www.binils.com Anna University | Polytechnic | Schools



Reg. No.:											
-----------	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code: X10301

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2021 Sixth Semester

Computer Science and Engineering
CS 8075 – DATA WAREHOUSING AND DATA MINING
(Common to Computer and Communication Engineering)
(Regulations 2017)

Time: Three Hours

Maximum: 100 Marks

Answer ALL questions

PART - A (10×2=20 Marks)

- 1. Define data warehouse metadata.
- 2. Name the schemas a multidimensional data model can adopt.
- 3. What is data integration?
- 4. Name the steps in the process of knowledge discovery.
- 5. What is an association rule? Give example.
- 6. Present an outline of frequent pattern growth method.
- 7. What is decision tree induction?
- 8. Outline how tree pruning algorithms attempt to improve accuracy.
- 9. How the Euclidean distance between two points or tuples, say, $X_1 = (x_{11}, x_{12}, ...x_{1n})$ and $X_2 = (x_{21}, x_{22}, ... x_{2n})$ is calculated?
- 10. What is a rule-based classifier?

11. a) What is lattice of cuboids? Construct a lattice of cuboids forming a data cube for the representation of sales data, according to the dimensions time, item, location and supplier. (13)

(OR)

b) Outline the typical OLAP operations for multidimensional data with an example and relevant diagrams. (13)

www.binils.com Anna University | Polytechnic | Schools

X10301 -2-

12. a) How is an attribute normalized? When is normalization particularly useful?

Outline min-max normalization with an example. (13)

(OR)

- b) What is data reduction? Outline how sampling can be used as a data reduction technique with an example. (13)
- 13. a) What is correlation analysis? Outline the steps in correlation analysis using lift with an example. (13)

(OR)

- b) Outline the steps in classification based on multiple association rules with an example. (13)
- 14. a) State Bayes' theorem of posterior probability. Outline the steps in Bayesian classification with an example. (13)

(OR)

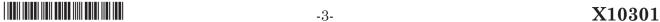
- b) What is a dendrogram? How to derive clusters from dendrogram? Outline with an example. (13)
- 15. a) What is Weka? Highlight the features of Weka. (13)

(OR)

b) Represent the data presented in the following table a IF-THEN rules, decision tree and neural network. (13)

Age	Income	Class
Youth	High	A
Youth	Low	В
Middle aged	High	С
Middle aged	Low	С
Senior	High	С
Senior	Low	С

www.binils.com Anna University | Polytechnic | Schools



PART – C (1×15=15 Marks)

16. a) Apply the Apriori algorithm for discovering frequent item sets to the following data set:

Trans ID	Items Purchased	
101	Mulberry, Raspberry, Cherry	
102	Mulberry, Papaya	
103	Papaya, Mango	
104	Mulberry, Raspberry, Cherry	
105	Passion Fruit, Cherry	
106	Passion Fruit	
107	Passion Fruit, Papaya	
108	Mulberry, Raspberry, Guava, Cherry	
109	Guava, Mango	
110	Mulberry, Raspberry	
Use 0.3 for the minimum supp	oort value.	(15)

(OR)

b) Consider five points $\{X_1, X_2, X_3, X_4, X_5\}$ with the following coordinates as a two-dimensional sample for clustering:

$$X_1 = (0, 2), X_2 = (1, 0), X_3 = (2, 1), X_4 = (4, 1)$$
 and $X_5 = (5, 3).$

Illustrate the K-means algorithm on the above data set. The required number of clusters is two and initially, clusters are formed from random distribution of samples: $C_1 = \{X_1, X_2, X_4\}$ and $C_2 = (X_3, X_5)$. (15)
