



Roll No:

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

MCA
(SEM IV) THEORY EXAMINATION 2021-22
DATA ANALYTICS

Time: 3 Hours**Total Marks: 100****Note:** Attempt all Sections. If you require any missing data, then choose suitably.**SECTION A****1. Attempt all questions in brief.****2*10 = 20**

Qno	Questions	CO
(a)	Explain various sources of data analytics.	1
(b)	What is Principal Component Analysis (PCM).	2
(c)	Describe types of data.	1
(d)	Explain the need of data analytics.	1
(e)	Describe data streaming.	3
(f)	Explain stochastic search method.	2
(g)	What is Competitive learning?	2
(h)	Explain ProCLUS.	4
(i)	Briefly explain filtering stream.	3
(j)	What are various analytical tools?	1

SECTION B**2. Attempt any three of the following:****10*3 = 30**

Qno	Questions	CO
(a)	Describe Hadoop and its architecture.	5
(b)	Explain all phases of data analytics lifecycle.	1
(c)	Describe the concept of extracting fuzzy models from data.	2
(d)	Explain K-means clustering with example.	4
(e)	Discuss stock market prediction case study.	3

SECTION C**3. Attempt any one part of the following:****10*1 = 10**

Qno	Questions	CO												
(a)	Apply Apriori algorithm for the following given transaction dataset, to generate rules. Consider the values as Support=50%, and Confidence=75%. <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Transaction ID</th> <th>Items Purchased</th> </tr> </thead> <tbody> <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> </tbody> </table>	Transaction ID	Items Purchased	1	Bread, Cheese, Egg, Juice	2	Bread, Cheese, Juice	3	Bread, Milk, Yogurt	4	Bread, Juice, Milk	5	Cheese, Juice, Milk	4
Transaction ID	Items Purchased													
1	Bread, Cheese, Egg, Juice													
2	Bread, Cheese, Juice													
3	Bread, Milk, Yogurt													
4	Bread, Juice, Milk													
5	Cheese, Juice, Milk													
(b)	What are NoSQL databases and its types?	5												

4. Attempt any one part of the following:**10 *1 = 10**

Qno	Questions	CO
(a)	Describe market based modeling with suitable example.	4
(b)	Apply Linear Regression algorithm to predict the salary with 10 years of experience using the method of least square. To solve the problem,	1



Roll No:

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

MCA
(SEM IV) THEORY EXAMINATION 2021-22
DATA ANALYTICS

	use the following data:	
	Experience (x) in years	Salary (y) in dollars
	3	30
	8	57
	9	64
	13	72
	3	36
	6	43
	11	59
	21	90
	1	20
	16	83

5. Attempt any *one* part of the following: 10*1 = 10

Qno	Questions	CO
(a)	What is Support Vector machine (SVM)? Explain it with suitable example.	2
(b)	Apply Bayes Theorem. Suppose you have to draw a single card from a standard deck of 52 cards. Now the probability that the card is a Queen is $P(\text{Queen}) = 4/52$. If you are given evidence that the card that you have picked is a face card, then calculate the posterior probability using Bayes theorem.	2

6. Attempt any *one* part of the following: 10*1 = 10

Qno	Questions	CO
(a)	What is Real time analytics platform? Explain it with "Real time sentiment analysis" case study.	3
(b)	Explain hierarchical clustering with suitable example.	4

7. Attempt any *one* part of the following: 10*1 = 10

Qno	Questions	CO
(a)	Explain MapReduce, Hive, Pig, and HBase in Data Analytics.	5
(b)	Apply PCY algorithm on following transactions on following transaction to find the candidate sets. Assume threshold value = 3, and hash function $h(x) = (i*j) \bmod 10$. T1 = {1, 2, 3} T2 = {2, 3, 4} T3 = {3, 4, 5} T4 = {4, 5, 6} T5 = {1, 3, 5} T6 = {2, 4, 6} T7 = {1, 3, 4} T8 = {2, 4, 5} T9 = {3, 4, 6} T10 = {1, 2, 4} T11 = {2, 3, 5} T12 = {3, 4, 6}	4