Total number of printed pages – 2

B. Tech

PEIT 5302

Fifth Semester Examination – 2013

DATA MINING AND DATA WAREHOUSING

BRANCH : IT, BIOTECH

QUESTION CODE : C- 342

Full Marks – 70

Time : 3 Hours

*Answer Question No. 1 which is compulsory and any **five** from the rest.*
The figures in the right-hand margin indicate marks.

1. Answer the following questions :   2×10

    (a) What is the need for a Data warehouse ? Explain.

    (b) What are multidimensional databases ? How do they play an important role in data warehouse ?

    (c) What do you mean by data mart ? How does the user can access data mart ?

    (d) What do you mean by association rule mining ?

    (e) List out the classification of Data Mining System.

    (f) What is a decision tree ? What are the measuring factors of data mining ?

    (g) Differentiate between classification and clustering.

    (h) What is K-means algorithm ? Explain.

    (i) Differentiate between spatial mining and temporal mining.

    (j) Why web mining is required ? Explain.

2. (a) What are the building blocks of data warehouse ? How do they contribute towards development of the data warehouse ? Explain.   5

    (b) Differentiate between OLAP and OLTP. Explain with example. Give an account of OLAP tools.   5

3. (a) Explain the process of data warehouse design and the three tier data warehouse architecture.   5

P.T.O.

(b)   What is the role of Datamart in a data warehouse ? Explain with example.

5

4.   (a)   Run a Apriori algorithm on the following transaction database with minimum support equal to 2 transaction. Explain step by step execution.

TID	ITEMS
T100	A, B, C
T200	A, B, C
T300	C, D
T400	C, D

Sketch a proof of correctness of the Apriori algorithm. It suffices to sketch a proof showing that all the frequent item sets of size k are among the candidates of size k.

5

(b)   Assume that each item in supermarket is bought by 1% of transactions. Assume that there are 10 million transactions and that items are statistically independent. Assume mid-sup = 10. What is the expected size of a frequent set ? What is the expected number of frequent sets ?

5

5.   (a)   Explain FP growth algorithms in association rule mining with an example. 5

(b)   What is Bayes theorem ? Explain about Naive Bayesian classification.   5

6.   (a)   Give an account of various clustering methods. Explain using examples. Give the strengths and limitations of different methods.

6

(b)   Differentiate between k-means and k-medoids. Illustrate the strength and weaknesses of k-means and k-medoids.

4

7.   (a)   What is special about spatial data mining relative to mining relational data ? Is it adequate to materialize spatial features to be used as input to classical data mining algorithms / models ?

5

(b)   Give an account of a data mining application that can run on a retail industry and produce better results for it.

5

8.   Answer any **two** of the following :   5×2

(a)   Standards in Data Warehouse

(b)   Correlation analysis

(c)   Outlier analysis.

_____