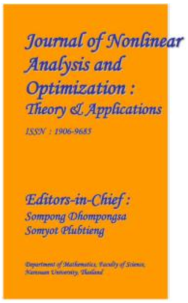
Journal of **N**onlinear **A**nalysis and **O**ptimization

Vol. 14, Issue. 2 : 2023

ISSN : **1906-9685**



Journal of Nonlinear
Analysis and
Optimization :
Theory & Applications

ISSN : 1906-9685

Editors-in-Chief :
Sompong Dhompongsa
Somyot Plubtieng

Department of Mathematics, Faculty of Science,
Naresuan University, Thailand

WHATSAPP CHAT ANALYSIS USING MACHINE LEARNING TECHNIQUES

[1]Hanumanthappa Mahesh, [2]Ch.Sreya, [3]D.Cherishma ,[4]G.Pooja, 5O.Kujitha
[1]Assistant Professor, [2,3,4,5]UG Students, [1,2,3,4,5]Dept. of CSE (AI & ML), Malla Reddy Engineering College for Women (Autonomous), Hyderabad, India. E-Mail: maheshh118@gmail.com

ABSTRACT

A programme called WhatsApp has emerged as the most popular and effective means of communication in recent years. WhatsApp chats are conversations between two people or a group of people that can take many different forms. There are several subjects discussed in this talk. This data can offer a wealth of information for cutting-edge technologies like machine learning. The correct learning experience is crucial for machine learning models, and this experience is indirectly influenced by the data we give the model. This tool seeks to offer thorough analysis of the information supplied by WhatsApp. No matter what the conversation's subject is, our generated code can be used to improve understanding of the data. The benefit of this tool is that it can easily be applied to larger datasets because it is implemented using straightforward Python modules like pandas, matplotlib, seaborn, streamlit, numpy, and a sentiment analysis technique. These modules are used to create data frames and plot various graphs, and the output is then displayed in the streamlit web application using efficient and resource-conserving algorithms.

INTRODUCTION

This technology is focused on the processing and analysis of data. Understanding the appropriate learning experience from which the model starts to improve is the first step in putting a machine learning algorithm into practise. An important aspect of machine learning is data preparation. We need a lot of data to improve the model's efficiency, therefore WhatsApp, one of Facebook's major data providers, is where we concentrated much of our attention. According to WhatsApp, almost 55 billion messages are sent daily.

The typical user logs on to WhatsApp 195 times per week and participates in many groups. We must set out on a mission to obtain understanding of the messages that our phones are compelled to record in light of the data treasure trove that is there in front of us. a list that displays the intriguing data it compiles after examining your WhatsApp messages using pie charts and infographics. The process is now familiar to you. You will save a copy of your chat and send it to one of the website's provided email addresses.

A statistical analysis tool for WhatsApp talks is called WhatsApp-Analyzer. It generates numerous graphs based on the chat files that can be exported from WhatsApp, such as which participant a user responds to the most. Internet-based communication between people becomes a regular aspect of their lives. In the past, people used the internet chat system to send messages to one another. To better understand the WhatsApp chat available on our phones, we suggest using dataset modification techniques. It displays the words and emojis that are used the most frequently. It keeps tabs on our exchange and gauges how much time we are putting in.

http://doi.org/10.36893/JNAO.2023.V14I2.0225-0232

LITERATURE SURVEY

A survey analysis of WhatsApp Messenger's usage and effects was undertaken, and a number of studies and analyses were discovered. These studies cover WhatsApp's effects on students and young people. According to the survey, people between the ages of 18 and 23 in the southern region of India use WhatsApp for about 8 hours per day and may occasionally spend close to 12 to 16 hours online. Most of them concurred that WhatsApp was the most frequently used website. They share videos, audio, and photos. This poll also demonstrated that WhatsApp is the app that is used on smart phones the most, compared to all other apps. The purpose of this poll was to determine the advantages and disadvantages of using WhatsApp.

Because WhatsApp is the most popular app among young people and other generations, according to this survey, our project can give them insights into their discussions and reveal unknown information to them.

