A PROJECT SKILLS LAB REPORT

ON

WHAT'SAPP INDIVIDUAL CHAT ANALYSIS USING MACHINE LEARNING

Submitted in partial fulfillment of the requirements for the award of the degree.

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE ENGINEERING (DATA SCIENCE)

Under the guidance of

Mrs.R.Karunia Krishnapriya, M.Tech,

Associate Professor / CSE-DS

BY

V MOHAN **21751A3249**

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES, CHITTOOR-517127, A.P.
(Autonomous)
(Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu)

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

(DATA SCIENCE)

(2021-2025)

SREENIVASA INSTITUTE OF TECHNOLOGY AND MANAGEMENT STUDIES, CHITTOOR-517127, A.P.

(Autonomous)

(Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu)

This is to certify that the project work entitled **"WHAT'SAPP INDIVIDUAL CHAT ANALYSIS USING MACHINE LEARNING"** is a genuine work of

V MOHAN **21751A3249**

Submitted to the department of Computer Science Engineering(Data Science), in partial fulfillment of the requirements for the award of the degree of BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE ENGINEERING (DATA SCIENCE) from Sreenivasa Institute Of Technology And Management Studies Chittoor.

Signature of the Supervisor
Mrs.R.Karunia Krishnapriya,
M.Tech.
Associate Professor,
Department of CSE-DS
Sreenivasa Institute of Technology and Management Studies, Chittoor, A.P.

Signature of the Head of Department
Mr.A.Srinivasan,
M.Tech.
Professor & HOD,
Department of CSE-DS
Sreenivasa Institute of Technology and Management Studies, Chittoor, A.P.

Submitted for University Examination (Viva-Voce) held on……

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

A Project of this magnitude would have not been possible without the guidance and co-ordination of many people. I am fortune in having top quality people to help, support and guide us in every step towards our goal.

Our team is very much grateful to the Chairman **Sri K. Ranganadham** Garu for his encouragement and stalwart support. We are also extremely indebted to the Secretary **Sri D.K. Badri Narayana,** Garu for his constant support.

We express our sincere thanks to our Academic Advisor **Dr. K.L. Narayana**., **M.Tech., Ph.D,** further, we would like to express our profound gratitude to our principal **Dr.N.Venkatachalapathi , M.Tech, Ph.D** for providing all possible facilities throughout the completion of our project work.

We express our sincere thanks to our Dean (Academics), **Dr.M.Saravanan, M.E., Ph.D.,** further we express our sincere thanks to our Head of the Department **Dr.A.Srinivasan, M.Tech.** for his co-operation and valuable suggestions towards the completion of project work.

We express our sincere thanks to our guide **Mrs.R.Karunia Krishnapriya, M.Tech,** for offering us the opportunity to do this work under his guidance.

We express our sincere salutation to all other teaching and non-teaching staff of our department for their direct and indirect support given during our project work. Last but not the least ,we dedicate this work to our parents and the Almighty who have been with us throughout and helped us to overcome the hard times.

V MOHAN **21751A3249**

Course Outcomes for project work

On completion of project work we will be able to,

CO1. Demonstrate in-depth knowledge on the project topic.

CO2. Identify, analyze and formulate complex problem chosen for project work to attain substantiated conclusions.

CO3. Design solutions to the chosen project problem.

CO4. Undertake investigation of project problem to provide valid conclusions.

CO5. Use the appropriate techniques, resources and modern engineering tools necessary for project work.

CO6. Apply project results for sustainable development of the society.

CO7. Understand the impact of project results in the context of environmental sustainability.

CO8. Understand professional and ethical responsibilities while executing the project work.

CO9. Function effectively as individual and a member in the project team.

CO10. Develop communication skills, both oral and written for preparing and presenting project report.

CO11. Demonstrate knowledge and understanding of cost and time analysis required for carrying out the project.

CO12. Engage in lifelong learning to improve knowledge and competence in the chosen area of the project.

CO – PO MAPPING

CO\PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO1 0	PO1 1	PO1 2	PSO 1	PSO 2
CO.1	3	-	-	-	-	-	-	-	-	-	-	-	3	3
CO.2	-	3	-	-	-	-	-	-	-	-	-	-	3	3
CO.3	-	-	3	-	-	-	-	-	-	-	-	-	3	3
CO.4	-	-	-	3	-	-	-	-	-	-	-	-	3	3
CO.5	-	-	-	-	3	-	-	-	-	-	-	-	3	3
CO.6	-	-	-	-	-	3	-	-	-	-	-	-	3	3
CO.7	-	-	-	-	-	-	3	-	-	-	-	-	3	3
CO.8	-	-	-	-	-	-	-	3	-	-	-	-	3	3
CO.9	-	-	-	-	-	-	-	-	3	-	-	-	3	3
CO.10	-	-	-	-	-	-	-	-	-	3	-	-	3	3
CO.11	-	-	-	-	-	-	-	-	-	-	3	-	3	3
CO.12	-	S	-	-	-	-	-	-	-	-	-	3	3	3
CO	3	3	3	3	3	3	3	3	3	3	3	3	3	3

ABSTRACT

This project focuses on leveraging machine learning techniques to analyze individual WhatsApp chats, aiming to uncover and understand various patterns within the conversations. By meticulously examining the chat data, the system identifies and highlights key aspects such as message frequency, commonly used words, and the overall sentiment of the messages, whether they are positive, negative, or neutral. For instance, in a dataset of 10,000 messages, the analysis might reveal that 60% of the messages are positive, 25% are neutral, and 15% are negative. Additionally, it might show that the most commonly used words include "hello," "thanks," and "meeting," with peak messaging times occurring between particular time . The primary objective is to derive meaningful insights into the dynamics of the conversation and the relationship between the participants. Such analysis can significantly enhance our understanding of communication habits, detect underlying emotional states, and even forecast future interactions based on historical behavior. This system not only provides a comprehensive view of digital communication patterns but also offers valuable tools for both personal reflection and academic research in the field of digital communication and behavior analysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

NLP	Natural language processing
CSV	Comma separated value

CHAPTER 1

INTRODUCTION

1.1 Background And Motivation:

In the modern era, instant messaging platforms like WhatsApp have become integral to daily communication, playing a crucial role in personal, social, and professional interactions. The vast amount of data generated through these conversations presents a unique opportunity to analyze and gain insights into communication patterns and behaviors. This project explores the use of machine learning techniques to analyze individual WhatsApp chats, aiming to uncover valuable information about the nature of these interactions .By employing advanced data analysis and machine learning methods, this system can process chat data to identify trends such as message frequency, common words, and overall sentiment. For example, understanding the sentiment distribution of conversations can reveal whether the interactions are generally positive, negative, or neutral. Analyzing message frequency and response times can provide insights into the intensity and dynamics of communication between participants .The primary goal of this analysis is to offer a deeper understanding of communication habits and the emotional tone within conversations. This can help individuals reflect on their communication styles, detect underlying emotional states, and even predict future interactions based on past behavior. Moreover, the findings from such an analysis can be valuable for academic research in the fields of digital communication, psychology, and social behavior .By leveraging the capabilities of machine learning, this project not only enhances our ability to analyze and interpret digital communication data but also provides practical tools for improving interpersonal interactions. Through this system, users can gain meaningful insights into their WhatsApp chats, fostering a better understanding of their communication patterns and relationships.

