WHATSAPP CHATS ANALYZER USING MACHINE LEARNING

Vedika Jadhav	514
Sarthak Kochrekar	517
Azeem Ansari	501

Introduction

Welcome to WhatsApp-Analyzer: Understanding Your Chats with Data In this presentation, we will explore how machine learning can help us extract insights from our WhatsApp chats. So, let's get started!

LITERATURE REVIEW

Ravishankara K, Dhanush in 2016 e xp lain that the purpose of this study is to employ the WhatsApp Chat Analyzer tool, developed was to gain insights into communication dynamics within WhatsApp groups. The study utilized the WhatsApp Chat Analyzer, a Python-based tool, to process and analyze WhatsApp chat data. The tool leverages Python modules like pandas, matplotlib, seaborn, and sentiment analysis to create data frames, perform exploratory data analysis, and generate visual representations of the data. Through the application of the WhatsApp Chat Analyzer, several key findings emerged. The most active date was identified as June 15, 2018, with 190 messages sent on that day. The analysis also revealed the most frequently used word, "the," appearing 43,313 times.

Additionally, the tool highlighted the top 10 and top 20 active users in the chat group, providing percentages of their contributions. The analysis also revealed the most frequently used word, "the," appearing 43,313 times. The results signify a robust utilization of the WhatsApp platform within the analyzed group. The high message count on June 15, 2018, suggests a significant event or discussion on that day. The identification of top contributors sheds light on group dynamics, indicating who plays pivotal roles in the communication flow. This study demonstrates the efficiency of the WhatsApp Chat Analyzer in providing valuable insights into WhatsApp group communication.

WHAT IS MACHINE LEARNING?

Machine learning is a subset of artificial intelligence (AI) that allows computers to learn and improve from experience without being explicitly programmed. It involves the use of algorithms and statistical models to enable computers to perform specific tasks without being explicitly programmed.

WHY USE MACHINE LEARNING FOR WHATSAPP CHATS ANALYSIS?

- Data Richness
- Efficiency
- Insight Discovery
- Accuracy
- Automation
- Scalability

Python Libraries Used

- **Streamlit**: Streamlit is a free and open-source python framework. [2] We can quickly develop web apps for Machine Learning and Data Science by using Streamlit.

- **Matplotlib**: Matplotlib is a popular Python packages used for data visualization. It is a cross-platform library for making plots from data in arrays. It helps in creating static, animated and interactive visualizations in python.

- **Seaborn**: Seaborn is the data visualization library. It is used for making statistical graphs. Visualization is the central part of seaborn. Seaborn provides exploration and better understanding of data. Seaborn closely integrates into the data structures from python.

- **Pandas**: Pandas is an open-source python library. Pandas used to convert string data into Data frame. Data frame is the representation of data into 2-dimensional table of rows and columns. We can work with large data sets using Pandas library. Pandas library has many built-in functions for data analysis, data cleaning, data exploration and data manipulation

Python Libraries Used

URLExtract : python class for collecting (extracting) URLs from given text based on locating TLD.

Word Cloud : A data visualization technique used for representing text data in which the size of each word indicates its frequency or importance.

Collections : containers used for storing data and are commonly known as data structures, such as lists, tuples, arrays, dictionaries, etc.

SVM algorithm

Support vector machines (SVM) is a supervised machine learning algorithm used for classification and regression tasks. SVM finds a hyperplane in an N-dimensional space that best separates the data points into different classes. In classification, SVM aims to find the optimal decision boundary that maximizes the margin (distance) between the classes. The data points closest to the decision boundary are called support vectors. SVM can handle both linearly separable and non-linearly separable data through the use of kernels.

DATA COLLECTION

Input: WhatsApp chat data files exported from the application.

Process: Users export their chat data and provide the exported files to the system.

Output: A collection of chat data files for analysis.

File Formats: Ensure that the chat data is saved in a format that can be processed, such as plain text.

Backup and Save: Back up the chat files to prevent data loss or overwriting.

