A logo of a company

AI-generated content may be incorrect.

National Textile University

**Department of Computer Science**

A PROJECT REPORT ON

**Whatsapp App Analyzer**

**-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

**-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------**

**Submitted By**

**Aalyan Riasat (23-ntu-cs-1121) Abid Ali (23-ntu-cs-1124) Ali Hussnain (23-ntu-cs-1134)**

**Submitted To**  **Sir Waqar Ahmad**

**Subject Introduction To AI**

**ABSTRACT**

The most used and efficient method of communication in recent times is an application called WhatsApp. WhatsApp chats consist of various kinds of conversations held among groups of people. This chat consists of various topics. This information can provide lots of data for the latest technologies such as machine learning. The most important thing for machine learning models is to provide the right learning experience which is indirectly affected by the data that we provide to the model. This tool aims to provide in depth analysis of this data which is provided by WhatsApp. Irrespective of whichever topic the conversation is based on our developed code can be applied to obtain a better understanding of the data. The advantage of this tool is that it is implemented using simple python modules such as pandas, matplotlib, seaborn and sentiment analysis which are used to create data frames and plot different graphs, where then it is displayed in the flutter application which is efficient and less resources consuming algorithm, therefore it can be easily applied to largest dataset.

1. **INTRODUCTION**

**INTRODUCTION**

* 1. **Introduction**

This tool is based on data analysis and processing. The first step in implementing a machine learning algorithm is to understand the right learning experience from which the model starts improving on. Data pre-processing plays a major role when it comes to machine learning. To make the model more efficient we need lots of data, so we turned our focus primarily on one of the largescale data producers owned by Facebook which is nothing but WhatsApp. WhatsApp claims that nearly 55 billion messages are sent each day. The average user spends 195 minutes per week on WhatsApp and is a member of plenty of groups. With this treasure house of data right under our very noses, it is but imperative that we embark on a mission to gain insights into the messages which our phones are forced to bear witness.

* 1. **Problem Statement**

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

* 1. **Existing Systems**
* **Chat Stats**
* **Whatsanalyze**
* **Chatilyzer**
* **Chat analyzer**
  1. **Proposed system**

Data pre-processing, the initial part of the project is to understand implementation and usage of various python-built modules. The above process helps us to understand why different modules are helpful rather than implementing those functions from scratch by the developer. These various modules provide better code representation and user understandability. The following libraries are used such as NumPy, scipy pandas, csv, sklearn, matplotlib, sys, re, emoji, nltk seaborn etc. Consider the impact of journey time on fare prediction. Longer or shorter flights might have different pricing dynamics, which could be factored into the model to enhance accuracy.

Exploratory data analysis, the first step in applying a sentiment analysis algorithm which provides positive negative and neutral part of th chat and is used to plot pie chart based on these parameters. To plot a line graph which shows author and message count of each date, to plot a line graph which shows author and message count of each author, ordered graph of date vs message count, media sent by authors and their count, Display the message, which is di not have authors, plot graph of hour vs message count.

* 1. **Objectives**
* This project aims to provide a better understanding of various types of chats. This analysis proves to be better input to machine learning models which essentially explore the chat data. It requires proper learning instances which provide better accuracy for these models. Our project ensures to prove an in-depth exploratory data analysis on various types of WhatsApp chats.
* Sentiment Analysis: Determine the overall sentiment of the conversations, whether they are positive, negative, or neutral. This can help in understanding the emotional tone of the discussions.
* Topic Modeling: Identify and categorize the main topics or themes discussed in conversations. This can be achieved through techniques such as topic modeling algorithms (e.g., Latent Dirichlet Allocation).
* User Behavior Analysis: Analyze the behavior of individual users within the chat, including the frequency of messages, response times, and participation levels. This can provide insights into user engagement.
* Keyword Extraction: Extract relevant keywords or phrases that are frequently used in conversations. This can help in identifying key topics or trends.
* Anomaly Detection: Identify unusual patterns or outliers in the chat data. This can be useful for detecting potential issues or abnormal behavior within the group.
* Named Entity Recognition (NER): Recognize and classify named entities such as people, locations, organizations, and dates mentioned in the conversations. This can assist in understanding the context of the discussions.
* Language Understanding: Develop a model that understands the language nuances, abbreviations, and emojis commonly used in WhatsApp chats to improve the accuracy of analysis.
* Time Series Analysis: Explore how conversation patterns change over time, identifying peak activity periods, and understanding the dynamics of the group or individual interactions.
* User Profiling: Create profiles for individual users based on their language use, topics of interest, and overall behavior in the chat. This can be valuable for personalized insights.
* Privacy Considerations: Implement measures to respect privacy and ensure that sensitive information is not exposed during the analysis process.

**Diagrammatic representation of the “Project Work Flow”**

**A diagram of a machine learning model

AI-generated content may be incorrect.**

# 

# REQUIREMENT ANALYSIS

* 1. **Software Requirement Specification**

**Requirement Specification:**

Conceptually every SRS should have the component:

* Functionality
* Performance
* Design Constraints imposed on
* Implementation External Interface
  1. **Use Case Model**
     + 1. In the use case diagram the actor is User.
       2. User can make use of chat upload use cases to give input to the system.
       3. Select time format use case describe that user can input the time for part of the file in the system select user use case is to select whose analysis result is desired.
       4. Users can make use of Show analysis use cases to see the result of the entire analysis done by the system.

A screenshot of a computer

AI-generated content may be incorrect.