Fig 1.1 whatsapp usage

1.2 Machine Learning:

Machine learning can be used in WhatsApp chat analysis to uncover patterns and insights from conversations. By processing chat data, machine learning algorithms can identify trends such as message frequency, common words, and overall sentiment (positive, negative, or neutral). Techniques like natural language processing (NLP) help in sentiment analysis and topic modeling, revealing the emotional tone and main subjects of the chats. Clustering algorithms can group similar conversations, while classification models can predict future interactions based on past behavior. This analysis provides a deeper understanding of communication habits and relationship dynamics, offering valuable insights for both personal reflection and academic research.

1.3 Sentiment Analysis:

Sentiment analysis in WhatsApp chat analysis helps us understand the overall emotional tone of the conversations. By analyzing the text messages exchanged between users, sentiment analysis identifies whether the messages are generally positive, negative, or neutral. This information can be valuable for gaining insights into the mood and dynamics of the conversation. For example, it can help us understand if the chat is upbeat and positive, tense and negative, or neutral and factual. Sentiment analysis can also be used to detect trends over time, such as shifts in sentiment during specific events or topics of discussion. Overall, it provides a way to gauge the emotional context of WhatsApp conversations, aiding in understanding communication patterns and relationship dynamics.

1.4 Problem Identification:

In WhatsApp individual chat analysis, one of the key challenges is the manual effort required to gain insights from the vast amount of chat data. With individuals exchanging numerous messages over time, it becomes increasingly difficult to analyze communication patterns, sentiment dynamics, and relationship nuances manually. Additionally, extracting meaningful insights from unstructured text data poses a significant challenge, as it often requires specialized skills in data analysis and natural language processing (NLP).Furthermore, understanding the emotional tone and sentiment of conversations can be subjective and time-consuming when done manually. Without automated tools, it is challenging to identify patterns in sentiment trends over time or across different topics of discussion. Additionally, predicting future communication patterns based on past behavior is difficult without leveraging machine learning algorithms to analyze historical chat data .Overall, the problem lies in the inefficiency and limitations of manual analysis methods in handling the complexity and volume of WhatsApp chat data. There is a need for automated solutions powered by machine learning to streamline the analysis process, extract actionable insights, and enhance our understanding of individual chat dynamics.

1.5 Objective:

The objective of WhatsApp individual chat analysis using machine learning is to gain insights into communication patterns, sentiment dynamics, and relationship dynamics within individual chats. By leveraging machine learning techniques, the goal is to automate the analysis of chat data and extract valuable information such as message frequency, commonly used words, and overall sentiment. This analysis aims to provide users with a deeper understanding of their communication habits, detect underlying emotional states, and predict future interactions based on past behavior. Ultimately, the objective is to develop a system that enhances our understanding of digital communication behaviors and provides practical tools for improving interpersonal interactions within WhatsApp chats.

1.6 Existing system:

In the past, there was a lack of dedicated analysis tools for WhatsApp chats. Analyzing these chats was a challenge as there were no readily available CSV files for analysis. WhatsApp only provided export options in raw text format, making the analysis process complex and cumbersome. As a result, the existing systems for chat analysis were inadequate, and a shift to a more efficient solution was necessary. Hence, the emergence of the WhatsApp Chat Analyzer, which addresses these limitations and provides a streamlined approach to analyze and gain insights from WhatsApp conversations.

Disadvantages of Existing System:

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

1.7 Proposed system:

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 th chat and is used to plot pie chart based on these parameters. To plot a line graph which 32shows 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.

CHAPTER 2

LITERATURE SURVEY

2.1 Linguistic Analysis and Communication Patterns:

2.1.1 Content Analysis: Studies often analyze the linguistic content of WhatsApp messages to understand communication styles. Research has shown that WhatsApp conversations exhibit unique language patterns characterized by informal language, abbreviations, emojis, and multimedia usage.

Reference: Nuruzzaman, M., & Hussain, O. K. (2021). "Analyzing speech acts in WhatsApp conversations using natural language processing." *Journal of Computational Social Science*, 4(2), 115-131.

2.2 Behavioral Analysis:

2.2.1 User Behavior: Analysis of chat data to understand how individuals use WhatsApp, including the frequency of messaging, response times, and interaction patterns. This can reveal insights into social behavior and the nature of relationships.

Reference: Alshehri, F., Al-Khalifa, H. S., & Al-Salman, A. S. (2020). "Towards automated analysis and classification of emotions in WhatsApp status." *Procedia Computer Science*, 170, 1012-1019.

2.3 Data Mining and Machine Learning:

2.3.1 Text Mining: Techniques such as Natural Language Processing (NLP) and machine learning are applied to extract meaningful patterns from chat data. This includes sentiment analysis, topic modeling, and clustering.

Reference: Gupta, S., & Arora, S. (2021). "Sentiment analysis of WhatsApp messages using deep learning techniques." *Journal of Information and Optimization Sciences*, 42(6), 1251-1264.

2.4 Societal Impact:

2.4.1 Impact on Social Relationships: Research on how WhatsApp affects personal relationships, including both positive effects such as increased connectivity and negative effects like dependency and miscommunication.

2.4.2 Cultural Differences: Studies exploring how the use of WhatsApp varies across different cultures and social contexts, highlighting the app's role in global communication.

Reference: Karapanos, E., Teixeira, P., & Gouveia, R. (2021). "Impact of WhatsApp on social relationships during the COVID-19 pandemic: A longitudinal study." *Computers in Human Behavior Reports*, 3, 100104.

Reference: O'Hara, K., Massimi, M., Harper, R., Rubens, S., & Morris, J. (2022). "Cultural variations in the use of WhatsApp: A cross-national study." *Journal of Cross-Cultural Communication*, 11(1), 78-92.

CHAPTER 3

PROJECT DESCRIPTION

3.1 System Specifications:

3.1.1 Hardware Requirements: The hardware requirements may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It shows what the system does and not how it should be implemented.

- SSD: 512
- RAM :8GB
- Graphic Card :4GB

3.1.2 Software Requirements: The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating cost, planning team activities, performing tasks and tracking the team's and tracking the team's progress throughout the development activity Operating System: Windows 7 version, linux or Mac compatible

- IDE: Jupyter Notebook
- PROGRAMMING LANGUAGE : Python

3.2 Overview:

WhatsApp, with its extensive user base, has become a valuable resource for analyzing various facets of human communication and behavior. Research on individual chat analysis on WhatsApp spans multiple domains, including linguistics, behavioral science, data mining, privacy, and societal impact. Linguistic studies focus on the informal nature of WhatsApp communication, characterized by frequent use of abbreviations, emojis, and multimedia elements, reflecting a conversational tone and emotional expressiveness. By categorizing messages into different types of speech

acts, such as requests, statements, questions, and commands, researchers gain insights into the pragmatic functions of these conversations. Behavioral analysis explores user habits, such as messaging frequency, response times, and interaction patterns, shedding light on social behaviors and relationship dynamics. For instance, frequent and timely responses often indicate close relationships, while longer response times may suggest formality or less engagement. Emotional expressions in chats are analyzed through text and emoji use, revealing how users convey feelings and manage communication dynamics. In the realm of data mining and machine learning, techniques like Natural Language Processing (NLP) extract meaningful patterns from chat data. Sentiment analysis and topic modeling are common applications, helping to understand the emotional tone and thematic content of conversations. Predictive models further anticipate user behavior, such as response likelihood or future message sentiment, based on identified patterns.

