Indian Institute of Information Technology, Nagpur, India
Department of Electronics and Communication
Session 2021-22

NLP Project Report

Collaborators:-

BT19CSE121 ANIKET RANSING

BT19CSE128 GAURAV BHANDARKAR

Course Name: NATURAL LANGUAGE PROCESSING

Course Coordinator: Dr. Pooja Jain

Introduction

WhatsApp Chat Analysis is a **Natural Language Processing (NLP)** based project in which we try to create a summary using textual inputs like chat messages, emojis etc.

We analyse different group texts (from WhatsApp, Telegram etc.) data analysis tools like Pandas and Numpy. We analyse the type of messages being sent in the group.

In this world, where almost everyone is using social media, many are added to such groups which are not that useful for that particular user. But, by the time he comes to know about it, there are already many messages received. Some unnecessary media sent in that group also occupies space in the device storage.

But if we already have some data of that group, we can immediately leave that group if necessary. That is why this project is really important. It analyses the group chats using the textual inputs and gives a summary of the type of group. In this summary it gives the information of the types of messages (positive or negative), the type of emojis used, the activity of members etc.

Dataset acquisition

The Dataset we used to train our Deep Learning Model is aarline twitter sentiment dataset. While the dataset used to perform Data Analysis was created by us by exporting our WhatsApp and Telegram Chats. Here we have converted the exported chats to Datasets with the help of python scripts.

These scripts were written by us and optimised as such to get desired output.

Similarly, we acquired dataset from some other WhatsApp groups as well as some Telegram groups. The link for the same is provided below.

```
df.tail()
```

	Date	Time	Message	Messenger
103	2022-03-29	12:34 pm	In fact there won't be class on Thursday and ...	Pooja Jain
104	2022-03-29	12:40 pm	Ok ma'am	+91 70136 41413
105	2022-03-30	2:37 pm	No class tomorrow and day after	Pooja Jain
106	2022-08-04	1:05 pm	I have uploaded all the lectures and slides i...	Pooja Jain
107	2022-09-04	9:31 am	https //nlp.stanford.edu/IR-book/html/htmledi...	Pooja Jain

Problem Statement

Task Definition:

Sentiment analysis is a technique for determining whether input is good, bad, or neutral using natural language processing (NLP). Sentiment analysis is frequently used on textual data to assist organisations in tracking brand and product sentiment in consumer feedback and better understanding customer demands.

Algorithm Definition:

We have trained our own Deep Learning based Sentiment Analysis Model and compared it to pre-existing models like VADER and TextBlob.

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a popular rule-based sentiment analyzer. It uses a list of lexical features (e.g. word) which are labeled as positive or negative according to their semantic orientation to calculate the text sentiment. Vader sentiment returns the probability of a given input sentence to be Positive, negative, and neutral.

TextBlob is a simple python library that offers API access to different NLP tasks such as sentiment analysis, spelling correction, etc. Textblob sentiment analyzer returns two properties for a given input sentence i.e. Polarity and Subjectivity

Deep Learning Implementation of NLP based sentiment analysis is done with the help of **LSTM layers** and **word embedding.** We have used tweets.csv dataset which has 10000+ rows each with labels either Positive or Negative label. We have used sigmoid activation function. We have managed to achieve 95.76% **Prediction Accuracy** and 94.46% **Validation Accuracy** after running 5 EPOCHs.
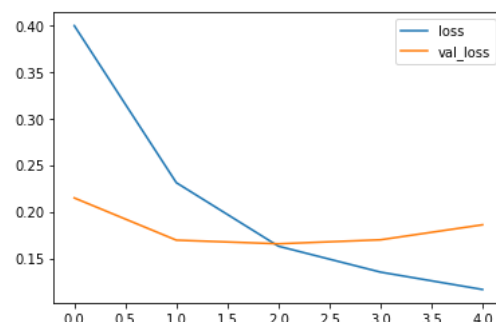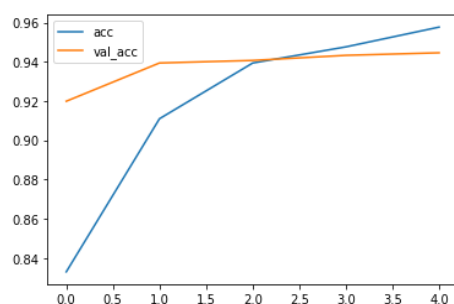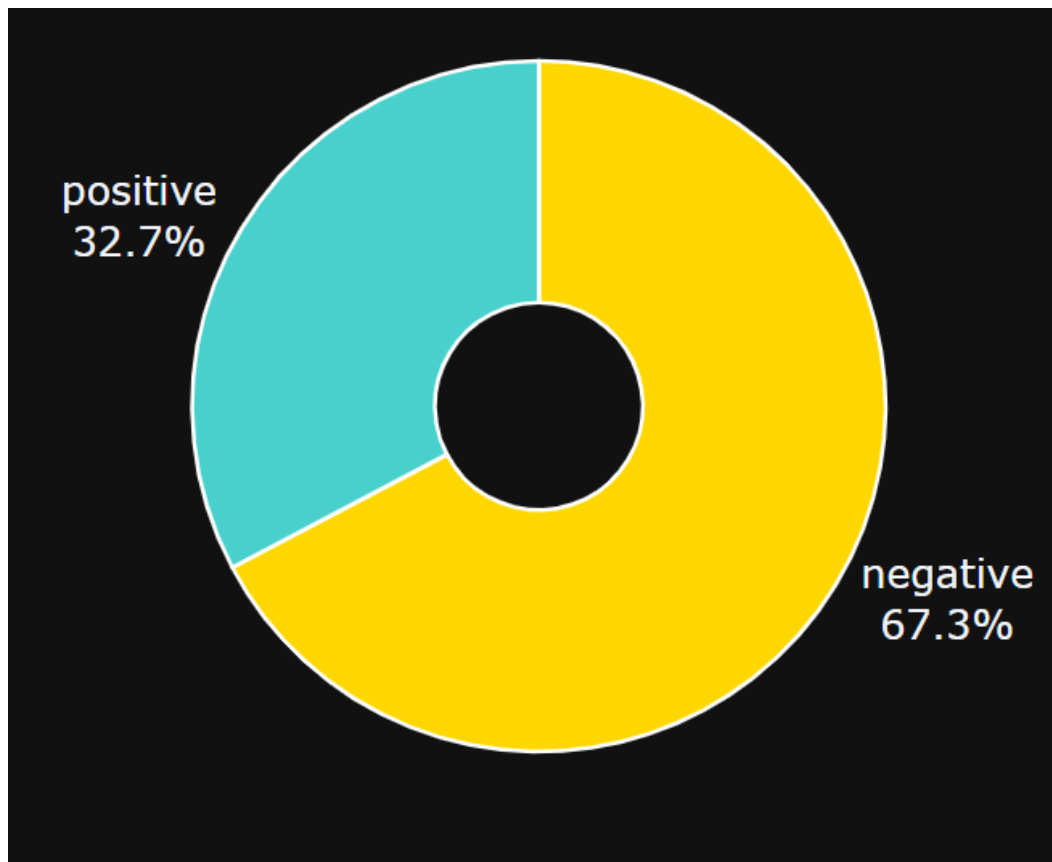
Data Analysis and Visualisation:

