



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

PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations – October 2025	
Date: 29-10-2025	Time: 02.30pm to 04.00pm

School: SOCSE	Program: B.Tech. Big Data(CST)	
Course Code : CBD3410	Course Name: Natural Language Processing for Big Data	
Semester: VII	Max Marks:50	Weightage: 25%

CO - Levels	C01	C02	C03	C04	C05
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	What is the role of text preprocessing in NLP?	2 Marks	L1	C01
2	Why is Stop Word Removal important in text preprocessing?	2 Marks	L1	C01
3	Define Normalization and give one example of how it changes text data.	2 Marks	L2	C01
4	What is an N-gram model? Give an example of a Bigram.	2 Marks	L2	C02
5	Give two real-life applications of Named Entity Recognition (NER).	2 Marks	L1	C02

Part B

Answer the Questions. Total Marks 40M

6.	a.	Describe the importance of text preprocessing in Natural Language Processing.Explain each step of the pipeline using the sentence “Data Science is exciting.”	10 Marks	L2	C01
Or					
7.	a.	What are Regular Expressions (RegEx) in NLP? Explain their importance in text preprocessing and give examples of at least three RegEx patterns used for tasks such as cleaning data or extracting information.	10 Marks	L3	C01

8.	a.	<p>What is the Bag-of-Words (BoW) model in NLP? Explain its working, advantages.</p> <p>Construct a BoW representation for the following two sentences and interpret the result:</p> <p>“NLP is interesting.”</p> <p>“NLP is challenging.”</p>	10 Marks	L3	C01
Or					
9.	a.	<p>What is Part-of-Speech (POS) Tagging in NLP? Explain its importance and describe how POS tagging helps in various NLP applications. Illustrate the working of POS tagging with a suitable example sentence.</p>	10 Marks	L2	C01
10.	a.	<p>Explain the Trigram Language Model with an example. Using the given corpus, estimate Trigram probabilities and find the most probable next word for the sequence:</p> <p><s> Students learn ___ ?</p> <p>Corpus:</p> <p><s> Students learn NLP </s></p> <p><s> Students learn coding </s></p> <p><s> Students learn Python </s></p> <p><s> Students learn NLP </s></p> <p><s> We learn NLP </s></p> <p><s> Students love NLP </s></p> <p><s> Students learn programming </s></p>	10 Marks	L3	C02
Or					
11.	a.	<p>What is Syntax Parsing in NLP? Explain its importance and differentiate between Constituency Parsing and Dependency Parsing with suitable examples.</p>	10 Marks	L3	C02
12.	a.	<p>Define N-gram Language Model. Explain its importance in NLP. Describe Unigram, Bigram, and Trigram Models with suitable examples.</p>	10 Marks	L3	C02
Or					
13.	a.	<p>Explain the concept and working of Named Entity Recognition (NER). Discuss various entity types, the steps involved in NER processing, and common challenges faced</p>	10 Marks	L2	C02