His work was focused on estimating the amount of addiction of an individual to the WhatsApp group with respect to the age group and gender using a survey of WhatsApp group analysis. R statistics software was used by him. The research gave us a foundational understanding of statistical analysis. The research was conducted on a specific WhatsApp group's data in order to ascertain the most popular form of communication in these groups, as well as the most active day of the week and the most active age group. Whether male users of the WhatsApp group are more addicted to it than female users is another question. Additionally, they included a research in their work that looked at how high school students and instructional staff used WhatsApp to communicate in the classroom.

EXISTING SYSTEM

There isn't a dashboard for visualisation that would display various metrics taken from the exported chat file and data frames to study chat insights. Status display, document sharing, and location sharing were absent from earlier iterations. The current version has all of these features. Images in document format could not be shared in earlier versions. Through the QR code, the solution enables users to remotely access their WhatsApp on any web application. The raw formatted text file can be exported from the WhatsApp application. It is really difficult to analyse. Therefore, we must abandon that system and use WhatsApp instead.

PROPOSED SYSTEM

Quite a bit of work has gone into the current WhatsApp programme. Status display, document sharing, and location sharing were absent from earlier iterations. The current version has all of these features. Images in document format could not be shared in earlier versions. Through the QR code, the solution enables users to remotely access their WhatsApp on any web application.

We have developed a visualisation dashboard for the Whatsapp Chat Sentiment Analyzer that will display several parameters taken from the exported chat file. Prior to processing, the exported chat is first cleaned and formatted using numpy. Additionally, a data frame is created using the pandas package, which is then utilised to analyse the data and produce insightful findings. The sentiments of the group discussion or a particular person are then analysed using NLTK, notably the Vader library, and the data is then visualised using statistical methods.

System Architecture

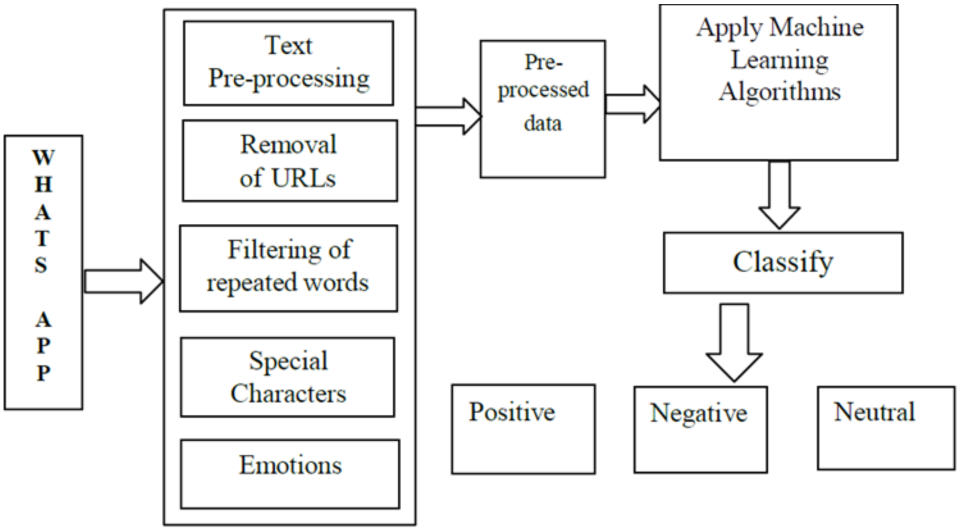


Fig.1. Block Diagram of Whats App chat analysis

WhatsApp is a popular messaging application that operates on multiple platforms, including mobile devices and web browsers. Its system architecture typically involves various components and technologies working together to provide seamless messaging services.

WhatsApp employs a range of network protocols to enable secure communication and data transfer, including HTTPS (HTTP Secure) for client-server communication, XMPP (Extensible Messaging and Presence Protocol) for message routing, and other encryption protocols to protect user data.

It's important to note that the actual architecture and infrastructure of WhatsApp may be more complex and involve additional components to ensure scalability, security, and high availability. The description provided here offers a general understanding of the system architecture of WhatsApp based on available information.

