

USN

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

13MCA442

Fourth Semester MCA Degree Examination, Jun/July 2015
Data Warehousing and Data Mining

Time: 3 hrs.

Max. Marks: 100

Note: Answer any FIVE full questions.

- 1**
- Describe 3 – tier dataware architecture using a neat diagram. (10 Marks)
 - What are the different Datacube operation? Explain with examples. (05 Marks)
 - Compare OLTP and OLAP system. (05 Marks)
- 2**
- What is Data Mining? Explain motivation and challenges in Data mining. (06 Marks)
 - Write a note on ROLAP and MOLAP. (04 Marks)
 - Explain the different type of data sets in data mining with example for each. (10 Marks)
- 3**
- What is Data preprocessing? Explain Aggregation and sampling techniques. (10 Marks)
 - Consider the following two binary vectors.
 $X = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$
 $Y = (0, 0, 0, 0, 0, 0, 1, 0, 0, 1)$
 Calculate SMC and Jaccard Co – efficient. (06 Marks)
 - For the following Vector X and Y calculate Cosine Co – efficient.
 $X = (1, 1, 0, 1, 0, 1)$
 $Y = (1, 1, 1, 0, 0, 1)$ (04 Marks)
- 4**
- Write and explain Apriori principle with an example and Algorithm. (10 Marks)
 - Draw F – P Tree for the following transaction data.
- | Tid | Items |
|-----|----------------|
| 1 | { a, b } |
| 2 | { b, c, d } |
| 3 | { a, c, d, e } |
| 4 | { a, d, e } |
| 5 | { a, b, c } |
| 6 | { a, b, c, d } |
| 7 | { a } |
| 8 | { a, b, c } |
| 9 | { a, b, d } |
| 10 | { b, c, e } |
- (10 Marks)
- 5**
- Describe different methods used for estimating predictive accuracy of classification. (10 Marks)
 - Briefly explain characteristics nearest neighbor classifiers. (05 Marks)
 - Briefly Discuss, How Rule based classifier work using an example. (05 Marks)
- 6**
- Write a note on Multiclass problem using an example. (05 Marks)
 - Write a note an Bayesian classifiers. (05 Marks)
 - What is a decision Tree? Using a training set of your choice illustrate building a decision tree using hunts Algorithm. (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.

13MCA442

- 7 a. What is cluster Analysis? Explain different types of cluster analysis methods. (10 Marks)
b. Briefly explain basic K – means algorithm using pseudocode. (10 Marks)
- 8 a. Define outlier and Discuss proximity based outlier detection method. (10 Marks)
b. Explain Briefly the terms with their mathematical representations.
i) Gini Index ii) Entropy iii) Information Gain. (10 Marks)

2 of 2

Highly confidential document EDC - 034, @ 09-06-2015 12:53:59