A Project Report on

WHATSAPP GROUP CHAT ANALYSIS AND PREDICTION USING NAÏVE BAYES CLASSIFICATION

Submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

Submitted by

Shovon Raul	**(19P31A0550)**
Sathi DivyaJnana Rama Koti Reddy	**(19P31A0546)**
Thibitisetti Sai Kiran	**(19P31A0556)**
Kondapalli Veera Ganesh	**(20P35A0505)**

Under the esteemed supervisor of

Mr. Chakka S.V.V.S.N. Murty, M. Tech, (Ph.D.)

Assistant Professor

ADITYA COLLEGE OF ENGINEERING & TECHNOLOGY

Approved by AICTE, New Delhi & Affiliated to JNTUK, Kakinada

Accredited by NAAC (A+) and NBA

Surampalem, Kakinada District, Andhra Pradesh - 533 437

2019-2023

VISION

To induce higher planes of learning by imparting technical education with

- ✓ International standards
- ✓ Applied research
- ✓ Creative Ability
- ✓ Value based instruction and to emerge as a premiere institute.

MISSION

Achieving academic excellence by providing globally acceptable technical education by forecasting technology through

- ✓ Innovative Research And development
- ✓ Industry Institute Interaction
- ✓ Empowered Manpower

Principal

ADITYA COLLEGE OF ENGINEERING & TECHNOLOGY SURAMPALEM

PRINCIPAL
Aditya College of
Engineering & Technology
SURAMPALEM- 533 437

Aditya College of Engineering & Technology

Aditya Nagar, ADB Road, Surampalem - 533437

Department of Computer Science and Engineering

VISION

- ✓ To become a center for excellence in Computer Science and Engineering education and innovation.

MISSION

- ✓ Provide state of art infrastructure.
- ✓ Adapt Skill based learner centric teaching methodology.
- ✓ Organize socio-cultural events for better society.
- ✓ Undertake collaborative works with academia and industry.
- ✓ Encourage students and staff self-motivated, problem-solving individuals using Artificial Intelligence.
- ✓ Encourage entrepreneurship in young minds.

Head of the Department

Head of the Department
Dept.of CSE
Aditya College of Engineering
& Technology
SURAMPALEM-533437

Aditya College of Engineering & Technology

Aditya Nagar, ADB Road, Surampalem - 533437

Department of Computer Science and Engineering

Program Educational Objectives

PEO 1	Capability to design and develop new software products as per requirements of the various domains and eligible to take the roles in various government, research organizations and industry
PEO 2	More enthusiastic to adopt new technologies and to improve existing solutions by reducing complexity which serves society requirements as per timeline changes
PEO 3	With good hands-on basic knowledge and ready improve academic qualifications in India or abroad.
PEO 4	Ability to build and lead the team to achieve organizational goals

Head of the Department

Head of the Department
Dept.of CSE
Aditya College of Engineering
& Technology
SURAMPALEM-533437

Department of Computer Science and Engineering

PROGRAM SPECIFIC OUTCOMES

PSO 1: The ability to design and develop computer programs for analyzing the data.

PSO 2: The ability to analyze data & develop Innovative ideas and provide solution by adopting emerging technologies for real time problems of software industry.

PSO 3: To encourage the research in software field that contribute to enhance the standards of human life style and maintain ethical values.

Head of the Department

Head of the Department
Dept.of CSE
Aditya College of Engineering
& Technology
SURAMPALEM-533437

ADITYA COLLEGE OF ENGINEERING & TECHNOLOGY

Approved by AICTE, New Delhi & Affiliated to JNTUK, Kakinada

Accredited by NAAC (A+) and NBA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project work entitled, "**Whatsapp Group Chat Analysis And Prediction Using Naïve Bayes Classification**", is a bonafide work carried out by **Shovon Raul (19P31A0550), Sathi DivyaJnana Rama Koti Reddy(19P31A0546), Thibirisetti Sai Kiran (19P31A0556), Kondapalli Veera Ganesh (20P35A0505)**, in partial fulfillment of the requirements for the award **of the degree of BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE AND ENGINEERING** from **Aditya College of Engineering and Technology**, Surampalem, during the academic year 2022-2023.

This project work has not been submitted in full or part to any other University or educational institute for the award of any degree or diploma.

PROJECT SUPERVISOR

Mr. Chakka S.V.V.S.N. Murty,

M. Tech., (Ph.D.)

Assistant Professor

HEAD OF THE DEPARTMENT

Dr. M. Anil Kumar,

M. Tech.,Ph.D.

Professor

EXTERNAL EXAMINER

DECLARATION

We hereby declare that this project entitled "**Whatsapp Group Chat Analysis And Prediction Using Naïve Bayes Classification**" has been undertaken by us and this work has been submitted to **ADITYA COLLEGE OF ENGINEERING AND TECHONOLOGY,** Surampalem affiliated to JNTUK, Kakinada, in partial fulfillment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE AND ENGINEERING.**

We further declare that this project work has not been submitted in full or part to any other University or educational institute for the award of any degree or diploma.

PROJECT ASSOCIATES

Shovon Raul	**(19P31A0550)**
Sathi DivyaJnana RamaKoti Reddy	**(19P31A0546)**
Thibirisetti Sai Kiran	**(19P31A0556)**
Kondapalli Veera Ganesh	**(20P35A0505)**

ACKNOWLEDGEMENT

It is with immense pleasure that we would like to express our indebted gratitude to my **project supervisor**, **Mr. CHAKKA S.V.V.S.N MURTY, M.Tech.,(Ph.D.),** who has guided us a lot and encouraged us in every step of project work, his valuable moral support and guidance has been helpful in successful completion of this Project.

We wish to express our sincere thanks to **Dr. M. ANIL KUMAR M.Tech.,Ph.D., Head of the Department of CSE**, for his valuable guidance given to us throughout the period of the project work.

We feel elated to thank **Principal**, **Dr. DOLA SANJAY S M.Tech.,Ph.D.**, of Aditya College of Engineering and Technology for his cooperation in completion of our project and throughout our course.

We feel elated to thank **Dr. A. RAMA KRISHNA M.Tech.,Ph.D., Dean (Academics & Administration)** of Aditya College of Engineering and Technology for his cooperation in completion of our project work.

We wish to express our sincere thanks to all faculty members, and lab programmers for their valuable guidance given to us throughout the period of the project.

We avail this opportunity to express our deep sense and heart full thanks to the **Management** of **Aditya College of Engineering & Technology** for providing a great support for us by arranging the trainees, and facilities needed to complete our project and for giving us the opportunity for doing this work.

