Report On

Whatsapp Group Chat Analysis

Submitted in partial fulfillment of the requirements of the Course project in

Semester VIII of Final Year Computer Engineering

by

Anand Tiwari (Roll No. 79)

Mihir Thakur (Roll No. 78)

Pushkar Save (Roll No. 72)

Mentor

Dr. Tatwadarshi Nagarhalli



**University of Mumbai**

**Vidyavardhini's College of Engineering & Technology**

**Department of Computer Engineering**

****

**(A.Y. 2021-22)**

**Vidyavardhini's College of Engineering & Technology**

**Department of Computer Engineering**

**CERTIFICATE**

This is to certify that the Mini Project entitled **“Whatsapp Group Chat Analysis”** is a bonafide work of **Anand Tiwari (Roll no. 79), Mihir Thakur (Roll no. 78), Pushkar Save (Roll no. 72)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **“Bachelor of Engineering”** in Semester VIII of Final Year **“Computer Engineering” .**

|  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Dr. Tatwadarshi Nagarhalli  Mentor |  |
| --- | --- | --- |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Dr Megha Trivedi  Head of Department |  | Dr. H.V. Vankudre  Principal |

**Vidyavardhini's College of Engineering & Technology**

**Department of Computer Engineering**

# Course Project Approval

This Mini Project entitled **“Whatsapp Group Chat Analysis”** by **Anand Tiwari (Roll no. 79), Mihir Thakur (Roll no. 78), Pushkar Save (Roll no. 72)** is approved for the degree of **Bachelor of Engineering** in in Semester VIII of Final Year **Computer Engineering.**

**Examiners**

**1……………………………………**

(Internal Examiner Name & Sign)

## 2………………………………………

(External Examiner name & Sign)

Date:

Place:

# Contents

# Abstract

## Introduction

* 1. Introduction
  2. Problem Statement & Objectives
  3. Scope

## Proposed System

* 1. Introduction
  2. Architecture/ Framework/Block diagram
  3. Algorithm and Process Design
  4. Details of Hardware & Software
  5. Experiment and Results for Validation and Verification
  6. Conclusion and Future Work.
  7. References

**ABSTRACT**

The most used and efficient method of communication in recent times is an application called

WhatsApp. WhatsApp chats consist of various kinds of conversations held among groups of people. This chat consists of various topics. This information can provide lots of data for the latest technologies such as machine learning. The most important thing for a machine learning model is to provide the right learning experience which is indirectly affected by the data that we provide to the model. This tool aims to provide a depth analysis of the data which is provided by WhatsApp. Irrespective of whichever topic the conversation is based our developed code can be applied to obtain a better understanding of the data.

* 1. **INTRODUCTION**

WhatsApp is an instant messaging application that allows users to send text messages, chat and share media files like images, audio, and video files. Users can also share documents and applications. With WhatsApp, users have the opportunity to communicate with several other users at the same time in a group. In addition, a user can send a broadcast message to up to two hundred and fifty-six (256) users at a single message stance. This feature makes the message appear as though it was sent to each individual alone, describing the application as a proprietary, instant messaging application for smartphones, that cuts across all platforms. Apart from text messaging, users can also send images, video, and audio media messages to each other.

This tool is based on data analysis and processing. The first step in implementing a machine learning algorithm is to understand the right learning experience from which the model starts improving. Data preprocessing plays a major role when it comes to machine learning. In order to make the model more efficient, we need lots of data, so we turned our focus primarily on one of the largescale data producers owned by Facebook which is nothing but WhatsApp. WhatsApp claims that nearly 55 billion messages are sent each day. The average user spends 195 minutes per week on WhatsApp and is a member of plenty of groups. With this treasure house of data right under our very noses, it is imperative that we embark on a mission to gain insights into the messages to which our phones are forced to bear witness.

