USN | | | | | | | | | | |

10MCA542

Fifth Semester MCA Degree Examination, Dec.2013/Jan.2014

Data Warehouse and Data Mining

Time: 3 hrs. Max. Marks:100

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

1 a. The transforming data phase of ETL is very complex. Justify this statement. **(10 Marks)**

b. What is data warehouse? List out the steps involved in building a data warehouse. **(10 Marks)**

2 a. Differentiate OLTP and OLAP systems. **(05 Marks)**

b. List out Codd's desirable characteristics of an OLAP systems. **(15 Marks)**

3 a. Briefly explain the specific challenges that motivated the development of data mining. **(10 Marks)**

b. Calculate the SMC and Jaccard similarity values for the binary vectors:
X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0); Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1) **(05 Marks)**

c. Compute cosine similarity between the document vectors:
X = (3, 2, 0, 5, 0, 0, 0, 2, 0, 0); Y = (1, 0, 0, 0, 0, 0, 0, 1, 0, 2) **(05 Marks)**

4 a. With a suitable example illustrate the process of reducing the number of candidate item sets using apriori principle. **(10 Marks)**

b. Develop a pseudo code to generate frequent item set using apriori algorithm. **(05 Marks)**

c. Define maximal frequent item set and closed frequent item sets. **(05 Marks)**

5 a. What is a decision tree? Using a training set of your choice illustrate how one can build a decision tree using Hunt's algorithm. **(10 Marks)**

b. What is the impurity for the node N having one element in class 1 and five elements in class 2 as measured by entropy and Gini index? **(05 Marks)**

c. Write a note on Bayesian classifier. **(05 Marks)**

6 a. List out the problems associated with centroid selection in k-means clustering technique. Is there any possibilities to handle the problems of initial centroid selection? Give details. **(12 Marks)**

b. Explain different ways of finding interclass cluster distance in hierarchical clustering approach. **(08 Marks)**

7 a. Write notes on text mining. **(10 Marks)**

b. Explain the applications of mining spatial data and temporal data. **(10 Marks)**

8 a. How is data mining used in hospital management systems? **(07 Marks)**

b. Briefly explain the role of data clustering in web search application. **(07 Marks)**

c. Defend the significant importance of rule based classifiers in mining information from the data available in an academic institution. **(06 Marks)**

* * * * *