Total No. of Questions : 8]

SEAT No. :

P3337

[Total No. of Pages : 3

[5461]-597

B.E. (Computer Engineering)

DATA MINING AND WAREHOUSING

(2015 Course) (Semester - I) (End Sem.) (410244D)

Time : 2½ Hours] *[Max. Marks : 70*

Instructions to the candidates:

1) ***Answer Q1 or Q2, Q3 or Q4, Q5 or Q6, Q7 or Q8.***
2) ***Assume suitable data if necessary.***
3) ***Neat diagrams must be drawn wherever necessary.***
4) ***Figures to the right indicate full marks.***

Q1) a) For the given attribute AGE values : 16, 16, 180, 4, 12, 24, 26, 28, apply following Binning technique for smoothing the noise. **[6]**

i) Bin Medians

ii) Bin Boundaries

iii) Bin Means

b) Differentiate between Star schema and Snowflake schema. **[6]**

c) Calculate the Jaccard coefficient between Ram and Hari assuming that all binary attributes are a symmetric and for each pair values for an attribute, first one is more frequent than the second. **[8]**

Object	Gender	Food	Caste	Education	Hobby	Job
Hari	M(1)	V(1)	M(0)	L(1)	C(0)	N (0)
Ram	M(1)	N(0)	M(0)	I(0)	T(1)	N (0)
Tomi	F(0)	N(0)	H(1)	L(1)	C(0)	Y (1)

OR

Q2) a) Explain following attribute types with example. **[6]**

i) Ordinal

ii) Binary

iii) Nominal

b) Differentiate between OLTP and OLAP with example. **[6]**

P.T.O.

c) Calculate the Euclidean distance matrix for given Data points. **[8]**

point	x	y
p1	0	2
p2	2	0
p3	3	1
p4	5	1

Q3) a) A database has 6 transactions. Let minimum support = 60% and Minimum confidence = 70% **[8]**

Transaction ID	Items Bought
T1	{A, B, C, E}
T2	{A, C, D, E}
T3	{B, C, E}
T4	{A, C, D, E}
T5	{C, D, E}
T6	{A, D, E}

i) Find Closed frequent Itemsets
ii) Find Maximal frequent itemsets
iii) Design FP Tree using FP growth algorithm

b) Explain with example Multi level and Constraint based association Rule mining. **[5]**

c) How can we improve the efficiency of a-priori algorithm. **[4]**

OR

Q4) a) Consider the Market basket transactions shown below. Assuming the minimum support = 50% and Minimum confidence = 80% **[8]**

i) Find all frequent item sets using Apriori algorithm
ii) Find all association rules using Apriori algorithm

Transaction ID	Items Bought
T1	{Mango, Apple, Banana, Dates}
T2	{Apple, Dates, Coconut, Banana, Fig}
T3	{Apple, Coconut, Banana, Fig}
T4	{Apple, Banana, Dates}

b) Explain FP growth algorithm with example. **[5]**

c) Explain following measures used in association Rule mining **[4]**

i) Minimum Support
ii) Minimum Confidence
iii) Support
iv) Confidence

Q5) a) Explain the training and testing phase using Decision Tree in detail. Support your answer with relevant example. **[8]**

b) Apply KNN algorithm to find class of new tissue paper (X1 = 3, X2 = 7). Assume K = 3 **[5]**

X1 = Acid Durability (secs)	X2 = Strength(kg/sq.meter)	Y = Classification
7	7	Bad
7	4	Bad
3	4	Good
1	4	Good

c) Explain the use of regression model in prediction of real estate prices. **[4]**

OR

Q6) a) What is Bayesian Belief Network. Elaborate the training process of a Bayesian Belief Network with suitable example. **[8]**

b) Explain K-nearest neighbor classifier algorithm with suitable application. **[5]**

c) Elaborate on Associative Classification with appropriate applications. **[4]**

Q7) a) Discuss the Sequential Covering algorithm in detail. **[8]**

b) Explain following measures for evaluating classifier accuracy **[4]**

i) Specificity

ii) Sensitivity

c) Differentiate between Wholistic learning and Multi perspective learning. **[4]**

OR

Q8) a) How is the performance of Classifiers algorithms evaluated. Discuss in detail. **[8]**

b) Discuss Reinforcement learning relevance and its applications in real time environment. **[4]**

c) Explain following measures for evaluating classifier accuracy **[4]**

i) Recall

ii) Precision

♦ ♦ ♦