Define Problem

The problem at hand is to develop a machine learning solution for analyzing WhatsApp chats. WhatsApp is a popular messaging platform used by millions of people worldwide, generating vast amounts of textual data. The objective is to leverage ML technology to extract valuable insights, patterns, and trends from these chats, enabling users to gain a deeper understanding of the conversations and derive meaningful information

Python

It is a programming language with several uses. It offers various library types that give projects various functionalities. In order to make predictions and identify patterns, Python is employed. Python has a large number of libraries that offer mathematical and statistical functions that aid in deriving insights from data.

Pandas

This Python library is open-source and primarily used in fields related to data science and machine learning. This library offers analysis tools for altering data, and these tools are used to analyse time series analysis and numerical data utilising its data structures.

Numpy

The term Numpy may have originated from Numeric Python; it is a data analysis library for Python that has a variety of numerical functions and techniques for numerical analysis as well as multi-dimensional array objects and routines to process them.

Matplotlib

http://doi.org/10.36893/JNAO.2023.V14I2.0225-0232

Python's Matplotlib is a fantastic visualising package that is simple to use. It uses the larger SciPy stack and is based on NumPy arrays. It has a variety of plots, including pie, line, bar, graph, scatter, histogram, etc. Matplotlib is used in this project for a number of visualisations.

Seaborn

Python's Seaborn package is mostly used for statistical graphing. It offers lovely colour palettes and default styles to make statistics charts more aesthetically pleasing. Seaborn is utilised in this project to create a heatmap that displays 24 hours, 7 days, and various colour gradations for messages ranging from the most important to the least important.

Streamlit

This library is used in this project to provide stunning web items and objects for displaying Whatsapp chat analysis with various charts and visualisations on Streamlit.

NLP

In this project, NLP features like text parsing, stop word elimination, and text analysis are applied. Text is parsed to separate messages into words for analysis, such as word counts and frequently used words. The Python programme is instructed to present only significant words by removing all stop words from a file that contains all stop words. To determine how many media files and URLs are shared, text analysis is employed.

IMPLEMENTATION

NLTK (Natural Language Toolkit):

A well-liked library for tasks involving natural language processing is NLTK. For tasks like tokenization, stemming, part-of-speech tagging, and sentiment analysis, it offers a wide range of tools and resources.

PLOTLY:

Using Plotly in WhatsApp chat analysis with machine learning (ML) involves leveraging ML techniques to extract insights from the chat data and using Plotly to visualize and present those insights.

NUMPY:

While NumPy is primarily a numerical computing library, it can be utilized in conjunction with machine learning (ML) for WhatsApp chat analysis.

REGEX:

Use regex or other text processing techniques to extract emojis from the chat data. Regular expressions can help identify emoji patterns based on Unicode ranges or specific character sequences.

ALGORITHMS

Sentiment Analysis

The NLP technique most frequently employed is sentiment analysis. In situations where consumers provide their opinions and suggestions, such as polls, reviews, and discussions on social media, emotion analysis is very helpful.

A three-point scale (positive/negative/neutral) is the easiest to construct in emotion analysis. The output in more complicated circumstances may be a statistical score that can be broken down into as many categories as required.

Text Summarization

As the name suggests, NLP techniques can help with the summary of large amounts of text. In instances like news headlines and research projects, text summary is frequently used.

There are two methods for text summarization: extraction and abstraction. The rundown is produced by extraction methods by removing portions of the text. Through the creation of new text that captures the core of the original content, abstraction techniques create summaries.

Bag of Words

This paradigm ignores syntax and even word order while maintaining multiplicity, representing a text as a bag (multiset) of words. The bag of words paradigm essentially creates an incidence matrix. The training of a classifier uses these word frequencies or instances as features.

Tokenization

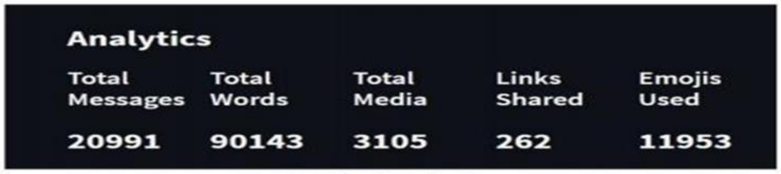
It's the process of breaking down the text into sentences and phrases. The work entails breaking down a text into smaller chunks (known as tokens) while discarding some characters, such as punctuation.

