Question Paper

Exam Date & Time: 25-Jul-2022 (09:00 AM - 12:00 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

VI SEMESTER B. TECH (Data Warehousing and Data Mining)
MAKE UP EXAMINATIONS, JULY 2022
[ICT 3253]

DATA WAREHOUSING AND DATA MINING [ICT 3253]

Marks: 50 **Duration: 180 mins.**

DESCRIPTIVE TYPE

Answer all the questions.

Answer all questions
Missing data, if any may be suitably assumed

1) If Hopkins Statistic H>0.5, then the dataset D is significantly a clusterable data. Support your answer with the help of an example. (5)

2) Compare and contrast between OLAP and OLTP systems. (3)

3) Design and explain three tier data warehousing architecture. (2)

4) Consider the data points given in Figure Q.2A and min_points=2. Answer the following questions with respect to DBSCAN algorithm. (5)

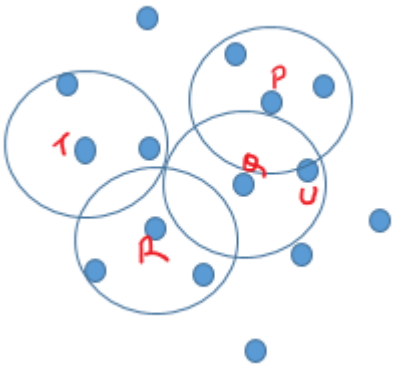

Fig Q.2A

i. Which among the named points are core points? Why?

ii. Is 'U' directly density reachable from 'Q'. Defend your answer

iii. Are there any named points which are density reachable from any other point. Defend your answer.

5) Show with an example the effect of outliers on k-means algorithm. (3)

6) Compare and contrast CLARA and CLARANS (2)

7) Obtain clusters by applying k-means algorithm for the data shown in Table Q.2B. Assume k=2 and initial centroids as (1,1) and (5,7) (5)

and (5,7)

Table Q.2B

Individual	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

8) Table Q.3B shows ticket prices (in $) for Padman and Black Panther movies respectively in Big Cinemas, Manipal. Find the covariance between the two movies and also state the type of covariance between the two movies (3)

Table Q.3B

Days of the week	Padman	Black Panther
1	7	21
2	6	11
3	5	15
4	4	6
5	3	6

9) Classify the following attributes as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation, so briefly indicate your reasoning. (**Example:** Age in years. **Answer:** Discrete, quantitative, ratio) (2)

i. Brightness as measured by a light meter.

ii. Angles as measured in degrees between 0◦ and 360◦.

iii. Bronze, Silver, and Gold medals as awarded at the Olympics.

10) A database has five transactions as given in Table Q.4A. Let min sup = 60% and min conf = 80%. Find all frequent itemsets using Apriori and FP_Growth algorithms, respectively. (5)

Table Q.4A

TID	Items_bought
T100	M,O,N,K,E,Y

T100	M,O,N,K,E,Y
T200	D,O,N,K,E,Y
T300	M,A,K,E
T400	M,U,C,K,Y
T500	C,O,O,K,I,E

11) Consider attributes associated with Objects A and B as given in theTable Q.4B. Find the distances between then using Euclidean, Minkowski and CityBlock distance methods. Also, mention the distance between the above three distance methods. (3)

Table Q.4B

	Cost	Time	Weight	Incentive
A	0	3	4	5
B	7	6	3	-1

12) Use the two methods below to *normalize* the following group of data: (2)

200*;* 300*;* 400*;* 600*;* 1000

(a) min-max normalization by setting *min* = 0 and *max* = 1

(b) z-score normalization

13) Why is naive Bayesian classification called "naive"? Briefly outline the major ideas of naive Bayesian classification. (5)

14) Suppose that a group of 1320 students were surveyed. The location of each student was noted. Each student was polled as to whether his or her stay is near university or not. Observed values are given in Contingency Table as in Table Q.5B. Find the correlation at 0.001 significance level. (3)

Assume: For 1 degree of freedom, the $\chi 2$ value needed to reject the hypothesis at the 0.001 significance level is 10.828

Table Q.5B

	Graduated Student	Current Student	Sum
Near University	140	190	330
Not Near University	40	950	990
Sum	180	1140	1320

15) Identify whether the following task requires data mining or not (2)

i) By looking at a CT scan, a doctor wants to classify if a patient id covid +ve or not. He uses many labeled CT scans for making the decision.

ii) Monitoring heart rate of a patient for abnormalities

iii) Predicting the outcome of tossing a fair pair of dice

iv) Extracting the frequencies of a sound wave

-----End-----