Roll No.....

MCA- Examination November 2022
Final Year

Subject Code: 680313 **Subject Name: Data Mining and Warehousing**

Time: 3 Hours Maximum Marks: 70 Minimum Pass Marks: 28

Note :		1. *Answer all five questions. All questions carry equal marks.* 2. *In each question part a, b, c are compulsory and part d has internal choice. Out of which part a & b (Max. 50 words) carry 2 marks, part c (Max. 100 words) carry 3 marks and part d (Max. 400 words) carry 7 marks. Word limit would not be followed for (except) diagram, numerical, derivation etc.* 3. *All Parts of each question are to be attempted at one place.* 4. *Assume suitable value for missing data, if any.*		
Question No.			**Marks**	
1.	(a)	Describe basic characteristics of Data Warehouse.	02	CO1
	(b)	Explain the term "Data Mart".	02	CO2
	(c)	Discuss the advantages and disadvantages of Data Warehousing in an organization.	03	CO2
	(d)	For the Dominos explore the dimension table, fact table. and the three classes of schemas. Also write the query to find out the total sales of classic pizza during year 2021.	07	CO6
		OR		
	(e)	Justify your view about -"Availability of Metadata repository can affect recovery of data warehouse".	07	CO6
2.	(a)	Define Virtual Warehouse.	02	CO1
	(b)	Describe process of Materialization in warehouse.	02	CO1
	(c)	Analyse roll-up, drill-down, slice and dice operation using a cube for dimension time, location, supplier and item.	03	CO4
	(d)	Select appropriate features to clearly compare the On Line Transactional Processing with On Line Analytical Processing.	07	CO5
		OR		
	(e)	Suppose that Maruti sales database includes the following attributes: customer, address, income, profile-credit (excellent, good, bad) based on income range. a) Propose a concept hierarchy for each of the attribute b) For each concept hierarchy proposed in a) what type of concept hierarchy is it?	07	CO5
3.	(a)	Define Data Mining.	02	CO1
	(b)	Discuss the use of correlation in establishing a relationship between data variables.	02	CO2
	(c)	Discuss whether or not each of the following activities is a data mining task. Support your view with proper explanation: 1. Dividing the customer of a company according to their gender 2. Extracting the frequencies of a sound wave. 3. Sorting a student database based on student identification number	03	CO2
	(d)	Smooth the given data which reflects the frequency of electricity problem during a cricket match on various days. 4,8,15,21,21,24,25,28,34.	07	CO3
		OR		

	(e)	For the following vectors , X and Y, calculate the indicated similarity or distance measures 1. X=(1,1,1,1), Y= (2,2,2,2) cosine, correlation, Euclidean 1. X=(1.0.0.0.0.0.0.0.0.0), Y= (0,0,0,0,0,0,1,0,0,1) cosine, correlation, Euclidean, Jaccard 2. X=(0,-1,0,1), Y= (1,0,-1,0) cosine, correlation, Euclidean 3. X=(3,2,0,5,0,0,0,2,0,0), Y= (1,0,0,0,0,0,0,1,0,2) cosine, correlation, Jaccard	07	CO3
4.	(a)	List two differences between Hierarchical algorithm and divisive algorithm	02	CO1
	(b)	Describe neural network as a technique of data mining.	02	CO1
	(c)	Depict the issues faced by decision tree when used for classification?	03	CO2
	(d)	Analyse the problems that could occur when clustering is applied to real world databases. Also write and explain K-Mean Clustering algorithm for clustering the data set D= { 2, 4, 10, 12, 3, 20, 30, 11, 25} and sets of clusters to be 2.	07	CO4
		OR		
	(e)	Analyze the difference between ID3 and C4.5	07	CO4
5.	(a)	Describe PCA	02	CO1
	(b)	Illustrate applications of Web mining	02	CO1
	(c)	Discuss how will you measure the utility of a rule?	03	CO2
	(d)	Discuss thoroughly the impact of curse of dimensionality. Investigate the mechanism used to reduce the dimensions.	07	CO2
		OR		
	(e)	Discuss the frequent itemsets? Illustrate Apriori algorithm taking example of your choice.	07	CO2
		* * * * * * *		