REPORT ON ALGORITHM USED IN THE TASK

Name-Siddharth Pareek

The first task is to convert each text in the data set into a vector. A vector here refers to each and every word in each text with their respective occurrences . This can be done with the help of RE library and Counter from the collection module. After we have converted the texts to vectors then our task is to find the similarity between the vector pairs we have got after conversion. This will be done by using co-sine similarity method . In this method we find out the cos(angle) between the 2 vectors. Since we know that the dot product of the 2 vectors is given by vector1.vector2=|vector1|\*|vector2|\*cos(angle).

After doing this we will get the similarity score of each pair which will be in the range of 0 to 1. Where 0 is extremely un-similar and 1 is extremely similar. I did this all using Jupyter Notebook using Python language.

Thankyou.