Printed Pages:02 **Sub Code: NCS 066**

Paper Id: | 1 | 1 | 0 | 6 | 2 | 6 | **Roll No.** | | | | | | | | | | |

B.Tech

(SEM VI) THEORY EXAMINATION 2017-18

DATAWAREHOUSING AND DATA MINING

Time: 3 Hours ***Total Marks: 100***

Note: 1. Attempt all Sections. If require any missing data; then choose suitably.

SECTION A

1. Attempt *all* questions in brief. 2 x 10 = 20

a. Draw the diagram for key steps of data mining.
b. Define the term Support and Confidence.
c. What are attribute selection measures? What is the drawback of information gain?
d. Differentiate between classification and clustering
e. Write the statement for Apriori Algorithm.
f. What are the drawbacks of k-mean algorithm?
g. What is Chi Square test?
h. Compare Roll up, Drill down operation.
i. What are Hierarchal methods for clustering?
j. Name main features of Genetic Algorithm.

SECTION B

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

a. Explain the data mining / knowledge extraction process in detail?
b. Differentiate between OLAP and OLTP.
c. Find frequent patterns and the association rules by using Apriori Algorithm for the following transactional database:

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

Let Minimum support= 60% and Minimum Confidence= 80%

d. What are different database schemas .shows with an example?
e. How data back- up and data recovery is managed in data warehouse?

SECTION C

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

(a) Draw the 3-tier data warehouse architecture. Explain ETL process.

(b) Elaborate the different strategies for data cleaning.

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

(a) What are different clustering methods? Explain STING in detail.

(b) What are the applicatio Explain web mining and spatial mining.

5. Attempt any *one* part of th 10 x 1 = 10

(a) Define data warehouse be taken care while designing a warehouse?

quescol.com

(b) Write short notes on the following:
(i) Concept Hierarchy
(ii) ROLAP vs MOLAP
(iii) Gain Ratio
(iv) Classification Vs Clustering

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

(a) Compute the decision rules by deriving a decision tree classifier and information gain as selection measure for the given database in table.

Table 6

Age	Income	Student	Credit rating	Class : buys computer
youth	high	No	Fair	No
youth	high	No	Excellent	No
middle aged	high	No	Fair	Yes
senior	medium	No	Fair	Yes
senior	low	Yes	Fair	Yes
senior	low	Yes	Excellent	No
middle aged	low	Yes	Excellent	Yes
youth	medium	No	Fair	No
youth	low	Yes	Fair	Yes
senior	medium	Yes	Fair	Yes
youth	medium	Yes	Excellent	Yes
middle aged	medium	No	Excellent	Yes
middle aged	high	Yes	Fair	Yes
senior	medium	No	Excellent	No

Given: Gain (age) = 0.246, Gain (student) =0.151 and Gain (Credit Rating) =0.048

b) What is Laplacian Correction in Bayesian Classifier? Compute the class of the for following tuple by using Bayesian classification for given database in **table 6**.
X= (Age = senior, Credit rating = fair, Income= medium, student= no)

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

(a) Write the k- mean algorithm. Suppose that the data mining task is to cluster points (with (x,y) representing location) into three clusters , where the points are:
A1 (2, 10), A2 (2, 5) A3 (8, 4)
B1 (5, 8), B2 (7, 5) B3 (6, 4)
C1 (1, 2), C2 (4, 9)
The distance function is Euclidian distance. Suppose initially we assign A1, B1, and C1 as the center of y. Use the k- means algorithm to show only
The three cluste und of execution.

(b) What is Hierarch g? Explain BIRCH method.