We performed various data visualisation operations on the Group Chat Dataset to better understand our dataset. These operations include Activity Measurement, Emoji Recurrence, Word Cloud Exportation etc.
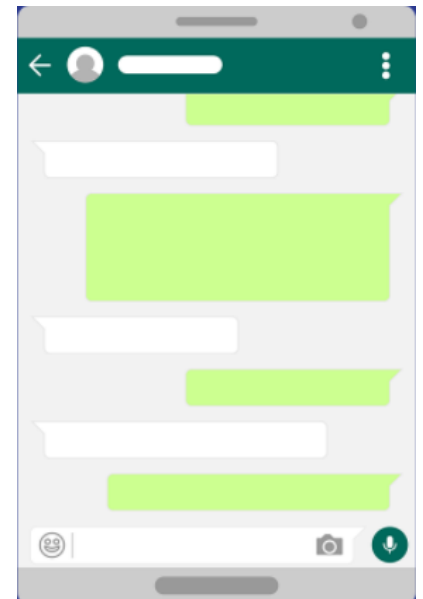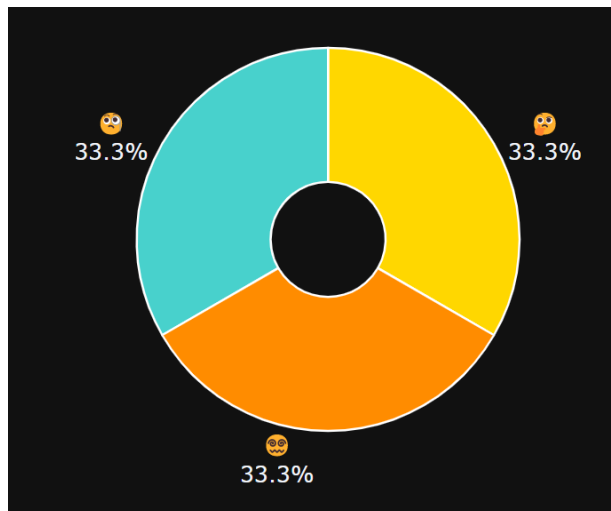
Results:

This is the result obtained after passing the Group Chat Throught our Deep Learning Model

Important Links:

1. Main Notebook:

   https://colab.research.google.com/drive/1tfS57xi65JyUIcV-jDzHUIgx0BjWxyGd

2. Deep Learning:

   https://colab.research.google.com/drive/1BBDfbn5nb0f3T7LV1x4fkNIBbsgDAfXJ

3. Scripts:

   https://colab.research.google.com/drive/1r2gheNENRZMJ9IOJohmLT0ZsULRqcrKI?usp=sharing

   https://colab.research.google.com/drive/1Nz61d1Udb27pXCf1BHtD-NGqLFsc1zAc?usp=sharing

   https://colab.research.google.com/drive/1bFDC60aCwPRTwYAnWIJFW8MeG5AhbwbK?usp=sharing

4. Datasets Used:

   https://drive.google.com/drive/folders/1W_Sym1n2Icrmq-LR3cb0BVk8pCYGPLaG?usp=sharing

Reference

- [People's Behaviour Analysis in Chat Message using Natural Language Processing | IEEE Conference Publication](#)

- [Behavioral analysis of internet messaging and malicious activity detection | IEEE Conference Publication](#)

- [Text Classification based Behavioural Analysis of WhatsApp Chats | IEEE Conference Publication](#)

- [Social Communication Database - Mendeley Data](#)

- [Twitter US Airline Sentiment | Kaggle](#)