TASK-1 SUMMARY:

Date: 23-05-22

**Task name:** Web Scraping

**Task leader:** Esha Priyadarshi & Ameen Shaikh

**Task weekly meeting:** Tuesday 9 PM IST

**Task objective:** Data collection for chatbot & Dashboard

*Completed;*

* Scraped data for 6 diseases. With the following columns/details.

Symptoms(common),

Symptoms(babies/infants),

Symptoms(old people)

Symptoms(male)

Symptoms(Female)

Symptoms (critical\_stage)

Medication (general)

Medication(babies/infants)

Medication(old people)

Treatment(general)

Treatment(babies/infants)

Treatment(old people)

Treatment(critical\_stage)

Expenses

Facts\_or\_Myths

Survival\_Rate

Funding

Prevention

Diagnosis

FAQ

FAQ Answers

* Created different csv for each disease with above mentioned columns.
* Created a merged csv with diseases arranged row wise.
* Annotating the tweets (a small sample)
* Performed Active learning for classifying relevant and irrelevant tweets using small text & Transformer model
* Provided relevant tweets for NLP sentiment analysis

*In progress*

* Trying to improve the results of active learning

*Future work:*

* Add more data for different diseases.
* Add the resources used to scrape the information.
* Add informative medical pages link too.
* If needed scrape data for dashboard