## **GUJARAT TECHNOLOGICAL UNIVERSITY** MCA - SEMESTER-IV • EXAMINATION - WINTER 2015

Subject Code:640005 Date:09							
	:10.30 tions: 1. At 2. M	Ime: Data Warehousing & Data Mining         ) a.m. To 01.00 p.m.       Total Marks:         Itempt all questions.       Total Marks:         Take suitable assumptions wherever necessary.       Total Marks.         gures to the right indicate full marks.       Total Marks.	70				
Q. 01	(a)	<ul> <li>Answer The following Questions <ul> <li>(i) Define The Term "Data Warehouse".</li> <li>(ii) What is ROLAP?</li> <li>(iii) What is Data Mining?</li> <li>(iv) Give The Full-Form of "OLAP".</li> </ul> </li> </ul>					
	(b)	<ul> <li>Answer The following Questions <ul> <li>(i) Define The Term "Virtual Warehouse".</li> <li>(ii) What is Data cleaning in Data Preprocessing?</li> <li>(iii) What is Data Generalization?</li> <li>(iv) What is the accuracy of classifier?</li> </ul> </li> </ul>					
Q. 02	(a) (b)	<ul> <li>Answer The following Questions <ul> <li>(i) Define The Term "Supervised Learning".</li> <li>(ii) What is Learning by Observation?</li> </ul> </li> <li>(iii) What is Lift-ratio for Association Rule Mining?</li> <li>(iv) What is Prediction?</li> <li>Differentiate Operational Database Systems and Data Warehouses.</li> </ul>					
		OR					
	<b>(b)</b>	Explain Three-Tier Data Warehouse Architecture.	[07]				
Q. 03	(a)	Find out the Association Rules from Given Dataset with respect to [( Minimum Support = 60% and Minimum Confidence = 80%. Obtain the Rules with Degree 1 only. i.e. for any transaction, $buys(X, item1) \rightarrow buys(X, item3)$ Transection ID Items Bought T001 {M, O, N, K, E, Y}					

Transection 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] [03]
- (ii) What is Attribute Generalization?

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

	Hot – Dog	$\overline{Hot-Dog}$	$\sum_{row}$
Hamburgers	2000	500	2500
Hamburgers	1000	1500	2500
$\sum_{coloum}$	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?

#### **Answer The following Questions (b)**

- (i) Explain the Attribute Oriented Induction. [04]
- (ii) What is Class Characterization?
- Q. 04 Predict the class label "Defaulted Borrower" of the test record [07] (a) X = (Home Owner = No; Marital Status = Married; Income = 1,20,000) using Naïve Bayes classifier on following data:

Sayes classifier on following data.					
ID	Home	Marital	Annual	Defaulted	
	Owner	Status	Income	Borrower	
1	Yes	Single	1,25,00 0	No	
2	No	Married	1,00,00 0	No	
3	No	Single	70,000	No	
4	Yes	Married	1,20,00 0	No	
5	No	Divorced	95,000	Yes	
6	No	Married	60,000	No	
7	Yes	Divorced	2,20,00 0	No	
8	No	Single	85,000	Yes	
9	No	Married	75,000	No	
10	No	Single	90,000	Yes	

#### **Answer The following Questions (b)**

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

OR

[03]

the predicating attribute.						
ID	Owns Home	Gender	Employed	Risk Class		
1	Yes	Male	Yes	В		
2	No	Female	Yes	Α		
3	Yes	Female	Yes	С		
4	Yes	Male	No	В		
5	No	Female	Yes	С		
6	No	Female	Yes	Α		
7	No	Male	No	В		
8	Yes	Female	Yes	Α		
9	No	Female	Yes	С		
10	Yes	Female	Yes	С		

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 predicating attribute.

### (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-Mea	ns Algorithn	n 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) (b)	What is an outlier? Explain Statistical Distribution-Based Outlier Detection. Answer The following Questions		
	(~)	(i) List out Four Typical cases of Data Mining in Telecommunication	[04]	
		<ul><li>Industry.</li><li>(ii) Explain: Constraint-based Clustering method.</li></ul>	[03]	

\*\*\*\*\*