

Seat No. : _____

AG-166

April-2015

**Fourth Year M.Sc. (CA & IT) Integrated
Data Warehousing and Data Mining**

Time : 3 Hours]

[Max. Marks : 100

1. Answer **all** : **4 × 5 = 20**

- (1) Write down five points of difference between ODS and Data Warehouse.
- (2) Write a brief note on MOLAP and ROLAP
- (3) Explain the steps of Data Warehouse Implementation.
- (4) Explain the attribute oriented induction for class comparisons.

2. Answer **all** : **4 × 5 = 20**

- (1) Explain the different data cube operations.
- (2) Explain briefly data mining. Explain the iterative sequence of knowledge discovery process.
- (3) Write a note on Data Cleaning with the methods of filling the missing values.
- (4) Explain the application of data mining in the field of Science and engineering.

3. (a) Consider an example of five transactions of six items. Find association rules using Apriori algorithm with 50% support and 75% confidence. (First Create the item list table along with their frequencies). **15**

Transaction ID	Items
100	Shampoo, Facewash, Handwash, Conditioner
200	Shampoo, Facewash, Conditioner
300	Shampoo, Soap, Facepack
400	Shampoo, Conditioner, Soap
500	Facewash, Conditioner, Soap

(b) Explain the concept of “Lift” for correlation measure with its possible values. **5**

4. (a) Answer any **two** :

2 × 5 = 10

- (1) Write the decision tree induction algorithm.
- (2) Explain the concept of tree pruning.
- (3) What is supervised classification ? In what situations can this technique be useful ? Also explain the concept of training sample.

(b) Using the following data compute the three posterior probabilities for the three class, namely that the person with attribute values X has credit risk Class A or Class B or Class C i.e. $P(X | C_i)$ (where X is the attributes and C_i are the classes) using Bayes method. Compute for $P(\{ \text{yes, yes, male, yes, A} \} | C_i)$ for each of the classes.

10

Own's home	Married	Gender	Employed	Credit rating	Risk class
Yes	Yes	Male	Yes	A	B
No	No	Female	Yes	A	A
Yes	Yes	Female	Yes	B	C
Yes	No	Male	No	B	B
No	Yes	Female	Yes	B	C
No	No	Female	Yes	B	A
No	No	Male	No	B	B
Yes	No	Female	Yes	A	A
No	Yes	Female	Yes	A	C
Yes	Yes	Female	Yes	A	C

5. (a) Write a short note along with proper examples on different types of attributes.
(Nominal, Binary, Ordinal, Numeric, Discrete, Continuous)

10

(b) Answer any **two** :

2 × 5 = 10

- (1) Briefly explain any three requirements for cluster analysis.
- (2) Explain the Hierarchical and Density based methods for clustering.
- (3) Explain the concept of outliers with proper example. Explain contextual outlier.