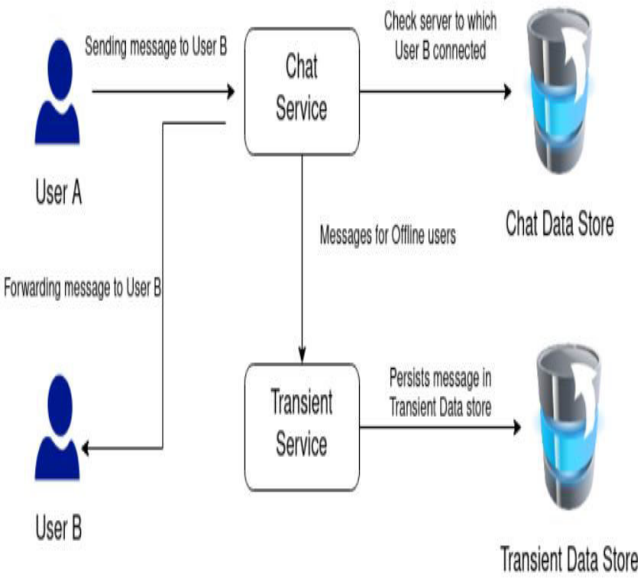
deverse.

Software Design

In designing the software following principles are followed:

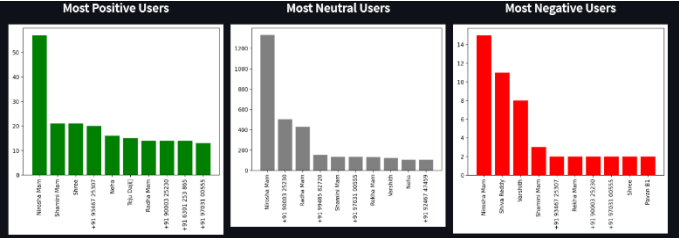
- 1.Modularity and partitioning: Software is designed such that, each system should consists of hierarchy of modules and serve to partition into separate function.
- 2.Coupling: Modules should have little dependence on other modules of a system.
- a.Cohesion: Modules should carry out in a single processing function.
- b.Shared use: Avoid duplication by allowing a single module be called by other that need the function it provides.

System Design

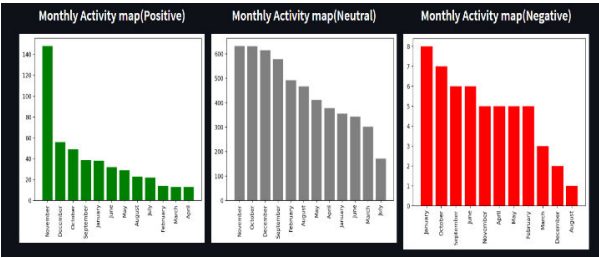


Output screenshots

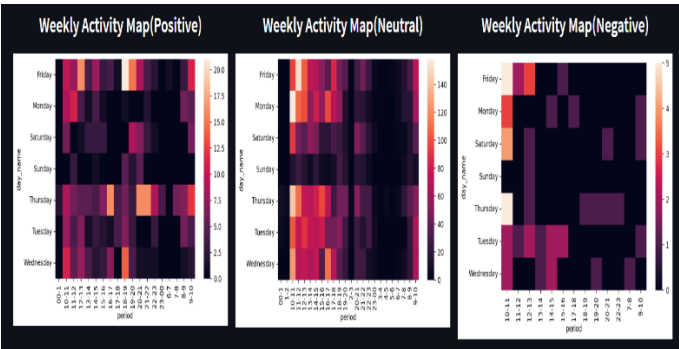
Most Busy User



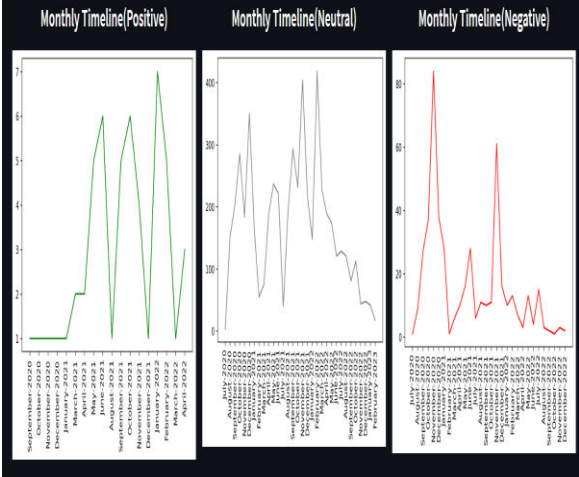
Monthly Activity User



WEEKLY ACTIVITY MAP



MONTHLY TIMELINE



CONCLUSION

In conclusion, It may be argued that the strength of the Python programming language in accomplishing whatever network data analysis required, as well as the capabilities of the WhatsApp application, cannot be overemphasised. The WhatsApp programme and its libraries were discussed in this paper, along with an analysis of a WhatsApp group conversation and a visual representation of the top 10 and top 20 members in the chat groups. The story was described in pseudocode, and a visual depiction of the plot was added in the end. Additionally, the top 10 and top 20 users were examined. The system was written in Python, and NumPy, Pandas, Matplotlib, and Seaborn were among the Python libraries used. The analysis was able to illustrate the amount of participation of the various people on the specified WhatsApp group at the conclusion of the task, and the findings were as anticipated. On a more serious note, this technology is capable of analysing any WhatsApp group data that is inputted. Successful completion of the key target set at the initial phase of the requirement analysis. After installation, the system delivers accurate outcomes. The proposed system is completely menu-based and user-friendly, making it simple 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 approach. Successful completion of the

key target set at the initial phase of the requirement analysis. After installation, the system delivers accurate outcomes. The proposed system is completely menu-based and user-friendly, making it simple 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 approach. The analysis was able to demonstrate the amount of engagement of the various people on the provided WhatsApp group at the conclusion of the task, and the findings were as anticipated. On a more serious note, this technology is capable of analysing any data input from a WhatsApp group.

REFERENCES

- Available from: <http://www.statista.com/statistics/260819/number-of-monthly-active-WhatsApp-users>. Number of monthly active WhatsApp users worldwide from April 2013 to February 2016(in millions).
- 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).
- 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).
- Access Data Corporation. FTK Imager, 2013. Available at

<http://www.accessdata.com/support/product-downloads>.

8887) Volume 139 No.7, pp. 23- 26, April 2016.

- D.Radha, R. Jayaparvathy, D. Yamini, Analysis on Social Media Addiction using Data Mining Technique, International Journal of Computer Applications (0975

- Jessica Ho, Ping Ji, Weifang Chen, Raymond Hsieh, Identifying google talk, IEEE International Conference on Intelligence and Security Informatics, ISI 09, pp. 285-290, 2009.