Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – December 2025

Date: 13 – 12- 2025 **Time:** 09:30am – 12:30pm

School: SOIS-PG	**Program:** MCA	
Course Code: CSA4040	**Course Name:** Natural Language Processing	
Semester: III	**Max Marks**: 100	**Weightage**: 50%

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

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. **10Q x 2M=20M**

1.	What is ambiguity? Give example.	**2 Marks**	**L1**	**CO1**
2.	Explain the importance of word tokenization.	**2 Marks**	**L2**	**CO1**
3.	Describe the softmax function.	**2 Marks**	**L2**	**CO2**
4.	Give the full form or CNN, RNN.	**2 Marks**	**L1**	**CO2**
5.	What is CFG?	**2 Marks**	**L1**	**CO3**
6.	Discuss the importance of NER for NLP.	**2 Marks**	**L2**	**CO3**
7.	Explain BIO tagging.	**2 Marks**	**L2**	**CO3**
8.	Describe rule based MT.	**2 Marks**	**L2**	**CO4**
9.	Define Machine Translation in your own words.	**2 Marks**	**L1**	**CO4**
10.	Name two types of Agreement measures in NLP.	**2 Marks**	**L2**	**CO4**

Part B

Answer the Questions. **Total Marks 80M**

<table>
<tr><td>11.</td><td>a.</td><td>Develop a small program to demonstrate corpus cleaning.</td><td>10 Marks</td><td>L3</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>What is syntax? Explain syntactic analysis using the sentence ‘I like mangoes’.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>12.</td><td>a.</td><td>Calculate the parameters such as accuracy, precision, recall and F1-score from the following matrix:
<table><tr><td>150</td><td>Predicted No</td><td>Predicted Yes</td></tr><tr><td>Actual No</td><td>100</td><td>15</td></tr><tr><td>Actual Yes</td><td>10</td><td>25</td></tr></table></td><td>10 Marks</td><td>L3</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>What is full form of POS? Describe the importance of POS tagging with 5 examples.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
</table>

<table>
<tr><td>13.</td><td>a.</td><td>Vector are used to represent the word pineapple = [3, 8, 10, 11] and word orange = [2, 2, 15, 13]. Apply the cosine similarity for the two words.</td><td>10 Marks</td><td>L3</td><td>CO2</td></tr>
<tr><td></td><td>b.</td><td>Write a simple program to demonstrate word relationship using Word2Vec.</td><td>10 Marks</td><td>L3</td><td>CO2</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>14.</td><td>a.</td><td>Apply the TF-IDF principles for the query ‘the cat’ with respect to a database of three documents
D1: ‘The cat is on the table.’
D2: “My dog and cat are the best in town.”
D3: “The locals are not playing.”</td><td>10 Marks</td><td>L3</td><td>CO2</td></tr>
<tr><td></td><td>b.</td><td>Analyze how a computer applies word embeddings to identify semantic relationships among words.</td><td>10 Marks</td><td>L3</td><td>CO2</td></tr>
</table>

<table>
<tr><td>15.</td><td>a.</td><td>Use the principles of constituency parsing to analyze a given sentence and build a phrase structure tree that represents its hierarchical syntactic organization. Illustrate your response with a clear and relevant example.</td><td>20 Marks</td><td>L3</td><td>CO3</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>16.</td><td>a.</td><td>Use the Viterbi algorithm to compute and determine the most probable sequence of hidden states for a given observation</td><td>20 Marks</td><td>L3</td><td>CO3</td></tr>
</table>

		sequence. Show each step of the process clearly and illustrate your explanation with an appropriate example.			

17.	**a.**	Given the following 3×3 rating matrix, compute the Linear Weighted Kappa **20** | **5** | **0** **3** | **25** | **2** **0** | **5** | **40**	**10 Marks**	**L3**	**CO4**
	b.	Given the reference sentence: "the cat is on the mat" and the candidate translation: "the cat the cat on the mat" Compute the BLEU score (up to 2-gram precision) step by step, assuming no smoothing and a brevity penalty of 1.	**10 Marks**	**L3**	**CO4**
		Or			
18.	**a.**	Apply your understanding of NLP evaluation methods to differentiate between BLEU and Percentage Agreement metrics. Demonstrate how each metric is calculated, illustrate the type of task it is used to evaluate, and interpret what the resulting scores signify in their respective NLP contexts.	**10 Marks**	**L3**	**CO4**
		Illustrate how automatic evaluation can be combined with human evaluation to improve the reliability of machine translation performance measurement. Provide a simple example to support your answer.	**10 Marks**	**L3**	**CO4**