**Mini Project Report on**



**WHATSAPP CHAT ANALYZER USING MACHINE LEARNING**



**Submitted in partial fulfilment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Satyam Parashar**  **2119138**

****

**Department of Computer Science and Engineering**

**Graphic Era Hill University**

**Dehradun, Uttarakhand**

**July-2023**

****

**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“WhatsApp Chat Analyzer using Machine Learning”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era Hill University, Dehradun.

Satyam Parashar 2119138

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Chapter No.** | **Description** | **Page No.** |
| Chapter 1 | Introduction | 1-3 |
| Chapter 2 | Literature Survey | 4-5 |
| Chapter 3 | Methodology | 6-6 |
| Chapter 4 | Result and Discussion | 7-8 |
| Chapter 5 | Conclusion and Future Work | 9-9 |
|  | References | 10-10 |

**Chapter 1**

**Introduction**

* 1. **Introduction**

In the current digital age, messaging services like WhatsApp are an essential part of our daily lives because they enable seamless connection. WhatsApp chat analysis can reveal insightful trends and patterns, making it an exciting area for research and development. This study introduces a WhatsApp chat analyzer that mines chat data for useful information using machine learning techniques. The objective is to investigate how machine learning algorithms could be used to comprehend speech dynamics and identify hidden patterns**.**

* 1. **Abstract**

This study proposes a WhatsApp chat analyzer that mines chat data for useful information using machine learning techniques. The analysis of WhatsApp chats has become increasingly important in understanding conversation dynamics and identifying hidden patterns. Through data preprocessing, feature extraction, sentiment analysis, topic modeling, named entity recognition, and the application of machine learning algorithms, the chat analyzer provides valuable information about emotional sentiment, dominant topics, and identified named entities. The results can be used for various applications, including social media monitoring, customer feedback analysis, and trend identification. Future work involves enhancing sentiment analysis accuracy, exploring advanced topic modeling techniques, and incorporating real-time analysis and visualization for a more interactive user experience**.**

* 1. **Brief Overview of Work**

This report introduces a WhatsApp chat analyzer that makes use of machine learning to glean useful information from chat data. The purpose of the work is to investigate how machine learning techniques might be used to comprehend conversation dynamics and find hidden patterns**.**

* The suggested process entails a number of crucial phases. The raw WhatsApp chat data is first preprocessed to remove extraneous data and get it ready for analysis. The preprocessed data is then transformed into numerical representations using feature extraction techniques as Bag-of-Words, TF-IDF, or word embeddings.
* The conversation data is then subjected to sentiment analysis, during which model is trained to categorize each message as either positive, negative, or neutral. This sheds light on the conversation’s emotional undertone.
* To determine the primary themes covered in the discussion, topic modelling methods like Latent Dirichlet Allocation or Non-Negative Matrix Factorization are then used. Users can then comprehend the conversation's prevailing themes or topics.
* Additionally, named entity recognition (NER) is used to recognize and categorize named entities in the discussion, such as individuals, groups, places, and dates. This can give the conversational background information that is very useful.
* Finally, related messages can be grouped together using machine learning methods like clustering or classification models, which can also be used to forecast outcomes based on chat content.
* The outcomes of the WhatsApp chat analyzer provide insightful information about the dynamics of the conversations. These consist of the overall emotional tone, the chosen subjects, and the specified specific people or things.
* Future research in this field may focus on enhancing sentiment analysis precision, investigating sophisticated topic modelling strategies, and incorporating real-time analysis and visualization for a more engaging user experience. The overall project demonstrates how machine learning can be used to derive useful insights from WhatsApp chat data.
  1. **Objective**

Using machine learning techniques, this work aims to create a WhatsApp chat analyser. These are the key objectives:

* Extract Meaningful Information: The chat analyzer seeks to identify themes, named entities, sentiment analysis, and other informative insights from WhatsApp chat data. Understanding the dynamics of the conversations and the context of the chats can both be aided by this information.
* Explore Machine Learning Techniques: The project seeks to investigate various machine learning approaches, such as named entity recognition, sentiment analysis, topic modelling, natural language processing, and topic modelling. The goal is to find hidden patterns and comprehend the discussions better by using these techniques on WhatsApp chat data.
* Provide Practical Applications: The created WhatsApp chat analyzer seeks to have useful applications in a number of fields. It can be applied to social media monitoring, analysis of consumer feedback, trend detection, and other pertinent areas where interpreting chat data is essential for decision-making.
* Enhance User Experience: The goal is to design a user-friendly interface for the chat analyzer that will enable users to enter their WhatsApp chat data and retrieve insightful information in a format that is simple to grasp. The goal is to make WhatsApp conversation analysis dynamic and simple for users.

