माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA
Deemed to be University
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE

A Skill Based Mini Project Report
on

"WhatsApp Chat Analyzer Tool"

Submitted by

Pratham Bajpai (0901EO211043)

Submitted to

Dr. Gaurav Khare
Assistant Professor

Centre for Internet of Things
Madhav Institute of Technology & Science, Gwalior
Gole ka Mandir, Gwalior - 474005, M.P., India

Jan – Jun 2024

Madhav Institute of Technology & Science, Gwalior

(Deemed to be University)

NAAC Accredited with A++ Grade

(Declared under Distinct Category by Ministry of Education, Government of India)

Centre for Internet of Things

DECLARATION

I/We hereby declare that the work being presented in this skill based mini project report, for the partial fulfilment of requirement for the award of the degree of Bachelor of Technology in Internet of Things at Madhav Institute of Technology & Science, Gwalior is an authenticated and original record of my work under the mentorship of **Dr. Gaurav Khare**, Assistant Professor, Centre for Internet of Things.

I/We declare that I/We have not submitted the matter embodied in this report for the award of any degree or diploma anywhere else.

Pratham Bajpai

(0901EO211043)

Centre for Internet of Things

Madhav Institute of Technology & Science, Gwalior

(Deemed to be University)

NAAC Accredited with A++ Grade

(Declared under Distinct Category by Ministry of Education, Government of India)

Centre for Internet of Things

CERTIFICATE

This is certified that **Pratham Bajpai (0901EO211043)** has submitted the skill based mini project report titled "**WhatsApp Chat Analyzer Tool"** under the mentorship of **Dr. Gaurav Khare**, in partial fulfilment of the requirement for the award of degree of Bachelor of Technology in Internet of Things from Madhav Institute of Technology and Science, Gwalior.

Dr. Gaurav Khare
Assistant Professor
Centre for Internet of Things

Madhav Institute of Technology & Science, Gwalior

(Deemed to be University)

NAAC Accredited with A++ Grade

(Declared under Distinct Category by Ministry of Education, Government of India)

Centre for Internet of Things

ACKNOWLEDGEMENT

The full semester project has proved to be pivotal to my career. I am thankful to my institute, **Madhav Institute of Technology & Science** to allow me to continue my disciplinary/interdisciplinary project as a curriculum requirement, under the provisions of the Flexible Curriculum Scheme approved by the Academic Council of the institute. I extend my gratitude to the Director of the institute, **Dr. R. K. Pandit** and Dean Academics, **Dr. Manjaree Pandit** for this.

I would sincerely like to thank my department, **Centre for Internet of Things,** for allowing me to explore this project. I humbly thank **Dr. Praveen Bansal**, Assistant Professor and Coordinator, Centre for Internet of Things, for his continued support during the course of this engagement, which eased the process and formalities involved.

I am sincerely thankful to my faculty mentors. I am grateful to the guidance of **Dr. Gaurav Khare**, Assistant Professor, and Centre for Internet of Things, for his continued support and guidance throughout the project. I am also very thankful to the faculty and staff of the department.

Pratham Bajpai

(0901EO211043)

Centre for Internet of Things

TABLE OF CONTENTS

ABSTRACT

The WhatsApp Chat Analyzer is a web-based tool designed to empower users with insights into their WhatsApp conversations. By uploading exported chat data, users can access comprehensive analyses encompassing various parameters, including message frequency, word usage patterns, media sharing trends, and emotive expressions through emojis. The Analyzer offers intuitive visualizations such as word clouds and monthly timelines, enabling users to uncover prevalent themes, communication habits, and relationship dynamics. With the addition of a feature to identify busy users within group chats, users gain deeper insights into group dynamics and individual contributions to conversations. The findings of the analysis provide valuable insights into the evolving nature of digital communication, with implications for individuals, businesses, and researchers. Future developments could include integrating advanced natural language processing techniques, expanding platform support, and incorporating social network analysis to further enhance the Analyzer's capabilities. Overall, the WhatsApp Chat Analyzer serves as a versatile tool for understanding and optimizing digital conversations in the digital age.

The WhatsApp Chat Analyzer serves as a valuable resource for individuals seeking self-reflection and self-awareness, as well as businesses aiming to inform marketing strategies and enhance team collaboration. Researchers also benefit from the Analyzer's capabilities, using it to study digital communication dynamics, linguistic trends, and social network structures in online communities. With continuous development and innovation, the Analyzer holds great promise for unlocking new possibilities in chat analysis and communication optimization. Its user-friendly interface, robust analytical capabilities, and potential for future expansion position it as a versatile tool for navigating the complexities of digital communication in the modern age.