* 1. **Class Diagram**

Description

The class diagram has the following two classes with their respective attribute and methods:

* Data Frame
  + Attributes : user, message, date, time, year, month, day, dayname, dayofweek, weeknum, hour, minute, meridian
  + Methods: separateDateTime
* Generate report
  + Attributes : selectedUser, message, dataFrame, timeFormat
  + Methods : fetch\_stats, chat\_form, most\_talkative, hourly\_timeline, daily\_timeline, weekly\_timeline, most\_busy\_day, most\_busy\_month, crete\_worldcloud

The class Dataframe is creating the class Generate report so Dataframe class include Generate report class.

A close-up of a document

AI-generated content may be incorrect.

#### Activity Diagram

* In the activity diagram as the initial activity starts user will upload the file as a input which is action and in the next action time format will be selected.
* The decision box check chat format represents the validity of the time format of the file.
* If the time format is correct then analysis will be done and process will end.
* If the time format is wrong user will have to again check for the correct format.

A diagram of a chat

AI-generated content may be incorrect.

# 

# SYSTEM MODELING

* 1. **Sequence Diagram**
* The Sequence diagram start with upload chat in front-end then check time format will be exe it will match time format of chat upload with time format user selected then it goes to server then server perform analysis operation and send back to result in user end.
* If time format of chat and user select time format not match it will display a invalid time format select error.

A diagram of a process flow

AI-generated content may be incorrect.

* 1. **Collaboration Diagram**
* This collaboration diagram shows the relationship between the objects in a system.
* An object consists of several features. Multiple objects present in the system are connected to each other.

A diagram of a data frame

AI-generated content may be incorrect.

**Figure: COLLABORATION DIAGRAM**

* 1. **Conceptual Level State Diagram**
* The State diagram start with the uploading of the file and after that in the next state time format will be selected if the time format is valid then in the next state analysis will be done. The analysis state will complete when the overall result will be shown on the user interface.
* In the analysis state the user can select the option of whose analysis he or she wants to see and this will give corresponding next state of display result.

A diagram with text and black text

AI-generated content may be incorrect.

**Figure: STATE CHAT**

* 1. **Conceptual Level Component Diagram**

Whatsapp chat analyzer has following component:

* Chat export which connected with other components via input file.
* The data of the input file will be accessed by following components:
  + Top stats
  + Message sent
  + Most busy
  + Most common word
  + Emoji analysis
  + Sentimental analysis

A diagram of a data flow

AI-generated content may be incorrect.

**Figure: Component Diagram**

* 1. **Deployment Diagram**
* In the deployment diagram it has two notes browser which is on the user end and server.
* Both the nodes will connect using http protocol.
* The browser node will certain user interface while all the operation such as managing and analysing will be done on the server node.
* The data generated will be sent on the browser node to user.

A diagram of a computer software

AI-generated content may be incorrect.

**Figure: Deployment Diagram**

* 1. **Implementation Phas****e**

**3.6.1 Implementation**

Python is a highly general purpose and a very popular programming language. Python programming language is being used in web development, machine learning application, along with all cutting-edge technology in the software industry. Python programming language is very well suited for beginners.

1. python is currently the most widely used multi-purpose, high level programming language.
2. Python allows programming in object oriented and procedural paradigm.
3. Python programs generally are smaller than other programming languages like Java. Programmers have to type a relatively less and indention requirement of the language makes them readable all the time.
4. Python language is being used by almost all technician companies like Google, Amazon, Facebook, Instagram, Dropbox, Uber…. Etc.
   * + 1. **Software requirements for developing application**
          1. Python
          2. VS code

**Technologies**

Python and its libraries (streamlit)

ML algorithm

NLTK

#### Testing

Testing is the major quality control that can be used during software development. Its basic function is to detect the errors in the software. During requirement analysis and design, the output is the document that is usually textual and non-executable. After the coding phase, a computer program is available that can be executed for testing process purposes.

#### Testing Objectives

* + - 1. To check if the application is working as expected.
      2. To check the errors of different scenarios by using different cases.

#### Testing Methods & Strategies used along with Test Data

**Software Testing Strategies :** Software testing is defined as an activity to check whether the actual results match the expected results and to ensure that the software system is defect free. It involves execution of a software component or system component to evaluate one or more properties of interest. Software testing also helps to identify errors, gaps or missing requirements in contrast to the actual requirements. It can be either done manually or using automated tools.

In simple terms, Software Testing means Verification of Application under Test (AUT).

1. Functional Testing
2. Non-Functional Testing

A diagram of software testing

AI-generated content may be incorrect.

1. **Functional Testing**

Functional testing is defined as a type of testing which verifies that each function of the software application operates in conformance with the requirement specification. This testing involves checking of User Interface, APIs, Database, security client or server application and functionality of the Application under Test. The testing can be done either manually or using automation.

1. **Non-Functional Testing**

Non-functional testing is defined as a type of software testing to check non- functional aspect of a software application. It is designed to test the readiness of a system as per non-functional parameters which are never addressed by functional testing. An excellent example of a non-functional test would be to check how many people can simultaneously login into a software. Non-functional testing is equally important as a functional testing and affects client satisfaction. Non-functional testing should increase usability, efficiency and maintainability.

# 

# CONCLUSION & FUTURE WORK

#### Conclusion

* 1. **Limitation of Project**
* Maximum file size to be uploaded is 200MB.
* Only supports English languages.
* Supports only txt extension.

#### Future Enhancement Suggestions

* Add multiple languages for analysis.

**Screenshots**

# 