

ALCCS – NEW SCHEME

Time: 3 Hours

AUGUST 2013

Max. Marks: 100

PLEASE WRITE YOUR ROLL NO. AT THE SPACE PROVIDED ON EACH PAGE IMMEDIATELY AFTER RECEIVING THE QUESTION PAPER.

NOTE:

- Question 1 is compulsory and carries 28 marks. Answer any FOUR questions from the rest. Marks are indicated against each question.
- Parts of a question should be answered at the same place.

-
- Q.1** a. Differentiate between data warehouse and operational database.
- b. List the applications of Data Mining.
- c. Differentiate between Data Warehouse and Data Mart.
- d. Explain Data granularity in Data Warehouse.
- e. Briefly explain the Data Mining Strategies.
- f. Explain with reference to Data Warehouse: “Data inconsistencies are removed; data from diverse operational applications is integrated”.
- g. What do you understand by information packages? (7×4)
- Q.2** a. What are the advantages of Star Schema? Explain. (9)
- b. What are the various Star Schema Keys? Explain With the help of an example. (9)
- Q.3** a. What are the characteristics of the OLTP and the basic data warehouse environments as they relate to information delivery needs? (9)
- b. There are mainly two models of OLAP – ROLAP and MOLAP. State the characteristics of these two models of OLAP. (9)
- Q.4** a. List and explain activities of ETL process. (9)
- b. What are the advantages and disadvantages of Top-Down Approach and Bottom-Up Approach for data warehouse design? (9)
- Q.5** a. Explain the differences between Regression Tree and Model Tree. (4)
- b. Why is data security a major concern for web-enabled data warehouse? (5)

- c. The average member technique is sometimes used to explain the results for an unsupervised neural network clustering
 (i) List the advantages and disadvantages of this approach.
 (ii) Is there any similarity between this explanation technique and the K-means algorithm? Explain. (9)

- Q.6 a. List the similarities and differences between bagging and boosting. State a situation where boosting is likely to outperform bagging. (4)
- b. Use the data contained in Table1 to fill in the counts and probabilities in Table 2. The output attribute is *life insurance promotion*.

Table1:

Magazine Promotion	Watch Promotion	Life insurance promotion	Credit card insurance	Gender
Yes	No	No	No	Male
Yes	Yes	Yes	Yes	Female
No	No	No	No	Male
Yes	Yes	Yes	Yes	Male
Yes	No	Yes	No	Female
No	No	No	No	Female
Yes	Yes	Yes	Yes	Male
No	No	No	No	Male
Yes	No	No	No	Male
Yes	Yes	Yes	No	Female

Table 2:

	Magazine Promotion		Watch Promotion		Credit Insurance		Gender
	Yes	No	Yes	Yes	No	Male	
Life Insurance Promotion							Female
Yes							
No							
Ratio: Yes/Total							
Ratio: No/Total							

- (i) Use the completed table together with Bayes classifier to determine the value of *life insurance promotion* for the following instance.

Magazine promotion=yes
 Watch promotion=yes
 Credit Card insurance=No
 Gender=female
 Life Insurance promotion=?

Code: CT75**Subject: DATA WAREHOUSING AND DATA MINING**

(ii) Repeat part (i), but assume that the gender of the customer is unknown. (7×2)

Q.7 a. Write short note on any **TWO**

(i) Data Transformation

(ii) Constraint Based Association mining

(iii) Requirements of clustering in data mining (6+6)

b. State the criteria for evaluation of classification and prediction methods. (6)