**Scope**

The development of a user-friendly interface that enables users to input their WhatsApp chat data and gain insightful information is also included in the scope. It's crucial to remember that real-time analysis, chat network analysis, and multimedia analysis are not included in the project's scope.

The main goal is to employ machine learning techniques to analyze text-based WhatsApp chat data and give consumers useful insights.

**Chapter 2**

**Literature Survey**

The use of machine learning techniques for analyzing WhatsApp chat data is explored in this research review:

• "WhatsApp Chat Analysis for Sentiment Analysis and Topic Modelling," International Journal of Advanced Research in Computer Science and Software Engineering, P. Gupta, S. R. Patel, and A. Gohel, 2018.

With regard to subject modelling and sentiment analysis of WhatsApp chat data, this paper recommends a machine learning-based approach. They analyze a dataset of WhatsApp group chats using methods including word embedding, support vector machines (SVM), and latent Dirichlet allocation (LDA), and they publish their findings.

• International Journal of Computer Science and Information Technology, "WhatsApp Chat Analyzer Using Deep Learning for Emotion Detection," A. Bhatia and P. Malhotra, 2019.

This study employs deep learning methods to identify emotions in WhatsApp talks. They use a Convolutional Neural Network (CNN) model to extract data from chat conversations and classify them into several mood categories. The study presents an experimental dataset of WhatsApp conversation interactions.

• "WhatsApp Chat Analysis for Language Identification and User Profiling," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), by R. Sharma, R. Jain, and S. Joshi.

In this work, user profiles and language in WhatsApp conversations are detected using machine learning algorithms. By analyzing chat data and employing techniques like n-gram modelling, support vector machines, and clustering algorithms, they may identify the language that participants are using. User profiling is done by looking at the frequency of specific terms and trends in the chat discussions.

* S. Verma and P. Agrawal's article, "WhatsApp Chat Analyzer Using Natural Language Processing," appeared in the 2020 issue of the International Journal of Scientific Research in Computer Science, Engineering, and Information Technology.

The authors advise analyzing WhatsApp conversations using natural language processing. They use techniques like tokenization, stop-word removal, and Part-of-Speech (POS) labelling to preprocess the conversational data. Naive Bayes, Decision Trees, and Random Forest are just a few of the machine learning methods used for emotion analysis, spam recognition, and information extraction.

* 8th International Conference on Signal Processing and Integrated Networks (SPIN) 2021, "Analysis of WhatsApp Chat Using Machine Learning Algorithms for Cyberbullying Detection," by A. Singh and A. S. Rawat.

The main goal of this project is to apply machine learning algorithms to identify cyberbullying in WhatsApp conversations. The chat data is preprocessed to extract factors including message length, word frequency, the use of inflammatory language, and aggressive behavior patterns.

