2004M158 Candidate's Seat No :_____

Integ. M.Sc DS Semester-8 Examination
CC-414
Advanced NLP

Time : 2-30 Hours] April-2024 [Max. Marks : 70

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

Q.1 **(a)** What is NLP? Describe major components of NLP.
(b) Discuss any two APIs of NLP.

OR

(a) List out and explain various stages of NLP pipeline.
(b) List out various NLP applications. Discuss any two.

Q.2 **(a)** What is language model? Write a note on N-gram language model.
(b) What is smoothing? Explain Laplace smoothing in N-gram language model using a suitable example.

OR

(a) Discuss advanced smoothing method for language modeling: Kneyser-Ney smoothing.
(b) Describe various types of evaluation to measure the performance of a language model.

Q.3 **(a)** Explain continuous bag of words.
(b) Write a note on word2vec.

OR

(a) Explain knowledge based and supervised word sense disambiguation.
(b) Explain recurrent neural net language model.

Q.4 **(a)** Discuss various problems of machine translation.
(b) Explain seq2seq model.

OR

(a) Explain rule-based machine translation.
(b) Write a note on Transformers.

Q.5 Attempt any **SEVEN** out of **TWELVE:**
(1) Mention merits and demerits of NLP.
(2) Mention various NLP libraries.
(3) Why NLP is difficult? – Explain in brief
(4) Discuss Good-Turing Discounting.
(5) Describe backoff and interpolation in brief.
(6) Discuss open versus closed vocabulary.
(7) Define the terms with suitable examples: hypernyms, meronomy
(8) Discuss one-hot encoding in brief.
(9) Describe skip-gram.
(10) Discuss the need of machine translation.
(11) Discuss DMT in brief.
(12) Discuss EBMT in brief.

**** ****