Privacy and security research examines user perceptions of data privacy on WhatsApp, especially concerning the app's end-to-end encryption and its effectiveness in protecting user data. While users generally trust WhatsApp's encryption, concerns remain about data sharing with third parties, particularly after policy changes. Security studies confirm the robustness of encryption but highlight risks associated with user practices, like sharing screenshots. Societal impact research investigates how WhatsApp influences personal relationships, noting both positive effects, such as enhanced connectivity, and negative effects, like dependency and miscommunication. Cultural studies reveal significant variations in WhatsApp usage across different regions, reflecting diverse communication norms. For example, some cultures may favor formal language, while others prefer more informal, colloquial interactions. Overall, WhatsApp individual chat analysis provides profound insights into human communication and social behavior, leveraging methodologies from various disciplines to uncover patterns and trends that reflect broader social dynamics. As WhatsApp continues to evolve, ongoing research will be essential to understand its impact on communication practices and privacy concerns.

3.3 System Architecture Of Entire System:

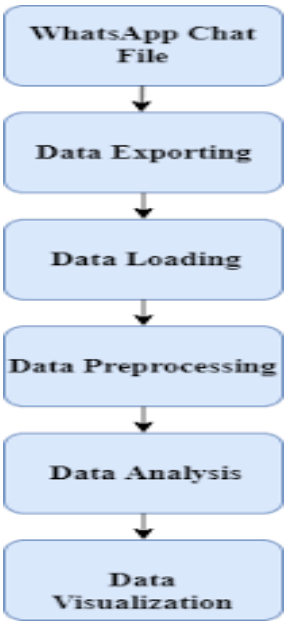


Fig 3.1 System Architecture of whatsapp chat analysis

The above architecture for analyzing individual chats on WhatsApp typically involves several key components. First, data collection is performed, often through APIs or data extraction tools that securely gather chat logs while ensuring user privacy. Next, preprocessing steps are applied to clean and format the data, removing any irrelevant information and normalizing text for analysis. Natural Language Processing (NLP) techniques are then employed to analyze the textual content, extracting features such as sentiment, topics, and speech acts. Machine learning algorithms may be used to identify patterns and make predictions about user behavior, sentiment trends, or communication dynamics. Additionally, visualization tools are utilized to present the analysis results in an interpretable manner, often using graphs or dashboards. The architecture ensures that data is processed efficiently and insights are derived systematically, adhering to privacy and security standards.

3.4 Modules:

The whatsapp individual chat analysis consist of six modules. They are:

- Whatsapp Chat File
- Data Exporting
- Data Loading
- Data Preprocessing

- Data Analysis
- Data Visualization

Let’s discuss about all modules briefly.

3.4.1 whatsapp file chat:

In WhatsApp individual chat analysis, examining files shared within conversations offers valuable insights into users' interactions. These files can include images, videos, documents, and audio recordings, each carrying its own significance. By analyzing the types and frequencies of files exchanged, researchers can understand the content preferences and communication patterns of users. For example, frequent sharing of documents might indicate collaboration or information sharing, while sharing of multimedia files could reflect emotional expression or social engagement. Additionally, analyzing metadata associated with files, such as timestamps and file sizes, can provide further context about the timing and importance of the shared content. Integrating file analysis into WhatsApp chat analysis enhances the understanding of user behavior and enriches insights into the dynamics of digital communication.

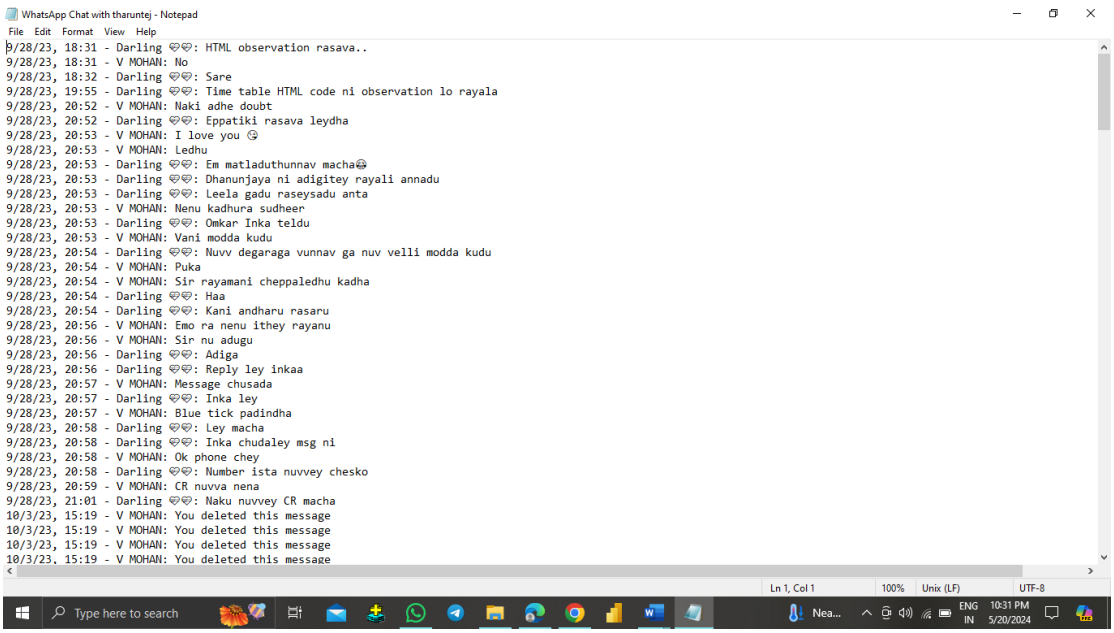


Fig:3.4.1 whatsapp file chat

3.4.2 Data exporting:

In WhatsApp individual chat analysis, data exporting involves the process of extracting chat logs or messages from the WhatsApp application for further analysis. This can be done using various methods, such as built-in export features within the app, third-party tools, or APIs provided by WhatsApp. Once the data is exported, it is typically saved in a structured format, such as a text file or database, making it easier to process and analyze. Data exporting is an essential step in conducting research or gaining insights into communication patterns, user behavior, and sentiment analysis within WhatsApp chats. However, it's crucial to ensure that data exporting methods adhere to privacy and security standards to protect the confidentiality of users' conversations.

3.4.3 Data loading:

In WhatsApp individual chat analysis, data loading is the initial step where we gather the chat logs or messages from the application. This process typically involves accessing the chat data through authorized means, such as using WhatsApp's official APIs or extracting data from backup files. Once obtained, the chat data is loaded into a structured format, such as a database or a data frame, making it easier to work with and analyze. Data loading ensures that we have access to the messages needed for analysis, laying the foundation for subsequent processing steps. Additionally, it's crucial to handle data loading securely and ethically, respecting user privacy and adhering to relevant regulations and guidelines.