LIST OF FIGURES

Chapter 1:
Introduction

1.1 About WhatsApp Chat Analyzer Tool

The WhatsApp Chat Analyzer is an innovative web-based tool designed to empower users with comprehensive insights into their WhatsApp conversations. By allowing users to upload their exported chat data, the Analyzer offers a thorough examination of key parameters such as message frequency, word usage patterns, media sharing trends, and emotive expressions through emojis. Through intuitive visualizations like word clouds and monthly timelines, users can effortlessly identify prevalent themes, communication habits, and relationship dynamics within their chats. The tool's user-friendly interface ensures seamless navigation, enabling users to explore analytics, generate reports, and derive actionable insights without requiring technical expertise.

With its robust features and accessibility across various devices and browsers, the WhatsApp Chat Analyzer caters to a wide range of users, from individuals seeking personal reflection to professionals conducting in-depth analysis or researchers investigating communication trends. By offering a platform for understanding and optimizing WhatsApp conversations, the Analyzer serves as a valuable resource for enhancing communication efficiency, fostering deeper connections, and uncovering hidden insights within chat data.

1.2 Parameters Considered for Analysis

The tool considers several parameters for analysis, including:

Total messages: The overall count of messages exchanged in the conversation.

Total words: The total number of words exchanged in the conversation.

Total media files: The number of media files shared, such as images, videos, or audio files.

Total links: The count of links shared within the conversation.

Wordcloud: A visual representation of the most commonly used words in the conversation.

Most 25 Common Words Used By User: A list of the top 25 most frequently used words by the user.

Most Emojis Used: A list of the most frequently used emojis in the conversation.

Monthly Timeline: A visualization of message activity over the months.

1.3 Objective of Project

The objective of the WhatsApp Chat Analyzer project is to provide users with a comprehensive tool to analyze their WhatsApp conversations. By uploading their chat data, users can gain insights into their communication patterns, including message volume, word usage, media sharing behavior, and emotional expression through emojis. The project aims to help users understand their communication habits, identify common topics of discussion, and visualize their conversation dynamics over time. Additionally, the project seeks to offer a user-friendly interface deployed on the web for easy accessibility to users.

Fig 1.1 Main Page of Tool

Chapter 2:
Literature Survey

The field of chat analysis has garnered significant attention from researchers and practitioners alike, driven by the increasing ubiquity of messaging platforms such as WhatsApp and the wealth of data they generate. In recent years, numerous studies have explored various aspects of chat data analysis, shedding light on communication patterns, linguistic trends, and social dynamics within digital conversations.

One prominent area of research within chat analysis is the study of linguistic features and patterns. Researchers have delved into the use of language in digital conversations, examining factors such as word choice, sentence structure, and linguistic diversity. Studies have revealed intriguing insights into how language evolves within digital contexts, including the emergence of new vocabulary, the influence of cultural factors, and the adaptation of linguistic norms to the constraints of messaging platforms.

Additionally, researchers have explored the social dynamics of chat interactions, investigating phenomena such as social network formation, group cohesion, and leadership dynamics within digital communities. By analyzing communication patterns and network structures within messaging platforms, studies have elucidated the mechanisms through which social relationships are formed, maintained, and dissolved in digital environments.

Furthermore, the literature on chat analysis encompasses a diverse range of methodologies and techniques for data collection, processing, and analysis. From manual coding and content analysis to automated text mining and machine learning approaches, researchers have employed a variety of tools and methodologies to extract meaningful insights from chat data. This interdisciplinary approach has facilitated a deeper understanding of digital communication dynamics and has paved the way for the development of innovative tools and technologies for chat analysis, such as the WhatsApp Chat Analyzer.

Overall, the literature survey underscores the growing importance of chat analysis as a means of understanding human communication in digital environments. By drawing upon insights from linguistic analysis, social network theory, and computational methods, researchers have made significant strides in unraveling the complexities of digital conversations, offering valuable insights into human behavior, social dynamics, and communication practices in the digital age.

