29/6/22
Evening

Exam.Code:0918
Sub. Code: 6798

2062
B.E. (Computer Science and Engineering)
Sixth Semester
Elective – I
CS-605C: Data Mining and Analysis

Time allowed: 3 Hours

Max. Marks: 50

NOTE: Attempt five questions in all, including Question No. I which is compulsory and selecting two questions from each Part.

x-x-x

I. Attempt the following:- (5x2)

a) Define attribute oriented induction.

b) Discuss the benefits of hybrid OLAP architecture.

c) Deliberate on the need for data reduction before data mining, given the tremendous growth in computational power of systems.

d) State some applications of association rule mining.

e) Differentiate between classification and prediction with the help of suitable example.

Part A

II What is a concept hierarchy? Describe any method to automatically generate concept hierarchy for 'income' attribute of a retail store customers. (10)

III Suppose a data mart contains 10 dimensions, each with 3 levels. Users are mainly interested in 2 dimensions with 3 levels for analysis. How would you design a data cube to support this feature efficiently? (10)

IV Discuss the types of metadata and its role in a datawarehouse architecture. (10)

Part B

V a) How does correlation analysis help in data mining? Discuss any 1 method for calculating correlation between attributes. (5)

b) Consider a binary classification problem with two classes C1 and C2. Class labels of ten other training set instances sorted in increasing order of their distance to an instance x is as follows: {C1, C2, C1, C2, C2, C2, C1, C2, C1, C2}. How will a K=7 nearest neighbor classifier classify x? (5)

P.T.O.

VI a) A bank classifies its customer into two classes "fraud" and "normal" based on their installment payment behavior. We know that the probability of a customer being being fraud is P(fraud) = 0.20, the probability of customer defaulting installment payment is P(default) = 0.40, and the probability that a fraud customer defaults in installment payment is P(default|fraud) = 0.80. What is the probability of a customer who defaults in payment being a fraud? (5)

b) Write a short note on mining World Wide Web along with some relevant applications. (5)

VII Present pseudocode for mining single level associations with candidate generation from transaction database. What metrics can be used to evaluate a rule mining algorithm? (10)

x-x-x