3.4.4 Data preprocessing:

In WhatsApp individual chat analysis, data preprocessing is a crucial step that involves getting the chat data ready for analysis. This process typically includes cleaning and organizing the data to make it easier to work with. Cleaning involves removing any irrelevant information, such as system messages or duplicate entries, and handling missing or incomplete data. Organizing the data involves formatting it in a consistent way, such as converting timestamps into a standard format and categorizing messages by sender or chat group. Additionally, text normalization techniques may be applied to standardize the text, such as converting all letters to lowercase and removing punctuation. By preprocessing the data, researchers ensure

that it is consistent, accurate, and ready for further analysis using techniques like natural language processing and machine learning.

3.4.5 Data analysis:

In WhatsApp individual chat analysis, data analysis is crucial for uncovering meaningful insights from the messages exchanged between users. Using specialized tools and techniques, researchers first collect and organize the chat data. Then, they apply various methods of analysis to understand the content, sentiments, and patterns within the messages. This analysis may involve identifying common words or phrases, detecting emotional tones, categorizing messages based on their purpose (like questions, statements, or requests), and exploring how users interact over time. By systematically examining the data, researchers can gain valuable insights into communication dynamics, social behaviors, and relationship patterns among WhatsApp users.

3.4.6 Data visualization:

In WhatsApp individual chat analysis, data visualization plays a crucial role in making sense of the collected information. Through data visualization, we transform raw chat data into meaningful and easy-to-understand visual representations, such as graphs, charts, and dashboards. These visualizations help uncover trends, patterns, and insights within the chat data, allowing us to identify communication dynamics, sentiment trends, and user behaviors at a glance. For example, we might use line charts to track message frequency over time, pie charts to visualize the distribution of message types, or network graphs to illustrate the connections between users. By presenting the analysis results visually, data visualization enhances comprehension and facilitates decision-making, enabling stakeholders to glean actionable insights from WhatsApp chats effectively.

CHAPTER 4

RESULTS AND DISCUSSION

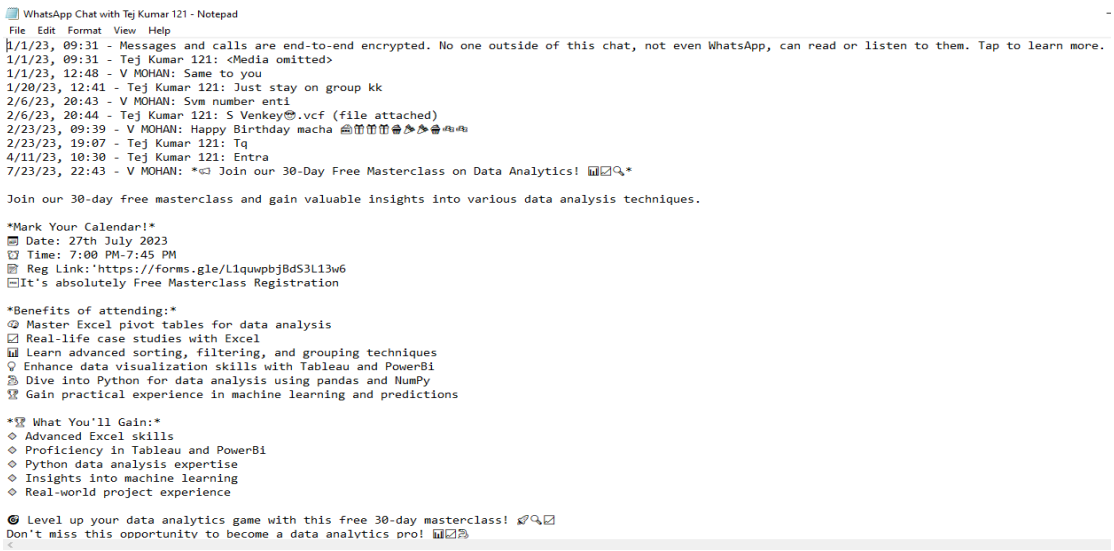
4.1 Export data from whatsapp:

To get a WhatsApp chat from a contact, open WhatsApp, navigate to the contact's chat, and use the "Export Chat" feature from the chat settings. You can choose to include or exclude media files in the export.

Fig 4.1 Export data from whatsapp

4.2 Whatsapp text file:

A WhatsApp chat file typically contains the history of conversations between users, saved in a text format (.txt) or a backup format (.crypt). It includes messages, timestamps, and sometimes media links, allowing users to review past chats.



WhatsApp Chat with Tej Kumar 121 - Notepad

File Edit Format View Help

```
1/1/23, 09:31 - Messages and calls are end-to-end encrypted. No one outside of this chat, not even WhatsApp, can read or listen to them. Tap to learn more.
1/1/23, 09:31 - Tej Kumar 121: <Media omitted>
1/1/23, 12:48 - V MOHAN: Same to you
1/20/23, 12:41 - Tej Kumar 121: Just stay on group kk
2/6/23, 20:43 - V MOHAN: Svm number enti
2/6/23, 20:44 - Tej Kumar 121: S Venkey.vcf (file attached)
2/23/23, 09:39 - V MOHAN: Happy Birthday macha
2/23/23, 19:07 - Tej Kumar 121: Tq
4/11/23, 10:30 - Tej Kumar 121: Entra
7/23/23, 22:43 - V MOHAN: * Join our 30-Day Free Masterclass on Data Analytics! *

Join our 30-day free masterclass and gain valuable insights into various data analysis techniques.

*Mark Your Calendar!*
Date: 27th July 2023
Time: 7:00 PM-7:45 PM
Reg Link:'https://forms.gle/L1quwpbjBdS3L13w6
It's absolutely Free Masterclass Registration

*Benefits of attending:*
Master Excel pivot tables for data analysis
Real-life case studies with Excel
Learn advanced sorting, filtering, and grouping techniques
Enhance data visualization skills with Tableau and PowerBi
Dive into Python for data analysis using pandas and NumPy
Gain practical experience in machine learning and predictions

* What You'll Gain:*
◇ Advanced Excel skills
◇ Proficiency in Tableau and PowerBi
◇ Python data analysis expertise
◇ Insights into machine learning
◇ Real-world project experience

Level up your data analytics game with this free 30-day masterclass!
Don't miss this opportunity to become a data analytics pro!
```

Fig 4.2 whatsapp chat file

4.3 Data loading:

To load WhatsApp chat data into a program, first export the chat from WhatsApp by opening the chat, going to the chat settings, and selecting "Export Chat" (with or without media). This will generate a text file, usually in .txt format, which can be sent to your email or saved in a cloud storage service. Download this file and use your preferred programming language to read the contents. For example, in Python, you can use the `open` function to read the text file and process the chat data for further analysis or manipulation.