Chapter 3:
Methodology

3.1 Design Structure

The methodology employed in developing the WhatsApp Chat Analyzer revolves around a structured design framework aimed at ensuring efficiency, scalability, and user-friendliness. The design structure follows a modular approach, comprising distinct components for data ingestion, processing, analysis, and visualization. Users interact with the Analyzer through a user-friendly web interface developed using Streamlit, which provides seamless navigation and interaction. The design structure prioritizes flexibility, allowing for easy integration of new features and functionalities to accommodate evolving user needs and technological advancements.

Fig. 3.1 Extracting Chat

3.2 Libraries Used

The WhatsApp Chat Analyzer project utilizes several libraries and frameworks to implement its functionality, including:

Pandas: Used for data manipulation and analysis, including loading and processing the WhatsApp chat data.

NLTK (Natural Language Toolkit): Employed for text processing tasks such as tokenization and stopwords removal.

Matplotlib and WordCloud: Utilized for data visualization, including the creation of charts and word clouds to represent the analysis results.

Streamlit: Used for developing the user interface of the application, allowing for seamless interaction with the analysis features.

Heroku CLI: Utilized for deploying the project on the Heroku platform.

Fig. 3.2 Extracting URL Using URLExtract

Fig. 3.3 Emoji Analysis Using Emoji Library

```
[35]: import matplotlib.pyplot as plt

[36]: name = x.index
      count = x.values

[37]: plt.bar(name, count)
      plt.title("Active Person Graph")
      plt.ylabel("No. of Messages")
      plt.xlabel("Users")
      plt.show()
```

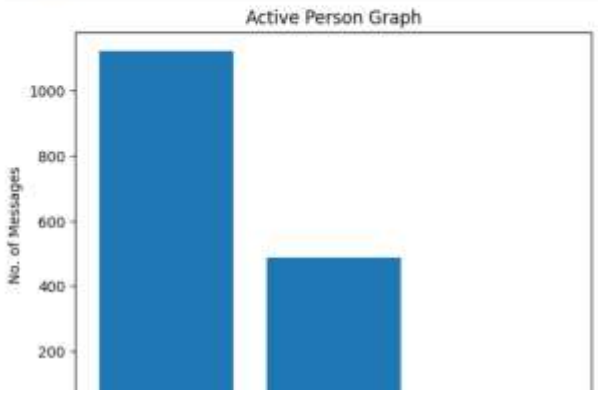

Fig. 3.4 Import Matplotlib Library

3.3 Algorithms Used

The WhatsApp Chat Analyzer employs a range of algorithms to analyze chat data and extract meaningful insights. These include algorithms for text preprocessing, sentiment analysis, word frequency analysis, and emoji detection. Text preprocessing algorithms clean and tokenize chat messages, removing noise and preparing the data for analysis. Sentiment analysis algorithms classify messages into positive, negative, or neutral categories based on the sentiment expressed. Word frequency analysis algorithms identify the most commonly used words in the conversation, while emoji detection algorithms identify and categorize emojis used by participants.

3.4 Software Used

The project relies on the following software tools and environments for its development and deployment:

Python: The primary programming language used for implementing the project's functionality.

Streamlit: Used for developing the user interface of the application.

Heroku: Employed as the platform for deploying the project and making it accessible to users via the web.

Text Editor/IDE: Such as Visual Studio Code or PyCharm, used for writing and editing the project code.

WhatsApp: The source of the chat data exported by users for analysis.