Vader

Valence Aware Dictionary for Sentiment Reasoning is known as VADER. It is an emotion analyzer with rules. It includes a list of lexical qualities that are typically classified as either positive or negative depending on their semantic orientation. It is a part of the NLTK package and can be used to process unlabeled text data right away. When given a string, VADER's Sentiment Intensity Analyzer() produces a dictionary with three categories:

a. Neutral
b. Positive
c. Negative

RESULTS



Analytics

Total Messages	Total Words	Total Media	Links Shared	Emojis Used
20991	90143	3105	262	11953

Fig.2. Analytics

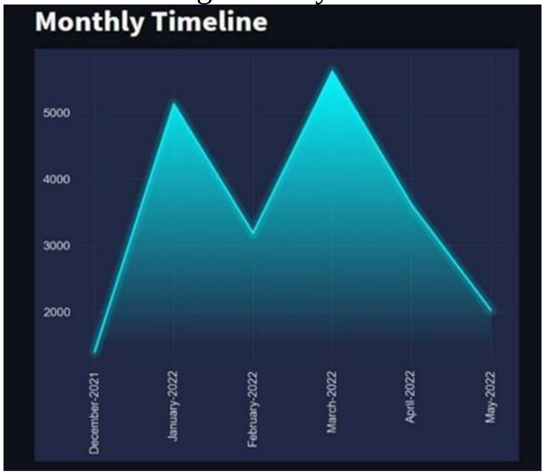


Fig.3. Monthly Time line

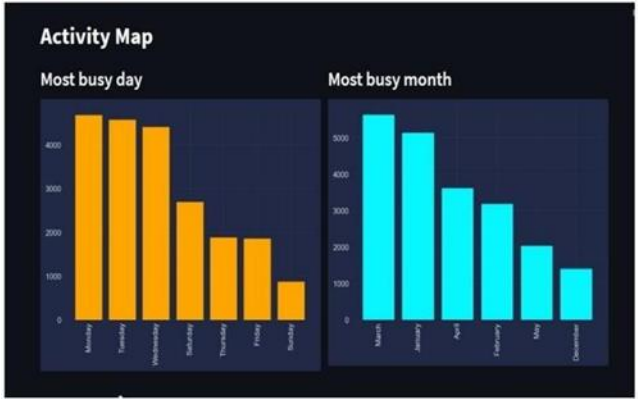
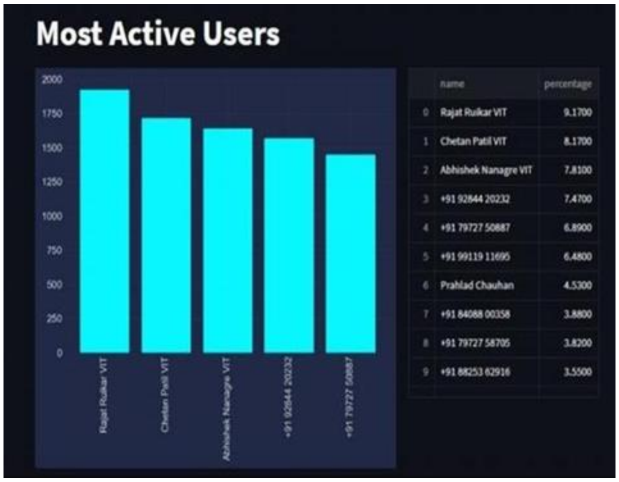
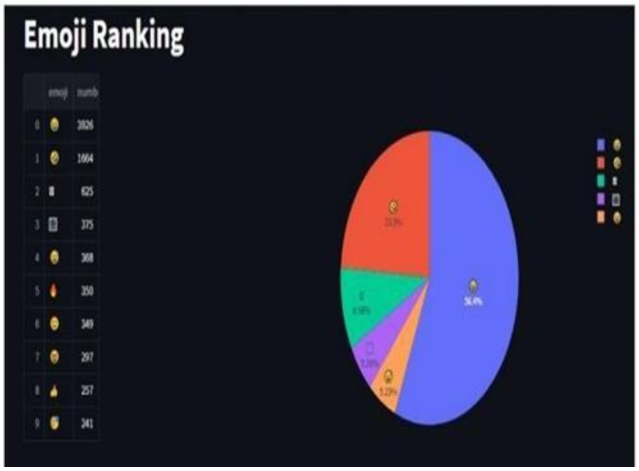


