**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. Created a basic framework for sentiment analysis using sklearn for classification and a movie review dataset as a dummy dataset.
   * <https://colab.research.google.com/drive/1Gw8gYLI7G8wnPdesRguDZ8HS9LTvW9hi>
2. Performed Sentiment analysis on the extracted dataset using VADER.
   * <https://colab.research.google.com/drive/1YxOW_GFEv5wbU_fMGVj8hZZMYgdCIsYt?usp=sharing>

**IN-PROGRESS:**

1. ACTIVE LEARNING:
   * Manually annotate 3000 rows from the dataset as Relevant and Irrelevant
   * Perform active Learning to label all the tweets
   * Drop all the tweets which are labeled as "Irrelevant", which will give us our final data for sentiment analysis. This will reduce our dataset size to a much smaller size as many tweets will be irrelevant as observed.
2. Pass the filtered dataset, through the VADER code snippet prepared.

**FUTURE SCOPE:**

1. Include a column of class for the tweets, for example, Medical staff, Treatment, etc., so that we can showw the word clouds of Positive, Negative, and Neutral tweets for each of the classes separately on the dashboard.
2. 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.
3. Include sentiment analysis in the chatbot for answering queries of the patients based on the sentiment of the previous comment/question