Total No. of Pages _02__ *Roll No. ……*

_Seventh__ SEMESTER **B.Tech [IT]**

CLASS TEST I *Sept-2020*

IT-425 NATURAL LANGUAGE PROCESSING

Time: 1:00 Hours ***Max. Marks : 20***

Note : Answer all questions in a precise manner.
Assume suitable missing data, if any.

Q.1 a) Consider the 6 x 5 feature matrix (one-hot encoding BoW representation) extracted from your respective CVs. (You can consider the feature matrix output in your programming assignment last week). For the last row (test vector), compute the cosine similarity with all other feature vectors (training set). Find the row with which the test sample is closest to using the COSINE SIMILARITY MEASURE by showing all the calculation steps in detail.

b) Does your experiment qualify as a text classification task? If yes, give reason. If no, then describe how to add onto your experiment to convert it to a text classification task.

(5+3)

Q. 2 For your CV, rewrite the 6 x 5 feature matrix (one-hot encoding BoW representation) using bigrams in feature columns (choose bigrams and trigrams of your choice), instead of unigrams. Analyse which is the better choice: the 6 x 5 matrix using unigrams or bigrams? (4)

Q.3 a) Explain using an example code, how to incorporate WordNet into your programs.

b) How will the 6 x 5 feature matrix in Q.1 change if you consider WordNet also for feature engineering?

(3+5)