Fig.4. Activity Map



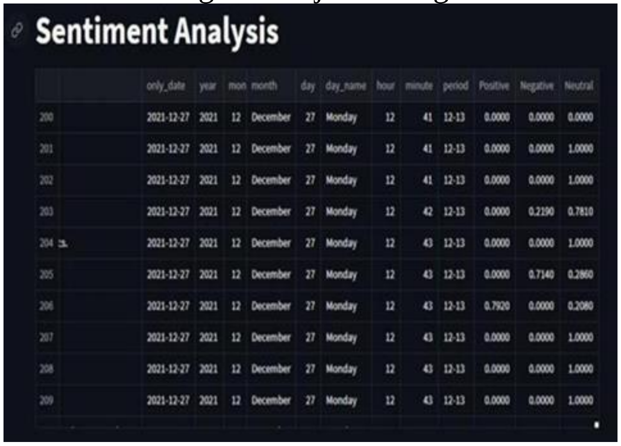
	name	percentage
0	Rajat Ruikar VIT	9.1700
1	Chetan Patil VIT	8.1700
2	Abhishek Nanagre VIT	7.8100
3	+91 92844 20232	7.4700
4	+91 79727 50887	6.8900
5	+91 99119 11695	6.4800
6	Prahlad Chauhan	4.5300
7	+91 84088 00358	3.8800
8	+91 79727 58705	3.8200
9	+91 88253 62916	3.5500

Fig.5. Most Active Users



	emoji	numb
0	●	2826
1	●	1664
2	▮	625
3	▯	375
4	●	368
5	🔥	350
6	●	349
7	●	297
8	👍	257
9	●	241

Fig.6. Emoji Ranking



		only_date	year	mon	month	day	day_name	hour	minute	period	Positive	Negative	Neutral
200		2021-12-27	2021	12	December	27	Monday	12	41	12-13	0.0000	0.0000	0.0000
201		2021-12-27	2021	12	December	27	Monday	12	41	12-13	0.0000	0.0000	1.0000
202		2021-12-27	2021	12	December	27	Monday	12	41	12-13	0.0000	0.0000	1.0000
203		2021-12-27	2021	12	December	27	Monday	12	42	12-13	0.0000	0.2190	0.7810
204	[illegible]	2021-12-27	2021	12	December	27	Monday	12	43	12-13	0.0000	0.0000	1.0000
205		2021-12-27	2021	12	December	27	Monday	12	43	12-13	0.0000	0.7140	0.2860
206		2021-12-27	2021	12	December	27	Monday	12	43	12-13	0.7920	0.0000	0.2080
207		2021-12-27	2021	12	December	27	Monday	12	43	12-13	0.0000	0.0000	1.0000
208		2021-12-27	2021	12	December	27	Monday	12	43	12-13	0.0000	0.0000	1.0000
209		2021-12-27	2021	12	December	27	Monday	12	43	12-13	0.0000	0.0000	1.0000

Fig.7. Sentiment Scores for each message

CONCLUSION

In conclusion, it can be concluded that the WhatsApp application's capabilities and the Python programming language's strength in realising the goals of data analysis cannot be overstated. In order to analyse a WhatsApp chat, this paper discussed the WhatsApp programme and python libraries.To better understand the WhatsApp chat available on our phones, we suggest using dataset modification techniques.It displays the words and emojis that are used the most frequently. It keeps tabs on our exchange and gauges how much time we are putting in.The system was written in Python, and NumPy,

http://doi.org/10.36893/JNAO.2023.V14I2.0225-0232

Pandas, Matplotlib, and Seaborn were among the Python libraries used. The analysis was able to reveal the amount of participation of the various people on the specified WhatsApp group chat at the conclusion of the work, and the findings were as anticipated. On a more serious note, this technology is capable of analysing any WhatsApp chat that is entered into it.

