

I.T/CBGS/VI/ DM & BI /14-12-16
Data mining & Business Intelligence.



Q.P. Code : 595001

(3 Hours)

[Total Marks : 80

- N.B. : (1) Q1 is compulsory.
(2) Attempt any **three** from the remaining.
(3) Assume suitable data.

1. (a) Define "Data Mining". Enumerate five example applications that can benefit by using Data Mining. 5
- (b) Clearly explain the data preprocessing phase for data mining. 5
- (c) Describe one hierarchical clustering algorithm using an example dendrogram. 5
- (d) Explain the concept of a decision support system with the help of an example application. 5
2. (a) Partition the given data into 4 bins using Equi-depth binning method and perform smoothing according to the following methods. 10
Smoothing by bin mean
Smoothing by bin median
Smoothing by bin boundaries

Data: 11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 21, 21, 22, 23, 24, 30, 40, 45, 45, 45, 71, 72, 73, 75
- (b) For the same set of data points in question 2. (a) 10
 - (a) Find Mean, Median and Mode.
 - (b) Show a boxplot of the data. Clearly indicating the five- number summary.

[TURN OVER

3. (a) The table below shows a sample dataset of whether a customer responds to a survey or not. "**Outcome**" is the class label. Construct a Naive Bayes' Classifier for the dataset. For a new example (**Rural, semidetached, low, No**), what will be the predicted class label? 10

District	House Type	Income	Previous Customer	Outcome
Suburban	Detached	High	No	Nothing
Suburban	Detached	High	Yes	Nothing
Rural	Detached	High	No	Responded
Urban	Semi-detached	High	No	Responded
Urban	Semi-detached	Low	No	Responded
Urban	Semi-detached	Low	Yes	Nothing
Rural	Semi-detached	Low	Yes	Responded
Suburban	Terrace	High	No	Nothing
Suburban	Semi-detached	Low	No	Responded
Urban	Terrace	Low	No	Responded
Suburban	Terrace	Low	Yes	Responded
Rural	Terrace	High	Yes	Responded
Rural	Detached	Low	No	Responded
Urban	Terrace	High	Yes	Nothing

- (b) Briefly explain Regression based Classifiers 10

4. (a) Use the Apriori algorithm to identify the frequent item-sets in the following database. Then extract the strong association rules from these sets. 10

Min. Support = 30% Min. Confidence = 75%

TID	Items
01	A, B, D, E, F
02	B, C, E
04	A, B, D, E
04	A, B, C, E,
05	A, B, C, D, E, F
06	B, C, D
07	A, B, D, E

- (b) Explain multidimensional and multi level Association rules with examples. 10
5. (a) What is clustering? Explain k-means clustering algorithm. Suppose the data for clustering is {2,4,10,12, 3,20,11,25} 10
Consider $k = 2$, cluster the given data using K-means algorithm
- (b) What is an outlier? Describe methods that can be used for outlier analysis. 10
6. (a) Consider the following case study: A telecom company wants to analyze and improve its performance by introducing a series of innovative mobile payment plans. For this case study design a BI system, clearly explaining all steps from data collection to decision making. 10
- (b) Clearly explain the working of the DBSCAN algorithm using appropriate diagrams. 10