QP Code: RD18001057 Reg. No □□□□□□□□□□ AR 18

GIET MAIN CAMPUS AUTONOMOUS GUNUPUR – 765022

B. Tech Degree Examinations, December – 2020

(Fifth Semester)

BCSPC5040 / BITPC 5040 – DATA MINING & DATA WAREHOUSING

(CSE & IT)

Time: 2 hrs Maximum: 50 Marks

The figures in the right hand margin indicate marks.

PART – A: (Multiple Choice Questions) (1 x 10 = 10 Marks)

Q.1. Answer ***ALL*** questions [CO#] [PO#]

a. The data mining primitives specifies 1 3

1) The set of task-relevant data to be mined

2) The expected representation for generating the pattern

3) The interestingness measures for limiting the performance.

4) The kind of knowledge to be mined

Which of the following is true about the above statement?

(i) 1,2 (ii) 2, 3

(iii) 1,4 (iv) 3, 4

b. In which of the following, data are stored, retrieved and updated? 1 2

(i) HTTP (ii) OLAP

(iii) OLTP (iv) MOLAP

c. Data warehouse is ________________________ 2 1

The actual discovery phase of a knowledge discovery process (ii)A subject-oriented integrated time variant non-volatile collection of data in support of management

(iii)The stage of selecting the right data for a KDD process (iv)None of these

d. The difference between supervised learning and unsupervised learning is given by: 2 2

(i) Unlike unsupervised learning, supervised learning needs labeled data (ii) Unlike unsupervised leaning, supervised learning can form new classes

(iii) Unlike unsupervised learning, supervised learning can be used to detect outliers (iv) Unlike supervised learning, unsupervised learning can predict the output class from among the known classes

e. An itemset whose support is greater than or equal to a minimum support threshold is ______ 3 1

(i) Itemset (ii) Frequent Itemset

(iii) Infrequent items (iv) Threshold values

f. In decision tree algorithms, attribute selection measures are used to 2 1

(i) Reduce the dimensionality (ii) Reduce the error rate

(iii) Select the splitting criteria which best separate the data (iv) Rank attributes

g. Which of the following is used to find inherent regularities in data? 4 2

(i) Clustering (ii) Regression analysis

(iii) Frequent pattern analysis (iv) Outlier analysis

h. Classification is a data mining task that maps the data into _________ . 3 3

(i) clusters (ii) real valued prediction variable

(iii) time series (iv) predefined group

i. To detect fraudulent usage of credit cards, the following data mining task should be used: 4 1

(i) Feature selection (ii) Prediction

(iii) Outlier Analysis (iv) All the above

j. The goal of clustering analysis is to: 1 2

(i) Maximize the inter-cluster similarity (ii) Maximize the number of clusters

(iii) Maximize the intra-cluster similarity (iv) Minimize the intra-cluster similarity

PART – B: (Short Answer Questions) (2 x 5 = 10 Marks)

Q.2. Answer ***ALL*** questions [CO#] [PO#]

a. Differentiate data, information and knowledge 1 1

b. State the major strengths of decision tree method. 1 3

c. Give two applications of frequent pattern analysis. 2 1

d. List the factors that affecting the complexity of Apriori algorithm. 3 3

e. How will you design a malware detection system using a data mining technique? 4 3

PART – C: (Long Answer Questions) (6 x 5 = 30 Marks)

Answer ***ANY FIVE*** questions Marks [CO#] [PO#]

3. How the data mining techniques helps an organization to improve the revenue generation in the competitive environment. (6) 1 2

4. Discuss Data Reduction and Data Discretization techniques in detail. (6) 2 1

5. Explain the architecture of a data warehouse with a neat illustration. (6) 4 1

6. Consider a data warehouse consists of three dimensions - time, doctor and patient .There are two measures to record the patient's visit namely count and charge, where charge is the fee that a doctor charges for a visit. Enumerate and draw all the schema diagrams for the given scenario. (6) 4 1

7. A data base has four transactions with min_sup=30% (6) 2 4

TID	Date	Items_bought
100	10/2/2010	{A,C,D,E}
210	12/2/2010	{A,B,E}
300	23/3/2010	{B,C,D}

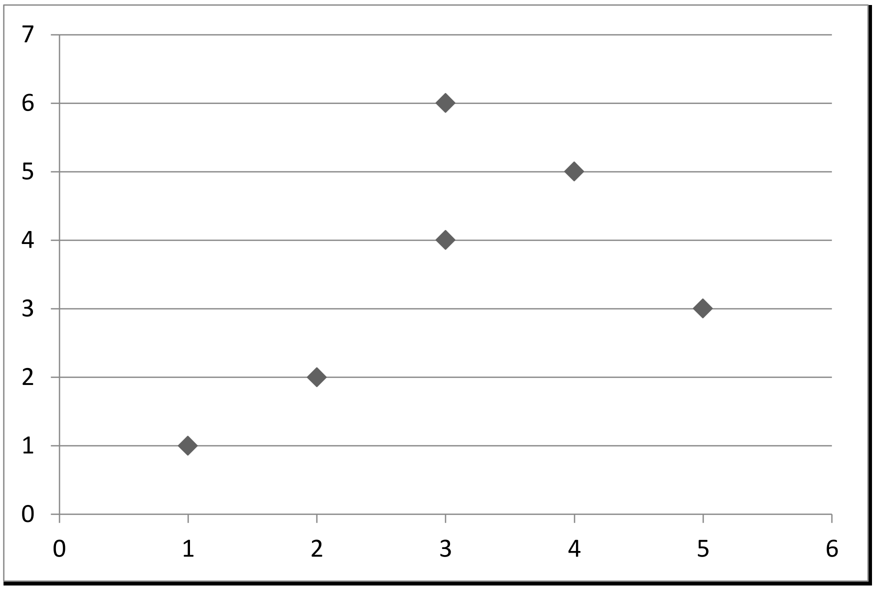
330	24/6/2011	{A,C}
400	10/1/2012	{C,E}
424	5/5/2015	{A,D}

Find all the frequent item set using Apriori Algorithm.

8. Design a genetic algorithm based classifier. (6) 2 4

9. With suitable example explain the model evaluation and selection (6) 3 2

10. Explain hierarchical clustering in detail. **Analyse** the below diagram and draw the dendrogram using hierarchical clustering algorithm . (6) 1 2



--- End of Paper ---