Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – MAY 2025

Date: 27-05-2025 **Time:** 09:30 am – 12:30 pm

School: SOCSE	**Program:** B. Tech-CBD	
Course Code: CSE3031	**Course Name:** Web Intelligence and Analytics	
Semester: VI	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	CO1	CO2	CO3	CO4	CO5
Marks	20	20	20	20	20

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 are the key steps involved in searching and indexing data?	**2 Marks**	**L1**	**CO1**
2.	What is an intelligent web application?	**2 Marks**	**L1**	**CO1**
3.	What is big data, and why is it important?	**2 Marks**	**L1**	**CO2**
4.	Mention two key trends in big data technology.	**2 Marks**	**L1**	**CO2**
5.	How does dataset size impact clustering performance?	**2 Marks**	**L1**	**CO3**
6.	What is the purpose of comparing multiple classifiers on the same data?	**2 Marks**	**L1**	**CO3**
7.	What is collective reasoning?	**2 Marks**	**L1**	**CO4**
8.	Explain the concept of resolution in logic.	**2 Marks**	**L1**	**CO4**
9.	Explain the concept of sequence memory.	**2 Marks**	**L1**	**CO5**
10.	What is feature selection in data analysis?	**2 Marks**	**L1**	**CO5**

Part B

Answer the Questions. **Total Marks 80M**

11.	**a.**	Presidency University College wanted to enhance the security and user experience of their online portal, where students and faculty log in to access academic resources, schedules, and personal data. Develop secure and efficient JavaScript-based login system	*20 Marks*	*L4*	*CO1*
		Or			
12.	**a.**	Explain in detail the architecture and functioning of a search engine, covering crawling, indexing, and ranking.	*10 Marks*	*L2*	*CO1*
	b.	Explain the four stages of data mining.	*10 Marks*	**L2**	*CO1*

13.	**a.**	A healthcare research institute is conducting a study on patient records and has collected a large dataset containing various medical attributes like age, blood pressure, cholesterol level, glucose level, and BMI. However, due to reasons such as manual data entry errors, incomplete surveys, or sensor malfunctions, some of the data entries are missing (NaN values). As a data scientist, your responsibility is to analyze the various imputation techniques available and implement them to fill missing values in a way that maintains the integrity of the dataset and supports accurate analysis. Write a detailed analysis of the most commonly used imputation techniques: mean, median and mode Write a Python program to demonstrate the three imputation methods using a synthetic or sample healthcare dataset	20 Marks	L4	CO2
		Or			
14.	**a.**	Mention the characteristics of Big data and explain the same.	**10 Marks**	**L1**	**CO2**
	b.	Classify the types of digital data and explain the same.	**10 Marks**	**L2**	**CO2**

15.	**a.**	Explain how a Decision Tree works. What are *splits*, *nodes*, *leaves*, and *entropy*? Write a python code for decision considering any example with necessary visualization.	**20 Marks**	**L2**	**CO3**
		Or			
16.	**a.**	List the clustering issues in large dataset and explain the briefly.	**10 Marks**	**L1**	**CO3**
	b.	Analyze the need of performance evaluation parameter such as Accuracy, recall, precision f1 score and ROC.	**10 Marks**	**L4**	**CO3**

17.	**a.**	I. Explain the basic architecture of an Artificial Neural Network, including input layer, hidden layers, and output layer. II. Describe the role of the activation function in an ANN. Name any two commonly used activation functions and explain their differences. III. What is backpropagation and how does it help in training a neural network?	**20 Marks**	**L2**	**CO5**
		Or			
18.	**a.**	What is Logics? List its types and explain the same, mentioning its limits.	10 Marks	**L2**	**CO4**
	b.	Analyze different methods for dealing with uncertainty in AI and reasoning systems.	10 Marks	**L4**	**CO4**