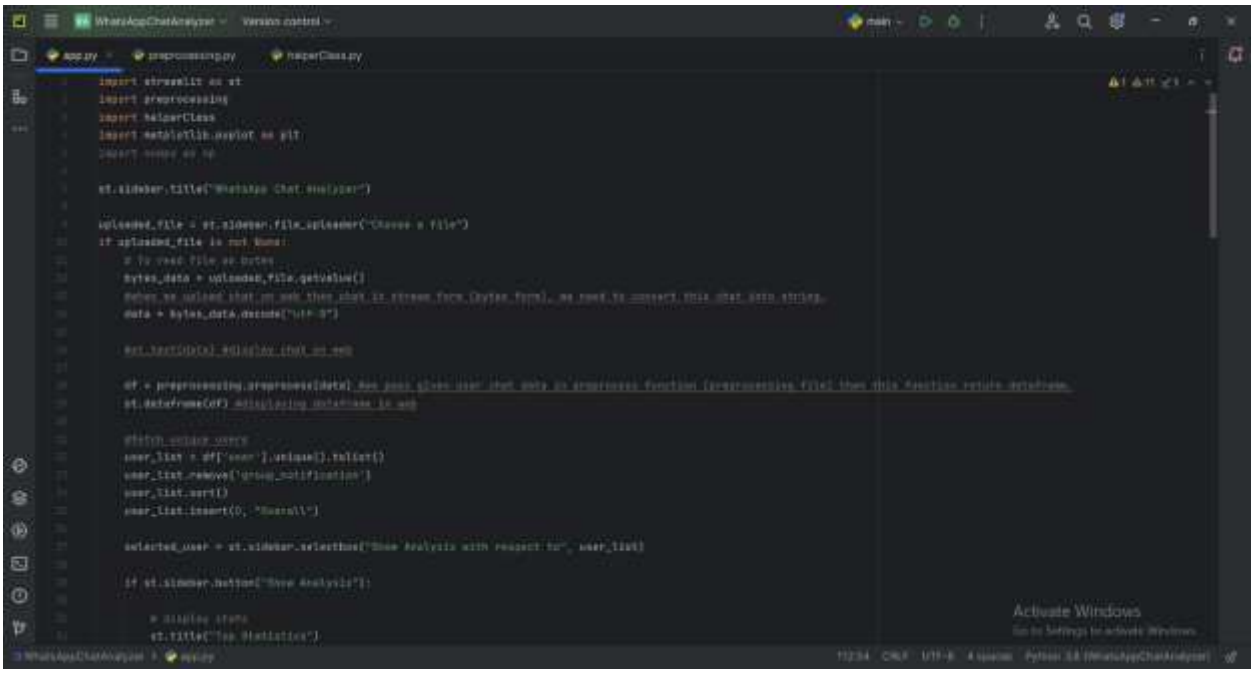

Fig. 3.4 Pycharm With Jupyter Notebook Extension

Chapter 4
Result & Discussions

The WhatsApp Chat Analyzer yields insightful results upon analyzing user-uploaded chat data, providing users with valuable insights into their communication patterns, content preferences, and social dynamics. This chapter presents the findings of the analysis and discusses their implications in the context of digital communication.

Upon uploading their chat data, users are presented with a comprehensive analysis of various parameters, including total messages, total words, total media files, total links, word cloud, most common words used by the user, most emojis used, and monthly timeline. The analysis reveals intriguing insights into the user's communication habits, highlighting key trends and patterns within their conversations. For instance, users can discover their most frequently used words, the distribution of emojis used, and the frequency of message exchanges over time.

Furthermore, with the addition of the new feature to identify busy users within group chats, users gain deeper insights into group dynamics and individual contributions to conversations. By highlighting the most active participants, the Analyzer facilitates a better understanding of group communication dynamics, leadership roles, and social interactions within the group.

