Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations – October 2025

Date: 11-10-2025 **Time:** 09.30am to 11.00am

School: SOCSE	**Program:** CSN (Computer Science and Engineering (Networks))	
Course Code: IST2001	**Course Name:** Fundamentals of Natural Language Processing	
Semester: V	**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	List the types of Text Representations.	**2 Marks**	**L1**	**CO1**
2	Define tokenization	**2 Marks**	**L1**	**CO1**
3	What are the major limitations of the Bag of Words model?	**2 Marks**	**L1**	**CO1**
4	What is Regression	**2 Marks**	**L1**	**CO2**
5	What is a Feed Forward Neural Network (FFNN), and why is it called "feed forward"?	**2 Marks**	**L1**	**CO2**

Part B

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

6.	a.	Given two text documents represented as vectors: • Document A = (1, 2, 3, 0, 2) • Document B = (2, 1, 0, 1, 1) Build the cosine similarity between the two documents. Show all steps of the calculation.	10 Marks	L3	CO1
	b.	Using the Histogram Intersection String Kernel (HISK), build the similarity between the two strings:	10 Marks	L3	CO1

		• String 1 = "banana" • String 2 = "bandana" Consider bigrams (k=2) as features for the kernel computation. Show all the steps of calculation			
		Or			
7.	a.	Construct the cosine similarity between the two document vectors: Document C = (0, 1, 3, 2) and Document D = (1, 0, 2, 1). Show all the steps of calculation	10 Marks	L3	CO1
	b.	Construct the Histogram Intersection String Kernel (HISK) between the two strings using trigrams (k = 3): • String X = "apple" • String Y = "pineapple" Show the raw HISK and the normalized HISK (so the value is between 0 and 1). Show all the steps.	10 Marks	L3	CO1

8.	a.	Consider a text classification problem where mails must be classified as Spam or Not Spam. The training dataset contains the following • Spam 1. Send us your Password 2. Review Us • Not Spam 1. Password Review 2. Send us the Review Using Naïve Bayes classifier , classify the new review: "Review us Now"	10 Marks	L2	CO2
	b.	Explain the working principle of the Naïve Bayes classifier in the context of text classification tasks in NLP. Why is it called "naïve"?	10 Marks	L2	CO2
		Or			
9.	a.	Explain the architecture and working principle of a Feed Forward Neural Network (FFNN). How does information flow from the input layer to the output layer?	10 Marks	L2	CO2
	b.	Explain the working principle of Logistic Regression for binary text classification. How does it differ from Linear Regression?	10 Marks	L2	CO2