iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.53596

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

WhatsApp Data Sentiment Analyzer

Nilanjan Jana[1], Priyesh Pandey[2], Prof. Shallu Bashambu[3]
[1, 2, 3]*Department of Information Technology, Maharaja Agrasen Institute of Technology Plot No 1 Rohini, Plot No 1, CH Bhim Singh Nambardar Marg, Sector 22, PSP Area, Delhi, 110086*

Abstract: ***WhatsApp, a globally used communication tool in today's society, is transforming into an application where individuals express their thoughts, feelings, and opinions. A significant portion of this communication occurs within the app's group conversations. To facilitate the analysis of WhatsApp conversations, a web application called the WhatsApp Conversation Analyzer has been developed. This application leverages various Python libraries, including matplotlib, re, seaborn, streamlit, pandas, as well as foundational knowledge of NLTK (Natural Language Toolkit) for a comprehensive understanding. By importing a chat file of whatsapp from a group or single user, this hybrid system combines NLP (Natural Language Processing), NLTK, and machine learning techniques to perform a detailed analysis and provide valuable insights***
Keywords: Whatsapp chat, seaborn, matplotlib, vader, NLTK, statistical analysis, streamlit.

I. INTRODUCTION

We presented a WhatsApp Data Sentiment Analyzer in this study. WhatsApp chat data contains several sorts of messages between group members and individual users. The exported conversation file can be used to train machine learning and natural language processing models. These technologies create the best learning environment. This application analyzes data extracted from exported WhatsApp chats. The main benefit of this system is that it is built with simple Python libraries such as seaborn, streamlit, numpy, matplotlib, and pandas. These are frequently used to generate data frames and graphs.

The overall goal of this research project is to provide a comprehensive platform that integrates statistical analysis, dataset modification methods, and sentiment analysis to offer insightful information about WhatsApp discussions. Researchers and analysts may better understand user behavior and communication dynamics in the digital age by utilizing the tool's capabilities to find patterns, trends, and sentiment variations within the chat data.

II. LITERATURE REVIEW

This paper introduces VADER (Valence Aware Dictionary and sEntiment Reasoner), a rule-based model specifically designed for sentiment analysis of social media text. VADER utilizes a combination of lexical and grammatical heuristics, along with a pretrained sentiment lexicon that incorporates both valence scores and lexical features. The model is trained on human-annotated data and is shown to outperform several state-of-the-art baselines on sentiment classification tasks. VADER's design principles prioritize simplicity, efficiency, and generalizability, making it suitable for real-time applications and large-scale social media analysis[1]. A study was conducted on sentiment analysis of WhatsApp group chats using VADER and machine learning algorithms. The experimental results show that VADER performs well in sentiment analysis but lags in accuracy[2].

The effectiveness of VADER in analyzing the sentiment of WhatsApp Conversation was shown by the experimental results from the study conducted by Singh , Kumar and Joshi[3]. The study was conducted by Ullah, Hassan and Malik to determine the sentiment polarity and evaluate the performance of VADER in comparison to NAIVE BAYES. The experimental results demonstrate that VADER achieves higher accuracy and precision compared to Naive Bayes in sentiment analysis of WhatsApp messages[4].

The study was conducted to analyze the sentiment polarity of group chat messages and evaluate the performance of VADER in comparison to Random Forest. The experimental results demonstrate the effectiveness of VADER in sentiment analysis and highlight the strengths of Random Forest in handling complex sentiment patterns[5].

The study was conducted to analyze the sentiment polarity of the chat messages and evaluate the performance of VADER in comparison to KNN. The experimental results demonstrate the effectiveness of VADER in sentiment analysis and highlight the advantages of the KNN algorithm in capturing similar sentiment patterns[6]. The study presents the Senti-N-Gram approach, which is a lexicon-based method for sentiment analysis incorporating n-grams. Traditional lexicon-based approaches for sentiment analysis often overlook the context and combinations of words by considering individual words alone. The Senti-N- 3 Gram approach addresses this limitation by treating n-grams, contiguous sequences of n words, as the fundamental units for sentiment analysis[7].

