ABES Engineering College, Ghaziabad
B. Tech Odd Semester Sessional Test-2

Printed Pages: 3
Session: 2022 - 2023

Course Code: KDS-501
Roll No:
Course Name: INTRODUCTION TO DATA ANALYTICS AND VISUALIZATION **Date of Exam:**
Maximum Marks: 75 **Time:**

Instructions:
1. **Attempt All sections.**
2. **If require any missing data, then choose suitably.**

Q. No.	Question	Marks	CO	KL	PI
Section-A		**Total Marks : 5*2 =10**			
1	**Attempt ALL Parts**				
a)	Compare the stratified and cluster sampling methods.	**2**	**CO3**	K2	2.2.4
b)	Interpret the Hierarchical clustering approach.	**2**	**CO4**	K2	1.3.1
c)	Explain the antecedent and consequent in Market Basket Analysis.	**2**	**CO4**	K2	1.2.1
d)	Summarize the concept of Human Vision.	**2**	**CO5**	K2	3.1.1
e)	Summarize the techniques used for Data Visualization.	**2**	**CO5**	K2	3.2.1
Section-B		**Total Marks : 3*5 = 15**			
2	**Attempt ANY ONE part from the following**				
a)	Introduce the concept of Bloom Stream Filtering. Justify the statement "Bloom filter always produce True Negatives and False Positives" by implementing an example.	**5**	**CO3**	K3	4.1.3
b)	Demonstrate the Estimating Moment in stream computing. $$Kth - moment = \sum_{i \in A} (mi)^k$$ Calculate 0^{th} Moment, 1^{st} Moment, and 2^{nd} Moment of given stream: {10,9,9,9,9,9,9,9,9,9,9}.	**5**	**CO3**	K3	4.1.3
3	**Attempt ANY ONE part from the following**				
a)	Implement the three main components of Market Basket Analysis, Confidence, Support, and Lift, using a favorable dataset.	**5**	**CO4**	K3	3.2.1

b)	Show the steps involved in the Apriori algorithm for Market Basket Analysis.	**5**	**CO4**	K3	3.2.1
4	**Attempt ANY ONE part from the following**				
a)	Summarize the different tools available for effective Data Visualization.	**5**	**CO5**	K2	5.1.1
b)	Explain Human Vision and compare the human vision with computer vision.	**5**	**CO5**	K2	5.1.1
Section-C		**Total Marks : 5*10 = 50**			
5	**Attempt ANY ONE part from the following**				
a)	Determine the distinct element in the stream using the Flajolet Martin algorithm. Input stream X: {1,3,2,1,2,3,4,3,1,2,3,1} Hash function, h(x) = 6x + 1 mod 5.	**10**	**CO3**	K3	1.3.1
b)	Elaborate the process of the Alon, Matias, Szegedy (AMS) algorithm used for limited space consumption by solving the given problem: **Stream:** {a, b, c, b, d, a, c, d, a, b, d, c, a, a, b} **Length of Stream:** 15 **Random 3 Positions:** c, d, a	**10**	**CO3**	K3	1.3.1
6	**Attempt ANY ONE part from the following**				
a)	Explain the PCY algorithm for handling the extensive data in the main memory by taking suitable data items.	**10**	**CO4**	K2	2.1.3
b)	Summarize the clustering approach. Explain different kinds of clustering methods to achieve better-segregated groups.	**10**	**CO4**	K2	2.1.3
7	**Attempt ANY ONE part from the following**				
a)	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9) Initial cluster centers are: A1(2, 10), A4(5, 8) and A7(1, 2). The distance function between two points a = (x1, y1) and b = (x2, y2) is defined as- P(a, b) = |x2 – x1| + |y2 – y1| Use K-Means Algorithm to find the three cluster centers after the second iteration.	**10**	**CO4**	K3	2.4.1

<table>
<tr><td>b)</td><td>For the following given Transaction Data-set, Generate Rules using Apriori Algorithm.
Consider the values as Support=50% and Confidence=75%
<table><tr><th>Transaction ID</th><th>Items Purchased</th></tr><tr><td>1</td><td>Bread, Cheese, Egg, Juice</td></tr><tr><td>2</td><td>Bread, Cheese, Juice</td></tr><tr><td>3</td><td>Bread, Milk, Yogurt</td></tr><tr><td>4</td><td>Bread, Juice, Milk</td></tr><tr><td>5</td><td>Cheese, Juice, Milk</td></tr></table>Given Support = 50% and Confidence = 75%</td><td>10</td><td>CO4</td><td>K3</td><td>2.4.1</td></tr>
<tr><td>8</td><td>Attempt ANY ONE part from the following</td><td></td><td></td><td></td><td></td></tr>
<tr><td>a)</td><td>Visualization is needed to present the facts available in the unstructured datasets. Explain, in brief, the most challenging issues that occur during effective data visualization in real life.</td><td>10</td><td>CO5</td><td>K2</td><td>2.4.4</td></tr>
<tr><td>b)</td><td>Although data visualization is a popular mechanism to get insight from data, it has various limitations. Discuss.</td><td>10</td><td>CO5</td><td>K2</td><td>2.4.4</td></tr>
<tr><td>9</td><td>Attempt ANY ONE part from the following</td><td></td><td></td><td></td><td></td></tr>
<tr><td>a)</td><td>Interpret the Design Exploration of Complex Information Space to provide a foundation for making design decisions.</td><td>10</td><td>CO5</td><td>K2</td><td>2.3.1</td></tr>
<tr><td>b)</td><td>Explain the Space Perception and Data in Space in human vision.</td><td>10</td><td>CO5</td><td>K2</td><td>2.3.1</td></tr>
</table>

CO Course Outcomes mapped with respective question
KL Bloom's knowledge Level (K1, K2, K3, K4, K5, K6)
K1- Remember, K2- Understand, K3-Apply, K4- Analyze, K5: Evaluate, K6- Create