Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 07/10/2026 **Time:** 02:00 PM - 03:30 PM

Course Code: ISE2508 **Course Name:** Natural Language Processing

Semester: Fifth Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	50	40

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	State true or false. The sentence "I say a boy with a telescope" is not ambiguous.	2	CO1	BT1
2	State the NLP application which uses a parallel corpus to learn alignments between words and phrases in different languages.	2	CO1	BT1
3	State the phenomenon where 2 words sound the same, but which may have different meanings.	2	CO1	BT1
4	State the range of Normalized HISK	2	CO1	BT1
5	State the formula for ReLU(x)	2	CO1	BT1

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	For the given confusion matrix, **calculate** the **accuracy** of the classifier, as well as the **precision, recall and F1 score** of **each class**. **Predicted C1 Predicted C2 Predicted C3** **Actual C1** 50 30 20 **Actual C2** 30 50 30 **Actual C3** 20 30 50	10	CO2	BT3
7	**Calculate** the edit distance between "sandeep" and "sammy". Show the working.	10	CO2	BT3

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	Consider a **bag-of-characters** representation for a word, where each word is represented as a 26-dimension vector denoting the count of each character. Eg. The word language = l + 2a + n + 2g + u + e. **Calculate** the **cosine similarity** between the pairs of words [natural, language], [language, processing] and [natural, processing]. Then, **find out** the most similar pair.	10	CO2	BT3
9	Given Z1 = [x_1, x_2, x_3] and Z2 = [a + x_1, a + x_2, a + x_3], **rewrite** Softmax(Z2) in terms of Softmax(Z1). Then, compute the value of Softmax([2024, 2025, 2026]).	10	CO2	BT2

PART D: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
10	Consider the **bag-of-character** representation as shown in Q8. Based on that, **compute** the **mean pooled** vector for the sentence "I love natural language processing"	10	CO1	BT2
11	Consider the **bag-of-character** representation as shown in Q8. Based on that, **compute** the **max pooled** vector for the sentence "I love natural language processing"	10	CO1	BT2

PART E: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
12	**Associate** the entries in **Column A** with those of **Column B** and **Column C**. **Column A** / **Column B** / **Column C** A. Sentiment Analysis / F. Syntactic Grammars / K. 1954 B. Part-of-Speech Tagging / G. Document Classification / L. Colourless green ideas sleep furiously C. Noam Chomsky / H. Machine Translation / M. Turing Test D. Alan Turing / I. Word Classification / N. Penn Treebank E. Georgetown Experiment / J. Imitation Game / O. Binary Polarity (Eg. Positive / Negative) NOTE: For your answers, you **ONLY NEED TO WRITE** the letters (Eg. AFK). No need to write all 3 entities in your group.	10	CO1	BT2
13	Consider a corpus of **N** documents. A stop word is a frequent word which is present in all the N documents. Explain how TF-IDF will penalize a word for being a stop word. Demonstrate this for a corpus which has N = 6 documents for the stop word ***there***. The counts for the word appearing in each of the documents are given below. **Term d1 d2 d3 d4 d5 d6** **there** 10 20 15 12 50 25 Use log to base 10 for the calcualtions	10	CO1	BT2