©IJRASET: All Rights are Reserved

III. METHODOLOGY

A tool for statistical analysis of WhatsApp talks is the WhatsApp Data Sentiment Analyzer. Working with exported conversation files will aid in producing various plots of analysis.The methodology for the WhatsApp data analyzer project involves several steps:

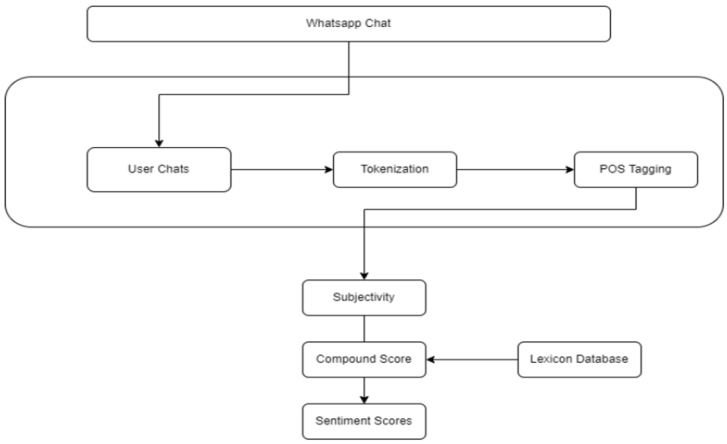
1) Data Collection:- Select the desired WhatsApp chat, go to chat settings, select "Export Chat," choose the export format (with or without media), select the export destination (email, cloud storage, etc.), and then follow the on-screen instructions to finish the export process.
2) Data Preprocessing:- The next step is to pre-process the data by cleaning and transforming it into a structured format that can be easily analyzed. This involves removing irrelevant information such as system messages and formatting the data into a tabular format. This step is implemented using Python libraries such as pandas and regular expressions. Our code imports the necessary libraries, ‘re’ for regular expressions and ‘pandas’ for data manipulation and analysis.
3) Data Visualization:- The next step is to visualize the pre-processed data using Python packages like matplotlib and seaborn. Bar charts, line charts, and scatter plots are just a few of the different graphs and charts that may be produced. Message statistics, such as the quantity of messages, total number of words, media messages, and shared URLs, are important visualizations. By adding up their message contributions, one can identify the users who are the most active. To display commonly used terms, word clouds can be created. Emoji analysis records and counts the emojis used in the discussion. Message exchanges can be tracked over time using timelines and activity maps, which show how messages are distributed over the course of a week and a month.
4) Streamlit:- The Python Streamlit module is used by the WhatsApp data analyzer project to build a web-based application for displaying the analysis' findings. Through an intuitive interface, users may engage with the chat data by choosing specific users or by seeing global statistics, word clouds, common words, and emoji analysis. Streamlit makes it possible to create dynamic visualizations that change in response to user inputs, offering a tailored analytical experience. It is simple to deploy the web application, enabling others to access and engage with the visualizations and analysis.

IV. SENTIMENT ANALYSIS USING VADER

The method of identifying whether a piece of writing is good, negative, or neutral is known as sentimental analysis. The Algorithm below is intended for use in financial texts. It consists of the following steps:

1) Cleaning using Stop Words Lists: Stop Words Lists which are presented in the stop word folder are used to omit those words which have irrelevant meaning and are not useful in sentimental analysis.
2) Creating a dictionary of Positive and Negative words:. We have a master lexicon which is inside the master dictionary folder . They are used to create lexicons of both positive and negative words. These are those words that were not present in the stop list.
3) Extracting Derived variables: Now the text we receive is converted into token list by using nltk tokenize module
4) Positive Score: This score is computed by assigning a value of +1 to each word discovered in the Positive Dictionary and then adding all of the values together.
5) Negative Score: This score is computed by assigning a value of -1 to each word discovered in the Negative Dictionary and then adding all of the values together. We multiply the score by -1 to make it a positive value.
6) VADER (Valence Aware Dictionary and Sentiment Reasoner) is a rule-based sentiment analysis model for social media text. It analyzes the sentiment of individual words and phrases by combining lexical and grammatical heuristics with a predetermined sentiment lexicon. Because of the following reasons, VADER is particularly well-suited for sentiment analysis in WhatsApp chat data: Lexicon for a Specific Domain: VADER's sentiment lexicon is designed exclusively for social media content, including casual language, slang, and emoticons found on services such as WhatsApp. This improves its ability to comprehend and analyze the sentiment communicated in WhatsApp talks. Handling Context and Intensity: VADER takes the context and intensity of sentiment expressions into account. . It considers words with modifiers, negations, capitalizations, and punctuation, allowing it to capture the sentiment nuances in chat messages more accurately.
7) Rule-Based Approach: VADER utilizes a rule-based approach, which means it doesn't require extensive training data for sentiment classification. It relies on predefined rules and heuristics to assign sentiment scores to 7 individual words and phrases. This makes it easy to implement and doesn't require large labeled datasets for training.
8) Speed and Efficiency: VADER is computationally efficient and provides real time sentiment analysis. It doesn't require

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com



Framework of the Vader Sentiment Analysis System

V. WORKING

To begin, navigate to the sidebar and select the "Browse files" button. We should have a Whatsapp chat exported file on the system, which we may pick as indicated in the figure.

