CLASS: BE   SEMESTER : VII
BRANCH: IT   SESSION : MO/19

SUBJECT: IT7021 DATA MINING CONCEPTS AND TECHNIQUES

TIME: 3:00 HOURS   FULL MARKS: 60

INSTRUCTIONS:
1. The question paper contains 7 questions each of 12 marks and total 84 marks.
2. Candidates may attempt any 5 questions maximum of 60 marks.
3. The missing data, if any, may be assumed suitably.
4. Before attempting the question paper, be sure that you have got the correct question paper.
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

Q.1(a) What are the data mining functionalities?   [2]
Q.1(b) What are the different types of attribute? Explain with examples.   [4]
Q.1(c) What are the different kinds of Data that can be Mined? What kind of data are used in multi-level and   [6]
        multi-dimensional association? Briefly explain Information Retrieval.

Q.2(a) What are the different types of OLAP server? Briefly explain   [2]
Q.2(b) State the motivations of data reduction? Write a short note on Data reduction strategies   [4]
Q.2(c) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are   [6]
        (in increasing order) 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.
        1) Use smoothing by bin means to smooth these data, using a bin depth of 3. Illustrate your steps.
        Comment on the effect of this technique for the given data.
        2) What other methods are there for data smoothing?

Q.3(a) Briefly explain Data Mining Engine in DM Architecture.   [2]
Q.3(b) Write a note on DMQL for Specifying the Kind of Knowledge.   [4]
Q.3(c) Write a note on the types of data mining architecture?   [6]

Q.4(a) Differentiate between data characterization and data discrimination.   [2]
Q.4(b) Write a note on Five-Number Summary with example.   [4]
Q.4(c) Suppose that a hospital tested the age and body fat data for 18 randomly selected adults   [6]
        with the following results:

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

        1) Calculate the median, and standard deviation of *age* and *%fat*.
        2) Draw the box plots for *age* and identify outlier.
        3) Draw a *quantile plot* based on these two variables

Q.5(a) What is Association rule mining? What is the major drawback of Apriori algorithm and how FP-growth   [2]
        overcome that?
Q.5(b) Explain how Support-Confidence Rule is sometimes misleading with example. Define the two   [4]
        correlation measures. Consider given contingency table of sales transactions for computer games and
        videos and find correlation using lift and χ^2:

	game	**$\overline{game}$**
video	4000 (4500)	3500 (3000)
$\overline{video}$	2000 (1500)	500 (1000)

Q.5(c) Write a note on join and prune step of Apriori Algorithm. Given the following transations (T1-T9):< T1: A, B, E >, < T2: B, D >< T3: B, C >< T4: A, B, D >< T5: A, C >< T6: B, C >< T7: A, C >< T8: A, B, C, E >< T9: A, B, C > . Generate frequent patterns using FP-Growth Algorithm. Consider minimum support threshold as 2. [6]

Q.6(a) Differentiate between classifier and predictor. Describes the general approach to classification as a two-step process. [2]

Q.6(b) What does splitting criterion mean in decision tree induction? Explain what it determines with example. What is tree pruning? [4]

Q.6(c) How neural network's topology designed? What are the advantages and disadvantages of neural network? What are the inputs to Back propagation algorithm? How does back propagation algorithm work? [6]

Q.7(a) What is Cluster Analysis? What are the requirements for Cluster Analysis? [2]

Q.7(b) State General Characteristics of partitioning methods, hierarchical methods and density-based methods. [4]

Q.7(c) Write a note on k-medoids and BIRCH. Explain with example. [6]

::::::25/11/2019::::::E