



**REPORT**

# **FAQ CHAT BOT**

**MACHINE LEARNING HACKATHON  
REPORT**

# REPORT

FAQ CHAT BOT

## PROBLEM STATEMENT DESCRIPTION :

This is a FAQ chat bot. It answers any question given by the customer about the topic previously defined in the dataset described in any possible manner(non-erratic manner).

## APPROACH :

The dataset is divided into questions and answers periodically. The questions are saved as a key with the corresponding answers as value. The questions are vectorised using an open source python API, indico.io. The vectorised questions are saved in a pickle file. When the user gives a question, it is vectorised and is compared with the vectorised questions in the dataset for finding the distance between the two questions using a python module, cosine distance(cdist).

"1 - distance" shows the similarity between the two sentences. if the similarity is more than some threshold it will list out most m=similar sentences in order, the most similar question's corresponding value is the required answer.

## **DATASET :**

The dataset is obtained from the public web page of the government of India(.gov.in).

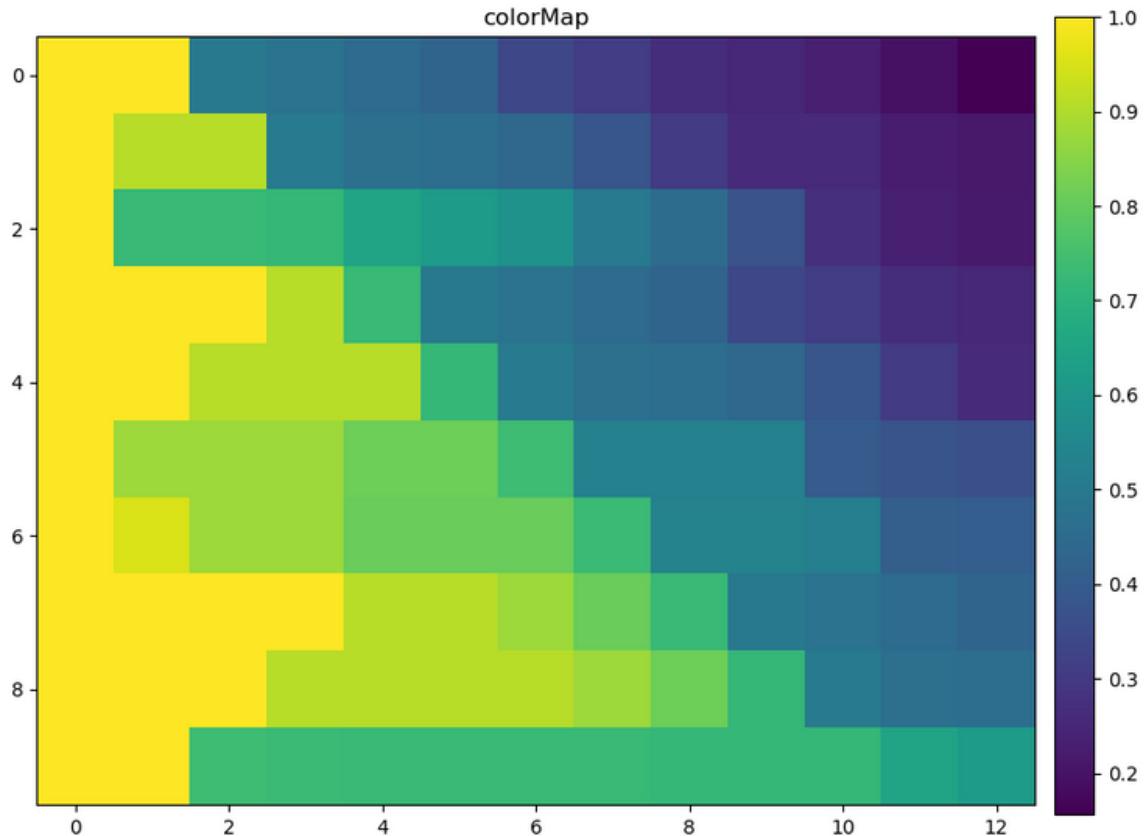
URL:

<https://portal2.passportindia.gov.in/AppOnlineProject/online/faqsServicesAvailable>

The data is organised in such a way that the questions and answer occur in a single line one after other periodically without skipping lines.

## ACCURACY :

Since this is an application level working model the accuracy is determined by the confusion matrix with the dataset pitted against itself. The measuring factor is the similarity between two sentences as shown in the graph. This is the colour graph of the confusion matrix.



## **NOTE :**

1. Prerequisites: install TQDM and indicoo
2. While running the python file use a required question as an argument.
3. Since this is a completely working module if there are no similarities it will give a predefined 'sorry' statement.

**TEAM NAME: TEAM\_STARK**

**TEAM MEMBERS:**

**AKHILANK M J      IMT2015004**

**SAPPARAPU RAHUL      IMT2016036**

**SUMANTH K      IMT2016041**