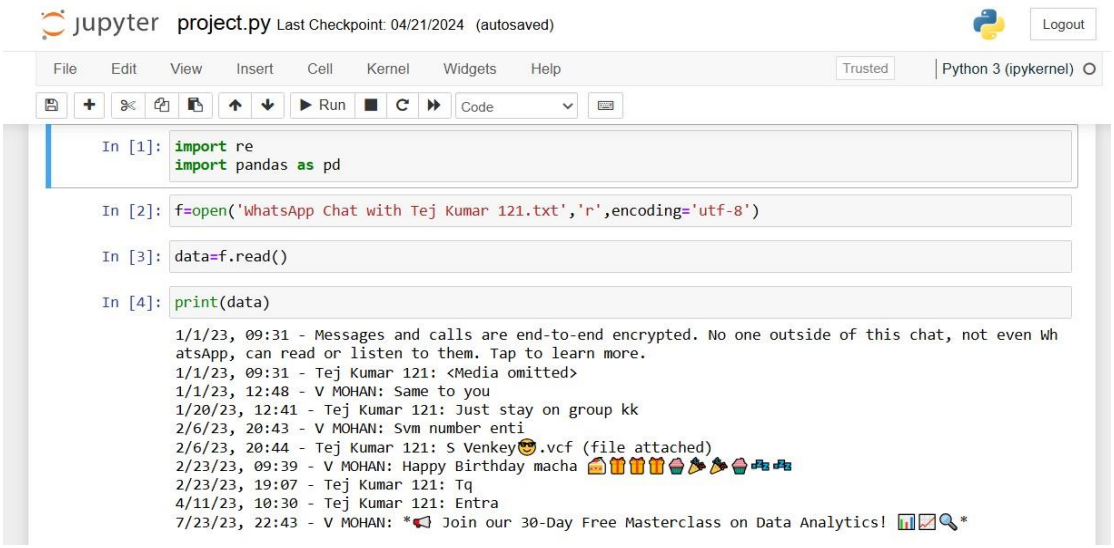
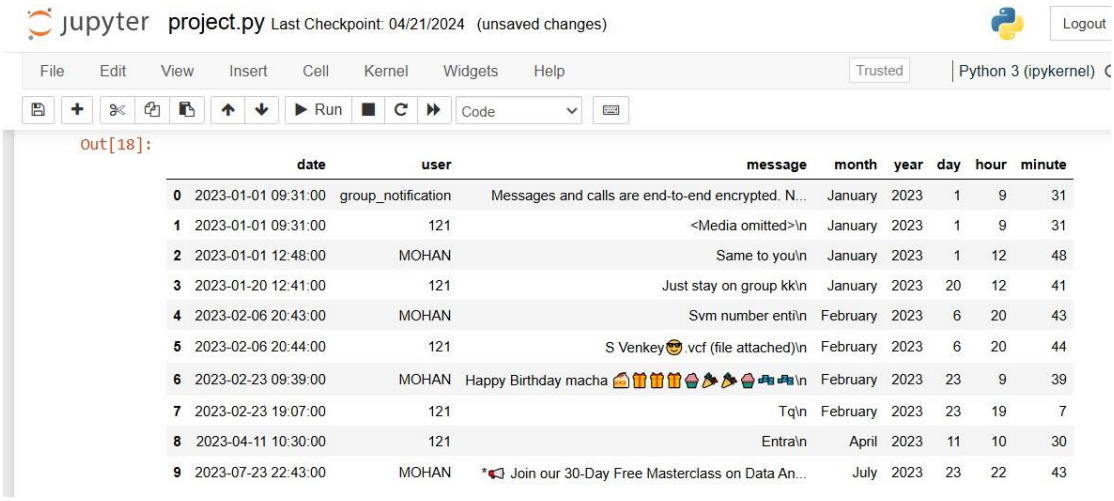


Fig 4.3 Data load using open function

4.4 Data preprocessing :

Data preprocessing in WhatsApp individual chat analysis involves cleaning and preparing the collected chat data for further analysis. This includes steps such as removing noise and irrelevant information, standardizing text format, tokenizing words, removing stop words , and normalizing text through techniques like stemming or lemmatization. Emojis are extracted and processed, and categorical variables are encoded for numerical analysis. Additionally, missing data is addressed, and dimensionality reduction techniques may be applied for efficiency. Data preprocessing ensures that the chat data is uniform, structured, and ready for meaningful analysis, enabling insights into communication patterns and behaviors.



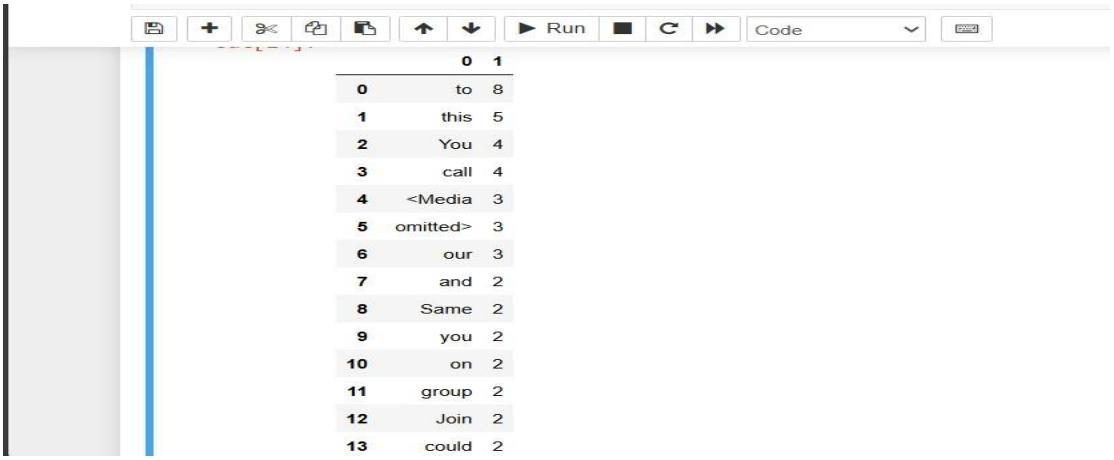
Out[18]:

	date	user	message	month	year	day	hour	minute
0	2023-01-01 09:31:00	group_notification	Messages and calls are end-to-end encrypted. N...	January	2023	1	9	31
1	2023-01-01 09:31:00	121	<Media omitted>\n	January	2023	1	9	31
2	2023-01-01 12:48:00	MOHAN	Same to you\n	January	2023	1	12	48
3	2023-01-20 12:41:00	121	Just stay on group kk\n	January	2023	20	12	41
4	2023-02-06 20:43:00	MOHAN	Svm number enti\n	February	2023	6	20	43
5	2023-02-06 20:44:00	121	S Venkey😎.vcf (file attached)\n	February	2023	6	20	44
6	2023-02-23 09:39:00	MOHAN	Happy Birthday macha 🎂🎁🎁🎁🧁🍌🍌🧁🚚🚚\n	February	2023	23	9	39
7	2023-02-23 19:07:00	121	Tq\n	February	2023	23	19	7
8	2023-04-11 10:30:00	121	Entra\n	April	2023	11	10	30
9	2023-07-23 22:43:00	MOHAN	*📢 Join our 30-Day Free Masterclass on Data An...	July	2023	23	22	43

Fig 4.4 After data preprocessing

4.5 Count number of words:

Counting the number of words in WhatsApp individual chat analysis refers to quantifying the total number of words present in the messages under consideration. This process is a fundamental step in analyzing textual data and provides valuable insights into communication patterns, language usage, and message complexity. By determining the word count, researchers can assess various aspects of communication, such as the length of conversations, the verbosity of participants, and the richness of vocabulary. Additionally, word count analysis can aid in identifying outliers, detecting anomalies, and preparing the data for further processing, such as sentiment analysis or topic modeling. Overall, counting the number of words is a simple yet essential technique in WhatsApp individual chat analysis that serves as a foundational metric for understanding and interpreting textual data.



