



# PRESIDENCY UNIVERSITY

BENGALURU

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

## Mid - Term Examinations - MARCH 2026

Date: 11-03- 2026

Time: 11.45am to 01.15pm

<b>School:</b> SOCSE	<b>Program:</b> B. Tech	
<b>Course Code:</b> CAI2504	<b>Course Name:</b> Natural Language Processing	
<b>Semester:</b> VI	<b>Max Marks:</b> 50	<b>Weightage:</b> 25%

CO - Levels	C01	C02	C03	C04	C05
Marks	24	14	12		

### 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	Explain the relationship between a confusion matrix and evaluation measure precision.	2 Marks	L2	C01
2	Explain word embedding.	2 Marks	L2	C02
3	Infer TF component in TF-IDF.	2 Marks	L2	C02
4	Outline what is a word- context matrix?	2 Marks	L2	C02
5	What is GloVe?	2 Marks	L2	C03

## Part B

### Answer the Questions.

**Total Marks 40M**

<b>6.</b>	<b>a.</b>	<table border="1" style="width: 100%; border-collapse: collapse; margin-bottom: 10px;"> <thead> <tr style="background-color: #e67e22; color: white;"> <th style="width: 30%;"></th> <th style="width: 15%;">pie</th> <th style="width: 15%;">data</th> <th style="width: 15%;">computer</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">cherry</td> <td style="text-align: center;">442</td> <td style="text-align: center;">8</td> <td style="text-align: center;">2</td> </tr> <tr> <td style="text-align: center;">digital</td> <td style="text-align: center;">5</td> <td style="text-align: center;">1683</td> <td style="text-align: center;">1670</td> </tr> <tr> <td style="text-align: center;">information</td> <td style="text-align: center;">5</td> <td style="text-align: center;">3982</td> <td style="text-align: center;">3325</td> </tr> </tbody> </table> <p style="text-align: center;">Use Cosine Similarity to show which of the words: <i>cherry</i> or <i>digital</i> is closer in meaning to <i>information</i></p>		pie	data	computer	cherry	442	8	2	digital	5	1683	1670	information	5	3982	3325	<b>10 Marks</b>	<b>L2</b>	<b>CO3</b>
	pie	data	computer																		
cherry	442	8	2																		
digital	5	1683	1670																		
information	5	3982	3325																		
	<b>b.</b>	<p>Illustrate the vector for the analogy:</p> <p style="text-align: center;">Paris – France + Italy</p> <p>given the following word vectors:</p> $\vec{Paris} = (0.9, 0.7), \vec{France} = (0.6, 0.4), \vec{Italy} = (0.5, 0.3)$	<b>10 Marks</b>	<b>L2</b>	<b>CO2</b>																
<b>Or</b>																					
<b>7.</b>	<b>a.</b>	<p>Given the following word-context matrix, Use the cosine similarity to show the similarity between neural and deep.</p> <p style="text-align: center;"><b>network data model</b></p> <table style="margin-left: auto; margin-right: auto;"> <tr> <td style="padding-right: 10px;">neural</td> <td style="padding-right: 10px;">2100</td> <td style="padding-right: 10px;">1800</td> <td>1950</td> </tr> <tr> <td style="padding-right: 10px;">deep</td> <td style="padding-right: 10px;">2050</td> <td style="padding-right: 10px;">1750</td> <td>1900</td> </tr> <tr> <td style="padding-right: 10px;">random</td> <td style="padding-right: 10px;">120</td> <td style="padding-right: 10px;">60</td> <td>80</td> </tr> </table>	neural	2100	1800	1950	deep	2050	1750	1900	random	120	60	80	<b>10 Marks</b>	<b>L2</b>	<b>CO3</b>				
neural	2100	1800	1950																		
deep	2050	1750	1900																		
random	120	60	80																		
	<b>b.</b>	<p>“Machine learning builds intelligent systems”. Assume a context window size of <math>\pm 1</math>.</p> <ul style="list-style-type: none"> <li>• List all the (target, context) word pairs generated from the sentence.</li> <li>• Show the word-context co-occurrence matrix for the words: Machine, learning, builds, intelligent, systems.</li> </ul>	<b>10 Marks</b>	<b>L2</b>	<b>CO2</b>																
<b>8.</b>	<b>a.</b>	<p>Consider two strings  <math>x = \text{“network”}</math> and <math>y = \text{“framework”}</math>.            List all common n-grams from unigram to 4-gram level.            Demonstrate the Histogram Intersection String Kernel value <math>HISK(x, y)</math>.            Normalize the kernel value using <math>HISK_{normal}(x, y)</math>.</p>	<b>10 Marks</b>	<b>L2</b>	<b>CO1</b>																

	<b>b.</b>	Apply the dynamic programming approach to Compute the Minimum Edit Distance between X = “distance” and Y = “editing” Assume the cost of: Insertion = 1; Deletion = 1; Substitution = 2	<b>10 Marks</b>	<b>L3</b>	<b>CO1</b>									
<b>Or</b>														
<b>9.</b>	<b>a.</b>	List and explain the stages of processing in NLP	<b>10 Marks</b>	<b>L2</b>	<b>CO1</b>									
	<b>b.</b>	Identify the values of TP, FP, FN, and TN. Also, Compute Accuracy, Precision, Recall and F1-score from the given confusion matrix:  <table style="margin-left: auto; margin-right: auto; border: none;"> <tr> <td></td> <td style="text-align: center;">Predicted Positive</td> <td style="text-align: center;">Predicted Negative</td> </tr> <tr> <td style="text-align: right;">Actual Positive</td> <td style="text-align: center;">180</td> <td style="text-align: center;">20</td> </tr> <tr> <td style="text-align: right;">Actual Negative</td> <td style="text-align: center;">30</td> <td style="text-align: center;">270</td> </tr> </table>		Predicted Positive	Predicted Negative	Actual Positive	180	20	Actual Negative	30	270	<b>10 Marks</b>	<b>L2</b>	<b>CO1</b>
	Predicted Positive	Predicted Negative												
Actual Positive	180	20												
Actual Negative	30	270												