International Journal on Science and Technology (IJSAT)



E-ISSN: 2229-7677 • Website: <u>www.ijsat.org</u> • Email: editor@ijsat.org

WhatsApp Chat Analysis

R. Harshavardhan¹, Soorej S², Manoj Kumar³

^{1,2,3}B. TECH CSE 3rd Year ^{1,2,3}srm Institute of Science And Technology, Ramapuram.

Abstract:

To identify complex patterns of behavior that users display in digital communication, this article provides a thorough analysis of WhatsApp chat data. Understanding the mechanics of human interaction within these digital environments is vital for several areas, including computer science, psychology, and sociology, given the widespread adoption of messaging apps like WhatsApp. Using a sizable dataset gathered from several user groups, this study analyzes communication patterns using both quantitative and qualitative methods. The study's findings provide insightful information on the nuances of WhatsApp users' digital communication habits. Comprehending these patterns can help with the development of more efficient messaging systems, improve user experience, and offer insightful information for applications like recommendation engines, content filtering, and targeted advertising.

Keywords:

WhatsApp, Chat Analysis, WhatsApp chat files, visualization, Sentiment Analysis, Emoji Analysis, NLP, etc.

1. INTRODUCTION

The goal of the WhatsApp Chat Analyzer project is to examine WhatsApp chat data and offer perceptions into user behavior and trends. The project takes information out of the chat files and puts it in an easily understood format. It offers several features that aid in comprehending chat patterns, emotions, and trends, such as word frequency analysis, chat sentiment analysis, and user activity analysis. The knowledge of this decade will be essential to the technology of the future. The only way to find this information is to conduct some research on the intended device. Since many people are interested in using machine learning to create models that assist in solving a range of problems, appropriate data is extremely necessary. These models need a thoughtfully created learning model that makes these models.

Our project offers deep data search across various WhatsApp chat types. Numerous fields, including social studies, psychology, and market research, can benefit from this project. The project can offer insightful information about political opinions, consumer behavior, and mental health. The purpose of the research paper is to examine the different facets of this project and its possible uses.



2. LITERATURE REVIEW

The field of sentiment analysis in WhatsApp chats is dynamic and ever-evolving, reflecting the growing significance of comprehending human emotions in the digital age. Researchers are always coming up with new ideas to. create sentiment analysis models that are more resilient and culturally aware in the particular environment of Talks on WhatsApp. This study emphasizes the research's multidisciplinary nature and emphasizes its potential for practical applications in a variety of fields, including marketing and the social sciences. Research on WhatsApp chat sentiment analysis has developed, with many approaches and concepts put out to address the particular difficulties of casual online discussions.

2.1 Machine Learning:

Machine learning (ML) is a subset of artificial intelligence (AI) that enables systems to learn from data and improve their performance over time, making predictions or decisions without being explicitly programmed. It can be categorized into three main types: supervised learning, unsupervised learning, and reinforcement learning. In supervised learning, models are trained on labeled data, using algorithms like linear regression and decision trees to tackle problems such as spam detection and image classification. Unsupervised learning, on the other hand, deals with unlabeled data, employing techniques like clustering and association rules to discover hidden patterns, often used in customer segmentation and anomaly detection. Reinforcement learning involves models learning by interacting with an environment, receiving feedback through rewards or penalties, with applications in areas such as game playing and robotic control.

Machine learning finds applications across various industries, including healthcare for predictive analytics and personalized medicine, finance for fraud detection and algorithmic trading, marketing for customer segmentation, and autonomous vehicles for object detection and decision-making systems.

However, challenges remain, such as overfitting, where a model learns noise from training data leading to poor generalization; data bias, which arises when training data is not representative; and the interpretability of complex models like deep learning.



2.2 Data Analytics:

Data analytics is the process of examining datasets to draw conclusions about the information they contain, using various techniques and tools to uncover patterns, correlations, and insights. It plays a crucial role in decision-making across industries by transforming raw data into actionable knowledge.



The field can be divided into four main types: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. Descriptive analytics focuses on summarizing historical data to provide insights into what has happened, using tools like dashboards and reports. Diagnostic analytics delves deeper to understand why certain events occurred, often employing statistical methods and data mining techniques to identify relationships and causations.

The data analytics process typically involves several stages, including data collection, data cleaning, data analysis, and data visualization. Data collection encompasses gathering relevant data from various sources, while data cleaning involves ensuring the accuracy and quality of the data. Once the data is prepared, analysts use statistical and computational methods to analyze it, often employing data visualization tools to create clear and informative representations of the findings.



