



LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION - STATISTICS

THIRD SEMESTER – NOVEMBER 2011

ST 3955 - DATA WAREHOUSING AND DATA MINING

Date : 08-11-2011
Time : 9:00 - 12:00

Dept. No.

Max. : 100 Marks

SECTION - A

Answer all the questions

(10x2=20)

1. Define Data Warehousing
2. Define Data Mart
3. Define OLTP
4. Define Data Mining
5. State any two uses of Multiple Linear regression model
6. Define dummy variable rule and explain the consequence of introducing m dummy variables for a categorical variable taking m categories in a multiple linear regression model with intercept
7. What is the use of a gains chart?
8. State the methods of hierarchical Clustering
9. Define kth principal component
10. Explain the two methods of model validation

SECTION B

Answer any five questions

(5x8=40)

11. State any five application of data mining and provide the steps involved in a data mining project
12. Explain order to cash process
13. Explain the purpose of the components for Decision support system
14. Explain the test for overall significance of the model and test for individual regression coefficients of a multiple linear regression model
15. Write a short note on factor analysis
16. Explain the steps involved in Naive Bayes classification
17. Write a short note on Artificial neural network

18. Factor Analysis Output (Varimax Rotation)

| Variable | Factor1 | Factor2 | Communality |
|---------------|---------|---------|-------------|
| Hours Studied | 0.953 | -0.198 | ? |
| Sports | -0.794 | 0.476 | ? |
| Film | -0.306 | 0.946 | ? |
| Mark | 0.868 | -0.428 | ? |
| | | | |
| Variance | ? | ? | ? |
| % Var | ? | ? | 0.932 |

- i) Determine the required values in the table
- ii) Name the factors and Interpret the variance explained by the factors
- iii) Interpret the communality values

SECTION C

Answer any two questions **(2x20=40)**

- 19. (i) Explain ETL Process (10)
 (ii) Explain three tier architecture (10)

- 20. (i) How would you load the data in flat file to a table using SQL loader (10)
 (ii) Explain purchase to pay process (10)

- 21. (i) Explain logistic regression model (6)
 (ii) Explain the steps involved in constructing a gains table (6)
 (iii) Define Decision rule, Classification table, Sensitivity, Specificity (4)
 (iv) Explain Gains chart and its use (4)

- 22. (i) Explain single linkage, Complete linkage, Average linkage hierarchical clustering (12)
 (ii) Explain K-Means clustering (8)
