Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – December 2025

Date: 15-12-2025 **Time:** 1.00pm to 04.00pm

School: SOIS	**Program:** BCA (AI & ML)	
Course Code: CSA3014	**Course Name:** Natural Language Processing	
Semester: V	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	CO1	CO2	CO3	CO4	CO5
Marks	30	28	22	20	NA

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. **10Q x 2M=20M**

1.	What is tokenization in NLP, and why is it important?	**2 Marks**	**L1**	**CO1**
2.	Define PoS tagging.	**2 Marks**	**L2**	**CO1**
3.	Explain one challenge in sentence boundary detection.	**2 Marks**	**L2**	**CO2**
4.	Explain the concept of n-grams in feature extraction.	**2 Marks**	**L1**	**CO2**
5.	Explain the difference between stemming and lemmatization.	**2 Marks**	**L1**	**CO1**
6.	What Tf-Idf and give its formula.	**2 Marks**	**L1**	**CO1**
7.	Consider the following two sentences: 1. "The cat sat on the mat." 2. "The dog sat on the mat." Generate BoW representation of these sentences.	**2 Marks**	**L3**	**CO3**

8.	What is Zipf's Law, and justify it using the following data: **Rank Word Frequency** 1 I 12,000 2 you 6,000 3 he 4,000 4 she 3,000 5 we 2,400	**2 Marks**	**L3**	**CO2**
9.	Mention 4 applications of sentiment analysis.	**2 Marks**	**L2**	**CO1**
10.	Given the following statements, determine whether each one is subjective or objective: 1. "The movie was thrilling, and I couldn't stop smiling throughout." **2.** "The Eiffel Tower is located in Paris, France."	**2 Marks**	**L3**	**CO2**

Part B

Answer the Questions **Total 80 Marks.**

11.	**a.**	Outline the different types of tokenizers used in NLP. What are the advantages of each?	**10 Marks**	**L1**	**CO1**
	b	Explain the different levels of Natural Language Processing (NLP) with example.	**10 Marks**	**L1**	**CO1**

or

12.	**a.**	Compare and contrast between stemming and lemmatization. Provide examples and discuss their importance in text preprocessing.	**10 Marks**	**L2**	**CO1**
	b.	Perform the basic text analytics on "The quick brown fox jumps over the lazy dog." 1. Check whether the word 'quick' belongs to that text. 2. Find out the index value of the word 'fox'. 3. Find out the rank of the word 'lazy'. 4. Print the third word of the given text. Concatenate the first and last words of the given sentence.	**10 Marks**	**L3**	**CO1**

13.	**a.**	Outline the role of Natural Language Processing (NLP) in language translation. Translate the following sentence from English to Spanish: "I am learning NLP."	**20 Marks**	**L2**	**CO2**

or

14.	**a.**	Demonstrate how n-grams are extracted from text, and analyze their application in feature extraction for NLP tasks. Perform bigram extraction for this sentence "The cute little boy is playing with the kitten." using Textblob library.	**20 Marks**	**L2**	**CO2**

15.	**a.**	Choose Python's CountVectorizer function to create a Bag of Words (BoW) model for the following paragraph: 'Artificial intelligence is revolutionizing industries by automating processes and boosting efficiency. In sectors like healthcare, finance, and manufacturing, AI assists companies in making data-driven decisions and optimizing operations. Its applications are broad, ranging from predictive analytics to natural language processing.' Generate the BoW model for the top 10 most frequent words. Provide the code implementation along with a brief interpretation of the results.	**20 Marks**	**L3**	**CO3**

Or

16.	**a.**	(i) Explain Zipf's law and its mathematical representation. How does it relate to word frequencies in a corpus? (ii) Apply Python programming to perform the following tasks using the 20newsgroups dataset: 1. Load the data and preprocess it by removing stop words and punctuation. 2. Calculate the frequency of tokens using the get_frequency() function for the top 5,000 words. Plot the actual ranks and expected ranks of word frequencies based on Zipf's law using matplotlib. Ensure the ranks are logarithmically scaled for better visualization.	**20 Marks**	**L3**	**CO3**

17.	**a.**	Analyze the differences among various types of sentiments that can be extracted from text, including polarity, intensity, and tone. Explain how these elements influence the interpretation of sentiment analysis results.	**20 Marks**	**L4**	**CO4**

		Perform sentiment analysis using Python to calculate and interpret the polarity and subjectivity scores for the sentences: "The food at this restaurant was absolutely delicious" (positive sentiment) and "The movie was boring and a waste of time" (negative sentiment). Discuss how these scores contribute to understanding the underlying sentiments in the text.			

Or

18.	**a.**	Analyze the given text data and find the frequency of various general features such as punctuation, uppercase and lowercase words, letters, digits, words, and whitespaces. 1. ['The interim budget for 2019 will be announced on 1st February.'], 2. ['Do you know how much expectation the middle-class working population is having from this budget?'], 3. ['February is the shortest month in a year.'], 4. ['This financial year will end on 31st March.']	**20 Marks**	**L4**	**CO4**