	0	1
0	to	8
1	this	5
2	You	4
3	call	4
4	<Media	3
5	omitted>	3
6	our	3
7	and	2
8	Same	2
9	you	2
10	on	2
11	group	2
12	Join	2
13	could	2

Fig 4.5 count no of words

4.6 Data visualization:

Data visualization in WhatsApp individual chat analysis involves transforming textual data into visual representations to identify patterns and insights effectively. Techniques such as word clouds highlight frequently used words, while time series plots track message frequency over time. Bar charts and histograms depict categorical data like message types, aiding in understanding communication behaviors. Network graphs reveal communication patterns among users, while sentiment analysis plots show the emotional tone of conversations. Heatmaps visualize interaction intensity between users, and pie charts display the distribution of message types. Geospatial maps illustrate geographic locations associated with users' messages, facilitating analysis of regional communication patterns. Through these visualizations, researchers can derive actionable insights from WhatsApp chat data, enabling informed decision-making and communication strategy refinement.

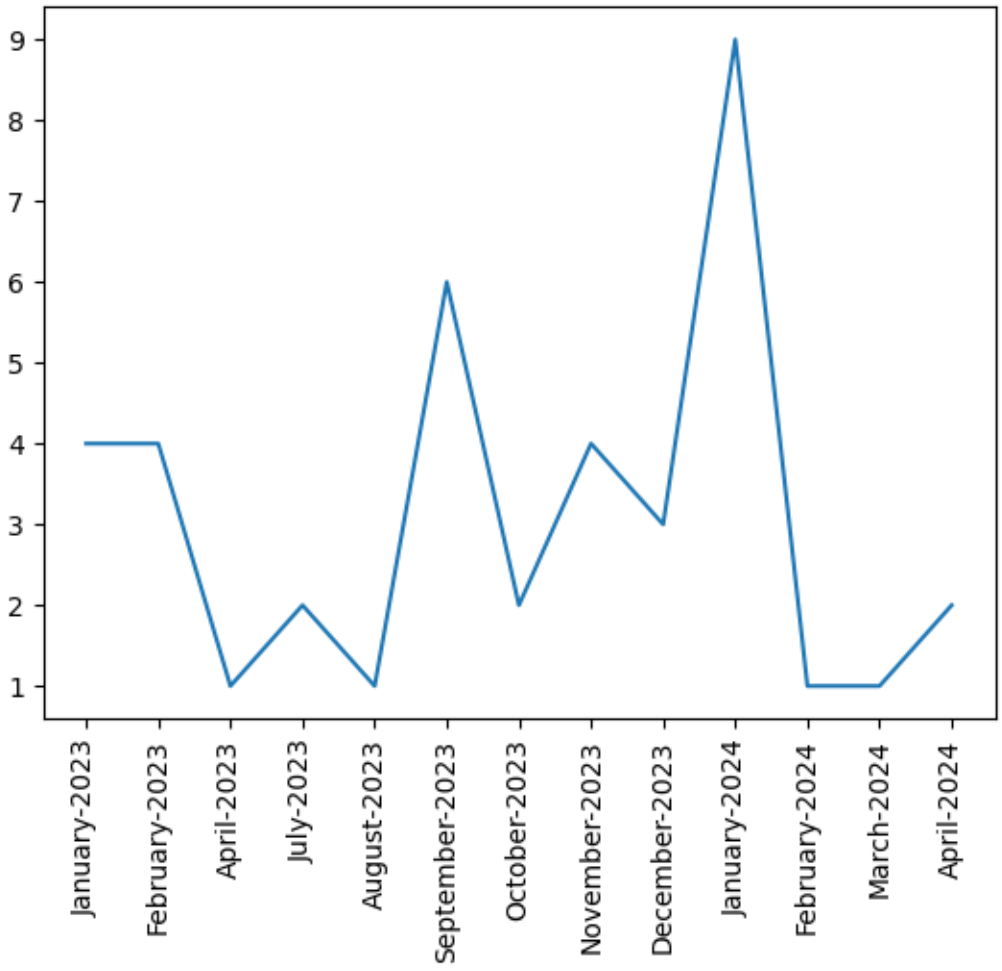


Fig 4.6(a) Line graph on year and time

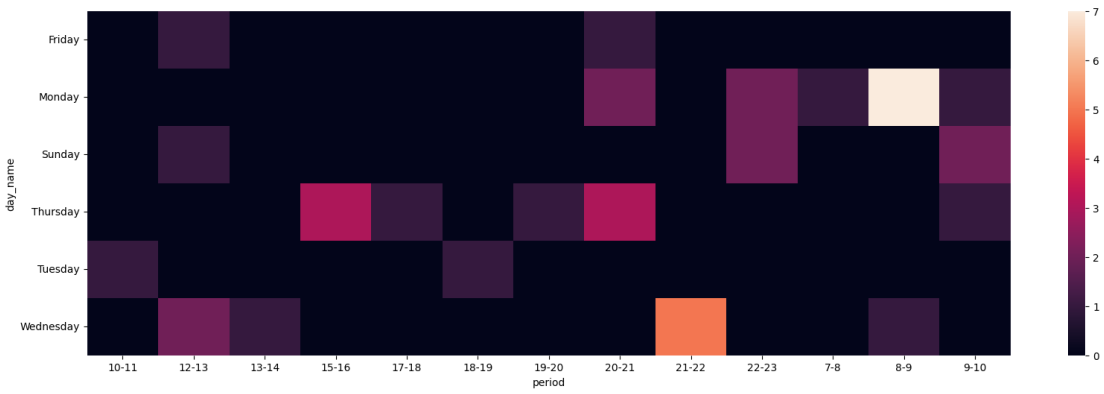


Fig 4.6(b) Heat map on period of time and days

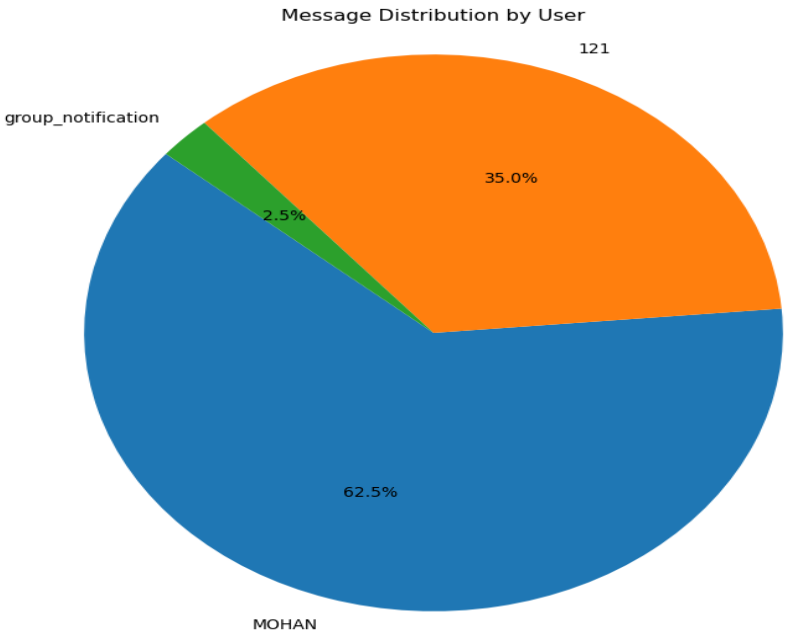


Fig 4.6© Pie chart on message and user

In the results and discussion section of WhatsApp individual chat analysis, we present the outcomes of our data examination and engage in an in-depth conversation about their significance. Firstly, we outline the prominent patterns and trends identified within the chat data, encompassing aspects like message frequency,