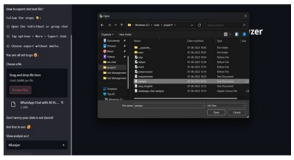


Fig 1.Select chat

Then, when you've uploaded the file. The user can acquire the analysis by selecting any user from the group or overall using the dropdown. The user will receive the total analysis by default..



Fig 2. Analyzed data

Following data processing, we obtain total messages, words, material shared, and links shared. A line chart is used to depict the monthly timeline. The activity map uses a bar graph to display the busiest days and months. The heatmap depicts weekly activity, as well as the most active users, and the wordcloud depicts the most often used word.

VI. OUTPUTS AND RESULTS



Fig 1. Statistics

©IJRASET: All Rights are Reserved

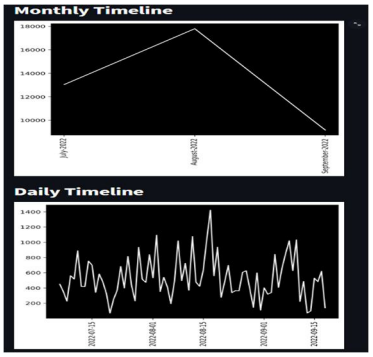


Fig 2. Monthly and Daily Timeline

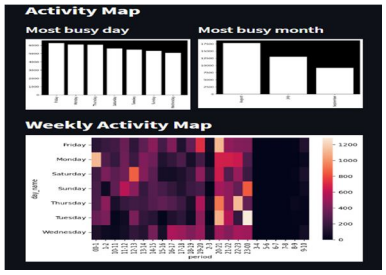


Fig 3. Activity Map

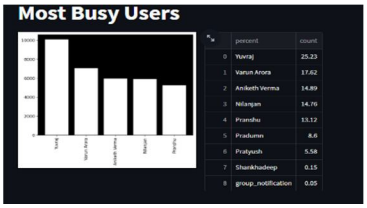


Fig 4. Most busy user

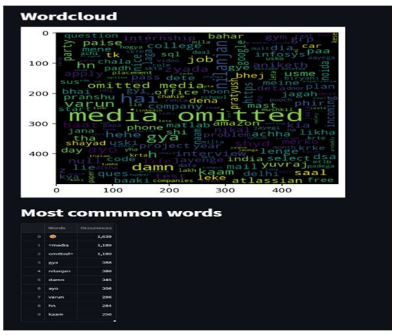


Fig 5. Wordcloud and most common words

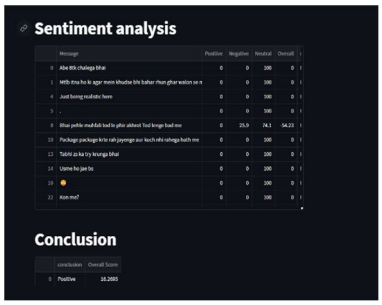


Fig 6. Sentiment Analysis

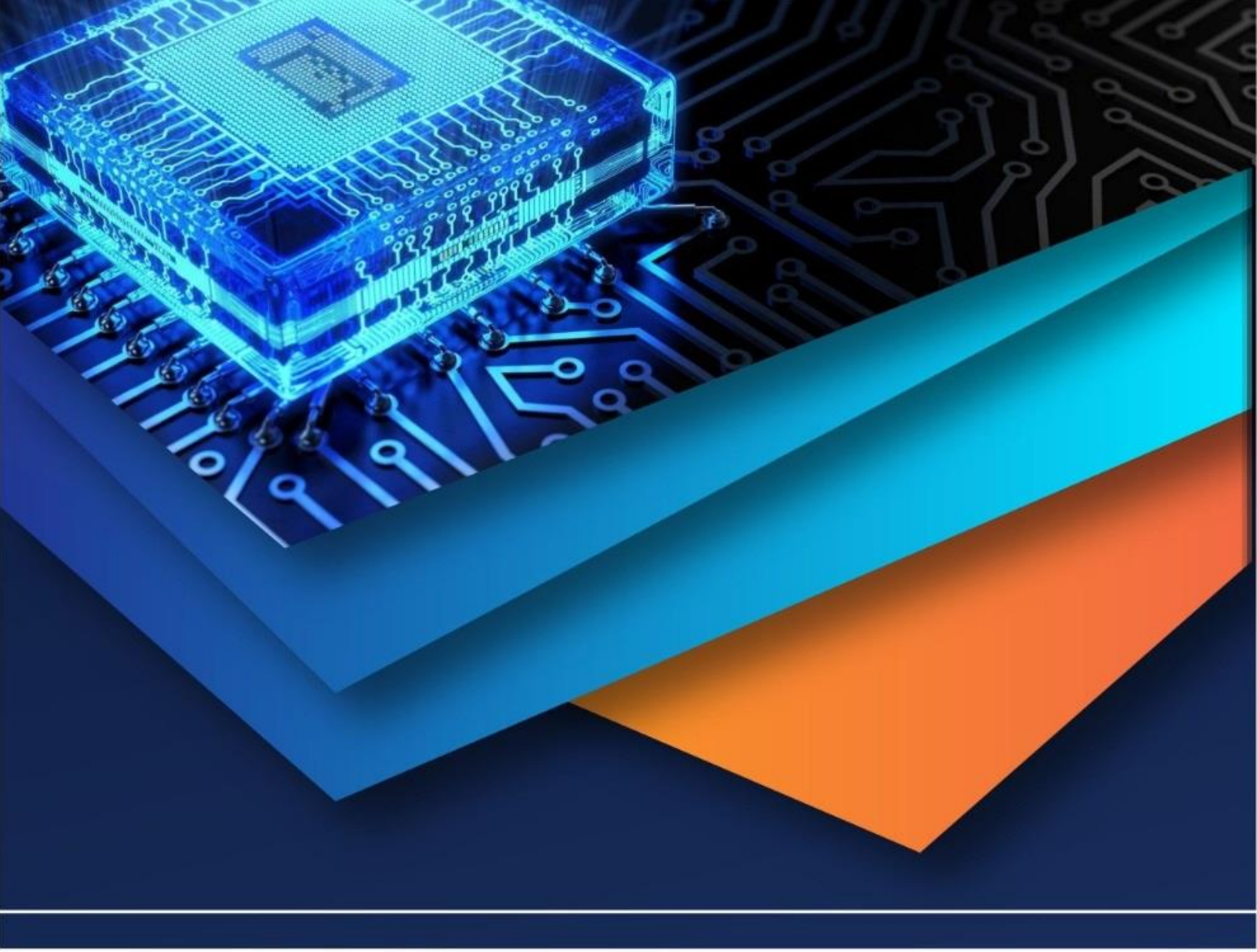
VII. CONCLUSION AND FUTURE WORK

The model's primary goal is to analyse WhatsApp messages and present the results of such analyses. The model also employs the Vader implementation for sentiment analysis. Following completion, the system generates consistent outcomes. The system is completely simple to use, allowing even those with limited computer knowledge to run the generated system. The model performs well with small data sets and lexicons. Future work will be to make it work for n-grams as Vader only works for unigrams, which could be accomplished by utilising other machine learning models or adding a dictionary in Vader for n-gram tokens. Additionally, because Vader is faster than other sentiment analysis tools, the goal is to maintain speed and efficiency.

REFERENCES