PROJECT ASSOCIATES

Shovon Raul	**(19P31A0550)**
Sathi DivyaJnana Rama Koti Reddy	**(19P31A0546)**
Thibirisetti Sai Kiran	**(19P31A0556)**
Kondapalli Veera Ganesh	**(20P35A0505)**

ABSTRACT

Recently WhatsApp has become the most used and efficient method of communication and people are more interested in doing important collaborate tasks using WhatsApp Group Chat facilities. As group can be created for various purposes in different field of work, these group chats can be sources of lot of information for further analysis and problem solving. Our project “WhatsApp group chat analysis and prediction using naïve bayes classification” can provide an in-depth statistical analysis based on the given group chat data and a better understanding about people’s behaviours through chats which can be further useful for sentiment analysis later. We have included some special features to existing systems like, our tool can future predict the most suitable hour of the day for admin’s reply in the group, Categorize the subject of the group, giving warnings after recognizing media overload and hate speech, Linguistic Manifestations of WhatsApp conversations, language detection etc. In brief the aim of the project is doing statistical analysis as well as future predictions to include all possible feature in a single application. The naïve bayes classification algorithm used here are easy to handle, efficient and less resource consuming algorithm, therefore it will succeed while applying to larger datasets also. This project will help individuals, group administrators and any kind of organizations for efficient analysis of WhatsApp group chats for their specific purposes.

INDEX

LIST OF FIGURES

LIST OF TABLES

1.INTRODUCTION

Today one of the trendy social media platforms is…. guess what? One and only WhatsApp. It is one of the favorite social media platforms among all of us because of its attractive features. It has more than 2B users worldwide and "According to one survey an average user spends more than 195 minutes per week on WhatsApp". How terrible the above statement is. Leave all these things and let's understand what actually WhatsApp analyzer means?

WhatsApp Analyzer means we are analyzing our WhatsApp group activities. It tracks our conversation and analyses how much time we are spending or saying it as "wasting" on WhatsApp. The aim of this project is to provide step by step guide to build our own WhatsApp analyzer using python. Here we used different python libraries which helps us to extract useful information from raw data. As group can be created for various purposes in different field of work, these group chats can be sources of lot of information for further analysis and problem solving.

In our society we come across problems based on the communications in social media where social violence occurs. As mentioned, WhatsApp is used by billions of people definitely we should have a system for detecting such events so that we can solve this before it is creating more trouble. If we build a system for sentiment analysis exclusively for WhatsApp chat then it will solve this problem.

We can add features like language detection in analysis as well so that admins can detect messages and labeled them with particular language for WhatsApp groups having connection with different linguistic persons.

This project will provide an in-depth statistical analysis based on the given group chat data and a better understanding about people's behavior through chats which can be further useful for sentiment analysis later.

1.1 Existing System

There are some basic applications available which can do basic analysis of WhatsApp chats by getting the input data from export chat feature of WhatsApp, but these apps are not so much popular due to limited analysis power, some third-party apps required so much of permissions of devices and trust issues and can't use for sentiment analysis. WhatsApp has in build end to end encryption features. So, we can't add any features without considering the security. WhatsApp has not such features available for their end users till now. The scope of the existing systems as followed.

1.2 Scope of the Existing System

The existing systems available till now is to perform only statistical analysis on the data given by the WhatsApp using the export data of any particular chat for a contact. But we are in the time of artificial intelligence and there is no automation available for this particular analysis. We can't do any kind of prediction about what the group member wants to tell without manually reading the whole WhatsApp chats which are very time consuming as well as hard working process. As mentioned earlier sentiment analysis is a very important step of evolving in digital communication regarding messaging applications like WhatsApp. As WhatsApp is used by millions of customers in a group communication, any kind of conceptual problems and hate speeches can create so much social issues. Sometimes they are using different languages while chatting and it is very crucial to understand what language they are talking about. Certainly, systems with intellectual abilities are urgent to society to solve this problem automatically.

Investigation organizations may have such trustworthy applications but they are used for very special purposes and not on demand of any particular individuals. We can not go directly to police station each time we came across any small conflict happens due to silly jokes or discussion on nonsense topics in any community group in WhatsApp. As well as we cannot share such confidential chats to some persons with data analytics skills for manual analysis or to some third-party applications where we don't know which kind of operations they are doing on our information.

They may not trustworthy and if we will do that there is no use of end-to-end encryption. Our proposed system can solve this problem preciously.

1.3 Proposed System

In our proposed system we will include two main features

- First of all, the initial part of the project is to understand implementation and usage of various python- built modules so that the input data can be formatted for future use and statistical analysis will be done using Numpy, Pandas modules.

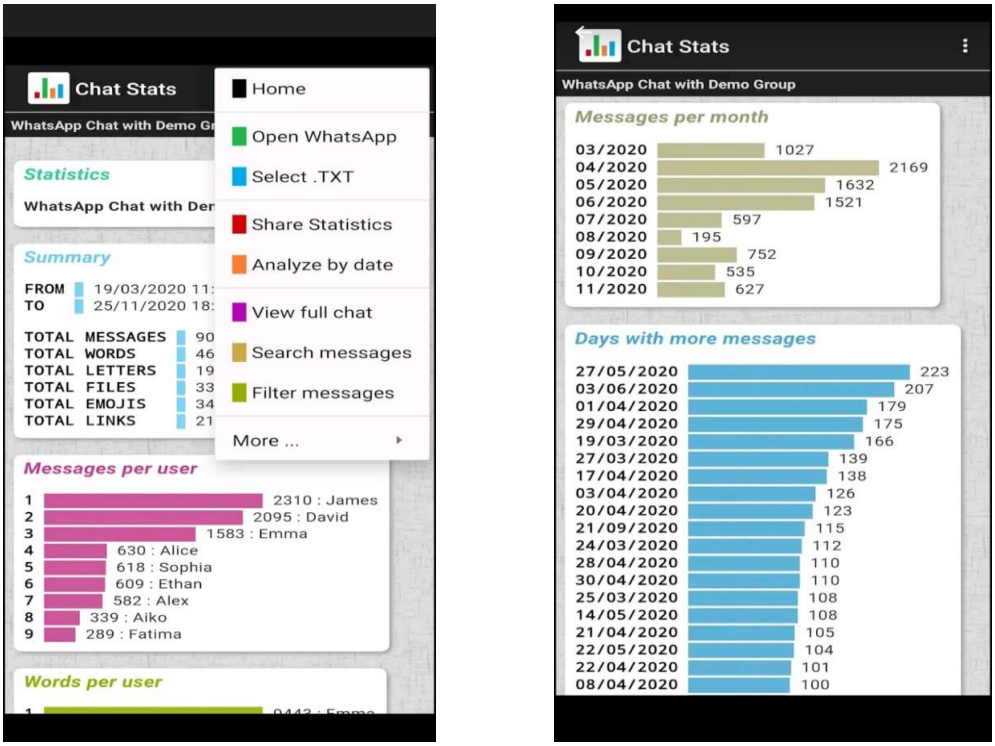

Fig:1.1 Demo Chat Analysis

- The above picture demonstrates the analysis made in one what's app group for a period of time i.e 19-03-2020 to 23-11-2020. From the picture we observed that total of 8446 messages are shared in the group by 9 different people, among which James shared 2310, David shared 2095, Brownie shared 1583, Alice, Sophie, Ethan, Alex, Alko, Fatima accounts for 630,618,609,582,339,289 respectively. Further it also represents the number of words , letters, files, emojis, links shared in the group. In the similar way our

project will be depicting the sentimental analysis of a WhatsApp group in a graphical manner using naive bayes algorithm.By representing the outcome in the above manner one can easily understand the various statistical views of our application.

