



PRESIDENCY UNIVERSITY

BENGALURU

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

End - Term Examinations – MAY 2025

Date: 20-05-2025

Time: 09:30 am – 12:30 pm

School: SOCSE	Program: B. Tech (CBD)	
Course Code: :CSE3032	Course Name: Streaming Data analytics	
Semester: VI	Max Marks: 100	Weightage: 50%

CO - Levels	C01	C02	C03	C04	C05
Marks	30	34	36		

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 Streaming data ?give an example	2 Marks	L1	C01
2.	Mention the resources for the data streaming	2 Marks	L1	C01
3.	Differentiate Structured and unstructured datasets	2 Marks	L1	C01
4.	How Hoeffding bound helps data clustering?	2 Marks	L1	C02
5.	What are the different types of Stream Windows?	2 Marks	L1	C01
6.	Mention the challenges faced in the data streaming.	2 Marks	L1	C01
7.	Define Poison's distribution with its equation.	2 Marks	L1	C02
8.	What is Frequent pattern mining? What is a pattern in the dataset?	2 Marks	L1	C03
9.	Mention the various algorithms used in Frequent Pattern mining?	2 Marks	L1	C03
10.	Mention various evaluation metrics.	2 Marks	L1	C03

Part B

Answer the Questions.

Total Marks 80M

11.	a.	Explain the data-stream-management system (DSMS) with architecture.	4 Marks	L1	C01
	b.	Define Entropy. Explain with an example	6 Marks	L1	C02
	c.	Explain Transactional data streams and Measurement data streams with an example	10 Marks	L2	C01
Or					
12.	a.	Give the formula to evaluate the Entropy of a given data sets	4 Marks	L1	C01
	b.	What are the characteristics of Stream Database?	6 Marks	L1	C02
	c.	Explain in detail architecture of Generic Data Streaming Management system taking each component	10 Marks	L1	C01

13.	a.	What is a very fast decision tree? How it is different from the general decision tree?	4 Marks	L1	C01
	b.	Mention the Applications of the streaming data	6 Marks	L2	C02
	c.	Write the algorithm for the FLAJOLET MARTIN to count the distinct elements in the streaming data. Apply the same for the streaming data $S = \{1,5,10,5,15,1\}$ with hash function $h(x) = x \bmod 11$ to find the unique/distinct element.	10 Marks	L3	C02

Or

14.	a.	Mention the Features of Data Streams	4 Marks	L1	C01												
	b.	A bank detects an average of 3 fraudulent transactions per day ($\lambda=3$). Using the Poisson distribution calculate the probability of exactly 5 fraudulent transactions occurring in a day.	6 Marks	L2	C02												
	c.	Define Sliding Window. A trader monitors a 5-minute sliding window for stock prices. Predict stock trends by computing the moving average of a stock price over a fixed period. <table><tr><td>t=1min</td><td>200</td></tr><tr><td>t=2min</td><td>202</td></tr><tr><td>t=3min</td><td>201</td></tr><tr><td>t=4min</td><td>203</td></tr><tr><td>t=5min</td><td>204</td></tr><tr><td>t=6min</td><td>206</td></tr></table>	t=1min	200	t=2min	202	t=3min	201	t=4min	203	t=5min	204	t=6min	206	10 Marks	L3	C02
t=1min	200																
t=2min	202																
t=3min	201																
t=4min	203																
t=5min	204																
t=6min	206																

15.	a.	Give the relation for the Hoeffding	04 Marks	L1	C01
	b.	Write the Hoeffding Tree Algorithm to construct the VFDT	06 Marks	L2	C02

	c.	Construct the decision tree using the following data sets by measuring the information gain.	10 Marks	L3	C03																		
		<table><tr><th>Salary</th><th>Property</th><th>Loan status</th></tr><tr><td>15000</td><td>Yes</td><td>approved</td></tr><tr><td>32000</td><td>No</td><td>approved</td></tr><tr><td>21000</td><td>Yes</td><td>approved</td></tr><tr><td>10000</td><td>No</td><td>Not approved</td></tr><tr><td>13500</td><td>no</td><td>Not approved</td></tr></table>	Salary	Property	Loan status	15000	Yes	approved	32000	No	approved	21000	Yes	approved	10000	No	Not approved	13500	no	Not approved			
Salary	Property	Loan status																					
15000	Yes	approved																					
32000	No	approved																					
21000	Yes	approved																					
10000	No	Not approved																					
13500	no	Not approved																					

Or

16.	a.	Define 0-1 Loss Function	4 Marks	L1	C01																					
	b.	Mention the properties of cluster algorithm	6 Marks	L2	C02																					
	c.	For a given dataset find the clustering using Average Linkage Technique. Use the Euclidean's Distance formula and draw the dendrogram <table><tr><td></td><td>X</td><td>Y</td></tr><tr><td>P1</td><td>0.40</td><td>0.53</td></tr><tr><td>P2</td><td>0.22</td><td>0.38</td></tr><tr><td>P3</td><td>0.35</td><td>0.32</td></tr><tr><td>P4</td><td>0.26</td><td>0.19</td></tr><tr><td>P5</td><td>0.08</td><td>0.41</td></tr><tr><td>P6</td><td>0.45</td><td>0.30</td></tr></table>		X	Y	P1	0.40	0.53	P2	0.22	0.38	P3	0.35	0.32	P4	0.26	0.19	P5	0.08	0.41	P6	0.45	0.30	10 Marks	L3	C03
	X	Y																								
P1	0.40	0.53																								
P2	0.22	0.38																								
P3	0.35	0.32																								
P4	0.26	0.19																								
P5	0.08	0.41																								
P6	0.45	0.30																								

17.	a.	Define sequence pattern mining.	4 Marks	L1	C03										
	b.	Explain the steps involved in mining sequential patterns using this algorithm.	6 Marks	L2	C03										
	c.	Consider the following sequence database: <table><tr><th>Customer</th><th>Sequences</th></tr><tr><td>C1</td><td>{A}->{B}->{C}</td></tr><tr><td>C2</td><td>{A}->{C}</td></tr><tr><td>C3</td><td>{A}->{B}->{D}</td></tr><tr><td>C4</td><td>{B}->{C}</td></tr></table> <p>i. Find all frequent 1-sequences. ii. Generate candidate 2-sequences. iii. Identify frequent 2-sequences from the candidate list. iv. List the final frequent sequential patterns.</p>	Customer	Sequences	C1	{A}->{B}->{C}	C2	{A}->{C}	C3	{A}->{B}->{D}	C4	{B}->{C}	10 Marks	L3	C03
Customer	Sequences														
C1	{A}->{B}->{C}														
C2	{A}->{C}														
C3	{A}->{B}->{D}														
C4	{B}->{C}														

Or

18.	a.	How do you validate the streaming algorithms for stationary and non-stationary environments?	4 Marks	L1	C03
	b.	Explain various Evaluation Metrics using Data Streaming.	6 Marks	L2	C03
	c.	<p>Write a program to implement Association rule mining algorithm to classify the Market Basket data set.</p> <ul style="list-style-type: none"> a. loading the dataset b. Data Pre-processing c. Model implementation d. model validation 	10 Marks	L3	C03