[1] VADER: A Parsimonious Rulebased Model for Sentiment Analysis of social media Text
[2] Chowdhury, A., Roy, A., & Dey, D. (2020). Sentiment analysis of WhatsApp group chats using VADER and machine learning algorithms. International Journal of Computer Science and Information Security, 18(12), 70- 77.
[3] Singh, A., Kumar, V., & Joshi, R. C. (2021). Sentiment analysis of WhatsApp conversations using VADER. International Journal of Advanced Research in Computer Science and Software Engineering, 11(6), 489-494
[4] Ullah, A., Hassan, M., & Malik, S. (2020). Sentiment analysis of WhatsApp messages using VADER and Naive Bayes. In Proceedings of the International Conference on Software Engineering, Mobile Computing and Media Informatics (pp. 100- 107).
[5] Saha, A., Islam, M. S., & Shahriar, H. (2021). Sentiment analysis of WhatsApp group chat using VADER and random forest. In Proceedings of the International Conference on Computer Science, Engineering and Applications (pp. 41-46).
[6] Ahmad, W., Khalid, S., & Bashir, M. A. (2020). Sentiment analysis of WhatsApp chat using VADER and K-Nearest Neighbor algorithm. In Proceedings of the International Conference on Artificial Intelligence and Sustainable Technologies (pp. 296-302).
[7] A. Dey, M. Jenamani, and J. J. Thakkar, "Senti-N-Gram: An ngram lexicon for sentiment analysis," Expert Systems with Applications, vol. 103, pp. 92-105, Aug. 2018.
[8] 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
[9] https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664
[10] https://www.analyticsvidhya.com/blog/2021/06/build-web-app-instantly-for-machine-learning-using-streamlit/
[11] 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.
[12] D.Radha, R. Jayaparvathy, D. Yamini, “Analysis on Social Media Addiction using Data Mining Technique”, International Journal of Computer Applications (0975 – 8887).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)