- In the next part, which is an exploratory data analysis part we will apply sentiment analysis algorithm based on naïve bias classification and advanced python modules which will differentiate messages into positive, negative and neutral labels so that we can understand the sentiment of the chat based on some mathematical value. We are building a language model based on naïve bayes classification to check the languages in the chat as well.

1.4 Novelty of Proposed System

Novelty of our proposed system are as followed

- Our proposed system works in such a way that any WhatsApp user can use and enjoy the power of analysis and they can get their required result in short amount of time.
- No such systems are implemented till now for society that can do both statistical and sentimental analysis for end users of WhatsApp and WhatsApp has not included any such features in it as well.
- Some of the systems which are specially designed for such analytical purpose of WhatsApp are usually limited to big organization and government agencies. Our system will make this available to all WhatsApp users, explicitly for the group admins having number of WhatsApp group with lot of daily conversation. So, they no need to depend on any kind of lawful actions for small problems.
- In case of countries like India there are people with different languages and cultures. For a group with different linguistic people, sometimes we face some messages with different languages. Our project can solve this by identifying people and labeling their languages using language detection model and make it east for group admins to do their work.

2.REQUIREMENT ANALYSIS

2.1 Functional Requirements

- There has to be a proper administrative interface to collect the WhatsApp data
- Proper Datasets should be available as training datasets for Language detection
- The system used the language datasets and sentimental analysis dataset for training the different models using Naïve Bayes Classifier.
- Imported chat data must be cleaned to remove inconsistent data and perform statistical analysis on it.
- After this the system will use the language model to detect the language used by the user.
- Later the data must be processed by sentimental analysis to give sentimental data outcomes.
- After Extracting all the outcomes from the analysis done a proper visual representation format should be available to show which is human understandable.

2.2 Non Functional Requirements

2.2.1 User Interface and Human Factors

User interface is actual blueprint and human factors is the study of human behavior as users interact with the physical world / with a system. The user interface must be simple and easy to understand. Admin should have the permissions to use that particular WhatsApp Data available and he needs to give the input data to the application interface from WhatsApp application built in export chat data The outcome of the input chat will be organized and diagrametical representation with the help of graphs increases readability and make user familiar to the software in less time.

Human Factors includes job factors,timeline,organizational. User have to input the data and the outcome will be visible to the user directly. The user interface for our system is very easy. User has only one work in user interface that ios inputting

WhatsApp data. The admin have the direct access to the code and he/she can update the training data for better accuracy.

2.2.2 Software Requirements

- Pandas
- streamlit
- Matplotlib
- numPy
- Seaborn
- Sklearn
- Wordcloud
- NLTK libraries
- Pycharm IDE
- Google colab

2.2.3 Hardware Requirements

- Windows: Windows 7/8/10
- Processor (CPU): Intel Core i3/i5 (5th generation or newer)
- Memory: Minimum 8GB RAM'

2.2.4 Usability

Sentiment analysis applications are the future of our society, as we are mainly focusing on artificial intelligenece especially natural language processing for automations. Human language understanding is a great topic of machine learning and have a better future. Our applications have usuabilities like

- This can be used to detect information such as most active user,most active time.
- Helpful in detecting hate speech and sentiments of the group members with their messages.
- This can be used for language detection as well.

- The report should have proper visualization graphs to understand in short amount of time.

2.2.5 Reliability

Reliability is a most important concept while considering the software development. Softwares should be trustworthy his particular application will be more secure and reliable because if the application is not trustworthy who will like to use that. The application will fail before implementation. The reliabilities of our applications are as followed.

- The person who is using giving the WhatsApp data has the ownership on the data.
- The data analysis is directly showed to that individual, nothing is stored in any database, third party servers or cloud.
- The application never stores the data locally also. So none other person can steal the information.

2.2.6 Performance

Performance makes one application successful. It will be more scalable and performance of the system will be more because

- We are using textual data which can be analyzed easily.
- Large amount of data can be imported in very short time so scalable.
- Any features relative to our system can be included in future
- It will never be a bulky application so execution speed will be more.

2.2.7 Supportability

The application will be able to process large amount of data in order to provide accurate results in terms of analyzing the goals of the group chat and classifying it as positive,negative or neutral.Cost of Implementation is very less. No need to have greater resources for execution. We can use this application in mobile devices as well.

2.2.8 Physical Environment

Now-a-days more number of people are using WhatsApp. It becomes the main messeging app. Getting the requiered data for analysis is easy for this system. The application also can be used in mobile devices and desktop systems as well. It requires a normal room temperature with proper power supply to the system and a well functioning system to run the software.

2.2.9 Security Requirements

In order to export the chat the user must be a member in the group and so no other person outside the group can export it. The in build end-to-end encryption of whatsapp is also can be added as security requirements as we are getting the data set which is confidential. We are doing the operation locally without using any third party application which means the security requirement of the system is preserved

2.2.10 Resource Requirements

Very less resources are required in this project. A single system with the mentioned hardware and software requirements is enough for this project.

Sentiment analysis Dataset:

# Index	review Stemmed and Lemmatized review using nltk	polarity Polarity score generated using TextBlob	division Categorical label generated using polarity score
0 4156	good 4% good product 2% Other (3868) 95%	-1 1	positive 75% neutral 13% Other (482) 12%
3870	able play youtube alexa	0.5	positive
62	able recognize indian accent really well drop function helpful call device talk person near device s...	0.2794	positive
487	absolute smart device amazon connect external sub woofer sound amaze recons voice even close room li	0.1827	positive

Fig 2.1: Sentiment Analysis Dataset

Source: https://www.kaggle.com/datasets/pradeeshprabhakar/preprocessed-dataset-sentiment-analysis

Language Detection Dataset: https://www.kaggle.com/datasets/basilb2s/language-detection

Text	Language
Text Details	Language
10267 unique values	English 13% French 10% Other (7938) 77%
Nature, in the broadest sense, is the natural, physical, material world or universe.	English
"Nature" can refer to the phenomena of the physical world, and also to life in general.	English
The study of nature is a large, if not the only, part of science.	English
Although humans are	English

