

SUBJECT CODE NO:- P-29
FACULTY OF ENGINEERING AND TECHNOLOGY
B.E.(CSE/IT) Examination MAY/JUNE-2016
Data Warehousing & Data Mining (CSE/IT)
(Revised)

[Time:ThreeHours]

[Max Marks:80]

“Please check whether you have got the right question paper.”

- N.B
- i) Q.No.1 and Q.No.6 are compulsory. Solve any two questions from the remaining ones in each section.
 - ii) Assume suitable data if necessary and state it clearly.
 - iii) Answers should be precise.

Section A

- | | | |
|-----|---|----|
| Q.1 | a) What is the role of Data Mining in web Search Engines? | 03 |
| | b) What is a KDD process? | 03 |
| | c) Describe possible integration (coupling) Methods of Data mining system with Database system. | 04 |
| Q.2 | a) Describe in brief any two schemas that are used for Multi-dimensional Data Model. Draw the necessary diagrams. | 08 |
| | b) What are typical OLAP operations performed of Data Cube? Explain any one operation with the data cube diagram. | 07 |
| Q.3 | a) What are the steps in designing a data warehouse? | 08 |
| | b) Which are the major issues in Data Mining? | 07 |
| Q.4 | a) With a suitable example, explain how statistical parameters are used to handle the data dissimilarity? | 08 |
| | b) What is the exact difference between Bitmap indexing and Join indexing used in OLAP data. Illustrate with an example | 07 |
| Q.5 | a) What is data dissimilarity? Two objects are represented by the tuples (32,1,22,6) and (20,0,12,8): | 08 |
| | i) Compute the Euclidean distance between the two objects. | |
| | ii) Compute the Manhattan distance between the two objects | |
| | iii) Compute the Minkowski distance between the two objects, using q=3. | |
| | b) What are the different methods of handling missing values in the tuples? | 07 |

Section B

Q.6	<ul style="list-style-type: none"> a) Precisely define Prediction, classification and clustering? b) Write a note on – BI Framework. c) What is Confusion Matrix? 	<ul style="list-style-type: none"> 03 03 04 												
Q.7	<ul style="list-style-type: none"> a) A database has five transactions. Take min support count as 2. Find all frequent itemsets using Apriori algorithm. <table border="0" style="margin-left: 20px;"> <thead> <tr> <th style="text-align: left;">TID</th> <th style="text-align: left;">Items_bought</th> </tr> </thead> <tbody> <tr> <td>T100</td> <td>{M,O,N,K,E,Y}</td> </tr> <tr> <td>T200</td> <td>{D,O,N,K,E,Y}</td> </tr> <tr> <td>T300</td> <td>{C,A,K,E}</td> </tr> <tr> <td>T400</td> <td>{M,I,C,K,Y}</td> </tr> <tr> <td>T500</td> <td>{C,O,O,K,I,E}</td> </tr> </tbody> </table> <p style="margin-left: 20px;">Generate Association Rules by taking minimum confidence as 80%.</p>	TID	Items_bought	T100	{M,O,N,K,E,Y}	T200	{D,O,N,K,E,Y}	T300	{C,A,K,E}	T400	{M,I,C,K,Y}	T500	{C,O,O,K,I,E}	10
TID	Items_bought													
T100	{M,O,N,K,E,Y}													
T200	{D,O,N,K,E,Y}													
T300	{C,A,K,E}													
T400	{M,I,C,K,Y}													
T500	{C,O,O,K,I,E}													
	<ul style="list-style-type: none"> b) What is the process of Market-Basket Analysis? How do you define ‘Support’ and ‘Confidence’? 	05												
Q.8	<ul style="list-style-type: none"> a) How does the k-means algorithm cluster or partition the data? Is there any limitation of this method? b) What are the different types of data on which cluster analysis is to be used? Explain any two data types/ representations in detail. 	<ul style="list-style-type: none"> 08 07 												
Q.9	<ul style="list-style-type: none"> a) By taking suitable database of 10 tuples, classify the tuples using the decision Tree Classifier algorithm ID3. b) What is the method of generating rules from Decision tree? 	<ul style="list-style-type: none"> 10 05 												
Q.10	<ul style="list-style-type: none"> a) What are the various components in Business Pressures-Responses-Support Model? b) Which are the Major Tools and Techniques used for business Intelligence? 	<ul style="list-style-type: none"> 08 07 												