

USN

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

10IS74

Seventh Semester B.E. Degree Examination, Dec.2016/Jan.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. Give the definition of Data warehousing. With a schematic diagram, explain the working of general Data warehousing architecture. (08 Marks)
 - b. List the major steps involved in ETL process. (08 Marks)
 - c. What is Meta data? Discuss different types of meta data used in Data warehouse. (04 Marks)
- 2
 - a. Explain Codd's OLAP characteristics. (10 Marks)
 - b. Describe data cube operations, with an example. (10 Marks)
- 3
 - a. Discuss the challenges that motivate the development of Data mining. (06 Marks)
 - b. Consider the following vectors. Find i) Simple matching co-efficient ii) Jaccard coefficient iii) Hamming distance. (04 Marks)
 $X = 0101010001$ $Y = 0100011000$.
 - c. Describe any five data preprocessing approaches. (10 Marks)
- 4
 - a. What is Association Analysis? Define Support and confidence, with an example. (04 Marks)
 - b. Develop the Apriori algorithm for frequent itemset generation, with an example. (08 Marks)
 - c. Explain the various measures of evaluating association patterns. (08 Marks)

PART – B

- 5
 - a. Explain the various measures for selecting the best split with an example. (06 Marks)
 - b. Give the difference between rule based ordering and class based ordering scheme. (04 Marks)
 - c. Consider a training set that contain 100 +ve examples and 400 –ve examples for each of the following candidate rule. Determine which is the best and worst candidate according to i) Rule accuracy ii) Foil Information gain.
 $R1 : A \rightarrow +$ (covers 4 + ve and 1 – ve examples)
 $R2 : B \rightarrow +$ (covers 30 +ve and 10 –ve examples)
 $R3 : C \rightarrow +$ (covers 100 +ve and 90 –ve examples). (10 Marks)
- 6
 - a. Explain Bayesian classification. (10 Marks)
 - b. Explain how the predictive accuracy of classification methods be estimated. (10 Marks)
- 7
 - a. Explain desired features of cluster analysis. (10 Marks)
 - b. Explain how distance between a pair of points can be computed. (05 Marks)
 - c. Write short note on density based methods. (05 Marks)
- 8
 - a. Explain web content mining and how it is used for discovering useful information from the web. (10 Marks)
 - b. Write short notes on : i) Spatial data mining ii) Text mining. (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 e.g. 4+4=8 = 20, will be treated as malpractice.