Fig 2.2: Language Dataset

WhatsApp Chat Analysis Dataset:

https://www.kaggle.com/datasets/sarthaknautiyal/whatsappsample

```
25/6/15, 1:42:12 AM: Vishnu Gaud created this group
25/6/15, 1:42:12 AM: You were added
18/12/16, 1:57:38 AM: Shahain: <image omitted>
21/12/16, 9:54:46 PM: Pankaj Sinha: <image omitted>
21/12/16, 9:57:45 PM: Shahain: Wow
21/12/16, 10:48:51 PM: Sakshi: <image omitted>
21/12/16, 10:49:00 PM: Sakshi: <image omitted>
21/12/16, 10:50:12 PM: Neha Wipro: Awsum😃😃👍🏻👍🏻
21/12/16, 10:51:21 PM: Sakshi: 🙇🏿
21/12/16, 10:57:01 PM: Ganguly: 🙂🙂👍🏻👍🏻
21/12/16, 11:28:51 PM: Vishnu Gaud: Waste out of wealth 😂
21/12/16, 11:48:42 PM: Venu Wipro: Fancy dress competition?
22/12/16, 12:08:04 AM: Kushbhu: 😂😂😂
22/12/16, 12:24:00 AM: Messages you send to this group are now secured with end-to-end enc
22/12/16, 12:25:02 AM: Nauty's phone: Superrrrrrr se bhiiiiiiii uperrrrrrrr
22/12/16, 12:36:54 AM: Sakshi: We were Divided into four groups..

Each group had to use newspapers only for dressing up one of their team members
22/12/16, 6:27:38 AM: Preeti: 👌👌👌👌
22/12/16, 12:10:31 PM: Kushbhu: Dunia ka sbse khatarnak proposal.. Must watch 😂😂😂😂😂
22/12/16. 12:10:45 PM: Kushbhu: <video omitted>
```

Fig 2.3: WhatsApp Dataset

3. SYSTEM ANALYSIS

3.1 Introduction

System analysis is aa process by which individual (s) studies a system such that an information system can be analyzed, modeled, and a logical alternative can be chosen. Systems analysis projects are initiated for three reasons: problems, opportunities, and directives. The people involved include systems analysts, sponsors, and users. Here in case of our project the system analysis is very easy. We will be more focusing on the use cases and structure of the system. System analysis represents the first diagrammatic representation for creation of any project. The use cases of our project are as followed.

3.2 Use cases

A use case is a methodology used in system analysis to identify, clarify and organize system requirements. The use case is made up of a set of possible sequences of interactions between systems and users in a particular environment and related to a particular goal. The method creates a document that describes all the steps taken by a user to complete an activity. In our project the working is based on the WhatsApp data so the use cases can be data extraction to the giving all kind of results, all functionalities. To describe use cases, we need 2 main components which are actors and use case. Brief explanations about all the components are as followed.

3.2.1 Actors

An Actor models a type of role played by an entity that interacts with the subject (e.g., by exchanging signals and data), but which is external to the subject. Actors may represent roles played by human users, external hardware, or other subjects. Actors do not necessarily represent specific physical entities but merely particular facets (i.e., “roles”) of some entities that are relevant to the specification of its associated use cases. A single physical instance may play the role of several different actors and a given actor may be played by multiple different instances.

Here our project can be described using mainly three actors, those are admin user and system.

Admin actor is administrator of the system who collects the datasets and upload to the system for creating all kind of models for sentiment analysis and language detection.

User actor refers to the end user who need some analysis from the application, so he/she decides to feed the system with proper format WhatsApp data to get the result.

System is another actor which is doing all kind of internal working for the chat analysis applying algorithm and build in methods and giving the output.Each actors have some use cases relevant to the projects. Those use cases are as followed.

3.2.2 List of Use Cases

A list of use cases is given in a tabular format for three actors.

Table 3.1: Use Cases of Admin

Use Cases	Description
Load dataset	Admin has use case as loading dataset to the system for analysis, datasets are of two types, so this use case has connection to sub use cases as followed. Sentiment dataset and language dataset includes here.
Sentiment dataset	Admin gives separate dataset for sentimental analysis for each user.
Language dataset	Admin gives a separate dataset for language detection of chats.
Performance analysis	Admin needs to check the performance and accuracy of the prepared detection models.

Table 3.2: Use Cases of actor user

Use Cases	Description
Input WhatsApp data	User has to input real time WhatsApp chat data into the system for further analysis for that particular chat data.
Getting result	User gets the report after analysis is over.

Table 3.3: Use Cases of actor system

Use cases	Description
WhatsApp Chat preprocessing	After getting the chats from the user system has to preprocess the chats in a proper format for further operation.

Building model	Sentiment analysis model and language detection model using the Naïve Bayes algorithm is build during the processing in the system.
Statistical analysis	System does statistical analysis about the given data.
Language detection	Language Detection for chats is also available in the system. It is applied on Chat of a person.
Sentimental analysis	System does sentimental analysis about the given data for users.
Showing result	Result is given to the user by the system.

