R16

SET - 1

III B. Tech II Semester Regular/Supplementary Examinations, August-2021
DATA WAREHOUSING AND DATA MINING
(Computer Science 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) Differentiate transactional data and data warehouse. [2M]
   b) How to handle missing values? Give examples. [2M]
   c) How to prepare data for classification? [2M]
   d) Write about priori and posteriori probabilities used in classification. [3M]
   e) What is strong association rule? Give example. [3M]
   f) Write a short note on type of clusters. [2M]

PART –B   (56 Marks)

2. a) Explain the following: [7M]
      i)   Classification and regression for predictive analysis;
      ii)   Cluster and outlier analysis.
   b) What is data mining? Why it is important? Explain various issues in data mining in detail. [7M]

3. a) How to improve the quality of data? Explain various approaches and tasks used in data preprocessing. [7M]
   b) Give the overview of data transformation strategies. And explain data transformation by normalization. [7M]

4. a) How to classify the data using decision tree induction algorithm? Explain what attributes can be considered for classifying email as spam or ham. [7M]
   b) Explain the role of attribute selection measures in decision tree construction with suitable examples. [7M]

5. a) What is Bayesian classifier? Explain Naïve Bayes classifier. [7M]
   b) Explain the working principle of Bayesian belief networks. How do they train on sample data? [7M]

6.   With an example explain, market-basket analysis. Use Aprioiri [14M]
      algorithm for the given transactions and perform market-basket
      analysis.

TID	Items
1	A, M
2	A, D, H, E
3	M, D, H, C
4	A, M, D, H
5	M, H, C

7. a) Discuss different variants of K-Means clustering algorithm and   [7M]
      discuss the issues they resolve in each case.
   b) How to implement Density Center-Based Approach using DBSCAN   [7M]
      algorithm? Explain.

*****

2 of 2