**TASK 3 REPORTING**

**DATE: 04/05/2022**

**TASK NAME:** NLP - SENTIMENT ANALYSIS

**TASK LEADER:** BHARATI PANIGRAHI

**TASK CO-LEADER:** SHREYA SHETYE

**TASK OBJECTIVE:** This task will focus on Performing NLP / Sentiment Analysis of the gathered database to make the bot more sophisticated. It's also used for the analysis of the data to gain insights and understand more about the FAQs in the chatbot.

**COMPLETED TASK:**

1. Collected and Pre-processed the tweets
2. Include a column of class for the tweets, for example, Medical staff, Treatment, etc., so that we can show the word clouds of Positive, Negative, and Neutral tweets for each of the classes separately on the dashboard.
3. Active Learning
   * Manually annotated 3000 rows from the dataset as Relevant and Irrelevant
   * Performed active Learning to label all the tweets
4. Performed Sentiment analysis on the filtered dataset, using VADER.
   * <https://colab.research.google.com/drive/1YxOW_GFEv5wbU_fMGVj8hZZMYgdCIsYt?usp=sharing>

**IN-PROGRESS:**

1. Perform Active Learning to assign classes to all the tweets
2. Trying to validate the model as well as Active Learning to ensure the best results.

**FUTURE SCOPE:**

1. Include a column in the dataset that shows a hit count of how frequently the chatbot has provided a specific piece of information as it will be helpful to justify why we chose specific diseases for our dataset from the administration perspective.
2. Include sentiment analysis in the chatbot for answering queries of the patients based on the sentiment of the previous comment/question