3.3 Use Case Diagram

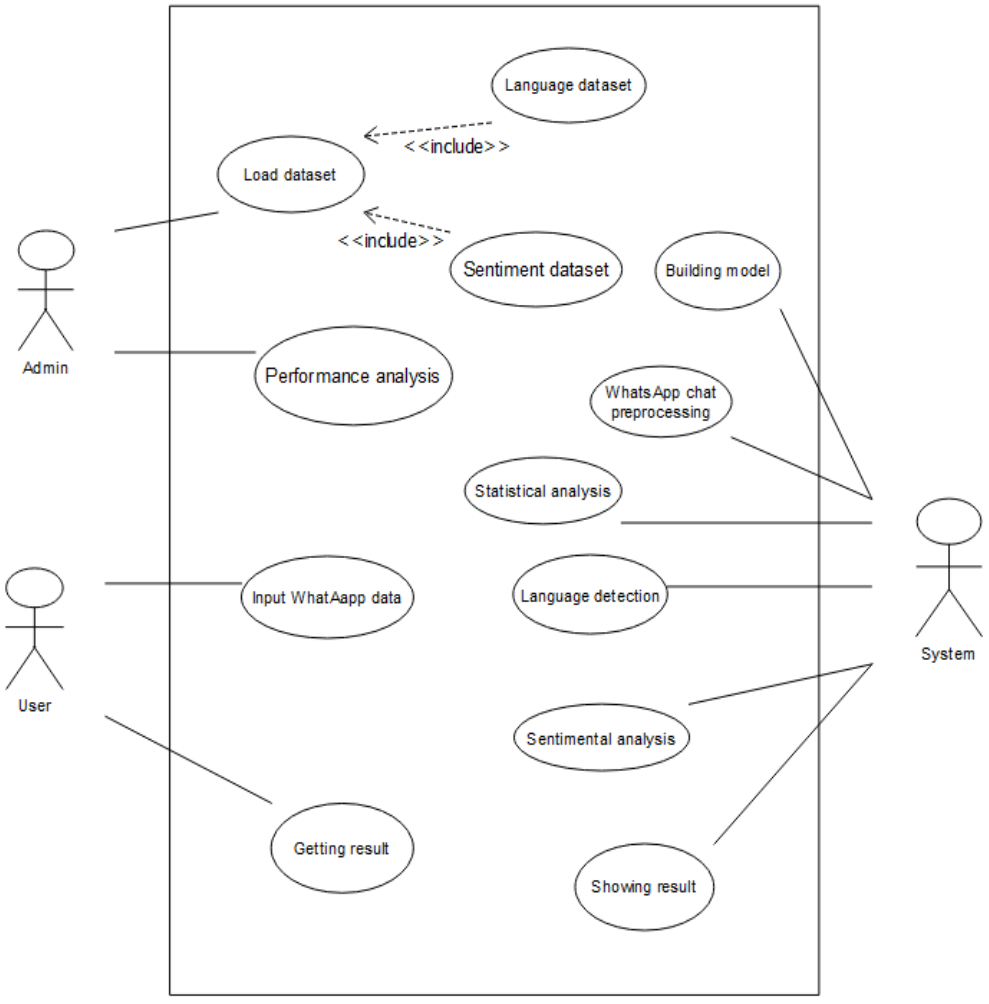

Fig 3.1: Use Case Diagram

4. SYSTEM DESIGN

4.1 Introduction

System Design is the process of designing the architecture, components, and interfaces for a system so that it meets the end-user requirements. Our project is based on WhatsApp Data Chat Analysis and in that we are doing both statistical analysis and sentimental analysis both together. We have added language detection feature to extend the actions of our application. Our Project is based on Natural Language Processing. Natural Language Processing is a part of machine learning which mainly focuses on human language, apply algorithms on that and come out with information from that input. Here we are using Naïve Bayes Classification algorithm for creating this project. We have used 3 datasets for building this project. So, language dataset and sentimental datasets are used for building the detection models and real time WhatsApp data is used for analysis of that particular data. The whole process is done using python modules like sklearn, pandas, numPy, matplotlib and many more. The detailed system design is as followed.

4.2 System Architecture

A system architecture is a representation of a system in which there is a mapping of functionality onto hardware and software components, a mapping of the software architecture onto the hardware architecture, and human interaction with these components. In case of this system, we are using a very easy system architecture. The architecture is as followed.

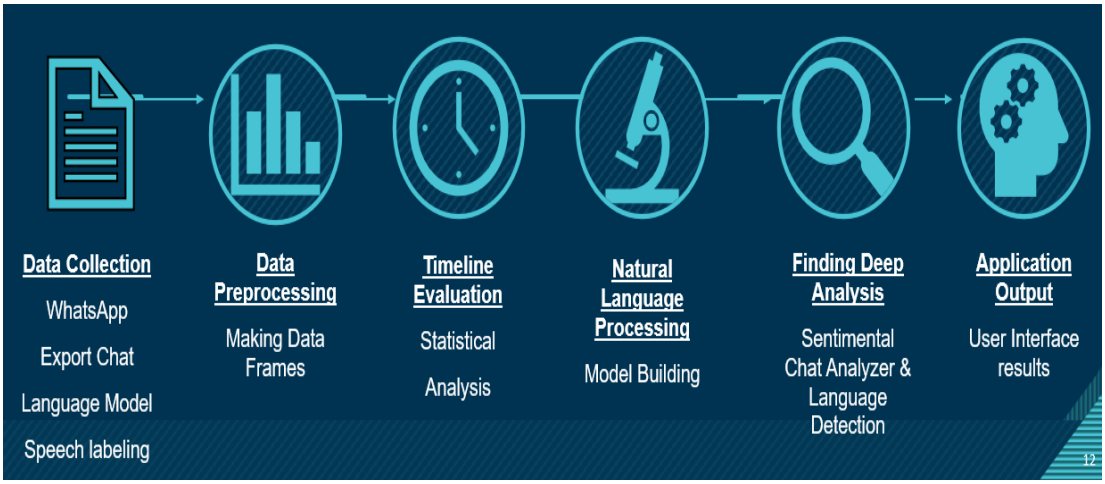

Fig 4.1: System Architecture

Here as given in the above diagram the system has to take input from outside, and in such case, there are mainly three kind of inputs which are WhatsApp text data for analysis, sentiment and language detection training data for training the detection model for doing the future operation.

In the second step the Data (WhatsApp Data) which is actually a real time data is to be preprocessed to form predefined data frames so the system can do further operations on it. There will be three kind of operations which are statistical analysis, sentiment analysis and language detection.

Timeline evaluation basically refers to the statistical analysis using build in modules like pandas and matplotlib.

In case of natural language processing a chat sentiment detection model and a language detection model is built. In the next step based on the model deep analysis is done on WhatsApp data and later it is shown to the user.

4.3 System Object Model

4.3.1 Introduction

It serves as an object-oriented model that can be distinguished from other models contained in object-oriented programming languages. SOM basically includes an interface definition language, a runtime environment with procedure calls and a set of enabling frameworks. In our system object model we are preparing our model using Naïve Bayes Classification algorithm. We have compared the WhatsApp text data with some regular expression and prepared data frames from it. Later in the case we are using transforms text to feature vectors that can be used as input to estimator. So that the sentiment analyzer and language detector models can use this input for further process.

4.3.2 Subsystems

Here our project has mainly three purposes which are Chat statistical analysis, Sentimental Analysis and Language detection. So, the three subsystems here are Statistical Analyze, Sentiment Analyze and Language Detection. For sentiment analyze mainly we are using vectorization. We are getting all kind of output a report with all kind of analysis results and visualizations graphs. A pictorial representation of subsystems as followed.

