



PRESIDENCY UNIVERSITY

BENGALURU

Roll No.														
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--

End - Term Examinations – May / June 2026

Date: 21- 05-2026

Time: 09:30am – 12:30pm

School: SOIS	Program: BCA (Data Science)	
Course Code: CSA3403	Course Name: Natural Language Processing	
Semester: IV	Max Marks: 100	Weightage: 50%

CO - Levels	C01	C02	C03	C04
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.	In a corpus: - Total periods (.) = 1000 - Sentence-ending periods = 700 - Periods used in abbreviations = 300 Identify: Probability that a randomly selected period marks end of sentence.	2 Marks	L1	C03
2.	A dataset contains: - Spam = 5 messages - Not Spam = 5 messages Word probabilities: $P(Offer Spam) = 0.6, P(Offer NotSpam) = 0.2$ $P(Win Spam) = 0.7, P(Win NotSpam) = 0.3$	2 Marks	L2	C01

	Question: Classify the message: " Offer Win "									
3.	Identify Probability using logistic regression if $z = 2$.	2 Marks	L1	C01						
4.	A dataset has: <table><tr><td>Class</td><td>Total Words</td><td>Vocabulary Size (V)</td></tr><tr><td>Positive</td><td>20</td><td>10</td></tr></table> Word " good " appears 4 times . 1. Calculate: $P(good \mid Positive)$	Class	Total Words	Vocabulary Size (V)	Positive	20	10	2 Marks	L3	C03
Class	Total Words	Vocabulary Size (V)								
Positive	20	10								
5.	Given: Initial Probabilities: $P(N) = 0.6$ $P(V) = 0.4$ Emission Probabilities: $P(dog \mid N) = 0.5$ $P(dog \mid V) = 0.1$ Calculate probability that first word is " dog "	2 Marks	L3	C03						
6.	State what is Machine Translation (MT)?	2 Marks	L1	C04						
7.	Define cell state?	2 Marks	L1	C02						
8.	Outline what is a Convolutional Neural Network (CNN)?	2 Marks	L1	C02						
9.	State why is SBD important in NLP tasks?	2 Marks	L1	C04						
10.	Define: - Transition probability - Emission probability	2 Marks	L1	C04						

Part B

Answer the Questions.

Total Marks 80M

11.	a. <i>Context words: "I" and "NLP"</i> <i>Word vectors:</i> $I \rightarrow [1, 0]$	10 Marks	L1	C02
-----	--	----------	----	-----

		$NLP \rightarrow [0, 1]$ Weights: $W = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}$ Identify hidden layer output (CBOW uses average of context vectors)			
Or					
12.	a.	Given: - Center word vector = $[0.6, 0.8]$ - Context word vector = $[0.5, 0.5]$ Identify similarity score using dot product.	10 Marks	L1	CO2

13.	a.	Given: Tags: Noun (N), Verb (V) Classify the tags of the Sentence: <i>Dogs bark</i> Transition Probabilities: $P(N Start) = 0.6$ $P(V Start) = 0.4$ $P(V N) = 0.7$ $P(N N) = 0.3$ Emission Probabilities: $P(Dogs N) = 0.8$ $P(Dogs V) = 0.2$ $P(bark V) = 0.6$ $P(bark N) = 0.4$	10 Marks	L3	CO2
Or					
14.	a.	States: N, V Transition: $P(N \rightarrow N) = 0.3$ $P(N \rightarrow V) = 0.7$ $P(V \rightarrow N) = 0.8$ $P(V \rightarrow V) = 0.2$	10 Marks	L3	CO2

		Emission: $P(\text{dog} \text{N}) = 0.5$ $P(\text{barks} \text{V}) = 0.6$ Initial: $P(\text{N}) = 0.6$ Calculate probability of state sequence: $\text{N} \rightarrow \text{V}$ for observation: dog barks			
--	--	--	--	--	--

15.	a.	Given: $B_0 = -1$ $B_1 = 1$ $B_2 = 2$ $X_1 = 1$ $X_2 = 2$ Determine the probability that Person D belongs to class 1. if $p \geq 0.75$ consider it as class 1 else class 0	10 Marks	L3	CO1
-----	----	---	----------	----	-----

Or

16.	a.	Given: • $P(A) = 0.5, P(B) = 0.5$ • $P(x_1 A) = 0.6, P(x_2 A) = 0.5$ • $P(x_1 B) = 0.3, P(x_2 B) = 0.4$ Question: Classify input: x_1, x_2	10 Marks	L3	CO1
-----	----	--	----------	----	-----

17.	a.	Describe types of Machine Translation with their applications in NLP briefly.	10 Marks	L1	CO1
Or					
18.	a.	Outline the ambiguity problems in translation (lexical, syntactic).	10 Marks	L1	CO1

19.	a.	A test set contains 500 sentences. An SBD system predicted 480 sentence boundaries correctly. Out of these, 450 were actually correct boundaries. Calculate: 1. Precision 2. Recall 3. F1-score	20 Marks	L3	CO3
------------	-----------	---	-----------------	-----------	------------

Or

20.	a.	An SBD system produced:	20 Marks	L3	CO3									
		<table><tr><td></td><td>Predicted Boundary</td><td>Predicted non-boundary</td></tr><tr><td>Actual Boundary</td><td>300</td><td>50</td></tr><tr><td>Actual non-boundary</td><td>40</td><td>610</td></tr></table>		Predicted Boundary	Predicted non-boundary	Actual Boundary	300	50	Actual non-boundary	40	610			
	Predicted Boundary	Predicted non-boundary												
Actual Boundary	300	50												
Actual non-boundary	40	610												
		Calculate:												
		1. Accuracy												
		2. Precision												
		3. Recall												

21.	a.	Explain the architecture of CNN with diagram	20 Marks	L2	CO4
Or					
22.	a.	Explain working of forget, input, and output gates.	20 Marks	L2	CO4