

CRASH COURSE

USN

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

10IS74

Seventh Semester B.E. Degree Examination, May 2017

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 ODS as proposed by Immon and Imhoff with a possible ODS structure. (08 Marks)
b. Explain the issues resolved by ETL system. (06 Marks)
c. What is Meta data? Explain it in the context of data warehouse. (06 Marks)
- 2 a. Explain the data cube implementation. (08 Marks)
b. List and explain the different data cube operations. (08 Marks)
c. Explain the FASMI characteristics of OLAP. (04 Marks)
- 3 a. What is Data Mining? Explain the core data mining tasks with example. (08 Marks)
b. What is Data preprocessing? List and explain the different preprocessing techniques. (08 Marks)
c. Calculate the SMC and Jaccard similarity coefficients for the binary vector given :
 $X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$; $Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$. (04 Marks)
- 4 a. Develop the Apriori algorithm for generating frequent item set. (08 Marks)
b. With transactional data set, explain different types of compact representation of frequent item set. (12 Marks)

PART – B

- 5 a. What is Classification? Explain the general approach for building classification model. (10 Marks)
b. What is Rule based Classifier? Explain how rule based classifier works. (10 Marks)
- 6 a. Explain Bayesian classification. (10 Marks)
b. Explain how the predictive accuracy of classification methods be estimated. (10 Marks)
- 7 a. What is Cluster Analysis? Explain the desired features of cluster analysis. (08 Marks)
b. Find the distance between following cluster using different distance metrics :
Cluster 1 : $\{(1, 5) (2, 4) (3, 3) (2, 1)\}$; Cluster 2 : $\{(5, 4) (6, 6) (7, 5) (8, 8)\}$. (08 Marks)
c. What are the advantages of Hierarchical approach? (04 Marks)
- 8 Write short notes on :
a. Text clustering.
b. Mining spatial and temporal databases.
c. Unstructured text.
d. Comparison of OLTP and data warehouse. (20 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank space.
2. Do not write anything in the space provided for questions marked as supplementary questions.