* 1. **PROBLEM STATEMENT**

In this decade the upcoming technologies are mainly dependent on data. This data can only be obtained if there is some research applied in the context of the requirements of the tool. Since a lot of machine learning enthusiasts develop models which help solve multiple problems the requirements of appropriate data are very large scale this project aims to provide a better understanding of various types of chats. This analysis proves to be better input to machine learning models which essentially explore the chat data. Analysis of chats can provide insights into the participation and engagement of an individual.

**OBJECTIVES**

1. Analyze WhatsApp group chat using various python libraries.
2. To use publicly available information from social media platforms (e.g., Whatsapp).
3. To apply natural language processing.
4. To present the analyzed data in various forms of graphs.
   1. **SCOPE**

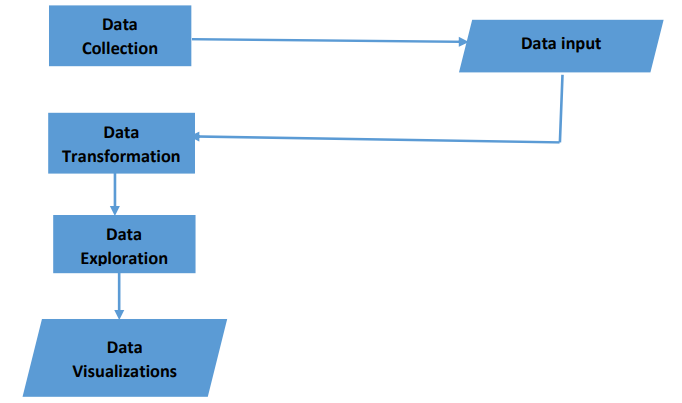
For our proposed model, we perform exploratory data analysis on WhatsApp group chats. In this regard, the system first filters out messages of a particular user and also differentiates the text messages and media messages of each user. The data from the chats of the WhatsApp group are preprocessed using different NLP techniques and stored in a data frame for further analysis. Each user’s statistics are collected based on the messages from him/her in the group along with stickers and emojis. Various data visualizations are made from the data and used to display the statistics for the user and the group as a whole. These visualizations can help to provide meaningful insights into the group activities and individual engagement. The system provides these visualizations and graphs as output to the WhatsApp chats.

* 1. **Introduction**

WhatsApp Analyzer means we are analyzing our WhatsApp group activities. It tracks our conversations and analyses how much time we are spending on WhatsApp. The aim of this article is to provide step by step guide to building our own WhatsApp analyzer using python. Here the system uses different python libraries which help me to extract useful information from raw data.

The system is a simple one using NLP for extracting information from the available raw data with the help of the ‘nltk’ python library and regular expressions and furthermore, storing it in a structured data frame in Python. Regex is used to distinguish different attributes in the text which is present in a normal Text file based on some characteristics of the data in the file. Date and author are some examples of the information extracted using Regex from the raw text file which will be used for further analysis.

* 1. **Architecture / Framework / Block diagram**



* 1. **Algorithm and Process Design**

Pandas and Numpy are python library and offers a simple syntax to access their methods and perform basic NLP tasks. Here, we use this library to convert the WhatsApp messages to data frames.

Initially, regex is used to convert the messages from the text file to desirable patterns and then each message is converted to individual data frames using pandas and NumPy.

After data frames are created data is analyzed on the basis of the following after filtering messages of each user, text, and media messages separately.

Using stopword analysis a wordCloud of most used words is created.

A bar graph of the most active users, most active days, and members who use max no of words in their messages, the time when this group is most active is created using matplotlib.pyplot a python library.

A pie chart of top-10 media contributors is created using matplotlib,pyplot.

