Reg. No. :

Question Paper Code: 52236

M.E. DEGREE EXAMINATION, JUNE 2016

Elective

Computer Science and Engineering

15PCS522 - DATA MINING TECHNIQUES

(Regulation 2015)

Duration: Three hours   Maximum: 100 Marks

Answer ALL Questions

PART A - (5 x 1 = 5 Marks)

1. The data mining process is to extract information from a data set and transform it into an

    (a) understandable structure   (b) reversible structure
    (c) reliable structure   (d) all the above

2. _____ takes a model and a dataset and produces a set of scores for the records in the dataset.

    (a) scoring data   (b) scoring engine
    (c) scoring structure   (d) none of these

3. Data cleaning involves removing the _____ and treatment of missing values.

    (a) Noise   (b) signal   (c) data   (d) all the above

4. The objective of cluster analysis is to partition a data set into a group of

    (a) Subject   (b) algorithms   (c) subsets   (d) none of these

5. Data mining application are

    (a) financial data analysis   (b) biological data analysis
    (c) both (a) and (b)   (d) none of these

PART B - (5 x 3 = 15 Marks)

6. List the issues in data mining.

7. Compare Predictive Vs Descriptive score functions.

8. Illustrate the role of Bayesian rule in classification process.

9. Explain Hierarchical clustering with example.

10. Classify the techniques for improving the efficiency of Apriori.

PART C - (5 x 16 = 80 Marks)

11. (a) Explain data mining architecture with neat architectural diagram.　　　(16)

Or

(b) Describe data preprocessing technique.　　　(16)

12. (a) Explain fundamentals of modelling.　　　(16)

Or

(b) Explain how will you evaluate models and patterns with data mining algorithms.　　　(16)

13. (a) Evaluate a decision tree classification model to classify bank loan applications by assigining applications to one of three classes.　　　(16)

Own Home?	Married	Gender	Employed	Class
Yes	Yes	Male	Yes	B
No	No	Female	Yes	A
Yes	Yes	Female	Yes	C
Yes	No	Male	No	B
No	Yes	Female	Yes	C
No	No	Female	Yes	A
No	No	Male	No	B
Yes	No	Female	Yes	A
No	Yes	Female	Yes	C
Yes	Yes	Female	Yes	C

52236

Or

(b) Examine the classification of back propagation method. (16)

14. (a) Solve the following clustering's:

$X1 = \{1, 3\}$ $X2 = \{0, 2\}$ $X3 = \{2, 2\}$ $X4 = \{1, 4\}$

Clusters: $C1 = \{X1, X3\}$ $C2 = \{X2, X4\}$

    (i) Apply one iteration of k-means partitioned clustering algorithm

    (ii) What is the change in total square errors (16)

Or

(b) Solve the following measurement for the variables

    56, 22, 76, 89, 11, 65, 78, 44, 98, 34, 52, 66

Standardize the variable by the following

    (i) Compute mean absolute deviation

    (ii) Compute Z-score for measurement (16)

15. (a) Evaluate the following set of transactions. Find the frequent item set and the association rules with at least 30% support and 60% confidence using Apriori algorithm. (16)

TID	Items Bought
10	B, M, T, Y
20	B, M
30	A, T, S, P
40	A, B, C, D
50	A, B
60	T, Y, E, M
70	A, B, M
80	B, C, D, T, P
90	D, T, S
100	A, B, M

Or

(b) Illustrate visual data mining with an example. (16)

———————————

3

52236