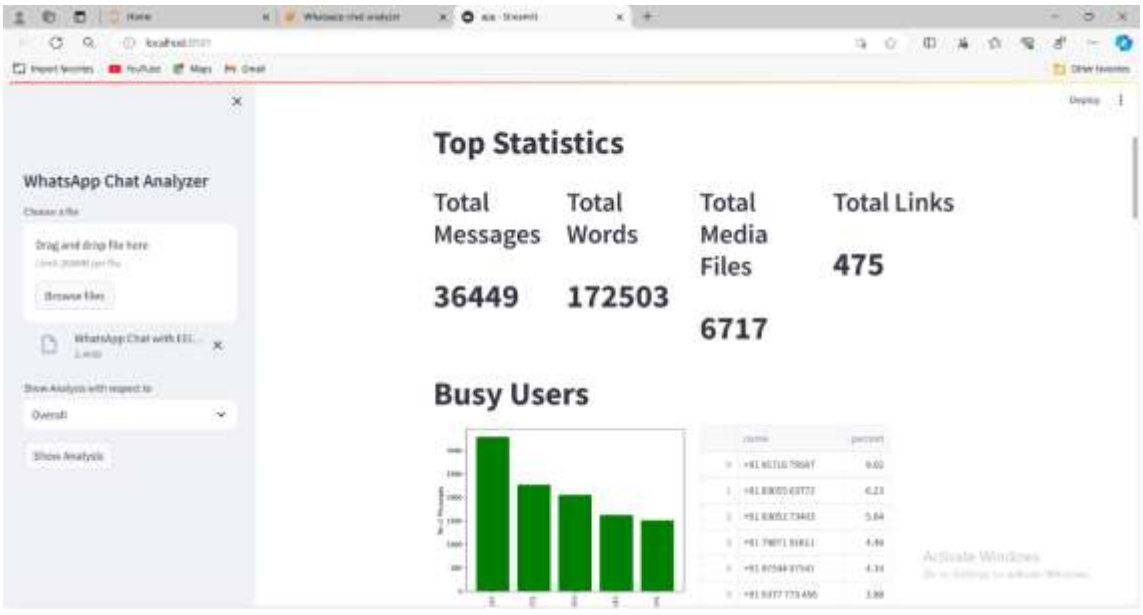

Fig. 4.1 Chat Analysis 1

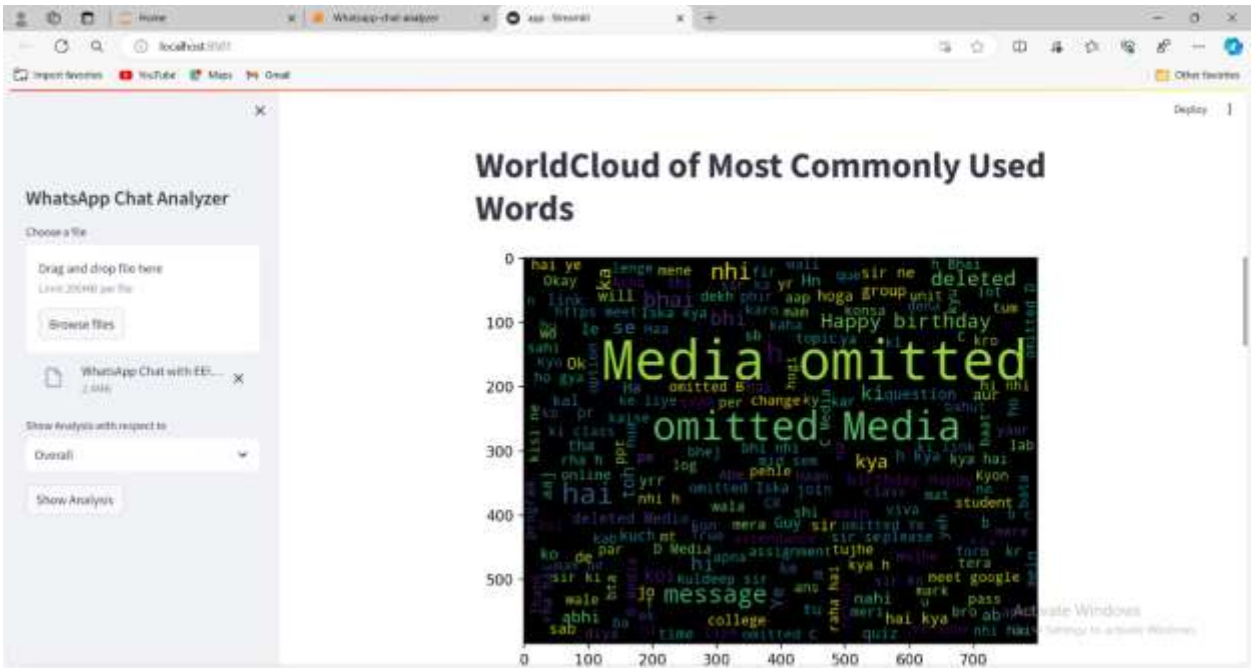

Fig. 4.2 Chat Analysis 2

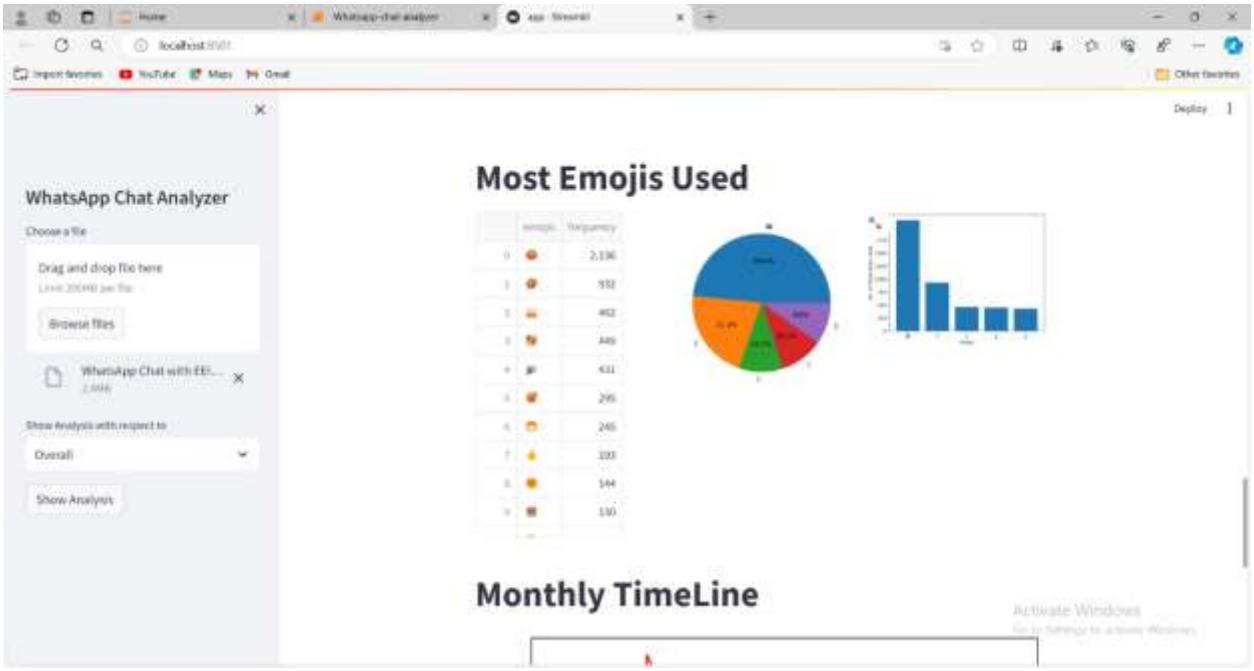

Fig. 4.3 Chat Analysis 3

Chapter 5:
Conclusion

The WhatsApp Chat Analyzer represents a significant advancement in the field of chat analysis, providing users with a powerful tool to gain insights into their digital conversations. Through a comprehensive analysis of chat data, users can uncover valuable insights into their communication patterns, content preferences, and social dynamics. The Analyzer offers a user-friendly interface, robust analytical capabilities, and intuitive visualizations, making it a valuable resource for individuals, businesses, and researchers alike.

5.1 Future Scope

Sentiment Analysis: Integrating sentiment analysis algorithms to determine the overall sentiment of conversations and individual messages, providing users with insights into the emotional tone of their communication.

Topic Modeling: Implementing topic modeling techniques to automatically identify and categorize the main topics of conversation within chat data, enabling users to explore and analyze the content more effectively.

User Engagement Metrics: Introducing additional metrics to measure user engagement and participation levels, such as response times, message frequency, and interaction patterns, to provide users with a deeper understanding of their communication habits.

Enhanced User Interface: Improving the user interface of the WhatsApp Chat Analyzer to enhance usability, interactivity, and visualization options, providing users with a more intuitive and engaging experience.

