

University of Mumbai

Examinations Commencing from 18th July 2022 to 25th July 2022

Program: MCA

Curriculum Scheme: MCA (2year – R-2020-21 Course)

Examination: M.C.A Semester I

Course Code: MCA13 and Course Name: Advanced Database Management System

Time: 2 Hrs 30 min

Max. Marks: 80

DATE: 22/07/2022

QP CODE:95943

Q1.	Choose the correct option for following questions. All the Questions are compulsory and carry equal marks
1.	Data mining is the process of finding _____, novel, useful and actionable patterns in large volume of data. Which of the following terms best fills the gap above?
Option A:	Voluminous
Option B:	Heterogeneous
Option C:	Valid
Option D:	Noisy
2.	Simple regression assumes a _____ relationship between the input attribute and output attribute.
Option A:	quadratic
Option B:	linear
Option C:	inverse
Option D:	reciprocal
3.	If dimensionality reduction is performed on a record data matrix, the transformed data matrix has _____
Option A:	Reduced number of rows
Option B:	Reduced number of columns
Option C:	Reduced number of both rows and columns
Option D:	Same number of rows and columns
4.	The process of viewing the cross-tab (Single dimensional) with a fixed value of one attribute is
Option A:	Slicing
Option B:	Dicing
Option C:	Pivoting
Option D:	Both Slicing and Dicing
5.	If a store has N items, the number of possible item sets is:
Option A:	2N-1
Option B:	2 ^N -1
Option C:	N/2
Option D:	N-1
6.	To increase the throughput and response time of the system especially when there are lots of smaller queries, which type of parallelism is most likely to perform better?

Option A:	Intra-query parallelism
Option B:	Inter-query parallelism
Option C:	Inter-operation parallelism
Option D:	Intra-operation parallelism
7.	Entropy is measure of _____ in set.
Option A:	order
Option B:	Attributes
Option C:	disorder
Option D:	data size
8.	The main purpose for structure mining is to extract previously unknown relationships between_____.
Option A:	Web pages
Option B:	Web hyperlinks
Option C:	Web data
Option D:	Web Contents
9.	Which of the following refers to the problem of finding abstracted patterns (or structures) in the unlabeled data?
Option A:	Supervised learning
Option B:	Unsupervised learning
Option C:	Hybrid learning
Option D:	Reinforcement learning
10.	In K-Means clustering if, there are total five elements 1,2,3,4 and 5. If they are classified into two clusters. C1{1,4,5}, and C2{2,3} in first iteration. And if the cluster allocation is same for second iteration, then what will be the next step of K-Means algorithm.
Option A:	stop the iteration
Option B:	Continue with the iteration
Option C:	Change the centroid
Option D:	Nothing

Q2	Solve any Two Questions out of Three. Each question carries 10 marks.
A	Define OLAP. Illustrate with an example different OLAP operations: Roll-Up, Drill-Down, Slice, Dice and Pivot.

	Apply naïve baye's algorithm and predict the Class for unseen sample {Can_Fly=Yes, Live_in_Water=No, Have_Legs=Yes}																																												
B	<table border="1"> <thead> <tr> <th>Can_Fly</th> <th>Live_in_Water</th> <th>Have_Legs</th> <th>Class</th> </tr> </thead> <tbody> <tr> <td>Yes</td> <td>Yes</td> <td>Yes</td> <td>Mammals</td> </tr> <tr> <td>Yes</td> <td>Yes</td> <td>Yes</td> <td>Non-Mammals</td> </tr> <tr> <td>Yes</td> <td>Yes</td> <td>Yes</td> <td>Mammals</td> </tr> <tr> <td>No</td> <td>Yes</td> <td>Yes</td> <td>Non-Mammals</td> </tr> <tr> <td>No</td> <td>Yes</td> <td>No</td> <td>Mammals</td> </tr> <tr> <td>No</td> <td>No</td> <td>No</td> <td>Non-Mammals</td> </tr> <tr> <td>No</td> <td>No</td> <td>No</td> <td>Mammals</td> </tr> <tr> <td>No</td> <td>No</td> <td>Yes</td> <td>Non-Mammals</td> </tr> <tr> <td>Yes</td> <td>No</td> <td>No</td> <td>Non-Mammals</td> </tr> <tr> <td>Yes</td> <td>Yes</td> <td>No</td> <td>Mammals</td> </tr> </tbody> </table>	Can_Fly	Live_in_Water	Have_Legs	Class	Yes	Yes	Yes	Mammals	Yes	Yes	Yes	Non-Mammals	Yes	Yes	Yes	Mammals	No	Yes	Yes	Non-Mammals	No	Yes	No	Mammals	No	No	No	Non-Mammals	No	No	No	Mammals	No	No	Yes	Non-Mammals	Yes	No	No	Non-Mammals	Yes	Yes	No	Mammals
Can_Fly	Live_in_Water	Have_Legs	Class																																										
Yes	Yes	Yes	Mammals																																										
Yes	Yes	Yes	Non-Mammals																																										
Yes	Yes	Yes	Mammals																																										
No	Yes	Yes	Non-Mammals																																										
No	Yes	No	Mammals																																										
No	No	No	Non-Mammals																																										
No	No	No	Mammals																																										
No	No	Yes	Non-Mammals																																										
Yes	No	No	Non-Mammals																																										
Yes	Yes	No	Mammals																																										
C	What is Clustering? Explain K mean Clustering in detail.																																												
Q3	Solve any Two Questions out of Three. Each question carries 10 marks.																																												
A	Define Classification. Explain Decision tree algorithm with suitable example																																												
B	<p>What is Market Basket Analysis? Find out frequent itemsets and strong association rule from the given transaction using Apriori Algorithm with Min_Support of 40% and Confidence of 70%.</p> <table border="1"> <thead> <tr> <th>TID</th> <th>Items</th> </tr> </thead> <tbody> <tr> <td>T1</td> <td>Bread, Cheese, Juice, Eggs</td> </tr> <tr> <td>T2</td> <td>Bread, Cheese, Juice</td> </tr> <tr> <td>T3</td> <td>Bread, Milk, Yogurt</td> </tr> <tr> <td>T4</td> <td>Bread, Juice, Milk</td> </tr> <tr> <td>T5</td> <td>Cheese, Juice, Milk</td> </tr> </tbody> </table>	TID	Items	T1	Bread, Cheese, Juice, Eggs	T2	Bread, Cheese, Juice	T3	Bread, Milk, Yogurt	T4	Bread, Juice, Milk	T5	Cheese, Juice, Milk																																
TID	Items																																												
T1	Bread, Cheese, Juice, Eggs																																												
T2	Bread, Cheese, Juice																																												
T3	Bread, Milk, Yogurt																																												
T4	Bread, Juice, Milk																																												
T5	Cheese, Juice, Milk																																												
C	Explain different data storing techniques available in distributed DBMS																																												
Q4	Solve any Two Questions out of Three. Each question carries 10 marks.																																												
A	Explain Knowledge Discovery Process (KDD) in detail. What is the role of data mining in KDD process?																																												
B	Describe in detail Data warehouse architecture and ETL process.																																												
C	Explain text mining and discuss in brief the Information retrieval methods.																																												