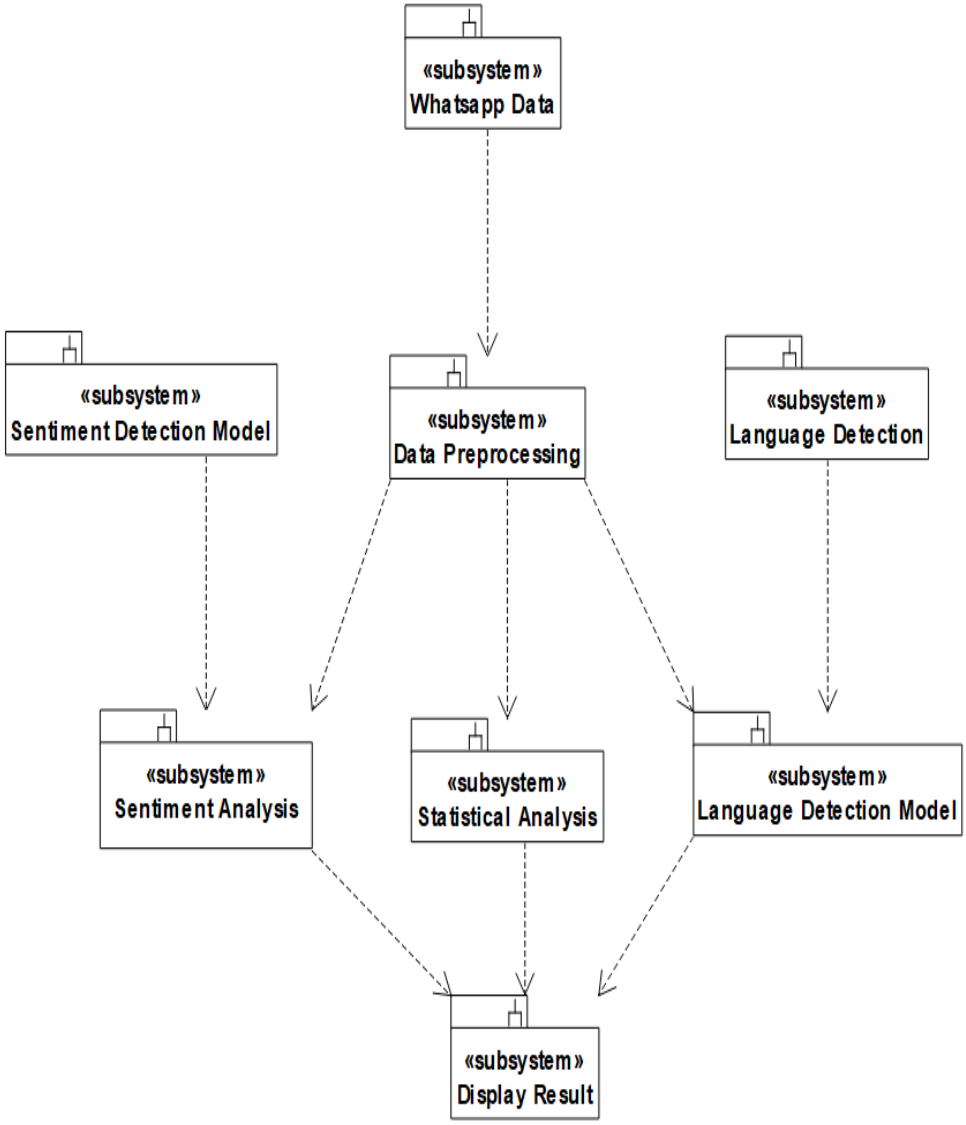

Fig 4.2: Subsystems

4.4 Object Description:

4.4.1 Objects

Objects play a major role in system design for any kind of software. Object basically refers to the runtime entity of class variable with proper functionalities and properties. Properties refer to the attributes and functionalities are refered to the operations. Objects have relations among themselves sometimes for their combinatorial working. Our WhatsApp Group Chat Analyzer and prediction application have relative objects for class variables for creating the whole system. The objects available are belongs to the classes which are as followed.

1. **Data processing** – object containing all the functionalities and attribute to work with data given by the user
2. **Statistics of chat –** functionalities required for all kind of statistical analysis with built in python modules are available here in this particular object
3. **Sentiment model –** It contains attributes like normalization vectors required for sentiment analysis and required datasets
4. **Language model –** It contains attributes like datasets required for language detection with available datasets for training and testing
5. **Sentiment of chat –** contains group of operations for sentimental analysis
6. **Language of chat –** contains group of operations for language detection
7. **Display Report –** showing a proper visualization about the whole analysis.

These objects which are actually the name of the particular classes are described below with a class diagram.

4.4.2 Class Diagram

A class diagram relative to our proposed system is as followed. Here the required classes are mentioned with the proper functionalities and attribute which should be available for building the total system. The diagram is very close to the coding implementation as the operations we are added for each class are the actual functions for the object in our software program. The class objects are described before in brief.

Classes are as followed

- Data Preprocessing
- Statistics Of Chats
- Language Model
- Sentiment Model
- Sentiment Of Chat
- Interface Report

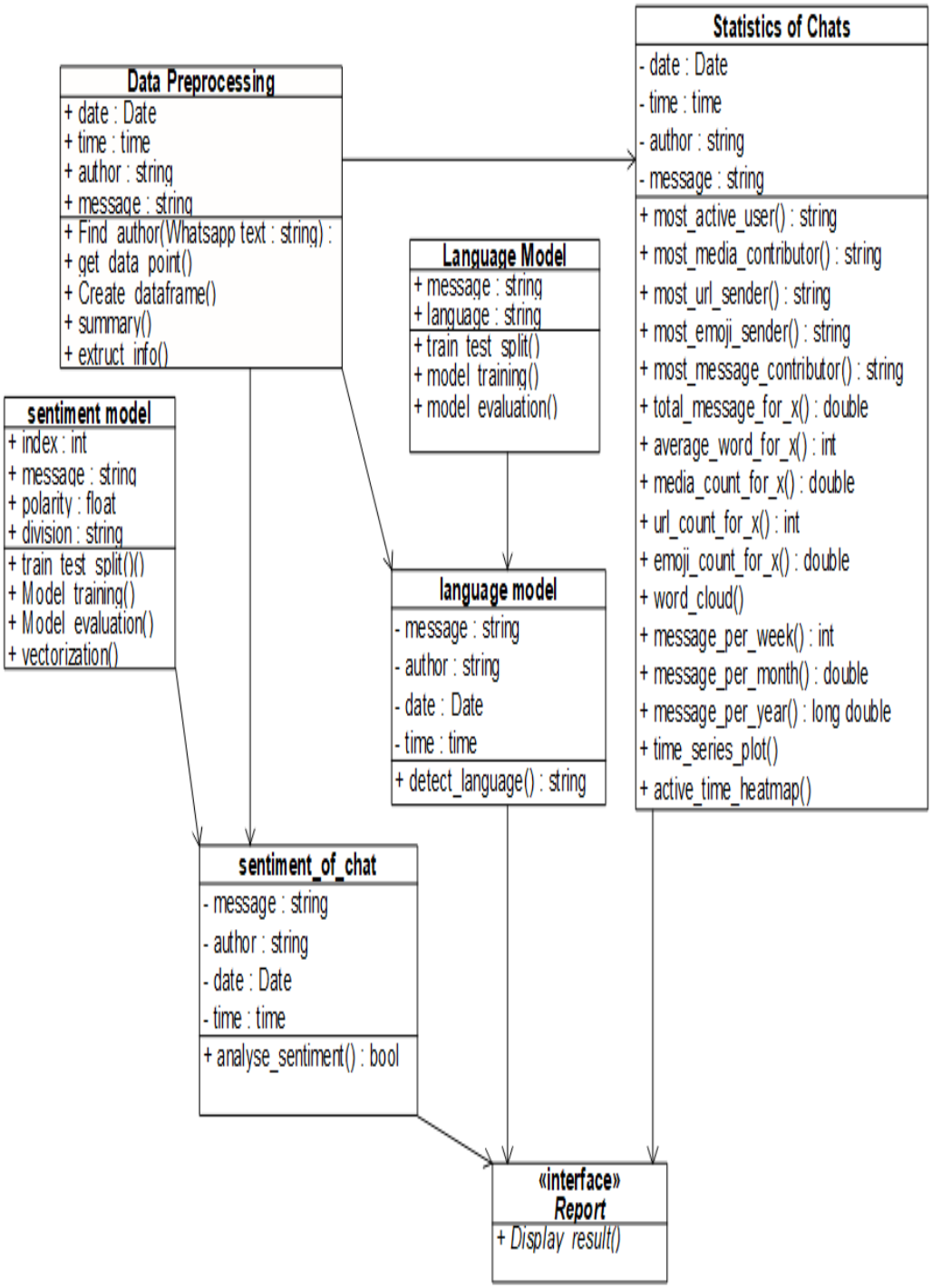

Fig 4.3: Class Diagram

4.5 Dynamic Model

The dynamic model is used to express and model the behavior of the system over time. The proposed system has been expressed by two dynamic model diagrams which are sequence diagram and activity diagram. The representation of both of them are as followed.

