**MAULANA AZAD**

**NATIONAL INSTITUTE OF TECHNOLOGY BHOPALINDIA, 462003**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

# WhatsApp Chat Analysis

## Minor Project Report

**Semester**

### Submitted by:

|  |  |
| --- | --- |
| Mridul Karan | 191112273 |
| Himanshu Kumar | 191112248 |
| Pawan Patel Priyanshu Mangal | 191112280  191112250 |

**Under the Guidance of**

Prof. Deepak Singh Tomar

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**Session: 2021-22**

**MAULANA AZAD**

**NATIONAL INSTITUTE OF TECHNOLOGY BHOPALINDIA, 462003**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

## CERTIFICATE

This is to certify that the project report carried out on “WhatsApp Chat Analysis by the 3rd year students:

|  |  |
| --- | --- |
| Mridul Karan | 191112273 |
| Himanshu Kumar | 191112248 |
| Pawan Patel  Priyanshu Mangal | 191112280  191112250 |

Have successfully completed their project in partial fulfilment of their Degree in Bachelor of Technology in Computer Science and Engineering.

**Prof. Deepak Singh Tomar**

### (Minor Project Mentor)

**DECLARATION**

We, hereby declare that the following report which is being presented in the Minor Project Documentation Entitled as “WhatsApp Chat Analysis” is an authentic documentation of our own original work and to best of our knowledge. The following project and its report, in part or whole, has not been presented or submitted by us for any purpose in any other institute or organization. Any contribution made to the research by others, with whom we have worked at Maulana Azad National Institute of Technology, Bhopal or elsewhere, is explicitly acknowledged in the report.

|  |  |
| --- | --- |
| Mridul Karan | 191112273 |
| Himanshu Kumar | 191112248 |
| Pawan Patel  Priyanshu Mangal | 191112280  191112250 |

### ACKNOWLEDGEMENT

With due respect, we express our deep sense of gratitude to our respected guide and coordinator Prof. Deepak Singh Tomar, for his valuable help and guidance. We are thankful for the encouragement that he has given us in completing this project successfully.

It is imperative for us to mention the fact that the report of minor project could not have been accomplished without the periodic suggestions and advice of our project guide Prof. Deepak Singh Tomar and project coordinators Dr. Dhirendra Pratap Singh and Dr. Jaytrilok Choudhary.

We are also grateful to our respected director Dr. N. S. Raghuwanshi for permitting us to utilize all the necessary facilities of the college.

We are also thankful to all the other faculty, staff members and laboratory attendants of our department for their kind cooperation and help. Last but certainly not the least; we would like to express our deep appreciation towards our family members and batch mates for providing the much-needed support and encouragement.

### ABSTRACT

The most used and efficient method of communication in recent times is an application called WhatsApp. WhatsApp chats consists of various kinds of conversations held among group of people. This chat consists of various topics. This information can provide lots of data for latest technologies such as machine learning. The most important thing for a machine learning models 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 in depth analysis of this 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. The advantage of this tool is that It is implemented using simple python modules such as pandas, matplotlib, seaborn and sentiment analysis which are used to create data frames and plot different graphs, where then it is displayed in the MERN webapp which is efficient and can be modified easily, therefore it can be easily applied to large dataset.

**TABLE OF CONTENTS**

[Certificate 2](#_bookmark0)

[Declaration 3](#_bookmark1)

[Acknowledgement 4](#_bookmark2)

[Abstract 5](#_bookmark3)

1. [Introduction 7](#_TOC_250006)
2. [Literature review and survey 8](#_TOC_250005)
3. [Gaps Identified… 9](#_TOC_250004)
4. [Proposed work and methodology 10](#_TOC_250003)
   1. Proposed Work
   2. Methodology
5. [Technology Used and System Requirement… 11](#_TOC_250002)
6. [Conclusion 13](#_TOC_250001)
7. [References… 14](#_TOC_250000)

# Introduction

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 on. Data pre-processing 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 but imperative that we embark on a mission to gain insights on the messages which our phones are forced to bear witness to.

## Literature Review and Survey

In a study of southern part of India was conducted on the age group of between 18 to 23 years to investigate the importance of WhatsApp among youth. Though this study, it was found that students spent 8 hours per day on using WhatsApp and remain online almost 16 hours a day. All the respondents agreed that they are using WhatsApp for communicating with their friends. They also exchange images, audio and video files with their friends using WhatsApp. It was also proved that the only application that the youth uses when they are spending time on their smart phone is WhatsApp. Methods used in this survey is to analyze the intensity of WhatsApp usage and its popular services and to identify the degree of positive or negative impacts of using WhatsApp.