PREPARING YOUR WHATSAPP CHATS FOR ANALYSIS

Cleaning

Formatting

Analysis Tokenization

Lowercasing

Stop Word Removal

Stemming or Lemmatization

Special Characters

TEXT PREPROCESSING FOR WHATSAPP CHAT ANALYZER

Input: Preprocessed chat data.

Stop Word Removal: Eliminate common words, special characters, and numbers.

Output: Cleaned and tokenized text data ready for analysis.

Tokenization: Split messages into individual words or tokens.

Stemming or Lemmatization: Reduce words to their root form.

Lowercasing: Convert text to lowercase for case-insensitivity.

Data Transformation: Create a bag of words or TF-IDF representation for messages.

VISUALIZATION FOR WHATSAPP CHAT ANALYZER:

- Input: Results from the analysis.

- Create Visualizations: Including charts, graphs and various visual aids.

- Generate Word Clouds: To represent word frequency.

- Produce Sentiment Plots: Depicting emotional tone.

- Build Network Graphs: To illustrate interactions.

- Output: Visual representations of the analysis results.

VISUALIZING THE RESULTS

- Clear Communication
- Concise Presentation
- Enhanced Understanding
- Impactful Insights
- Effective Communication

CASE STUDY: ANALYZING A GROUP CHAT

- Objective: Understand the capabilities of WhatsApp chat analysis and Python programming for data insights.
- Tools Used: Python, NumPy, Pandas, Matplotlib, Seaborn.
- Data Collection: Export WhatsApp chat as a text file.
- Data Preprocessing: Extract sender, message content, and timestamps; remove metadata.
- Text Preprocessing: Tokenize text, convert to lowercase, remove stop words, special characters, and numbers.
- Analysis Engine: Conducted basic statistics, word frequency, sentiment, emoticon, and emoji analysis, conversation trends, network analysis.
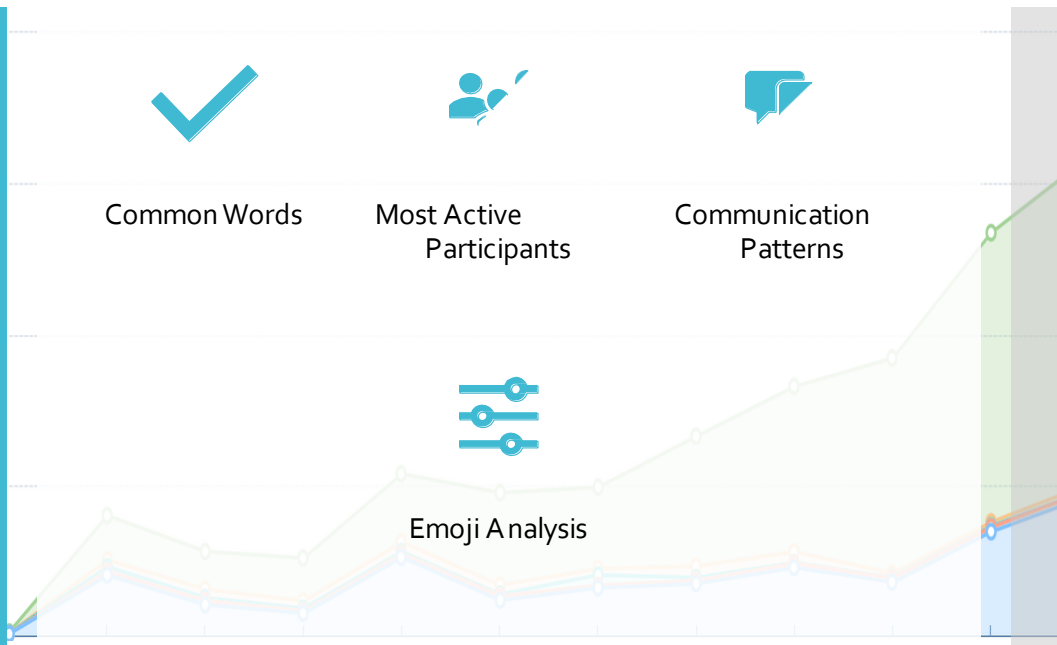
RESULTS OF THE CASE STUDY
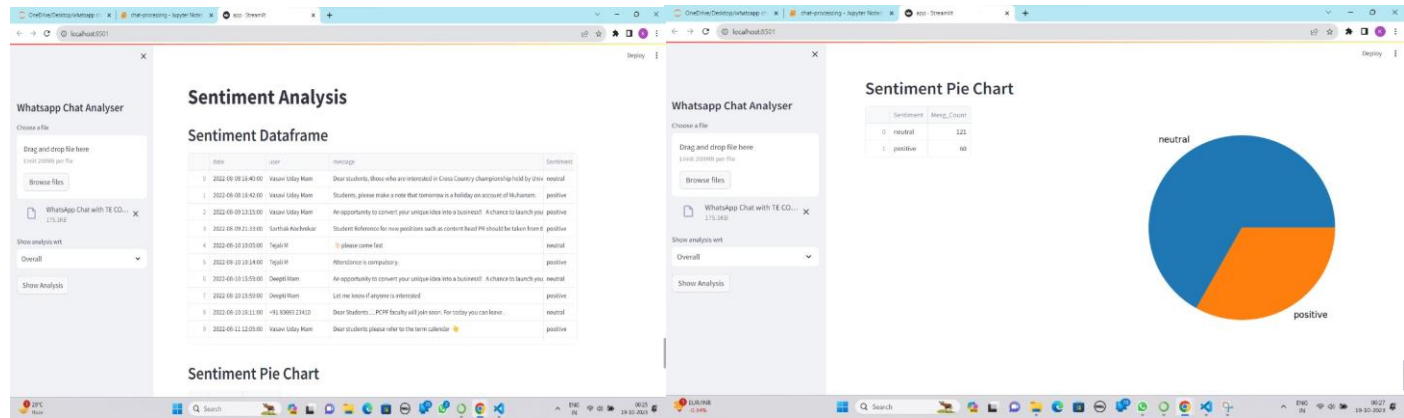
Common Words