Lexicon-Based Sentiment Analysis:

Numerous scholars have examined the sentiments in WhatsApp chats using sentiment lexicons, dictionaries, and rules. Using this technique, words or phrases in the chat are given sentiment scores, and an overall sentiment score for every communication or exchange. Critical Analysis: Lexicon- based analysis is simple to use and apply, but it may miss subtleties in complex language and context. As a result, it is less reliable for sentiment analysis in casual talks.

2.3 Methods for Machine Learning:

Sentiment classifiers have been developed using deep learning methods like Recurrent Neural Networks (RNNs) and machine learning models like Support Vector Machines (SVM) and Naive Bayes. These algorithms are able to categorize WhatsApp chat attitudes by learning from labelled data. Evaluate Critically: Machine learning techniques can lead to more precision and adjustment to the particular language and environment of WhatsApp conversations. They do, however, need a significant quantity of labelled data for training, and the caliber of the training data can affect how well the model performs.

2.4 Emotion and Emoji analysis:

Researchers have looked into how emoticons and emojis are used in WhatsApp conversations to convey emotion. To determine the emotional content, they examine the type and quantity of emojis Evaluate



Critically: Emojis and emoticons are useful indicators for sentiment analysis, particularly in casual talks. They can, however, be interpreted differently depending on the context, therefore consistent analysis requires a uniform lexicon of emoji sentiments.

3. PROBLEM STATEMENT

WhatsApp-Analyzer is a statistical analysis tool for WhatsApp chats. Working on the chat files that can be exported from WhatsApp it generates various plots showing, for example, which another participant a user responds to the most. We propose to employ dataset manipulation techniques to have a better understanding of WhatsApp chat present in our phones.

3.1 Existing System

There is a lot of development in the current system. In the older version there was no feature to display status, there was no feature to share documents and there was no feature to share location. In the current version, all of these features are available. In older version we couldn't share images through doc's format. In this system user is able to access WhatsApp in windows through WhatsApp web application, which can be connected through QR code. There is another feature called export chat where user can send or share or get the chat detail for data analysis through email, Facebook or some messenger application.

3.2 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 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 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 di not have authors, plot graph of hour vs message count.

4. OBJECTIVE

In this decade the upcoming technologies are mainly dependent on data. This data can only be obtained if there is some research applied on the context of the requirements of the tool. Since a lot of machine learning enthusiasts develop models which helps solve multiple problems the requirements of appropriate data are very large scale this project aims to provide a better understanding towards various types of chats. This analysis proves to be better input to machine learning models which essentially explore the chat data. These models require proper learning instances which provides better accuracy for these models. Our project ensures to provide an indepth exploratory data analysis on various types of WhatsApp chats

International Journal on Science and Technology (IJSAT)



5. SOFTWARE REQUIREMENT ANALYSIS

If software requirement analysis in the field of systems engineering and software engineering, encompasses those tasks that are used for a new or altered product or tool, taking account of the possibly conflicting requirements of the various stakeholders, documenting, validating and managing software or system requirement

5.1 Feasibility Study

The main objective of the feasibility study is to treat the technical operational and economic feasibility of developing the application. Feasibility is the determination of whether or not project is worth doing. The process followed in making this determination is called feasibility study. All systems are feasible, given unlimited resources and infinite time. The feasibility study to be conducted for this project involves:

- Technical Feasibility
- Operational Feasibility
- Economic Feasibility

6. METHODOLOGY

This section provides a brief overview of the technique used in our study, which examines the opinions expressed in WhatsApp conversation profiles. Python is the programming language used for the web-based interface used to do the analysis. To process the data and racetrack, this project makes use of a number of libraries, tools, and technologies, including as Matplotlib, Seaborn, NumPy, Text Blob, NLTK, Regular Expressions (re), and CSV.

6.1 Data Collection:

We began the crucial work of gathering data in order to start our research. We got in touch with the participants to get permission to view and examine the data from their WhatsApp conversations. The conversation export function of WhatsApp makes it simple to store chat history in CSV format. These data, which are the foundation of our analysis, include text messages, time logs, and user IDs.

6.2 Data Preprocessing:

Pre-processing data is crucial to ensuring that it is suitable and useful for statistical analysis. The data process needs to be cleaned up and modelled extensively during this step. The initial crucial actions consist of:

Subsequent cleaning: We take great effort to eliminate emoticons, special characters, and special characters that may produce noise on the fabric paper.

Tokenization: To make additional analysis easier, words are designated as single words. After removal: Prepositions, letters, etc. Take out any leftover words like.

Lowercase: To guarantee uniformity in text analysis, every text is transformed to lowercase. < br> **Date-time conversion**: To offer a standardized format for checking the time, we produced a date file.

