Seat No. : _____

AM-129

April-2016

M.Sc., Sem.-VIII (CA & IT)

Data Warehousing and Data Mining

Time: 3 Hours]

[Max. Marks : 100

1.	Answer the following questions (any four) :					
	(a)	Write a note on the characteristics of MOLAP and ROLAP.				
	(b)	Briefly explain Data Warehouse. Discuss the benefits of implementing data warehouse.				
	(c)	c) Describe the characteristics of ODS.				
	(d)) Explain with proper examples the data cube operations.				
	(e)	e) Discuss the various ways of implementing data cube.				
2.	Answer the following :					
	(1)	Briefly explain the following :				
		(a) Data cleaning				
		(b) ETL				
		(c) Data Mining				
		(d) Data Reduction				
		(e) Data characterization and discrimination				
	(2)	Explain the steps involved in data pre-processing.				
	(3)	Explain the various methods of filling the missing values.	5			
3.	Ans	Answer the following :				
	(1)	Explain with formula the concept of support and confidence.				
		Write the Apriori algorithm for finding association rules.				
	(2)	Define the following :	5			
		(a) k-itemset				
		(b) Lift				
		(c) Antecedent				
		(d) Maximal frequent itemset				
		(e) Support count				

4. (a) Answer the following :

PG Degree	Experience	Other	Touring	Category
	> 10 yrs.	certificates	Job	
Yes	Yes	Yes	Yes	А
Yes	Yes	Yes	No	А
Yes	Yes	No	No	В
Yes	No	Yes	Yes	А
Yes	No	No	Yes	В
No	Yes	No	Yes	В
No	No	No	No	C
No	Yes	No	Yes	C
No	Yes	Yes	Yes	В
Yes	No	No	No	C

Using the above data compute the three posterior probabilities for the three class, namely that the person with attribute values X has category Class A or Class B or Class C i.e. P(X|Ci) (where X is the attributes and Ci are the classes) using Bayes method. Compute for $P({\text{yes, yes, no}}|Ci)$ for each of the classes.

- (b) Answer any **two** :
 - Explain the different categories of attributes a decision tree node can have.
 - (2) Explain the method of rule extraction from a decision tree along with the characteristics a rule should have.
 - (3) Explain pre-pruning and post-pruning.
- 5. Answer the following :
 - (a) Explain with proper example the concept of clustering. List down the requirements for clustering analysis.
 - (b) Discuss interval scaled attributes and ratio scaled attributes with proper examples.
 - (c) Explain any three clustering methods.
 - (d) Explain contextual and collective outliers.

AM-129

 $2 \times 5 = 10$

20