4.5.1 Sequence Diagrams

Sequence diagrams are used to display the interaction between users, screens, objects and entities within the system. It provides a sequential map of message passing between objects over time.

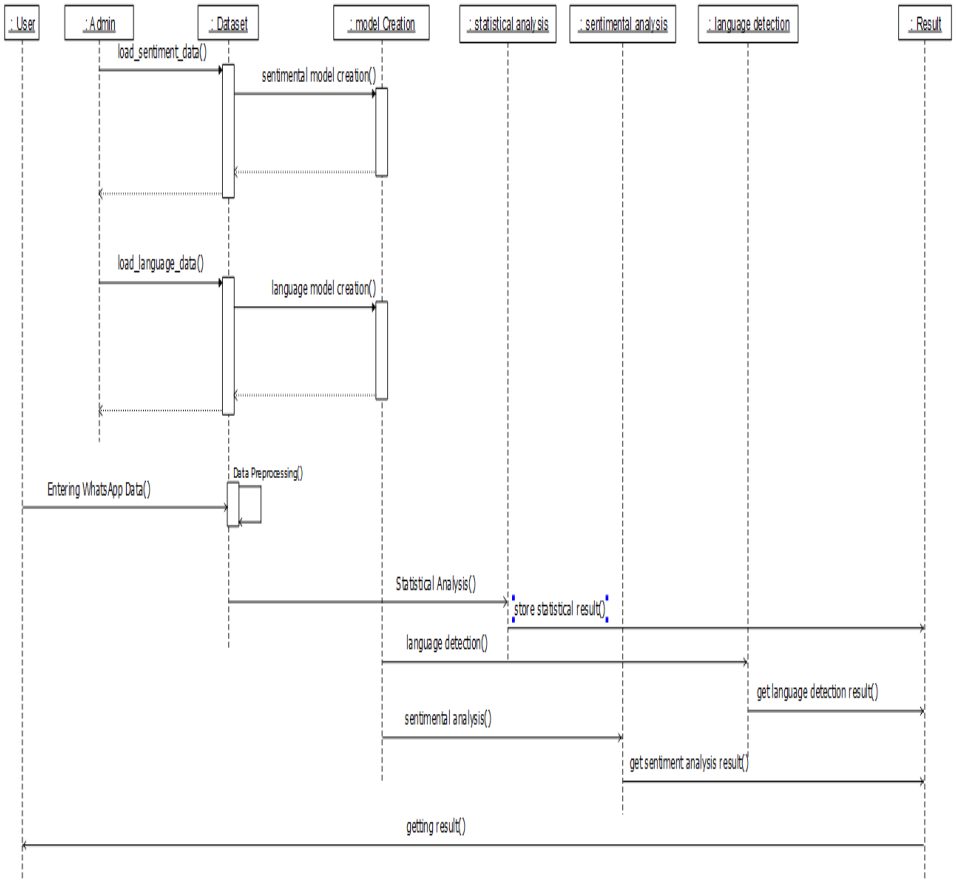

Fig 4.4: Sequence Diagram

As represented in the above sequence diagrams the objects are user, admin, dataset, model creation, statistical analysis, sentimental analysis, language detection and lastly the result.

In the first two dynamic processes the admin has to give sentimental analysis training dataset followed by language detection training dataset for creating the detection model for each individually. It should be synchronous signal as the model is created synchronously and we can update the dataset to update the model.

Coming to the working of the user, the user is giving WhatsApp data to the user for getting the output and after that all the processing is done by the system itself.

The system is taking doing statistical Analysis on the data followed by sentimental analysis and language detection using the created models and lastly the result is shown to the user.

4.5.2 Activity Diagram

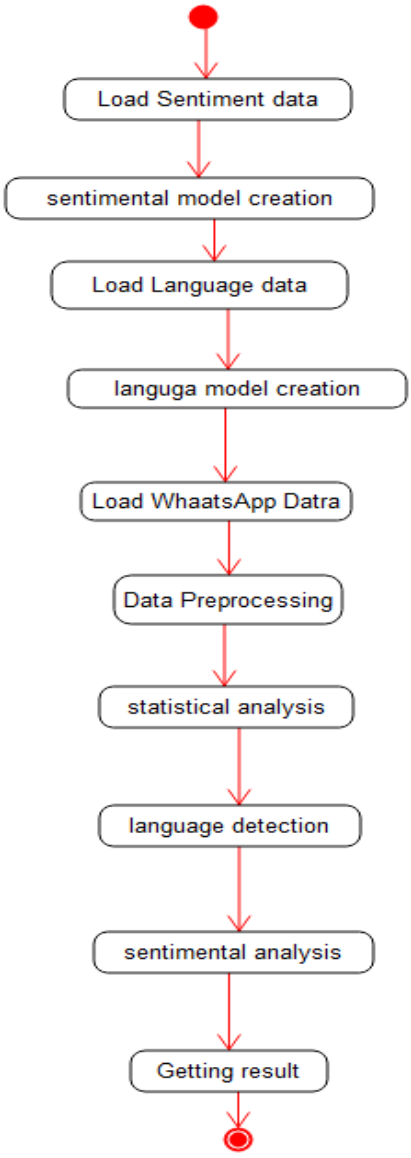

Fig 4.5: Activity Diagram

The above is the activity diagram of WhatsApp chat analysis and prediction application. It is the next step after the sequence diagram. Activity diagrams are used to show how different workflows in the system are constructed, how they start and the possibly many decision paths that can be taken from start to finish. The activity diagram is defined in a proper way showing the staring to the ending activities for the system So the model creation is easy. The datasets collected are one after another, firstly sentimental analysis training dataset followed by language labelling data and lasting the WhatsApp chats. The application terminates after showing the result to the user.

