

--	--	--	--	--	--	--	--	--	--

Seventh Semester B.E. Degree Examination, Dec.2015/Jan.2016
Data Warehousing and Data Mining

Time: 3 hrs.

Max. Marks:100

**Note: Answer any FIVE full questions, selecting
atleast TWO questions from each part.**

PART – A

- 1 a. Explain the structure of data warehouse with neat diagram. (08 Marks)
b. Explain star schema with example. (05 Marks)
c. What is metadata? Explain types of metadata. (07 Marks)
- 2 a. Explain Codd's OLAP characteristics. (08 Marks)
b. Explain the difference between MOLAP and ROLAP. (06 Marks)
c. Explain the following operations of Datacube with suitable example :
(i) Roll Up (ii) Drill Down (06 Marks)
- 3 a. What is data mining? Explain various data mining task with examples. (10 Marks)
b. List and explain general characteristics of data sets. (05 Marks)
c. Distinguish between categorical and numerical attributes. (05 Marks)
- 4 a. What is Apriori principle? Explain. (06 Marks)
b. Explain in detail frequent itemset generation and rule generation with reference to Apriori along with an example. (10 Marks)
c. Define following term:
i) Support ii) Confidence (04 Marks)

PART – B

- 5 a. Explain difference between nominal attribute and ordinal attribute. (04 Marks)
b. Explain rule based classifier and its characteristics. (08 Marks)
c. Write Hunts algorithm and illustrate its working. (08 Marks)
- 6 a. Consider the following data sets for a binary classification.

Tid	Refund	Marital status	Taxable income	Class
1	Yes	Single	125 K	No
2	No	Married	100 K	No
3	No	Single	70 K	No
4	Yes	Married	120 K	No
5	No	Divorced	95 K	Yes
6	No	Married	60 K	No
7	Yes	Divorced	220 K	No
8	No	Single	85 K	Yes
9	No	Married	75 K	No
10	No	Single	90 K	Yes

- i) Calculate the information gain for each attribute
- ii) Draw decision tree by selecting best split. (10 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and/or equations written eg. 42+8 = 50, will be treated as malpractice.

- b. What is Baye's theorem? Show how it is used for classification. **(05 Marks)**
 - c. What are the approaches used for extending binary classifier to handle multiclass problem? **(05 Marks)**
- 7**
- a. Explain cluster analysis method briefly. **(08 Marks)**
 - b. Write and explain basic K means algorithm. **(06 Marks)**
 - c. Explain DBSCAN clustering algorithm. **(06 Marks)**
- 8**
- a. Explain the need for mining World Wide Web. **(06 Marks)**
 - b. Explain different text mining approach. **(08 Marks)**
 - c. Write note on mining temporal databases. **(06 Marks)**

*** * * * ***