Forensic analysis of WhatsApp Messenger : WhatsApp provides its users with various forms of communications, namely user-to-user communications, broadcast messages, and group chats. When communicating, users may exchange plain text messages, as well as multimedia files (containing images, audio, and video), contact cards, and geolocation information. Each user is associated with its profile, set of information that includes his/her WhatsApp name, status line, and avatar (a graphic file, typically a picture). The profile of each user is stored on a central system, from which it is downloaded by other WhatsApp users that include that user in their contacts. The central systems provide also other services, like user registration, authentication, and message relay.

Impact of WhatsApp on students : Various Studies and analysis has been done on the usage and impact of WhatsApp. Some of these studies are for finding the impact of WhatsApp on the students and some are based on for the general public in a local region.

## Gaps Identified :

* + Adding a **Sentiment Analyser to the texts** — *Swear Words & Sentiments*!
  + Libraries called profanity\_check, TextBlob, VADER (Valence Aware Dictionary and Sentiment Reasoner) from NTLK and TextBlob.
  + Top 10 active/inactive users.

## Proposed work and methodology

* 1. Proposed Work : 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. Exploratory data analysis, first step in this to apply a sentiment analysis algorithm which provides positives negative and neutral part of the chat and is used to plot pie chart based on these parameters. To plot a line graph which shows author and message count of each date, to plot a line graph which shows author and message count of each author, Ordered graph of date vs message count, media sent by authors and their count, Display the message which is do not have authors, plot graph of hour vs message count.
  2. Methodology :

First we will collect the data in “.txt” format from the WhatsApp using export option in the chat. Then we will preprocess the data and get it in to appropriate data frame. Then accordingly we will add functionalities into out project such that data analysis is appropriate and correct.

1. **TECHNOLOGY USED AND SYSTEM REQUIREMENT**

### Technology Used :

Python: It is an interpreted, high-level general-purpose programming language. Created by Guido Van Rossum and first released in 1991. Its language constructs and objects-oriented approach aim to help programmer with clear, logical code for small and large-scale tools. Python is used for web development (server-side), software development, mathematics, it can be used alongside software to create workflows, it can connect to database systems, it can also read and modify files, it can be used to handle big data and perform complex mathematics and can be used for rapid prototyping, or for production-ready software development.

JSON: Java Script Object Notation is an open standard file format, and data interchange format, that uses human-readable text to store and transmit data objects consisting of attribute-value pairs and array data types. It is very common data format, with diverse range of applications. Such as serving as a replacement for xml in ajax systems. Json is a language-independent data format. It was derived from JavaScript, but many modern programming languages include code to generate and parse JSON-format data. The official Internet media type for Json is application/json. Json filenames use the extension (.json). When exchanging data between a browser and a server, the data can only be text. Json is text, and we can convert any JavaScript object into json and json to the server. We can also convert any json received from the server into JavaScript objects. This way we work with the data as JavaScript objects, with no complicated parsing and transactions.

MERN : MERN is a free and open-source JavaScript software stack for building dynamic web sites and web applications. Because all components of the

MERN stack support programs that are written in JavaScript, MERN applications can be written in one language for both server-side and client-side execution environments. Whereas MERN stands for MongoDB, Express.js, React and Node.js.

### System Requirement :

Minimum of 8GB RAM with i5 processor is required to run any MERN and data science project.

## Conclusion

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 intended, cannot be overemphasized. This work was able to discuss the WhatsApp application and its libraries, to create an analysis of a WhatsApp group chat and visually represent the top 10 and top 20 users in the chat groups. A pseudocode of the plot was given and at the end, visual representation of the plot was implemented. Also, an analysis of the top 10 and top 20 users were done. The system was done with python, and the python libraries that were implemented includes, NumPy, Pandas, Matplotlib and Seaborn. At the end of the work expected results were obtained and the analysis was able to show the level of participation of the various individuals on the given WhatsApp group. On serious note this system as the ability to analyze any WhatsApp group data input into it.

## References

1. Available from: https://[www.statista.com/statistics/260819/number-of-](http://www.statista.com/statistics/260819/number-of-) monthly-active-whatsapp-users/. Number of monthly active WhatsApp users worldwide rom April 2013 to March 2020 (in millions).
2. 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).
3. 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).
4. D. Radha, R. Jayaparvathy, D. Yamini, “Analysis on Social Media Addiction using Data Mining Technique”, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.7, pp. 23-26, April 2016