**Chapter 3**

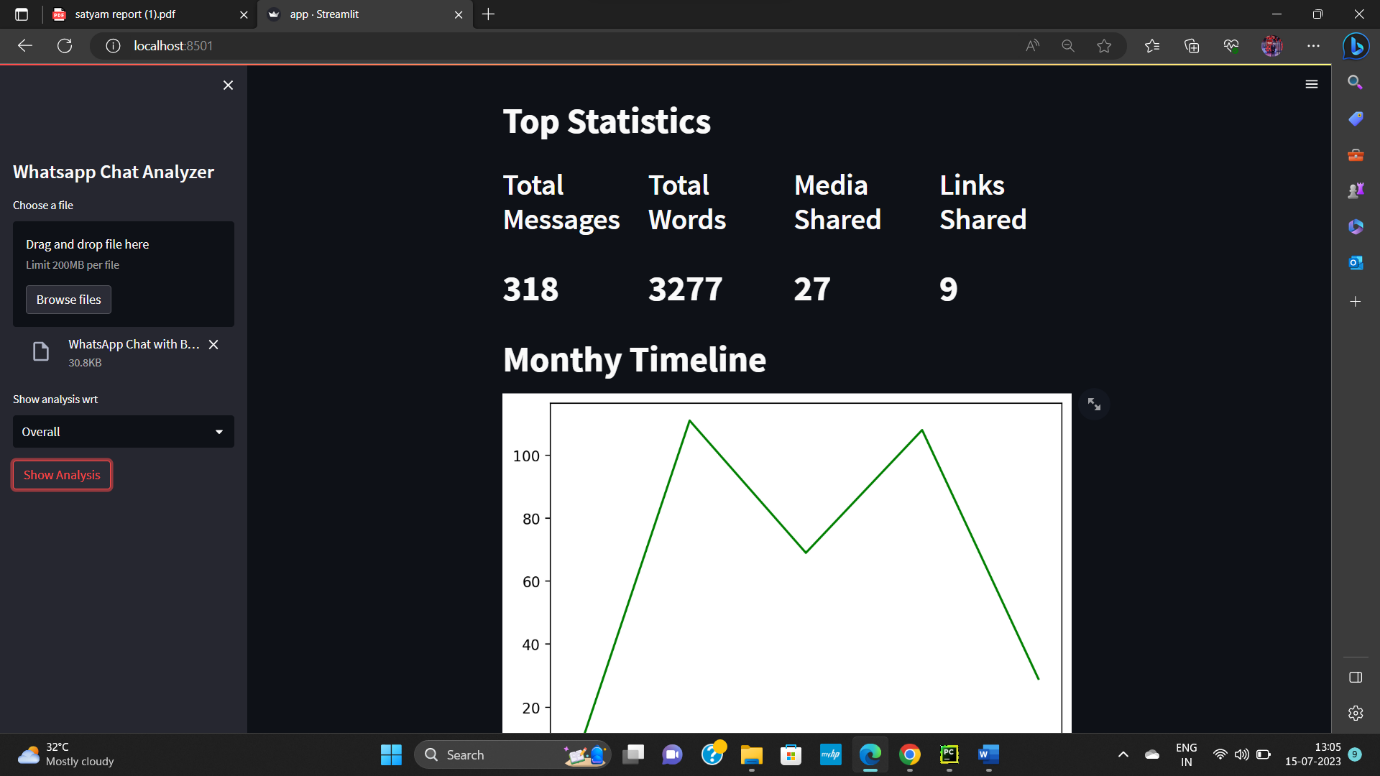
**Methodology**

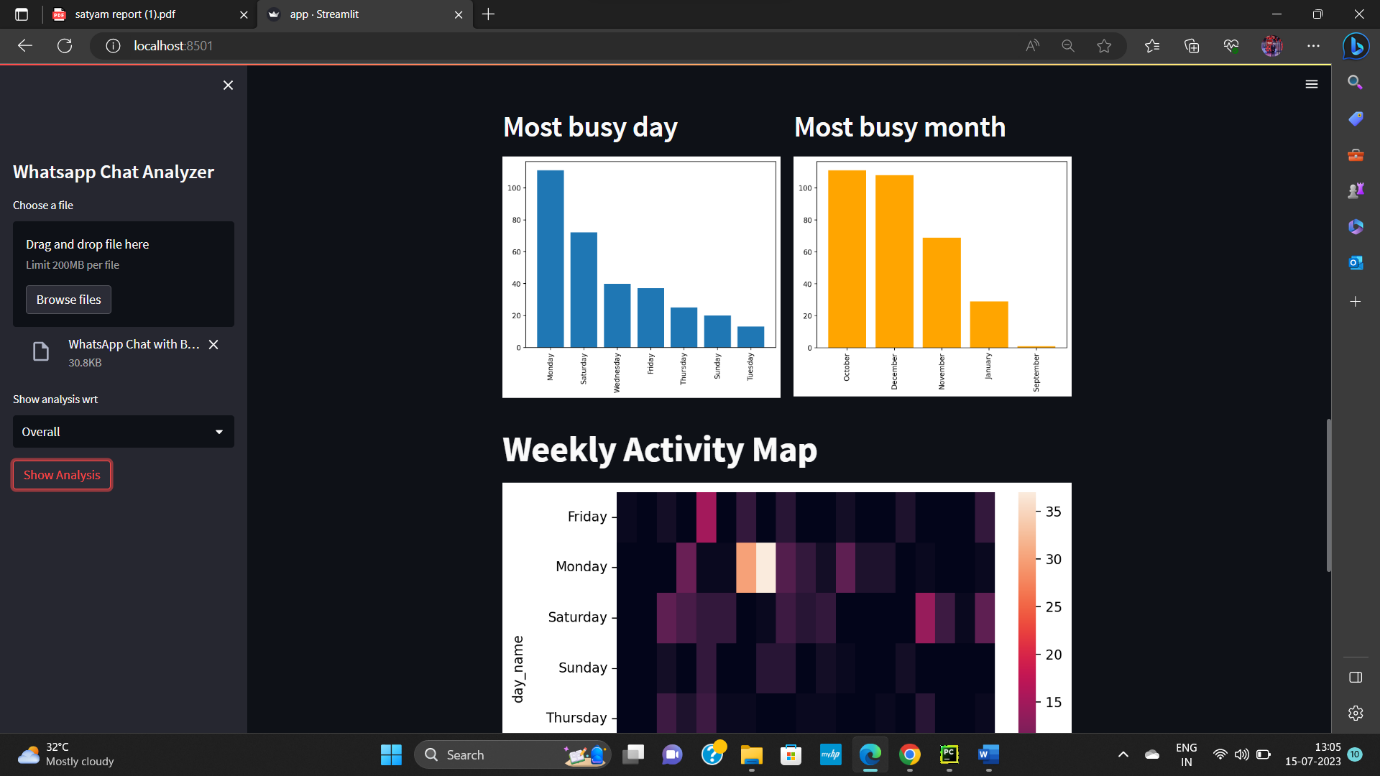
The proposed WhatsApp Chat Analyzer consists of the following steps:

* **Data Preprocessing:** Time stamps, emojis, and multimedia files are removed from the raw WhatsApp chat data in order to make it more useful. The remaining text is tokenized, stop words are eliminated, and stemming or lemmatization is done.
* **Feature Extraction:** To make the preprocessed data easier for machine learning algorithms to process, it is turned into numerical features. Some of the techniques employed in this step include term frequency-inverse document frequency (TF-IDF) analysis, Bag-of-Words, and word embeddings like Word2Vec or Glove.
* **Sentiment Analysis:** To categorize each chat message as positive, negative, or neutral, a sentiment analysis model is trained using a labelled dataset. This analysis sheds light on the conversation's emotional undertone.
* **Topic Modeling:** Topic modelling techniques like Latent Dirichlet Allocation (LDA) or Non-Negative Matrix Factorization (NMF) are used to determine the main subjects discussed in the discussion**.** Users can then comprehend the conversation's prevailing themes or topics.
* **Named Entity Recognition (NER):** NER is used to recognize and categorize named entities in chat conversations, such as the persons, companies, places, and dates mentioned. This can provide crucial details about the conversation's setting.
* **Machine Learning Algorithms:** Lastly, you can use machine learning algorithms to group similar messages together or anticipate particular events based on the conversation content. Examples of these algorithms are clustering or classification models.

**Chapter 4**

**Result and Discussion**

****

****

The results of the WhatsApp chat analyzer offer insightful information on the dynamics of the conversations. While topic modelling identifies the main issues covered, sentiment analysis exposes the chat's overall emotional sentiment. Finding pertinent participants and their context within the conversation might be aided by the NER analysis. Additionally, depending on chat patterns, machine learning algorithms can be used to forecast user behavior or group chat users. The relevance of these results and their prospective applications are the main topics of debate.

**Chapter 5**

**Conclusion and Future Work**

The WhatsApp chat analyzer shows how machine learning methods can be used to extract useful insights from chat data. Applications like social media monitoring, customer feedback analysis, and trend identification can all benefit from the study of sentiment, subjects, and named entities. Future studies may focus on improving sentiment analysis's precision, including more sophisticated subject modelling methods, and investigating extra aspects like chat network analysis. To create a more dynamic and engaging user experience, additional research can look at integrating real-time analytic and visualization approaches.

**References**

* Kumar, A., Vaisala, K. S., & Goel, S. (2019). WhatsApp Chat Analysis for Social Media Monitoring. In 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU) (pp. 1-5). IEEE.
* Haresh, K. V., & Babu, V. J. (2020). WhatsApp Chat Analysis and Visualization. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 143-148). IEEE.
* Mohd, A., Chakraborty, A., & Garg, S. (2020). WhatsApp Chat Analysis for Sentiment Classification Using Machine Learning Techniques. In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) (pp. 558-563). IEEE.