

**MSc (AIML) Sem.-3 Examination
Natural Language Processing
December-2025**

Time : 3.00 Hours]

[Max.Marks :100

Instructions:

- Write both the Sections in the separate answer book.
- Both Sections having equal weightage.
- Draw Diagrams wherever necessary.
- Make Assumptions wherever necessary.

SECTION – I

- Q-1 Answer the following questions (Any 3): 18
- a. What is Natural Language Processing and its components? What are the real-life applications and challenges it has in detail?
 - b. Explain ambiguity on all level of languages in NLP in detail?
 - c. What do you mean by n-grams in NLP? Explain how it can be used for language modelling?
 - d. What is the Text Classification and application of it in NLP? Explain how Naïve Bayes can be used to do same in detail?
 - e. What is Named Entity Recognition (NER) in NLP? List the common entity types and provide an example sentence with NER applied.
- Q-2 Answer the following questions (Any 2): 16
- a. Write all the text processing steps and explain each with their application. Explain text vectorization using BoW and TF-IDF and compare both?
 - b. How do Hidden Markov Models (HMM) contribute to Part-of-Speech (POS) tagging, and what is the role of the Viterbi algorithm with an example?
 - c. Compute the Levenstein, Damerau-Levenstein and hamming distance between the words "kareena" and "katrina". Construct the dynamic programming table for Levenstein distance. List the sequence of operations to transform "kareena" into "katrina" using all three methods. Explain how the edit distance can be used in spell-checking applications.

P.T.O

- Q-3 Answer the following questions (Any 2): 16
- a. Given a corpus of three documents
 D_1 : "the cat runs fast"
 D_2 : "the dog runs"
 D_3 : "cat and dog"
 calculate the Bag-of-Words (BoW) vector and TF-IDF vector for each document of corpus.
- b. Consider the following corpus.
- marry (N) jane (N) can (M) see (V) will (N)
 - will (M) jane (N) spot (V) marry (N)
 - spot (N) will (M) see (V) marry (N)
 - marry (N) will (M) pat (V) spot (N)
- Calculate Emission and transition probability of it.

- c. Consider the following data.

tag	Documents
-	just plain boring
-	entirely predictable and lacks energy
-	no surprises and very few laughs
+	very powerful
+	the most fun film of the summer

Check whether "predictable no fun" is positive or negative using naïve bayes

SECTION – II

- Q-1 Answer the following questions (Any 3): 18
- a. What is word sense and meaning in the context of NLP? Describe WordNet.
- b. What is word embedding and how it better than BOW and TF-IDF in vectorization, write comparison in table.
- c. Explain Word2Vec for word embedding in details with its two variants.
- d. What is parsing and explain constituent and dependency parsing in detail?
- e. What are the applications of feedforward networks in the field of NLP.
- Q-2 Answer the following question 16
- a. Explain RNN and how it is useful for sequence text classification. Explain LSTM in detail with proper diagram of gates and working flow.
- OR**
- a. Explain RNN and how it is useful language modelling. Explain GRU in detail with proper diagram of gates and working flow.

Q-3 Answer the following questions (Any 2):

16

a. What is Text summarization? Explain its types: Extractive and Abstractive.

b. Given the context grammar:

CFG: {
 $S \rightarrow NP VP$
 $S \rightarrow VP NP$
 $NP \rightarrow DET NP$
 $NP \rightarrow ADJ NP$
 $VP \rightarrow V NP$
 $VP \rightarrow AUX VP$
 $NP \rightarrow can \mid water$
 $VP \rightarrow hold$
 $AUX \rightarrow can$
 $DET \rightarrow a \mid an \mid the$
 $ADJ \rightarrow large$
 }

Implement constituent parsing using CKY algorithm and construct the respective table as well for "*the large can can hold water*".

c. Write the process in detail to develop an end-to-end language model using Sequential models starting from the corpus mention all the relevant steps of processing, data preparation and working flow of model.

~~—————~~