Code No: 57048

R09

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD
B. Tech IV Year I Semester Examinations, February/March – 2016
DATA WAREHOUSING AND DATA MINING
(Computer Science and Engineering)

Time: 3 Hours

Max. Marks: 75

Answer any Five Questions
All Questions Carry Equal Marks

- - -

1.a) What is data mining? Explain it as a phase in knowledge discovery process.
  b) Explain principal component analysis as a method of dimensionality reduction. [7+8]

2.a) Make comparison of OLTP with OLAP.
  b) Explain an approach for data warehouse development. [8+7]

3.a) Apply ECLAT algorithm to the following data to find frequent item sets. Support is 25%. Confidence is 60%. List at least two strong association rules.

TID	LIST OF ITEMS
1	Bread, Milk, Sugar, TeaPowder, Cheese, Tomato
2	Onion, Tomato, Chillies, Sugar, Milk
3	Milk, Cake, Biscuits, Cheese, Onion
4	Chillies, Potato, Milk, Cake, Sugar, Bread
5	Bread, Jam, Milk, Butter, Chilles
6	Butter, Cheese, Paneer, Curd, Milk, Biscuits
7	Onion, Paneer, Chilies, Garlic, Milk
8	Bread, Jam, Cake, Biscuits, Tomato

  b) Discuss correlation analysis with chi-square. [8+7]

4.a) Explain Naïve Bayesian classifier with illustrations.
  b) Discuss Boosting technique.
  c) What characteristics make a neural network a good classifier? [5+5+5]

5. What is an outlier? What is the necessity for outlier analysis? Explain various techniques for outlier detection. [15]

6.a) Write about BLAST tool for biosequence analysis.
  b) Explain the role of inductive logic programming in multirelational data mining. [7+8]

7.a) Discuss location preserving index for text mining.
  b) What is the significance of web usage mining?
  c) With an example discuss spatial data cube. [5+5+5]

8. Write technical notes on the following:
   a) Visual data mining
   b) Data mining for intrusion detection
   c) Collaborative filtering. [5+5+5]

--ooOoo--