Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations –October 2025

Date: 28-10-2025 **Time:** 11.00am to 12.30pm

School: SOCSE/SOE	**Program:**	
Course Code : CAI3427	**Course Name:** Language Models for Text Mining	
Semester: VII	**Max Marks**:50	**Weightage**: 25%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**25**	**25**			

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 importance of data preprocessing in Text Mining	2 Marks	L2	CO1
2	Define corpus in the context of NLP.	2 Marks	L1	CO1
3	Outline any two research paradigms in NLP.	2 Marks	L1	CO1
4	How does PoS tagging improve text analysis?	2 Marks	L1	CO2
5	Differentiate between stemming and lemmatization.	2 Marks	L2	CO2

Part B

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

6.	a.	Explain how string manipulation and data cleaning are performed in text preprocessing. Give examples of common cleaning techniques.	10 Marks	L2	CO1
	b.	Illustrate sequence labeling tasks such as POS tagging and Named Entity Recognition (NER). Explain their significance in NLP.	10 Marks	L3	CO1
		Or			
7.	a.	What is a Hidden Markov Model (HMM)? Explain the steps involved in building an HMM using a corpus and how it is used for sequence prediction.	10 Marks	L2	CO1
	b.	Interpret the research paradigms in NLP. Compare rule-based, statistical, and neural approaches with examples.	10 Marks	L3	CO1

8.	a.	Describe tokenization, stop word removal, stemming, and lemmatization. Discuss their impact on text mining accuracy.	10 Marks	L2	CO2
	b.	Interpret the importance of PoS tagging in Natural Language Processing. Explain how it supports downstream NLP tasks with examples.	10 Marks	L3	CO2
		Or			
9.	a.	Sketch an algorithmic workflow for text preprocessing including tokenization, stop word removal, and encoding techniques.	10 Marks	L3	CO2
	b.	Explain how preprocessing choices influence the performance of NLP models such as sentiment analysis or text classification.	10 Marks	L2	CO2