WHATSAPP CHAT ANALYSIS

Abstract

The most used and efficient method of communication in recent times is an application called WhatsApp .You can use your WhatsApp data for many data science tasks like sentiment analysis, keyword extraction, named entity recognition, text analysis and several other natural language processing tasks.WhatsApp chat analyzer is the application deployed on heroku web which provide analysis of WhatsApp group. .There are various methodologies available for analysis but here matplotlib, streamlit, seaborn, re, pandas libraries of python and some concept of NLP is used. This is the combination of machine learning and NLP. This Whatsapp chat analyzer take import whatsapp chat file from user and analyze it and give different Visualizations as a result.

1. Introduction

In this report I have proposed a WhatsApp Chat Analyzer.WhatsApp Analyzer means we are analyzing our WhatsApp group activities. 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. In order 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.. The advantage of this application is that

it is Implemented by simple python librarieses like seaborn, pandas, numpy, streamlit and matplotlib which are Commonly use for creating data frames and different graphs.

1.1 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 to employ dataset manipulation techniques to have a better understanding of WhatsApp chat present in our phones.

1.2 Existing System

There is a lot of development in the current system. In the older version there was no feature to display status, there was no feature to share documents and there was no feature to share location. In the current version, all of these features are available. In older version we couldn't share images through doc's format. In this system user is able to access WhatsApp in windows through WhatsApp web application, which can be connected through QR code. There is another feature called export chat where user can send or share or get the chat detail for data analysis through email, Facebook or some messenger application.

1.3 Proposed System

The "WhatsApp Chat Analyzer" provides a platform to the user which enables user to analyze whatsapp chats Online on heroku link. This application allows user to browse whatsapp exported (.txt) file and

import it to WhatsApp chat analyzer and get analysis according to that txt file. And user can Analyze by clicking Show Analysis button.

Advantages of Whatsapp Chat Analyzer

- **Runs on all devices.**
- **Shows based on whatsapp chat file.**
- **Shows different visualizations.**
- **Total Messages.**
- **Total words.**
- **Media shared.**
- **Link shared.**
- **Monthly timeline.**
- **Most busy day.**
- **Weekly activity.**
- **Most busy users.**
- **Most used words**
- **Emoji analysis.**

1.4 Objective

In this decade the upcoming technologies are mainly dependent on data. This data can only be obtained if there is some research applied on the context of the requirements of the tool. Since a lot of machine learning enthusiasts develop models which helps solve multiple problems the requirements of appropriate data are very large scale this project aims to provide a better understanding towards

various types of chats. This analysis proves to be better input to machine learning models which essentially explore the chat data. These models require proper learning instances which provides better accuracy for these models. Our project ensures to provide an in-depth exploratory data analysis on various types of WhatsApp chats.

2. Methodologies Using Technical Thinking

Python

It is an interpreted, high-level general-purpose programming language. Created by Guido Van Rossum and first released in 1991. Its language constructs and objects-oriented approach aim to help programmer with clear, logical code for small and large-scale tools. Python is used for web development (server-side), software development, mathematics, it can be used alongside software to create workflows, it can connect to database systems, it can also read and modify files, it can be used to handle big data and perform complex mathematics and can be used for rapid prototyping, or for production-ready software development.

Pandas

This is an open-source Python libraries which is mainly used in Data Science and machine learning subjects. This library provides analysis tool for data manipulation, using its data structures this are used for analyzing Data for manipulating time series analysis and numerical data.

Numpy

Numpy can be name come from Numeric Python, it is a data analysis library for Python that contains various Numerical functions and methods for numerical analysis and also having multi-dimensional array objects and to Process these arrays contains collection of routines.

3. Analyze Methodologies based upon Measures or Performance

Matplotlib

Matplotlib is easy to use and an amazing visualizing library in Python. It is built on NumPy arrays and it work With the broader SciPy stack and consists of several plots like pie, line, bar, graph, scatter, histogram, etc. In this Project, Matplotlib is used for various visualizations for analysis of whatsapp chats. Visualizations like bar Charts, line charts, pie charts are used.

Seaborn

Seaborn is a library mostly used for statistical plotting in Python. To make statistical plots more attractive it Provides beautiful color palettes and default styles. In this project, Seaborn is used for heatmap visualization for Showing 24 hours with 7 day with different scale of color for getting hour with max to min message.

Sentiment Analysis

Sentiment analysis is a natural language processing technique that identifies the polarity of a given text.Sentiment analysis is the practice of using algorithms to classify various samples of related text into overall positive and negative categories. With NLTK, you can employ

these algorithms through powerful built-in machine learning operations to obtain insights from linguistic data.

NLP

In this project, Features of NLP are used like Parsing Text, Eliminating stop words and Analyzing Text. Parsing Text is used for splitting messages into words for analysis like total words and mostly used words. A file is used That contains all stop words which is given to the python program to show meaningful words only by Eliminating all stop words. Text analysis is used to identify how many media are shared, how many links are Shared.

Result Analysis

This project is created in python using sentiment analysis and deployed on heroku web.

Working of project:

1. **User go to sidebar and click on browse file .**
2. **Select whatsapp chat text file and import it for analysis .**
3. **User have choice for overall analysis or specific user analysis from whole group.**
4. **After selecting user, User click on show analysis button to analyze imported file.**
5. **It shows analysis of imported whatsapp text file.**
6. **User can see Total messages, words, media and link shared in the group.**

This is the result of project and how project is working.

Conclusion

In conclusion, it can be said that the capabilities of the WhatsApp application and the power of the python programming language in implementing whatever network data analysis intended, cannot be overemphasized. The major objective that has been decided in the initial phase of the requirement analysis is achieved Successfully. After the implementation, the system provides reliable results.The system is totally menu and user friendly, which makes it easy for the users even with limited knowledge of Computer environment to operate the developed system. The system avoids the drawbacks of the existing Manual system and the validation facility of the system totally eliminates the chances of wrong data entry.The system was done with python, and the python libraries that were implemented includes, NumPy, Pandas, Matplotlib and Seaborn. At the end of the work expected results were obtained and the analysis was able to show the level of participation of the various individuals on the given WhatsApp group. On serious note this system has the ability to analyze any WhatsApp group data input into it.

It has following features:

- **User friendly**
- **Time saving**
- **Runs on any devices**
- **Analyzes any WhatsApp imported file**
- **Accuracy**
- **Reliability**
- **Easy to use.**