PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations – October 2025

Date: 31-10-2025   **Time:** 02:30Pm – 04:00Pm

School: SOIS-PG	**Program**: MCA
Course Code : CSA4040	**Course Name:** Natural Language Processing
Semester: III	**Max Marks**: 50 **Weightage**: 25%

CO - Levels	CO1	CO2	CO3	CO4	CO5
Marks	26	24	-	-	-

Instructions:

   (i) *Read all questions carefully and answer accordingly.*

   (ii) *Do not write anything on the question paper other than roll number.*

Part A

Answer ALL the Questions. Each question carries 2marks.   **5Q x 2M=10M**

1	What do you understand by NER?	2 Marks	L2	CO1
2	Why is text preprocessing important?	2 Marks	L2	CO1
3	Explain the concept of POS tagging.	2 Marks	L2	CO1
4	What is iDF in relation to NLP?	2 Marks	L2	CO2
5	Give the full form of RNN and CNN with respect to NLP.	2 Marks	L1	CO2

Part B

Answer the Questions.   **Total Marks 40M**

6.	**a.**	Given an AI model to test the collected test sample for covid-positive and covid negative classes, having the following confusion matrix. Evaluate the model as per its accuracy, Precision, recall and f-measure.	**10 Marks**	**L3**	**CO1**

N=100	Predicted No	Predicted Yes
Actual No 60	50	10
Actual Yes 40	5	35

	b.	Your team is given a text for translation but you have been asked to clean it first. What steps would you take to clean the text?	**10 Marks**	**L3**	**CO1**

Or

7.	**a.**	Explain the concept of the confusion matrix and its various derivatives like accuracy, Precision, recall and f-measure with formula and examples. When would you prefer F measure as compared to accuracy?	**10 Marks**	**L3**	**CO1**
	b.	Write a program to demonstrate corpus cleaning for the text "Amazing! My email id is not ad@sdf.com nor is my number 344563."	**10 Marks**	**L3**	**CO1**

8.	**a.**	Calculate the TF-IDF of the query 'the apple' with respect to a database of three documents D1: 'The apple is on the table.' D2: "My orange and apple are the best." D3: "The relatives are paying."	**10 Marks**	**L3**	**CO2**
	b.	Write a simple program to find the similarity between the words 'language' and 'text' using Word2Vec using the sample text corpus given below: "Natural Language Processing enables computers to understand human language. Machine learning algorithms can learn word meanings from text data. Word embeddings capture semantic relationships between words. Deep learning models like Word2Vec and GloVe create vector representations of words."	**10 Marks**	**L3**	**CO2**

		Or			
9.	**a.**	Vectors are used to represent the word cat = [3,2,0,5] and word dog = [1, 0,0,0]. Calculate the cosine similarity of the two words. What would be the dissimilarity?	**10 Marks**	**L3**	**CO2**
	b.	You are given the following small dataset of sentences: 1. "Machine learning is fun" 2. "Deep learning is used in AI applications" 3. "Python is a popular programming language" Using this dataset, demonstrate how a computer can analyze the words and establish relationships between them. Show how word embeddings could represent these relationships and give an example of a relationship or analogy between words.	**10 Marks**	**L3**	**CO2**