

**APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY**  
**FIRST SEMESTER M.TECH DEGREE EXAMINATION, DECEMBER 2018**

**Computer Science and Engineering**  
**(Computer Science and Engineering)**

**01CS6151: Data Warehousing & Mining**

**Max. Marks: 60**

**Duration: 3 Hours**

**Instructions: Answer any Two questions from each Part**

**PART A**

1. A data warehouse consists of three dimensions time, doctor and patient and two measures count and charge, where charge is the fee that a doctor charges a patient for a visit. (9)
  - (i) Enumerate three classes of schema that are popularly used for modeling data warehouses.
  - (ii) Draw a schema diagram for the above data warehouse using schema classes listed in.
2. a. When is data reduction used for preprocessing of data? How is data reduction done using principal component analysis? (5)  
b. Choose any normalization method to normalize the following data. Justify the choice of normalization method. (4)  
13,15,16,16,19,20,20,21,22,22,25,25,25,25,25,30,33,33,35.
3. a. Illustrate the 3 tier architecture of Data warehouse with a neat diagram. (5)  
b. Differentiate MOLAP and ROLAP. (4)

**PART B**

4. a. Illustrate a multi layer Feed Forward Neural Network. (5)  
b. List the stages involved in the training of Back Propagation Network (BPN). (4)
5. a. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8), compute the cosine similarity between the two tuples. (2)  
b. With the training data given, derive a regression equation to model the data (7)

and classify data as short (represented using 0) or medium (represented using 1). The data is: { (1.6,0), (1.9,1), (1.88,1), (1.7,0), (1.85,1), (1.6,0), (1.7,0), (1.8,1), (1.95,1), (1.9,1), 1.8,1), (1.75,1).

6. Briefly describe these approaches in clustering with example in each case. (9)

(i) Partitioning methods <http://www.ktuonline.com>

(ii) Hierarchical methods.

### PART C

7. a. Using a Bayesian Classifier, classify the tuple (Adam, M, 1.95) as short, medium or tall. The training data for the classifier is given below: (8)

| Name      | Gender | Height | Output |
|-----------|--------|--------|--------|
| Kristina  | F      | 1.6m   | Short  |
| Jim       | M      | 2m     | Tall   |
| Maggie    | F      | 1.9m   | Medium |
| Martha    | F      | 1.88m  | Medium |
| Stephanie | F      | 1.7m   | Short  |
| Bob       | M      | 1.85m  | Medium |
| Kathy     | F      | 1.6m   | Short  |
| Dave      | M      | 1.7m   | Short  |
| Worth     | M      | 2.2m   | Tall   |
| Steven    | M      | 2.1m   | Tall   |
| Debbie    | F      | 1.8m   | Medium |
| Todd      | M      | 1.95m  | Medium |
| Kim       | F      | 1.9m   | Medium |
| Amy       | F      | 1.8m   | Medium |
| Wynette   | F      | 1.75m  | Medium |

- b. How do context focused crawlers improve the performance of web search? (4)
8. a. A database has five transactions. Let min sup = 60% and min con f = 80%. (7)  
Find all frequent itemsets using Apriori algorithm.
- | TID  | items bought         |
|------|----------------------|
| T100 | { A, O, N, K, E, Y } |
| T200 | { D, O, N, K, E, Y } |
| T300 | { M, A, K, E, Y }    |
| T400 | { M, U, C, K }       |
| T500 | { C, O, U, K, I, E } |
- b. How are spatial query processing handled by R-trees? Illustrate, (5)
9. a. Spatial association mining can be implemented in at least two ways: (1) (8)  
dynamic computation of spatial association relationships among different spatial objects, based on the mining query, and (2) pre-computation of spatial distances between spatial objects, where the association mining is based on such pre-computed results. Discuss (1) how to implement each approach efficiently and (2) which approach is preferable under what situation.
- b. Write a note on temporal mining. (4)

http://www.ktuonline.com

Whatsapp @ 9300930012

Your old paper & get 10/-

पुराने पेपर्स भेजे और 10 रुपये पायें,

Paytm or Google Pay से