4.6 Static Model

4.6.1 Component Diagram

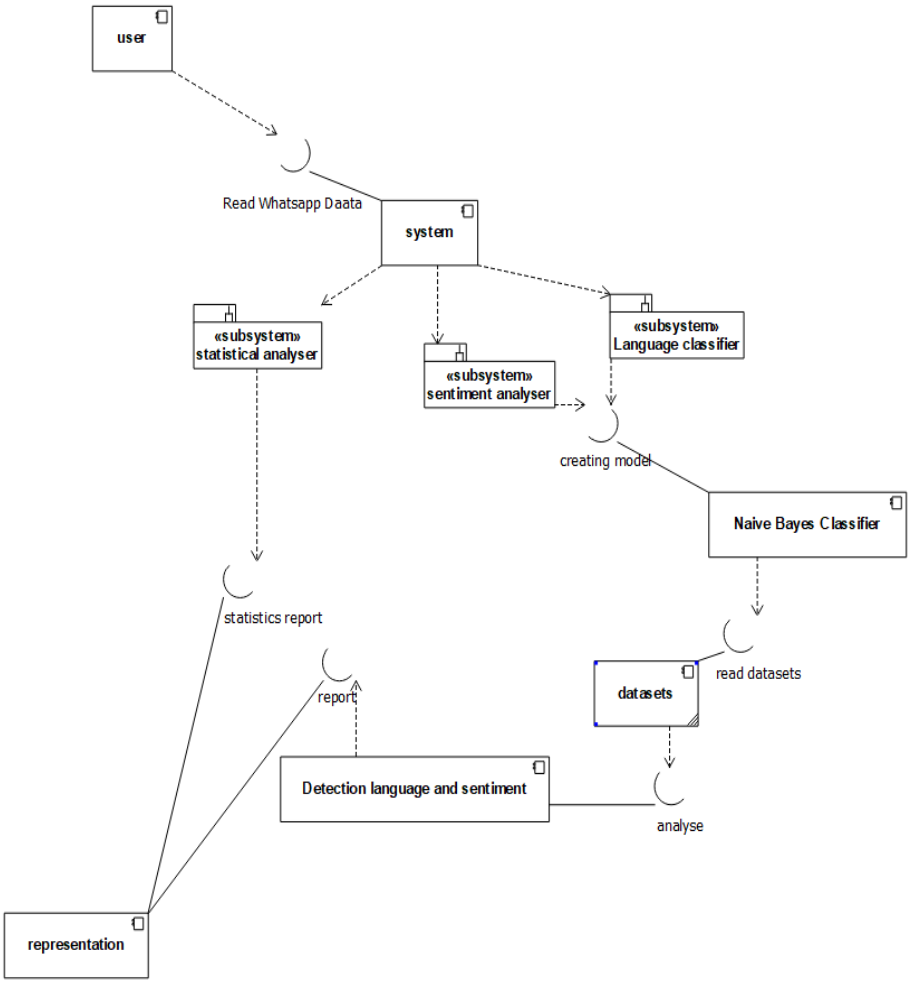

Fig 4.6: Component Diagram

Component diagrams are used to describe the physical artifacts of a system. This artifact includes files, executables, libraries, etc. In case of the proposed system

the components are user, system, naïve bayes classifier, datasets, detection language and sentiment and lastly representation. System has three subsystem as statistical analyzer, sentiment analyzer and language classifier. All the systems are connected with proper interfaces and they are named to denote the working.

4.6.2 Deployment Diagram

Deployment diagrams are used to visualize the hardware processors/ nodes/ devices of a system, the links of communication between them and the placement of software files on that hardware. The above describes the deployment diagram with the nodes as user system and the report. Deployment diagrams denotes the system is very easy to implement.

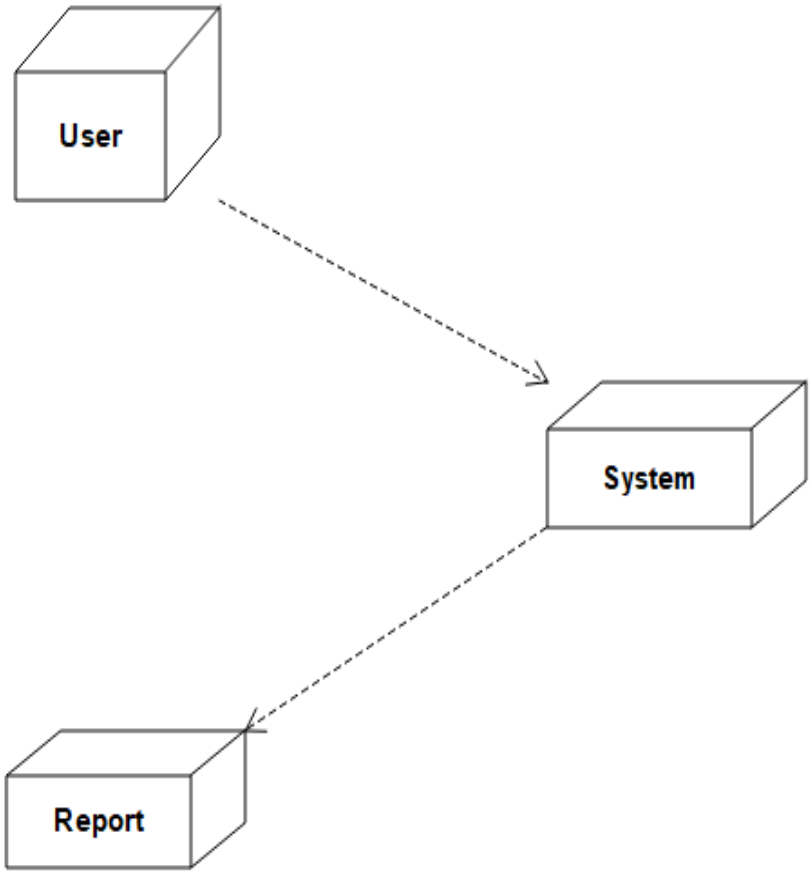

Fig 4.7: Deployment Diagram

CONCLUSION

In conclusion about our project, we can express that, there is not an existing system available with WhatsApp data sentiment analysis for end users. Conflicts in WhatsApp group conversations may create more social problems. Peoples are using delivering hate speeches and fake information to create panic in society through WhatsApp. People can use WhatsApp freely in developing countries so more security concerns is to be present for our society regarding this app. The used algorithms in the proposed system are very easy to implement and no need to have large infrastructures to implement this. We have implemented same Naïve bayes classification algorithm for creating both the models. Further any user can get a nice interaction with system with user friendly interfaces. After successful implementation of the system, it will help a large number of group admins who have so much daily work with WhatsApp chats. In future we can add other advanced features to the system as well. We hope this will solve the existing problems in certain extend and will express the power of artificial intelligence in society.

BIBLIOGRAPHY

[1] Dahiya, S., Mohta, A., & Jain, A. (2020, June). Text classification based behavioural analysis of whatsapp chats. In 2020 5th international conference on communication and electronics systems (ICCES) (pp. 717-724). IEEE.

[2] Mohta, A., Jain, A., Saluja, A., & Dahiya, S. (2020, October). Pre-Processing and Emoji Classification of WhatsApp Chats for Sentiment Analysis. In 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 514-519). IEEE.

[3] Kanakia, H., Raundale, P., Britto, R., & Sawardekar, R. (2019, May). Analysis of Social Networks using Naive Bayes. In 2019 International Conference on Intelligent Computing and Control Systems (ICCS) (pp. 88-91). IEEE.

[4] Shafiq, M., Yu, X., & Laghari, A. A. (2016, September). WeChat text messages service flow traffic classification using machine learning technique. In 2016 6th International Conference on IT Convergence and Security (ICITCS) (pp. 1-5). IEEE.

[5] Liu, B., Blasch, E., Chen, Y., Shen, D., & Chen, G. (2013, October). Scalable sentiment classification for big data analysis using naive bayes classifier. In 2013 IEEE international conference on big data (pp. 99-104). IEEE.