**Hackathon**

You are given a set of conversation between two individuals. Data has been collected based on some template conversation between individual and some context-based conversation. Based on the context of the conversation you are asked to create a Chabot or conversation-generating model that could reply to a set of starter conversation mentioned in test data.

**Data Description :**

There are two files.

**TrainData.txt** contains the conversion as answer and question separated by tab.

**TestData.csv** has two columns **I.D.** & **Question**.

**I.D.** 🡪 it contains the question number

**Question** 🡪 it contains starter of a conversation.

**Metric to measure :**

The metric to measure is similarity score between the Answers provided by the participant and the set of appropriate answer or reply to the conversation.

For approximate string matching concepts, participant may refer <https://en.wikipedia.org/wiki/Approximate_string_matching> .

The specific measure to calculate the distance between strings is token set ratio. To know further about the measure participants may refer the documentation of the fuzzywuzzy package in <https://github.com/seatgeek/fuzzywuzzy/blob/master/fuzzywuzzy/fuzz.py> created by **SeatGeek**.

**Marking Procedure :**

The score of the student will be converted into marks taking into consideration the relative score in the leader board and converting that into 20.

**Submission :**

Participant can submit their final file using

[**https://docs.google.com/forms/d/e/1FAIpQLSdsFiJy85Dph-m0bZitzS931iueGffHvRIIauKDUrmV-HXRQA/viewform?usp=sf\_link**](https://docs.google.com/forms/d/e/1FAIpQLSdsFiJy85Dph-m0bZitzS931iueGffHvRIIauKDUrmV-HXRQA/viewform?usp=sf_link)

**A student can submit a maximum of 10 times.**

File Format: You should submit a CSV file with exactly **543 entries** plus a header row. **The file should be saved with the participant name.** For example, a student with name Sutithi Chakraborty will submit their 4th submission file as SutithiChakraborty\_4.csv

The file should have exactly two columns ● I.D. ● Answer

**Score Board**

You can see the LeaderBoard using the link below

[**https://drive.google.com/file/d/1sM0zdk88ps1LoTnTQSpANKWix7KszVNT/view?usp=sharing**](https://drive.google.com/file/d/1sM0zdk88ps1LoTnTQSpANKWix7KszVNT/view?usp=sharing)

**The score board will be updated every hour.**