The overarching objectives established during the first phases of the requirements analysis are successfully met. After being put into use, the system produces accurate results. Users with little computer experience can easily utilise the designed system thanks to its complete user friendliness. Due to the system's validation capabilities, it entirely eliminates the chance of entering data incorrectly and avoids the drawbacks of existing manual solutions.

FUTURE SCOPE

The future scope of WhatsApp chat analysis using machine learning (ML) is quite promising and holds potential for various applications. Here are some potential areas of development and utilization:

Sentiment Analysis: ML algorithms can be applied to WhatsApp chat data to analyze the sentiment of conversations. This can be useful for understanding customer feedback, monitoring brand perception, and identifying emerging trends.

Chatbot Optimization: ML can help improve chatbot performance by analyzing WhatsApp conversations and identifying patterns in user queries and responses. This analysis can be used to refine the chatbot's natural language processing capabilities and enhance its ability to provide accurate and relevant information.

Personalized Recommendations: ML algorithms can analyze WhatsApp chats to understand users' preferences and interests. This data can be leveraged to provide personalized recommendations for products, services, or content, thereby enhancing the user experience.

Anomaly Detection: ML can be utilized to identify unusual or suspicious activities in WhatsApp chats, such as detecting spam messages, potential security threats, or abusive behavior. This can contribute to enhancing user safety and privacy.

Language Translation: WhatsApp conversations often involve users from different linguistic backgrounds. ML-powered language translation models can be integrated into WhatsApp to provide real-time translation of messages, facilitating communication between individuals who speak different languages.

Content Moderation: ML algorithms can assist in moderating WhatsApp chats by automatically identifying and flagging inappropriate or offensive content. This can help create a safer and more inclusive environment for users.

Predictive Analytics: By analyzing past WhatsApp conversations, ML models can predict future user behavior, such as purchase intent, engagement levels, or churn likelihood. This information can be valuable for businesses in optimizing their marketing strategies and making data-driven decisions.

Customer Support Optimization: ML algorithms can be trained on historical WhatsApp chat data to automate certain customer support tasks, such as answering frequently asked questions or routing inquiries to the appropriate departments. This can improve response times and overall customer satisfaction.

It's important to note that the development and deployment of such ML-powered applications should consider privacy and data security concerns, ensuring that user consent and data protection regulations are strictly adhered to.

REFERENCES

[1] Marada Pallavi, Meesala Nirmala, Modugaparapu Sravani, Mohammad Shameem. WhatsApp Chat Analysis. International Research Journal of Modernization in Engineering Technology and Science. Volume: 04/Issue:05/May-2022

[2] Shaikh Mohd Saqib. Whatsapp Chat Analyzer. International Research Journal of Modernization in Engineering Technology and Science. Volume: 04/Issue:05/May-2022

[3] K, Ravishankara & Dhanush, & Vaisakh, & S, Srajan. (2020). Whatsapp Chat Analyzer. International Journal of Engineering Research and. V9.10.17577/IJERTV9IS050676.

[4] D.Radha, R. Jayaparvathy, D. Yamini, "Analysis on Social Media Addiction using Data Mining Technique", International Journal of Computer Applications (0975 – 8887).

[5] https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664

[6]https://www.analyticsvidhya.com/blog/2021/06/build-web-app-instantly-for-machine- learning-using-streamlit/

[7] Meng Cai, "PubMed Central", PMCID: PMC7944036, PMID: 33732917

[8] E. Larson, "[Research Paper] Automatic Checking of Regular Expressions," 2018 IEEE 18th International Working Conference on Source Code Analysis and Manipulation (SCAM), 2018, pp. 225-234, doi: 10.1109/SCAM.2018.00034.