

**C.U.SHAH UNIVERSITY**

Wadhwan City

Subject Code : 5TE02BDA1

Summer Examination-2014

Date: 25/06/2014

Subject Name: Big Data and Analytics

Branch/Semester:- M.Tech(CE)/II

Time:02:00 To 5:00

Examination: Regular

**Instructions:-**

- (1) Attempt all Questions of both sections in same answer book / Supplementary
- (2) Use of Programmable calculator & any other electronic instrument is prohibited.
- (3) Instructions written on main answer Book are strictly to be obeyed.
- (4) Draw neat diagrams & figures (If necessary) at right places
- (5) Assume suitable & Perfect data if needed

**SECTION-I****Q-1 Attempt following Questions.**

- |    |   |          |
|----|---|----------|
| a) | Explain Data warehouse and Data mart.                 | <b>2</b> |
| b) | What is Binning? List and explain binning strategies. | <b>2</b> |
| c) | Discuss applications of "Fuzzy Logic".                | <b>2</b> |
| d) | List out methods for data normalization               | <b>1</b> |
- 
- |            |    |   |          |
|------------|----|---|----------|
| <b>Q-2</b> | a) | Describe the difference between the following approaches for the integration of a data mining system with a database or data warehouse system: no coupling, loose coupling, semi tight coupling and tight coupling. State which approach you think is most popular and why? | <b>5</b> |
|            | b) | A group of 12 sales price records has been sorted as follows:<br>5,10,11,13,15,35,50,55,72,92,204,215<br>Partition them into three bins by each of following methods.<br>1. Equal Frequency (Equi-depth) Partitioning<br>2. Equal Width Partitioning<br>3. Clustering       | <b>5</b> |
|            | c) | What are the challenges for effective resource and knowledge discovery in mining the world wide web?  | <b>4</b> |

**OR**

- |            |    |   |          |
|------------|----|---|----------|
| <b>Q-2</b> | a) | What is prediction? Explain various regression methods for it   | <b>5</b> |
|            | b) | Suppose a data warehouse for CUSU consists of following four dimensions: student, course, semester, faculty with two measures: count, avg_grade.<br>(i) Draw a snowflake schema for data warehouse.<br>(ii) Starting with base cuboid (student, course, semester, faculty), what specific OLAP operations should one perform in order to list average grade of CE course for each CUSU student. | <b>5</b> |
|            | c) | What do you mean by text mining? Explain various issues involved in it  | <b>4</b> |
- 
- |            |    |  |          |
|------------|----|--|----------|
| <b>Q-3</b> | a) | A base cuboid has three dimensions A, B, C with the following number of cells:  A  = 1,000,000,  B =100 and  C =1000. Suppose that each dimension is evenly partitioned into 10 portions for chunking.<br>1) Assuming each dimension has only one level, draw the complete lattice of the cube.<br>2) If each cube cell stores one measure with 4 bytes, what is the total size of the computed cube if the cube is dense? | <b>7</b> |
|            | b) | Describe the ID3 algorithm of the decision tree construction. Why it is unsuitable for Data mining applications  | <b>7</b> |

**OR**

- Q-3** a) The Apriori algorithm makes use of prior knowledge of subset support properties. 7  
 1) Prove that all non-empty subsets of a frequent item set must also be frequent.  
 2) Prove that the support of any non-empty subsets of item set s must be at least as great as the support of s.  
 b) Explain various OLAP operations performed on Data cube. 7

**SECTION-II**

**Q-4 Attempt following Questions.**

- a) What do you mean by concept hierarchy? Show its application. 2  
 b) What is difference between OLTP and OLAP? 2  
 c) What do you mean by closed frequent item set? What is its application? 2  
 d) Define: Clustering 1
- Q-5** a) Describe A Multilayer Feed-Forward Neural Network 5  
 b) What is Market Basket Analysis? Explain Association Rules with Confidence and support. 5  
 c) Compare the merits and demerits of Eager Classification Versus Lazy Classification. 4

**OR**

- Q-5** a) With the help of a neat diagram explain the 3-tier architecture of a data Warehouse. 5  
 b) Why is preprocessing required before any data mining method is applied? Give name of various data preprocessing methods 5  
 c) Explain Agglomerative and Divisive Hierarchical Clustering. 4
- Q-6** a) Discuss why analytical data characterization is needed and how it can be performed. 7  
 Compare the result of two induction methods.  
 1) With relevance Analysis  
 2) Without relevance Analysis  
 b) What is Data Mining? Write down short note on KDD process. 7

**OR**

- Q-6** a) Compare K-mean and K-medoid algorithms with suitable example. 7  
 b) List and describe five primitives for specifying data mining task. 7

\*\*\*\*\*25\*\*14\*\*\*\*S