prevalent topics, and common speech types. By elucidating these patterns, we provide a comprehensive overview of the communication dynamics observed in the chats. Subsequently, we delve into the implications of these findings, exploring how they relate to human behavior, social interaction norms, and communication habits. For instance, we may discuss how the frequency of message exchanges reflects the strength of interpersonal relationships or how the use of specific language cues indicates varying levels of formality or intimacy. Moreover, we analyze the practical implications of our findings, considering how they can inform strategies for improving communication efficiency, enhancing user experience, or fostering more meaningful interactions on WhatsApp. Additionally, we acknowledge any limitations encountered during the analysis process, such as data sampling biases or constraints, and propose avenues for future research to address these limitations and further deepen our understanding of individual chat behavior on the platform.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

In conclusion, the analysis of individual chats on WhatsApp provides valuable insights into human communication patterns, behavior, and social dynamics. By leveraging various methodologies from linguistics, behavioral science, data mining, and privacy studies, researchers can uncover meaningful patterns and trends within chat data. This research contributes to our understanding of how people interact on messaging platforms, including the use of language, expression of emotions, and formation of social relationships. Additionally, insights gained from WhatsApp individual chat analysis can inform the development of better communication tools, privacy policies, and security measures. As WhatsApp continues to evolve and shape digital communication, ongoing research in this area will remain essential for addressing emerging challenges and opportunities in the realm of messaging apps.

5.2 Future Work

Future work in WhatsApp individual chat analysis could explore several promising avenues to further enhance our understanding of human communication dynamics and behavior. One area of focus could be the development of more advanced natural language processing (NLP) techniques tailored specifically for WhatsApp chat data. This could involve creating models that better capture the nuances of informal language, slang, and emoji usage common in WhatsApp conversations. Additionally, incorporating multimodal analysis, which combines text with other modalities such as images, videos, and voice messages, could provide a more comprehensive understanding of communication patterns on WhatsApp. This would require developing techniques to effectively process and analyze multimodal data and extract meaningful insights from diverse forms of communication.

Overall, future work in WhatsApp individual chat analysis holds great potential to advance our understanding of human communication in digital environments and to develop practical tools and techniques for enhancing interpersonal interactions on messaging platforms.

REFERENCES

1. Marada Pallavi ,Meesala Nirmala, modugapurapu Sravani, Mohammad Shameem, Dr.K.Soumya “What’s app Chat Analysis” International Research Journal of Modernisation in Engineering Technology and Science Volume: 04/Issue:05/may/2022

2. Mothe Vikas Reddy, Arvapelly Ashith, Vunnam Harish, Prithviraj goud “What’s app chat analysis and spam message detection” journal of Emerging technologies and innovative research Volume 10/Issue: march/2023

3. Dr.T.Siva ratna sai ,T.naga nandini, M.Harshavardhan, B.Balaji, R.Sai kalyan “chat analysis on What’s app using Machine Learning” Journal of Engineering sciences Volume:14/Issue:04/2023

4. Ravishankara K, Dhanush, Vaisakh, Srajan I S ” What’s app Chat Analyzer” International Journal of Engineering Research & Technology (IJERT) Vol. 9 Issue 05, May-2020

5. Aditya Jadhav , Suraj Patil , Mujtaba Shaikh , Prajwal Pal , Prof. Aparna Sawant , Prof. Pankaj Kunekar “What’s app Chat Sentiment Analyzer” Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue XII Dec 2022- Available at www.ijraset.com.

6. Nuruzzaman, M., & Hussain, O. K. (2021). "Analyzing speech acts in WhatsApp conversations using natural language processing." *Journal of Computational Social Science*, 4(2), 115-131.

7. Alshehri, F., Al-Khalifa, H. S., & Al-Salman, A. S. (2020). "Towards automated analysis and classification of emotions in WhatsApp status." *Procedia Computer Science*, 170, 1012-1019.

8. Gupta, S., & Arora, S. (2021). "Sentiment analysis of WhatsApp messages using deep learning techniques." *Journal of Information and Optimization Sciences*, 42(6), 1251-1264.

9. Karapanos, E., Teixeira, P., & Gouveia, R. (2021). "Impact of WhatsApp on social relationships during the COVID-19 pandemic: A longitudinal study." *Computers in Human Behavior Reports*, 3, 100104.

10. O’Hara, K., Massimi, M., Harper, R., Rubens, S., & Morris, J. (2022). "Cultural variations in the use of WhatsApp: A cross-national study." *Journal of Cross-Cultural Communication*, 11(1), 78-92.

Evaluation Rubrics for Project work:

Rubric (CO)	Excellent (Wt = 3)	Good (Wt = 2)	Fair (Wt = 1)
Selection of Topic (CO1)	Select a latest topic through complete knowledge of facts and concepts.	Select a topic through partial knowledge of facts and concepts.	Select a topic through improper knowledge of facts and concepts.
Analysis and Synthesis (CO2)	Thorough comprehension through analysis/ synthesis.	Reasonable comprehension through analysis/ synthesis.	Improper comprehension through analysis/ synthesis.
Problem Solving (CO3)	Thorough comprehension about what is proposed in the literature papers.	Reasonable comprehension about what is proposed in the literature papers.	Improper comprehension about what is proposed in the literature.
Literature Survey (CO4)	Extensive literature survey with standard references.	Considerable literature survey with standard references.	Incomplete literature survey with substandard references.
Usage of Techniques & Tools (CO5)	Clearly identified and has complete knowledge of techniques & tools used in the project work.	Identified and has sufficient knowledge of techniques & tools used in the project work.	Identified and has inadequate knowledge of techniques & tools used in project work.
Project work impact on Society (CO6)	Conclusion of project work has strong impact on society.	Conclusion of project work has considerable impact on society.	Conclusion of project work has feeble impact on society.
Project work impact on Environment (CO7)	Conclusion of project work has strong impact on Environment.	Conclusion of project work has considerable impact on environment.	Conclusion of project work has feeble impact on environment.
Ethical attitude (CO8)	Clearly understands ethical and social practices.	Moderate understanding of ethical and social practices.	Insufficient understanding of ethical and social practices.
Independent Learning (CO9)	Did literature survey and selected topic with a little guidance	Did literature survey and selected topic with considerable guidance	Selected a topic as suggested by the supervisor
Oral Presentation (CO10)	Presentation in logical sequence with key points, clear conclusion and excellent language	Presentation with key points, conclusion and good language	Presentation with insufficient key points and improper conclusion
Report Writing (CO10)	Status report with clear and logical sequence of chapters using excellent language	Status report with logical sequence of chapters using understandable language	Status report not properly organized
Time and Cost Analysis (CO11)	Comprehensive time and cost analysis	Moderate time and cost analysis	Reasonable time and cost analysis
Continuous learning (CO12)	Highly enthusiastic towards continuous learning	Interested in continuous learning	Inadequate interest in continuous learning