Most Active Participants

Communication Patterns

Emoji Analysis

FUTURE SCOPE

The future scope for WhatsApp chat analysis is expansive and evolving, driven by advances in natural language processing (NLP), machine learning, and user demand for deeper insights into their communication patterns. Balancing innovation with privacy and security is crucial to ensure ethical and respectful chat analysis.

CONCLUSION

In conclusion, machine learning can help us unleash the secrets of our WhatsApp chats by enabling us to analyze large amounts of data quickly and accurately. By applying machine learning to our WhatsApp chats, we can gain insights into our conversations that might otherwise go unnoticed.

REFERENCES

1   R. E. Ziemer and W. H. Tranter, Principles of Communications, 7th ed. Hoboken, NJ: Wiley, 2015.
    [Online].
    Available: https://ebookcentral.proquest.com/lib/vu/reader.cation?docID=5106516&ppg=1

2   J. D. Bellamy et al., Computer Telephony Integration. New York: Wiley, 2010.

3   C. Jacks, High Rupturing Capacity (HRC) Fuses. New York: Penguin Random House, 2013, pp.
    175–225.

4   N. B. Vargafik, J. A. Wiebelt, and J. F. Malloy, "Radiative transfer,"

in Convective Heat. Melbourne: Engineering Education Australia, 2011, ch. 9, pp. 379–398.

[5]   H. C. Hottel and R. Siegel, "Film condensation," in Handbook of Heat

Transfer, 2nd ed. W. C. McAdams, Ed. New York: McGraw-Hill, 2011, ch. 9, pp. 78–99.

6   W. M. Rohsenow, "Heat transmission," in Thermal Radiation Properties, vol.

3, M. W. Catton and J. P. Hartnett, Eds. New York: Macmillan, 2012, ch. 9, pp. 37–62.

7   H. Schmidt-Walter and R. Kories, Electrical Engineering. A Pocket

Reference. Boston: Artech House, 2007. Accessed: Oct. 16, 2016. [Online]. Available: http://ebrary.com