Text Classification based Behavioural (Sentimental) Analysis of Chats

Chats conversations Analysis Platform based on NLP

Presented by :

Jaideep Yadav (0101IT211026)

Harshvardhan (0101IT211023)

Abhay Tripathi (0101IT211004)

Sanjeev Singh (0101IT211045)

01. Introduction

- Problem Statement
- Project idea

Major Problems

- Topic & Trend Analysis
- Group Engagement
- Anomaly Detection
- User Engagement Monitoring
- Sentiment Analysis
- Market Research

Project idea

The chat analyzer project is a tool that allows users to analyze their chat data to gain insights into their messaging habits, word frequency, sentiment analysis and more. This project can be implemented using various programming languages and libraries such as Python, NLTK, Pandas and Matplotlib.

Chat Analyzer is a data analysis-based engine where you can upload the chat in text format and generate a complete analysis report according to a group or an individual.

02.

Proposed Solution

- Project Design
- Detailed Solution

Project Design

1) Data Extraction.

2) Data Cleaning.

3) Applying the Natural Language Processing(NLP) to process the pre-processed data.

4) Creating the statistical and sentiment model

5) Testing and Evaluation of the projects.

6) Further evaluation and evolution of the project.

Data Extraction

Export chat document from social media platform and Upload:

Like, document is exported from WhatsApp,Discord,Telegram,etc.

- **Whatsapp:** Steps to export chat → Open individual/ Group chat → Tap Options → More → Export Chat → Choose export without media → Document is downloaded.

- **Telegram:** Log into your telegram account→Open the conversation you intend to download→Click on three dots on the top right corner and select Export Chat History.

Data Pre-processing

This system is based on data analysis and pre-processing. The first step is pre-processing.

The library that is used for preprocessing the data here is **Pandas** and **Regular expression**.

Pandas is an open-source python library. Pandas used to convert string data into Data frame. Data frame is the representation of data into 2-dimensional table of rows and columns. We can work with large data sets using Pandas library. Pandas library has many built-in functions for data analysis, data cleaning, data exploration and data manipulation

Apply Model and Analyse the data

Modules that have we will use to apply the model and analyse the data are:

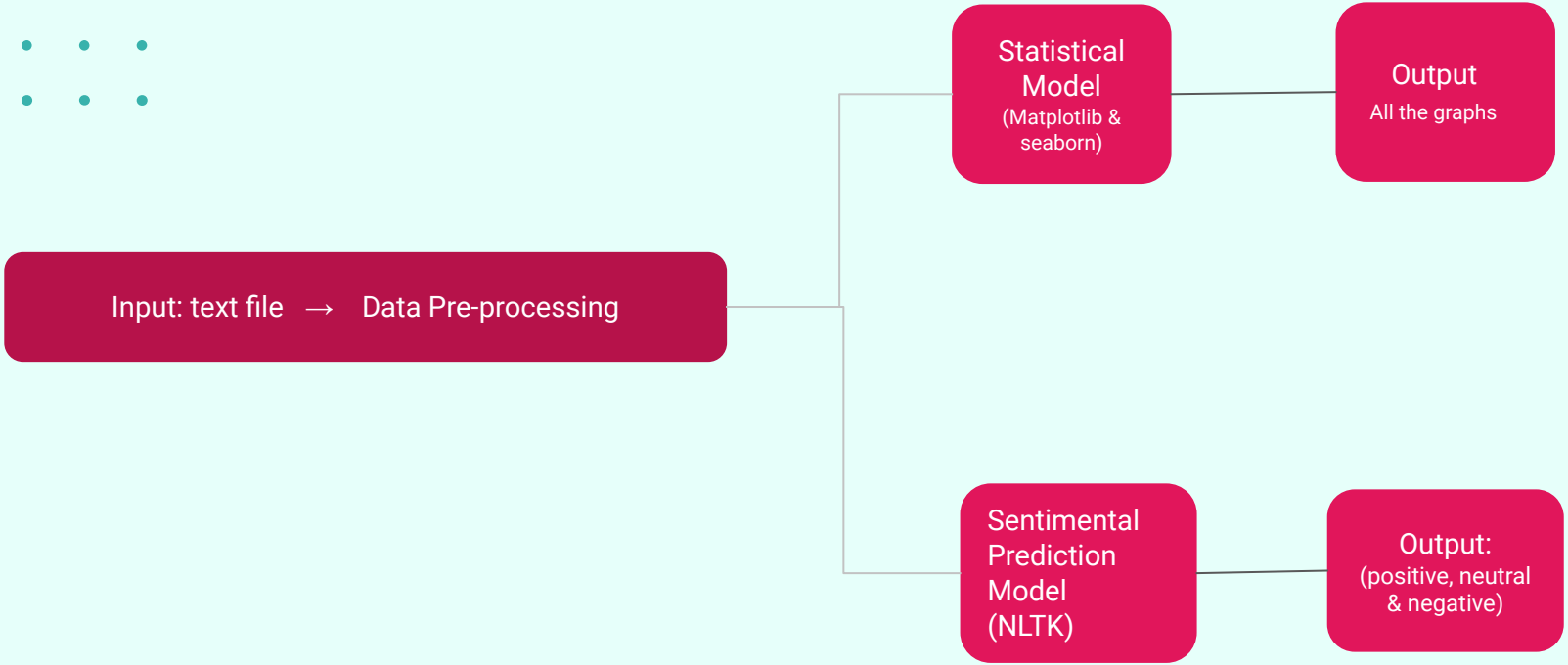
.a. **Streamlit:** Streamlit is a python framework. To develop web apps for Machine Learning and Data Science by using Streamlit.

b. **Matplotlib:** Matplotlib is a popular Python packages used for data visualization. It is a cross-platform library for making plots from data in arrays. It helps in creating static, animated and interactive visualizations in python.

c. **Seaborn:** Seaborn is the data visualization library. It is used for making statistical graphs. Visualization is the central part of seaborn.

d. **Word cloud:** Word Cloud is a data visualization library used for representing most frequently used words within a given text. Most frequent and important words are represented in bigger and bolder size

Proposed System



Advantages of Analyzer

Chat Analyzer do following things :

- Weekly activity
- Total Messages
- Total words
- Media shared
- Link shared
- Monthly timeline
- Most busy month
- Most busy day
- Most busy users
- Most used words
- Emoji analysis

References

https://sci-hub.se/downloads/2020-07-12/ee/10.1109@ICCES48766.2020.9137911.pdf

https://www.researchgate.net/publication/341876645_Whatsapp_Chat_Analyzer

https://www.irjmets.com/uploadedfiles/paper//issue_5_may_2022/22029/final/fin_irjmets1651575263.pdf

Thank You