

MCA.V/01.21.007

Reg.No

--	--	--	--	--	--	--	--

B

MCA DEGREE FIFTH SEMESTER EXAMINATION, JANUARY 2021
16-381-0570 NATURAL LANGUAGE PROCESSING
(Regular)

Time: 3 Hours

Maximum Marks: 50

(Answer **ANY FIVE** questions)
(All questions carry **EQUAL** marks)

- 1) Explain the concept of corpus in NLP in detail with examples. (10)
- 2) (a) What is ambiguity? Explain different types of ambiguity. (5)
(b) Explain the concept of homonymy and polysemy with suitable examples? (5)
- 3) Explain shift-reduce parser? Apply shift-reduce parsing on the sentence “the child ate the cake with the fork”. (10)
- 4) Explain TF-IDF model in NLP. Apply TF-IDF to the word “data scientist” with respect to the three documents given below:

Document 1: Ben studies about computers in Computer Lab
Document 2: Steve teaches at Brown University
Document 3: Data Scientists work on large datasets (10)
- 5) What is unsupervised machine learning? Explain k-means clustering with an example. (10)
- 6) Explain how NLP is used in sentiment Analysis clearly outlining the steps, taking a collection of 1000 tweets as an example. (10)
- 7) Explain Bag-of-words text vectorisation technique. Apply Bag-of-words to three documents given below:

Document 1:- It was the best of times
Document 2:- It was the worst of times
Document 3:- It was the age of wisdom
Document 4:- It was the age of foolishness

Create document vectors for each document . (10)
