

Reg. No. :

--	--	--	--	--	--	--	--	--	--

Question Paper Code: 92033

M.E. DEGREE EXAMINATION, MAY 2014.

Elective

Computer Science and Engineering

01PCS509 - DATA MINING AND DATA WAREHOUSING

(Regulation 2013)

Duration: Three hours

Maximum: 100 Marks

Answer ALL Questions.

PART A - (10 x 2 = 20 Marks)

1. List down the merits of a data warehouse.
2. What is the need for normalization?
3. Define Confidence.
4. What do you mean by numerosity reduction? Give an example.
5. What is the difference between supervised and unsupervised learning scheme?
6. State Bayes theorem.
7. What is a DBSCAN?
8. Define CF - tree.
9. What does audio data mining mean?
10. Can we construct a spatial data warehouse?

PART - B (5 x 14 = 70 Marks)

11. (a) Explain in detail the DBMS schema for decision support system with neat sketch. (14)

Or

(b) (i) Differentiate operational and informational systems. (6)

(ii) With suitable diagram, explain about three tier architecture of data warehouse. (8)

12. (a) Explain in detail about Apriori algorithm and discuss how the efficiency can be further improved in Apriori based mining? (14)

Or

(b) Explain the various data reduction techniques in detail. (14)

13. (a) State the principle of SVM and explain how the classification is done using SVM. (14)

Or

(b) (i) Discuss the various performance metrics considered in model selection of classifier. (10)

(ii) Write short notes on ensemble method. (4)

14. (a) (i) Explain density based method of cluster analysis. (7)

(ii) Write short notes on constraint based cluster analysis. (7)

Or

(b) (i) Describe the basic requirements for cluster analysis. (7)

(ii) Have a brief discussion on outlier analysis. (7)

15. (a) Explain in detail about text and web data mining. (14)

Or

- (b) Is web poses great challenges for effective resource and knowledge discovery? Justify your answer and also explain how the knowledge is discovered from social website. (14)

PART - C (1 x 10 = 10 Marks)

16. (a) Suppose that the data for analysis include the attribute the frequency of stop words in documents. the values are given in increasing order:13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
- (i) Use smoothing by bin means with a depth of 3.
 - (ii) Use min - max [0-1] normalization, normalization by decimal scaling to transform the value 35.
 - (iii) Plot equi _ width histogram of width 10.
 - (iv) Sketch examples of different sampling techniques using samples of size 5 and the strata low, medium and high. (10)

Or

- b) Build a decision tree using the training data in the given table. Divide the height attributes into ranges (0, 1.6), (1.6, 1.7), (1.7, 1.8), (1.8, 1.9), (1.9, 2.0), (2.0, 5.0).

Gender	Height	Class
F	1.6	Short
M	2	Tall
F	1.9	Medium
F	1.88	Medium
F	1.7	Short
M	1.85	Medium
F	1.6	Short
M	1.7	Short
M	2.2	Tall
M	2.1	Tall
F	1.8	Medium
M	1.95	Medium
F	1.9	Medium
F	1.8	Medium
F	1.75	Medium

(10)

