

USN

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

12SCS332

**Third Semester M.Tech. Degree Examination, Dec.2014/Jan.2015**  
**Data Mining and Data Warehousing**

Time: 3 hrs.

Max. Marks: 100

**Note: Answer any FIVE full questions.**

- 1 a. Define Data Warehouse. Differentiate OLTP and OLAP. (10 Marks)  
 b. Define Data Mining. Explain Data Mining functionalities. (10 Marks)
- 2 a. Explain schemas of Multi – Dimensional data model, with examples and also mention DMQL statements for these. (10 Marks)  
 b. What is the necessity of Data preprocessing? What are different data – preprocessing techniques? Explain Data cleaning with example. (10 Marks)
- 3 a. Explain different Data – Mining primitives. (10 Marks)  
 b. Write Top – level syntax of DMQL. (10 Marks)
- 4 a. Write Apriori Algorithm for finding frequent item sets using candidate generation. (10 Marks)  
 b. Apply the above Algorithm for below transactional data. Write association rules for frequent item set generated (any one frequent item set). (10 Marks)

TID	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>7</sub>	T <sub>8</sub>	T <sub>9</sub>
List of item ID's	I <sub>1</sub> , I <sub>2</sub> , I <sub>5</sub>	I <sub>2</sub> , I <sub>4</sub>	I <sub>2</sub> , I <sub>3</sub>	I <sub>1</sub> , I <sub>2</sub> , I <sub>4</sub>	I <sub>1</sub> , I <sub>2</sub> , I <sub>3</sub>	I <sub>1</sub> , I <sub>3</sub>	I <sub>1</sub> , I <sub>3</sub>	I <sub>1</sub> , I <sub>2</sub> , I <sub>3</sub> , I <sub>5</sub>	I <sub>1</sub> , I <sub>2</sub> , I <sub>3</sub>

- 5 a. What is Bayesian classification? Explain working procedure of Naive Bayesian classifier. (10 Marks)  
 b. Differentiate classification and prediction. Explain different prediction techniques. (10 Marks)
- 6 a. What are the typical requirements of clustering in data mining? Explain. (10 Marks)  
 b. Explain different types of data (or) variables in cluster analysis. (10 Marks)
- 7 a. Write Algorithm for following : (10 Marks)  
 i) K – Means    ii) CURE    iii) DB SCAN.  
 b. Write short note on Outlier Analysis. (05 Marks)
- 8 a. Explain different trends in Data Mining. (10 Marks)  
 b. Explain different applications of data mining for Bio – medical and DNA data analysis. (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.