DOE
03-02-16.

Semester: II (M.Tech., DA)
Sub & Code: DWDM, CS-6301
Branch (s): CSE (Data Analytics)

SPRING MID SEMESTER EXAMINATION-2016

Data Warehousing and Data Mining

[CS-6301]

Full Marks: 25 **Time: 2 Hours**

Answer any four questions including question No.1 which is compulsory.
The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.

1. Answer the following questions: [1 × 5]
   a) What is data mining?
   b) What are different ways to fill the missing attribute value?
   c) Define decision tree induction.
   d) Differentiate between supervised and unsupervised learning.
   e) Define support and confidence in Association rule mining with suitable example.

2. Describe the steps involved in data mining when viewed as a process of knowledge discovery in databases. 5

3. a) Use smoothing by bin means to smooth the given data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data. The data (in increasing order) for the attribute age is as follows: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. 3
   b) Use min-max normalization to transform the value 35 for age onto the range [0.0, 1.0]. (from the data shown in (a)). 2

4. The following database has five transactions. Let min_sup = 60% and min_conf = 80% Find all frequent itemsets using Apriori and FP-growth, respectively. Compare th efficiency of the two mining processes. 5

TID	Items_bought
T100	[M, O, N, K, E, Y]
T200	[D, O, N, K, E, Y]
T300	[M, A, K, E]
T400	[M, U, C, K, Y]
T500	[C, O, O, K, I, E]

5. Briefly outline the major steps of decision tree classification with a suitable example. 5

6. The following table shows the midterm and final exam grades obtained for students in a

database course.

Sl. no	X-Midterm Exam.	Y-Final Exam.	Sl. no	X-Midterm Exam.	Y-Final Exam.
1	72	84	7	59	49
2	50	63	8	83	79
3	81	77	9	65	77
4	74	78	10	33	52
5	94	90	11	88	73
6	86	75	12	81	90

a) Use the method of least squares to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course. 4

b) Predict the final exam grade of a student who received an 86 on the midterm exam. 1

7. Given a data tuple with the values "<=30", "medium", "yes" and "fair" for the attributes age, income, student, and credit_rating, respectively, what would a naive Bayesian classification of the buys_computer for the tuple be? 5

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	High	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

==========