So I took an online dataset and ran our decision tree classifier on it to train it. Then for testing, the user needs to input a string consisting of his/her symptoms. Our code pre-processes the data given by user by finding its synonyms, lemmatizing and tokenizing it. Then it compares the input with the dataset(using jacquard coefficient) and finds relevant and co-occuring symptoms of the input. The system asks user to enter the relevant ids of symptoms the user is facing and proceeds with the processing and calculates the final symptoms of the user. Then it calculates the top 5 diseases based on the final symptoms and finds the probability percentage of each disease. Then it asks user to input an id based on the disease and prints a description(diagnosis) of it.