[No. of Printed Pages – 4] **2120**

**CSE468**
[ET]

Enrol. No. A0231198 20027

END SEMESTER EXAMINATION : NOVEMBER-DECEMBER, 2023

# NATURAL LANGUAGE PROCESSING WITH DEEP LEARNING

*Time : 3 Hrs.* *Maximum Marks : 60*

**Note:** ***Attempt questions from all sections as directed.***

## SECTION – A (24 Marks)

*Attempt any* **four** *questions out of* **five.**
*Each question carries* **06** *marks.*

1. Why is it better to use GloVe than Word2Vec ? Explain the cost function used in GloVe?

2. Briefly describe the perplexity of the bigram model in terms of chain rule?

3. How do Naive Bayes' Classifier classifies amibiguity? Explain

P.T.O.

4. Draw a parse tree of a given sentence "John saw Mary and the boy with a telescope"

Rules give are :

S->NP VP
NP-> N
VP-> V NP PP
NP->N
PP-> PREP NP
NP-> DET N PP
NP-> DET N
PP-> PREP NP

5. The Text String is given to you is

Strl = "He is a good boy, but she is a very good girl, that girl is a good basketball player"

Show the representation of the text string,

the co-occurrence matrix, one hot encoding, word2vec using window size 2.

## SECTION – B (20 Marks)

*Attempt any* **two** *questions out of* **three.**

*Each question carries* **10** *marks.*

6. Explain word embedding functionality using common dimensionality reduction techniques such as PCA and t-SNE with suitable graphs assuming an appropriate example

7. Write short notes on

   (a) Large Language Models

   (b) LSTM Models

8. Briefly explain the purpose of Matrix Factorization? How to perform matrix factorization having some fixed column as output.

## SECTION – C (16 Marks)

*(Compulsory)*

9. (a) Statement : "RNN works as a sequence based model unlike CNNs". Justify this statement with an appropriate example (8)

(b) Calculate the PMI score, PMI(red, bull) of the following input:

| this is a red bull bull black horse red bull bull black horse red bull bull black horse bull black black |
|---|

(8)