

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA – SEMESTER-IV • EXAMINATION – WINTER 2015

Subject Code:640005

Date:09/12/2015

Subject Name: Data Warehousing & Data Mining

Time:10.30 a.m. To 01.00 p.m.

Total Marks: 70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

Q. 01 (a) Answer The following Questions

- (i) Define The Term “Data Warehouse”. [02]
- (ii) What is ROLAP? [02]
- (iii) What is Data Mining? [02]
- (iv) Give The Full-Form of “OLAP”. [01]

(b) Answer The following Questions

- (i) Define The Term “Virtual Warehouse”. [02]
- (ii) What is Data cleaning in Data Preprocessing? [02]
- (iii) What is Data Generalization? [02]
- (iv) What is the accuracy of classifier? [01]

Q. 02 (a) Answer The following Questions

- (i) Define The Term “Supervised Learning”. [02]
- (ii) What is Learning by Observation? [02]
- (iii) What is Lift-ratio for Association Rule Mining? [02]
- (iv) What is Prediction? [01]

(b) Differentiate Operational Database Systems and Data Warehouses. [07]

OR

(b) Explain Three-Tier Data Warehouse Architecture. [07]

Q. 03 (a) Find out the Association Rules from Given Dataset with respect to [07]
Minimum Support = 60% and Minimum Confidence = 80%.
Obtain the Rules with Degree 1 only.

i.e. for any transaction, $buys(X, item1) \rightarrow buys(X, item3)$

Transaction ID	Items Bought
T001	{M, O, N, K, E, Y}
T002	{D, O, N, K, E, Y}
T003	{M, A, K, E}
T004	{M, U, C, K, Y}
T005	{C, O, O, K, I, E}

(b) Answer The following Questions

- (i) How is class comparison performed? [04]
- (ii) What is Attribute Generalization? [03]

OR

- Q. 03 (a)** Using the Following contingency table of supermarket transaction data, [07]
answer the following.

	<i>Hot – Dog</i>	$\overline{Hot – Dog}$	$\sum_{row}$
<i>Hamburgers</i>	2000	500	2500
$\overline{Hamburgers}$	1000	1500	2500
$\sum_{column}$	3000	2000	5000

- (i) Is the association rule “Hot-dog $\rightarrow$ Hamburger” Strong?
Minimum Support = 25% and Minimum Confidence = 50%.
- (ii) What kind of correlation relationship exists between the purchase of hot dogs & the purchase of hamburgers?
- (b) Answer The following Questions**
- (i) Explain the Attribute Oriented Induction. [04]
- (ii) What is Class Characterization? [03]

- Q. 04 (a)** Predict the class label “Defaulted Borrower” of the test record [07]
X = (Home Owner = No; Marital Status = Married; Income = 1,20,000)
using Naïve Bayes classifier on following data:

ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	1,25,000	No
2	No	Married	1,00,000	No
3	No	Single	70,000	No
4	Yes	Married	1,20,000	No
5	No	Divorced	95,000	Yes
6	No	Married	60,000	No
7	Yes	Divorced	2,20,000	No
8	No	Single	85,000	Yes
9	No	Married	75,000	No
10	No	Single	90,000	Yes

- (b) Answer The following Questions**
- (i) How to compare the Accuracy of Classifier? [04]
- (ii) Explain: Concept Hierarchy Generation with Binning and Histogram Analysis. [03]

OR

- Q. 04 (a)** Find out which attribute is suitable to be the perfect split point from the [07]
Given data, use information gain measure. Calculate only first split point.
“Risk Class” is the predicated attribute.

ID	Owns Home	Gender	Employed	Risk Class
1	Yes	Male	Yes	B
2	No	Female	Yes	A
3	Yes	Female	Yes	C
4	Yes	Male	No	B
5	No	Female	Yes	C
6	No	Female	Yes	A
7	No	Male	No	B
8	Yes	Female	Yes	A
9	No	Female	Yes	C
10	Yes	Female	Yes	C

- (b) Answer The following Questions**

- (i) Discuss the essential data preparing steps for classification and [04]
prediction.
(ii) How to Generate Concept Hierarchy for Categorical data? [03]

- Q. 05 (a)** Explain K-Means Algorithm with its pseudo-code and example. [07]

- (b) Answer The following Questions**

- (i) List out Four Typical cases of Data Mining in Retail Industry. [04]
(ii) Explain: Density-based Clustering method. [03]

OR

- Q. 05 (a)** What is an outlier? Explain Statistical Distribution-Based Outlier Detection. [07]

- (b) Answer The following Questions**

- (i) List out Four Typical cases of Data Mining in Telecommunication [04]
Industry.
(ii) Explain: Constraint-based Clustering method. [03]