Integration with Messaging Platforms: Expanding the scope of the project to support chat data from other messaging platforms such as Facebook Messenger, Telegram, or Slack, allowing users to analyze their communication across multiple platforms.

Overall, the WhatsApp Chat Analyzer project has laid the groundwork for further exploration and development in the field of chat analysis and communication insights. By continuing to innovate and expand upon its capabilities, the project has the potential to become a valuable tool for users seeking to gain deeper insights into their digital communication habits and behaviors.

References

1. McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (pp. 51-56).
2. Bird, S., Loper, E., & Klein, E. (2009). Natural Language Processing with Python. O'Reilly Media Inc.
3. Streamlit Documentation. (n.d.). Retrieved from https://docs.streamlit.io/
4. Heroku Documentation. (n.d.). Retrieved from https://devcenter.heroku.com/
5. Matplotlib Documentation. (n.d.). Retrieved from https://matplotlib.org/stable/contents.html
6. NLTK Documentation. (n.d.). Retrieved from https://www.nltk.org/
7. WordCloud Documentation. (n.d.). Retrieved from https://amueller.github.io/word_cloud/
8. Emoji Data. (n.d.). Retrieved from https://unicode.org/emoji/charts/full-emoji-list.html
9. sklearn Documentation. (n.d.). Retrieved from https://scikit-learn.org/stable/documentation.html
10. WhatsApp. (n.d.). Retrieved from https://www.whatsapp.com/