* 1. **Details of Hardware & Software**

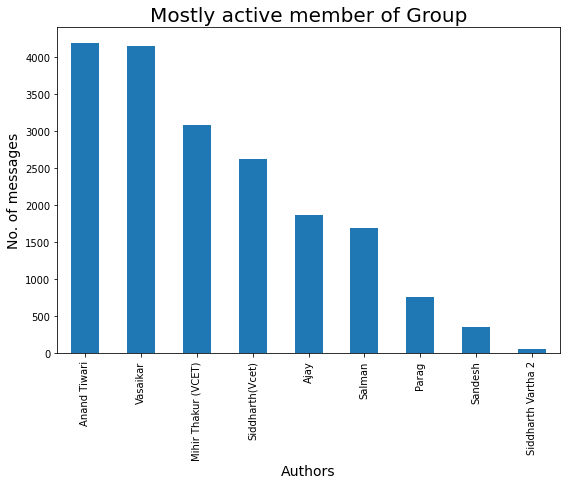
**Hardware**

* Intel i5 processor
* RAM – 8GB
* Hard disk – 10GB
* Web browser
* Internet Connection

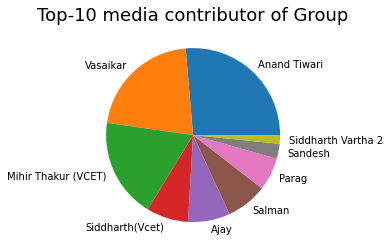
**Software**

* Jupyter Notebook
* Python — a programming language
* Tweepy — a type of RESTful API specifically for Twitter
* Textblob — processed textual data library tool (already trained on numerous textual data.)
* Pandas — data manipulation and analysis library
* NumPy — scientific computing library
* Matplotlib — plotting library
* Plotly — plotting library
  1. **Experiment and Results for Validation and Verification**

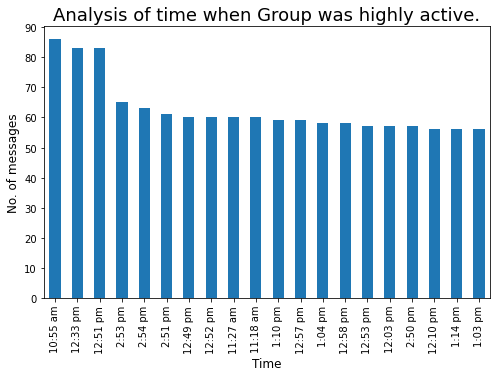
Visualizing data gives a clearer picture of what we are doing. Visualization of the various statistics from the chats provides proper insight into the data and the information that can be extracted from it.



The graph is a bar graph which shows information about the active users in the group based on the number of messages.



From the above figure, gives information about the top media contributors in the group and display the in the form of a pie chart.



Graph for the time at which the group is mostly active by the number of message

* 1. **Conclusion and Future Work.**

.In conclusion, it can be said that the capabilities of the WhatsApp application and the power of the Python programming language in implementing whatever network data analysis is intended, cannot be overemphasized. This work was able to discuss the WhatsApp application and its capabilities to a great extent, and it was able to use the Python programming language and its libraries, to create an analysis of a Whatsapp group chat and visually represent the top users in the chat group. A pseudocode of the plot was given and at the end, a visual representation of the plot was implemented. Also, an analysis of the top users was done. The system was done with Python, and the Python libraries that were implemented include, Numpy, Pandas, Matplotlib, and Seaborn. At the end of the work the expected results were obtained and the analysis was able to show the level of participation of the various individuals on the given WhatsApp group. One thing of note is that this system has the ability to analyze any WhatsApp data input into it.

* 1. **References**

[1] Nwamaka Iduh., “WhatsApp Network Group Chat Analysis Using Python Programming”, Academic Researcher, Computer Science Department, Nnamdi Azikiwe University, Awka, Nigeria.

[2]Ravishankara K, Dhanush, Vaisakh, Srajan I S et al., “Whatsapp Chat Analyzer”, Faculty of Department of Computer Science, Srinivas Institute of Technology Mangalore, Karnataka. 2,3Students of Department of Computer Science, Srinivas Institute of Technology Mangalore, Karnataka.