R16

SET - 1

III B. Tech II Semester Regular/Supplementary Examinations, October/November - 2020
DATA WAREHOUSING AND MINING
(Computer Science and Engineering)

Time: 3 hours   Max. Marks: 70

Note: 1. Question Paper consists of two parts (**Part-A** and **Part-B**)
2. Answer **ALL** the question in **Part-A**
3. Answer any **FOUR** Questions from **Part-B**
~~~~~~

PART –A   (14 Marks)

1. a) Justify the need for data mining. [2M]
   b) What are the principal components? How are principal components used in feature selection? [2M]
   c) What is a method used for classification that can be used even if some of the variables are categorical? [2M]
   d) What is the main idea of naïve Bayesian classification? [3M]
   e) Define frequent itemset, support and confidence. [3M]
   f) How would you measure the quality of clusters? [2M]

PART –B   (56 Marks)

2. a) What are the major issues in Data Mining? Explain briefly. [7M]
   b) Explain the steps involved in the Data Mining Process. Give the sketch of the KDD process. [7M]

3. Explain various data pre-processing methods with appropriate examples. [14M]

4. a) Explain the ID3 algorithm for the induction of decision trees. [6M]
   b) Given a decision tree, there exists an option of (i) converting the decision tree to rules and then pruning the resulting rules, or (ii) pruning the decision tree and then converting the pruned tree to rules? What advantage does (i) have over (ii)? [8M]

5. a) What is Bayes Theorem? Show how it is used for classification? [7M]
   b) Why Naïve Bayesian classification is called "naïve"? Explain. [7M]

6. a) Develop the Apriori Algorithm for generating frequent-item set. [8M]
   b) What is association analysis? Explain. [6M]

7. What are the advantages and disadvantages of k-means clustering against model-based clustering?
   You are given a set of numbers {2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377}.
   Use the following techniques to find two clusters from this data set.
   (i) K-Means with initial centroids {1} and {378}
   (ii) K-Means with initial centroids {21} and {34}. [14M]

   Explain the differences between K-means clustering and kernel K-means clustering.

*****