Code No: 07A60506 **R07** **Set No. 2**

III B.Tech II Semester Examinations,December 2010
DATA WAREHOUSING AND DATA MINING
Information Technology

Time: 3 hours **Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) Explain the design and construction process of data warehouses.
   (b) Explain the architecture of a typical data mining system. [8+8]
2. (a) Briefly discuss about data integration.
   (b) Briefly discuss about data transformation. [8+8]
3. Explain in detail the major steps of decision tree classification. [16]
4. (a) Describe cosine measure for similarity in documents.
   (b) Explain in detail similarity search in time-series analysis. [8+8]
5. (a) What are the desired architectures for Data mining systems.
   (b) Briefly explain about concept hierarchies. [8+8]
6. (a) Explain how COBWEB method is used for clustering.
   (b) Discuss in detail DENCLUE clustering methods. [8+8]
7. Compare and contrast Apriori algorithm with frequent pattern growth algorithm. Consider a data set apply both algorithms and explain the results. [16]
8. (a) How can we perform discrimination between different classes? Explain.
   (b) Explain the analytical characterization with an example. [8+8]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: 07A60506 **R07** **Set No. 4**

III B.Tech II Semester Examinations,December 2010
DATA WAREHOUSING AND DATA MINING
Information Technology

Time: 3 hours **Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. Write short notes for the following in detail:
   (a) Measuring the central tendency
   (b) Measuring the dispersion of data. [16]
2. (a) Explain K-means algorithm for clustering.
   (b) Given two objects represented by the tuples(22,1,42,10) and(20,0,36,8)
      i. Compute the Manhatten distance between the two objects.
      ii. Compute the Euchidean distance between the two objects.
3. (a) List and describe any four primitives for specifying a data mining task.
   (b) Describe why concept hierarchies are useful in data mining. [8+8]
4. (a) How can we smooth out noise in data cleaning process? Explain.
   (b) Why preprocessing of data is needed? [8+8]
5. (a) Discuss construction and mining of object cubes.
   (b) Give a detail note on trend analysis. [6+10]
6. (a) Explain the design and construction process of data warehouses.
   (b) Explain the architecture of a typical data mining system. [8+8]
7. (a) Explain how concept hierarchies are used in mining multilevel association rule?
   (b) Give the classification of association rules in detail. [8+8]
8. (a) Discuss the five criteria for the evaluation of classification and prediction methods.
   (b) Explain how rules can be extracted from training neural networks. [8+8]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: 07A60506 **R07** **Set No. 1**

III B.Tech II Semester Examinations,December 2010
DATA WAREHOUSING AND DATA MINING
Information Technology

Time: 3 hours **Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. (a) Discuss interval-scaled variable and binary variables.
   (b) Explain in detail K-Medoids algorithm for clustering. [8+8]
2. (a) Briefly discuss the data smoothing techniques.
   (b) Explain about concept hierarchy generation for categorical data. [8+8]
3. (a) Explain data mining as a step in the process of knowledge discovery.
   (b) Differentiate operational database systems and data warehousing. [8+8]
4. (a) Explain spatial datacube construction and spatial OLAP.
   (b) Give a note on item frequency matrix. [10+6]
5. Briefly discuss the following data mining primitives:
   (a) Task-relevant data
   (b) The kind of knowledge to be mined
   (c) Interestingness measures
   (d) Presentation and visualization of discovered patterns. [16]
6. (a) Explain with an example a measure of the goodness of split.
   (b) Write a detail note on genetic algorithms for classication. [8+8]
7. Suppose that the following table is derived by Attribute-oriented induction.

CLASS	BIRTH-PLACE	COUNT
Programmer	Canada	180
	others	120
DBA	Canada	20
	others	80

   (a) Transform the table into a crosstab showing the associated t-weights and d-weights.
   (b) Map the class programmer into a(Bi-directional) quantitative descriptive rule, for example, ∀X,Programmer(X) ⇔ (birth_place(X)= "Canada" ∧ ...) [t: x%, d:y%] ... ∧ (...) [t:w%, d:z%]

[8+8]

8. (a) Discribe mining multidimensional association rule using static discretization of quantitative attribute.
   (b) Expalin association rule generation from frequent itemsets. [8+8]

⋆ ⋆ ⋆ ⋆ ⋆

Code No: 07A60506 **R07** **Set No. 3**

III B.Tech II Semester Examinations,December 2010
DATA WAREHOUSING AND DATA MINING
Information Technology

Time: 3 hours **Max Marks: 80**

Answer any FIVE Questions
All Questions carry equal marks

⋆ ⋆ ⋆ ⋆ ⋆

1. Explain mining multilevel association rules from transaction databases. [16]
2. (a) Discuss distance based outlier detection.
   (b) Explain OPTICS algorithm for clustering. [8+8]
3. (a) How can you go about filling in the missing values in data cleaning process?
   (b) Discuss the data smoothing techniques. [8+8]
4. Write short notes on the following:
   (a) Association analysis
   (b) Classification and prediction
   (c) Cluster analysis
   (d) Outlier analysis. [16]
5. (a) Give a note on log-linear models.
   (b) Explain the hold out method for estimatin classifier accuracy.
   (c) Discuss Fuzzy set approach for classification. [5+5+6]
6. (a) Differentiate attribute generalization threshold control and generalized relation threshold control.
   (b) Differentiate between predictive and descriptive data mining. [8+8]
7. Explain the syntax for the following data mining primitives:
   (a) Task-relevant data
   (b) The kind of knowledge to be mined
   (c) Interestingness measures
   (d) Presentation and visualization of discovered patterns. [16]
8. (a) Discuss web content mining and web usage mining.
   (b) compare information retrieval with text mining. [8+8]

⋆ ⋆ ⋆ ⋆ ⋆