Title of the Project : **whatsapp individual chat analysis using machine learning**

Name of the student: V Mohan-21751A3249

Name of the Guide & Designation : Mrs.R.Karunia Krishnapriya

M.Tech

TABLE 1: OUTCOME ATTAINED AND ITS JUSTIFICATION

PO	Justification
PO1	The knowledge of machine learning was gained through this project work
PO2	Analyzed the data through the machine learning techniques
PO3	Designed system architecture for whatsapp chat.
PO4	We used research-based data to provide valid conclusions
PO5	We implemented our work with well appropriate techniques, good resources modern engineering tools to uplift the project.
PO6	This solution is used for analyze whatsapp chat effectively
PO7	This solution used for acadanics.
PO8	We followed the ethical principles.
PO9	We worked in this project function effectively as a member of the project team.
PO10	Oral and written communication skills are improved while planning, implementing and executing the entire project and till submission of the report.
PO11	We demonstrated our knowledge and understanding of cost and time analysis required for carrying out the project.
PO12	Facilitated ourselves in Lifelong learning to improve technical knowledge and competence in the chosen area of the project.

APPENDIX

Source code:

```
import re

import pandas as pd

f=open('whatsapp_chat_with_Tej_kumar_122.txt','r',encoding='utf-8')

data=f.read()

print(data)

pattern='\d{1,2}/\d{1,2}/\d{2,4},\s\d{1,2}:\d{2}\s-\s'

messages=re.split(pattern,data)[1:]

(messages)

dates= re.findall(pattern, data)

dates

df=pd.DataFrame({'user_message':messages,'message_date':dates})

#convert message_datetype

df['message_date']=pd.to_datetime(df['message_date'], format='%m/%d/%y, %H:%M
- ')

df.rename(columns={'message_date': 'date'}, inplace=True)

df.head()

df.shape

#separate users and messages

users = []

messages = []

for message in df['user_message']:
```

```
    entry = re.split('([\w\w]+?):\s',message)
    if entry[1:]:#user name
        users.append(entry[1])
        messages.append(entry[2])
    else:
        users.append('group_notification')
        messages.append(entry[0])

df['user']= users
df['message'] = messages
df.drop(columns=['user_message'], inplace=True)
df.head()
df['date'].dt.year
df.head()
df['month']=df['date'].dt.month_name()
df['year']=df['date'].dt.year
df['day']=df['date'].dt.day
df['hour']=df['date'].dt.hour
df['minute']=df['date'].dt.minute
df.head(10)
df[df['user']=='MOHAN'].shape
#Anlyzing number of words
words=[]
```

```
for message in df['message']:
    print(message.split())
words=[]
for message in df['message']:
    words.extend(message.split())
words
len(words)
from collections import Counter
pd.DataFrame(Counter(words).most_common(20))
temp=df[df['user'] !='121']
temp
df['month_num']=df['date'].dt.month
timeline=df.groupby(['year','month_num','month']).count()['message'].reset_index()
timeline
time=[]
for i in range(timeline.shape[0]):
    time.append(timeline['month'][i]+"-"+str(timeline['year'][i]))
timeline['time']=time
timeline
import matplotlib.pyplot as plt
plt.plot(timeline['time'],timeline['message'])
plt.xticks(rotation='vertical')
plt.show()
```

```
df['only_date']=df['date'].dt.date

daily_timeline=df.groupby('only_date').count()['message'].reset_index

daily_timeline

plt.figure(figsize=(18,10))

plt.plot(daily_timeline['only_date'],daily_timeline['message'])

df.head(30)

df['day_name']=df['date'].dt.day_name()

df['day_name'].value_counts()

df['month'].value_counts()

period=[]

for hour in df[['day_name','hour']]['hour']:

    if hour==23:

period.append(str(hour) + "-" +str('00'))

    elif hour==0:

        period.append(str('00') + "-" +str(hour+1))

    else:

        period.append(str(hour) + "-" +str(hour+1))


df['period']=period

df.sample(5)


import seaborn as sns

plt.figure(figsize=(20,6))
```

```
sns.heatmap(df.pivot_table(index='day_name',columns='period',values='message',aggfunc='count').fillna(0))

plt.yticks(rotation='horizontal')

plt.show()

import matplotlib.pyplot as plt


# Grouping data by 'user' and counting the number of messages

user_message_counts = df['user'].value_counts()


# Plotting a pie chart

plt.figure(figsize=(8, 8))

plt.pie(user_message_counts, labels=user_message_counts.index, autopct='%1.1f%%', startangle=140)

plt.title('Message Distribution by User')

plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle

plt.show()

import re

import pandas as pd

from collections import Counter


# Open and read the WhatsApp chat text file

with open('whatsapp_chat_with_Tej_kumar_122.txt', 'r', encoding='utf-8') as f:

    data = f.read()
```

```
# Define a regular expression pattern to match emojis

emoji_pattern = r'[\U0001F600-\U0001F64F\U0001F300-\U0001F5FF\U0001F680-
\U0001F6FF\U0001F700-\U0001F77F\U0001F780-\U0001F7FF\U0001F800-
\U0001F8FF\U0001F900-\U0001F9FF\U0001FA00-\U0001FA6F\U0001FA70-
\U0001FAFF\U00002702-\U000027B0\U000024C2-
\U0001F251\U0001F004\U0001F0CF\U0001F170-\U0001F251\U0001F300-
\U0001F5FF\U0001F600-\U0001F64F\U0001F680-\U0001F6FF\U0001F700-
\U0001F773\U0001F780-\U0001F7D8\U0001F7E0-\U0001F7EB\U0001F90D-
\U0001F93A\U0001F93C-\U0001F945\U0001F947-\U0001F971\U0001F973-
\U0001F976\U0001F97A\U0001F97C-\U0001F9A2\U0001F9B0-
\U0001F9B9\U0001F9C0-\U0001F9C2\U0001F9D0-\U0001F9FF\U0001FA70-
\U0001FA74\U0001FA78-\U0001FA7A\U0001FA80-\U0001FA86\U0001FA90-
\U0001FAA8\U0001FAB0-\U0001FAB6\U0001FAC0-\U0001FAC2\U0001FAD0-
\U0001FAD6]+'


# Extract emojis from each message using regular expressions

emojis = re.findall(emoji_pattern, data)


# Count the frequency of each emoji

emoji_freq = Counter(emojis)


# Display the most common emojis and their frequencies

print("Top 10 Most Common Emojis:")

print(pd.DataFrame(emoji_freq.most_common(10), columns=['Emoji', 'Frequency']))
```