Code No: R1632052 **R16** SET - 1

III B. Tech II Semester Supplementary Examinations, November - 2019

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) How to use Euclidian distance measure for similarity and dissimilarity? Give example. [2M]
   b) Write a short note on discretization. [2M]
   c) Write about induction and deduction tasks in classification. [2M]
   d) Explain Bayes theorem. [3M]
   e) Write a short note on market basket analysis. [3M]
   f) Describe various types of clusters. [2M]

PART –B (56 Marks)

2. a) What is data set? Describe different characteristics and types of data sets used in data mining. [7M]
   b) How to summarize the properties of data using statistical measures? Give example for each measure. [7M]

3. a) Describe the problem of data quality with some examples. Explain the usage of feature subset selection in data preprocessing. [7M]
   b) What is data consolidation? In detail discuss various techniques used to consolidate data. [7M]

4. a) Write and explain decision tree classifier with induction algorithm with an example. [7M]
   b) In continuation to Q. 4(a) explain various measures used to find the best split node in decision tree. [7M]

5. What is Naïve Bayes classifier? How to estimate the probabilities from data? Give example for classification. [14M]

6. Assume 5 transactions and explain the two step approach to generate frequent item sets and to mine association rules using Apriori algorithm. [14M]

7. a) Explain density based algorithm for clustering data. Discuss its merits and demerits. [7M]
   b) Discuss the working procedure of bisecting K-means algorithm. How it improves the performance of K-means? Explain. [7M]

*****
~~~~~~