6.3 NLTK (Natural Language Toolkit):

Our feature extraction process relies heavily on NLTK. Lemmatization and stemming are two common



tasks in natural language processing that are made easier by these techniques. We can further improve the reading material through this procedure so that it is appropriate for analysis.

7. MODELING AND ANALYSIS



Figure 5: Statistics and Graph

Figure 6: Timeline and bar Graph



8. RESULTS AND DISCUSSION



Figure 7: Back end Data Pre processing

Figure 8: Back end Data Pre processing

7.1 Data Collection:

Discover: Users can export their WhatsApp chat data without any media thanks to the smooth data collecting method. Chat histories are accessible to users, offering rich data for analysis.

7.2 Data Preprocessing:

Find out: The data session is cleaned and standardized during the data preprocessing stage, making it ready for analysis. To maintain consistency, this entails eliminating superfluous symbols and emoticons and changing text to lowercase.

7.3 Evaluation of Sentiment:

Examine: Robust libraries like NLTK and Text Blob enable sentiment analysis to produce precise outcomes. To enable dependable development, these function libraries are integrated with the Python language.

7.4 Web-based user interface:

Learn: The web-based conversation analytics interface for WhatsApp is easy to use and understand. The service is user-friendly, since users can effortlessly upload topics and choose when to leave comments.

7.5 Illustration:

Exploration: Interactive data is effectively exposed by Matplotlib, Seaborn, and other visualization tools. These displays feature various metrics, cloud messages, game maps, timelines for each month, daily timelines and analytics for emojis.



7.6 **Precision and Dependability:**

Discovering: The trustworthiness of the analysis results was ensured by the sentiment analysis model's high degree of accuracy and dependability in capturing chat sentiment.

Useful Applications:

Discovery: The sentiment analysis of WhatsApp conversations has useful applications in a number of fields, such as community management, marketing, and social research. Decision-making in various domains can benefit from an understanding of sentiment trends.

7.7 **Prospective Routes:**

Finding: Although the research has proven successful, much more work needs to be done. There is room for progress in processing various chat formats, applying sophisticated machine learning algorithms, and improving the accuracy of sentiment analysis.

7.8 Assessment of Measurements:

Measurement assessment Model: Assess the efficacy of the model assessment process by employing measures like F1 score Nia, accuracy, precision, and recall.

User Interface: Use surveys or written comments to analyse user happiness and ease of use in order to gauge how effective the web interface is.

9. CONCLUSION

Final Thoughts on the Topic of the WhatsApp Discussion In addition to NumPy, Seaborn, Matplotlib, and Python, etc. The debate takes place online and makes use of NLTK, Re, CSV, and other libraries, tools, and technologies. Tasks including data processing and time. There are multiple processes involved in implementing the WhatsApp chat analysis project: gathering data, setting priorities, analysing emotions, and visualizing the results. Launch the group chat on WhatsApp, select Messages, then More. Click Export Chat and then choose "No Messages." Chat analysis on WhatsApp is operational. First, choose the date mode on the WhatsApp Chat Analyzer website.

Upload the conversation file that was exported here. To analyse data, one uses the training model. Now, the training model handles the pre-processing of the data. Select each or every identifier at this point. displays analytics, such as rain pattern, statistics, word cloud, activity chart, monthly timeline, daily timeline.

REFERENCES

- 1. 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).
- 2. 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).
- 3. Access Data Corporation. FTK Imager, 2013. Available at http://www.accessdata.com/support/productdownloads.
- 4. WhatsApp Group Data Analysis with R. Available at https://www.ijcaonline.org/archives/volume154/number4/patil-2016-ijca-912116.pdf



- 5. Content Analysis of WhatsApp chats. Available at https://www.arcjournals.org/pdfs/ijmjmc/v4-i1/2.pdf
- 6. Available from: http://www.statista.com/statistics/260819/numberof-monthly-active-WhatsAppusers. Number of monthly active WhatsApp users worldwide from April 2013 to February 2016(in millions).
- D. Bouhnik and M. Deshen, "WhatsApp Goes to School: Mobile Instant Messaging Between Teachers and Students," Journal of Information Technology Education: Research, vol. 13, pp. 217-231, 25 August 2014.
- 8. Shaikh Mohd Saqib. Whatsapp Chat Analyzer. International Research Journal of Modernization in Engineering Technology and Science. Volume: 04/Issue:05/May-2022
- 9. K, Ravishankara & Dhanush, & Vaisakh, & S, Srajan. (2020). Whatsapp Chat Analyzer. International Journal of Engineering Research and. V9.10.17577/IJERTV9IS050676
- 10. Available from: http://www. statista.com/statistics/260819/numberof-monthly-active- WhatsAppusers. Number of monthly active WhatsApp users worldwide from April 2013 to February 2016(in millions).