QP Code : 5178

IT-VI CBGS   a2zSubjects.com

[Total Marks: 80]

(3 Hours)

N.B. : (1) Question **No. 1** is **compulsory**.

    (2) Answer any **three** out of the **remaining** questions.

Q1. **[a]** Describe the different types of attributes one may come across in a data mining data set with two examples of each type. [05]

    [b] Explain the different distance measures that can be used to compute distances between two clusters. [05]

[c] Define "Business Intelligence" and Decision Support System", with examples. [05]

[d] Define "Outlier". What are the different types of Outliers that occur in a dataset? [05]

Q2. **[a]** Consider the following data points: 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.

    (a) What is the *mean* of the data? What is the *median*?   a2zSubjects.com

    (b) What is the *mode* of the data?

    (c) What is the *midrange* of the data?

    (d) Can you find (roughly) the first quartile ($Q1$) and the third quartile ($Q3$) of the data? [10]

    (e) Show a *boxplot* of the data.

[b] Design a BI system for fraud detection. Describe all the steps from Data collection to Decision Making clearly [10]

Q 3. [a]

Id	Homeowner	Status	Income	Defaulted
1	Yes	Employed	High	No
2	No	Business	Average	No
3	No	Employed	Low	No
4	Yes	Business	High	No
5	No	Unemployed	Average	Yes
6	No	Business	Low	No
7	Yes	Unemployed	High	No
8	No	Employed	Average	Yes
9	No	Business	Low	No
10	No	Employed	Average	Yes

Illustrate any one classification technique for the above data set. Show how we can classify a new tuple, with (**Homeowner = Yes; Status = Employed; Income = Average**). [10]

[TURN OVER

JP-Con. : 11036-15.   a2zSubjects.com

[b] Why is Data Preprocessing required? Explain the different steps involved in Data Preprocessing.   **[10]**

Q 4. [a] Use **K-means** to cluster the following data set into 3 clusters.   **[10]**

Protein	20	21	15	22	20	25	26	20	18	20
Fat	9	9	7	17	8	12	14	9	9	9

[b] Describe the different visualization techniques that can be used in data Mining.   **[10]**

Q.5 [a] Consider the following transaction database:

TID	Items
01	A, B, C, D
02	A, B, C, D, E, G
03	A, C, G, H, K
04	B, C, D, E, K
05	D, E, F, H, L
06	A, B, C, D, L
07	B, I, E, K, L
08	A, B, D, E, K
09	A, E, F, H, L
10	B, C, D, F

Apply the **Apriori** algorithm with minimum support of 30% and minimum confidence of 70%, and find all the association rules in the data set.   **[10]**

[b] Explain different methods that can be used to evaluate and compare the accuracy of different classification algorithms.   **[10]**

Q 6. **Explain in brief:**
     [a] DBSCAN clustering algorithm with an example   **[10]**
     [b] Multilevel and Multidimensional Association rules   **[10]**

———— ✦ ————

JP-Con. : 11036-15.