Code No: R1632052



## 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) How to use Euclidian distance measure for similarity and dissimilarity? Give 1. a) [2M] example. Write a short note on discretization. **b**) [2M] Write about induction and deduction tasks in classification. c) [2M] Explain Bayes theorem. d) [3M] e) Write a short note on market basket analysis. [3M] Describe various types of clusters. f) [2M] PART –B (56 Marks) 2. What is data set? Describe different characteristics and types of data sets used in [7M] a) data mining. How to summarize the properties of data using statistical measures? Give example b) [7M] for each measure. Describe the problem of data quality with some examples. Explain the usage of 3. a) [7M] feature subset selection in data preprocessing. What is data consolidation? In detail discuss various techniques used to **b**) [7M] consolidate data. 4. Write and explain decision tree classifier with induction algorithm with an a) [7M] example. In continuation to Q. 4(a) explain various measures used to find the best split node **b**) [7M] in decision tree. 5. What is Naïve Bayes classifier? How to estimate the probabilities from data? Give [14M] example for classification. 6. Assume 5 transactions and explain the two step approach to generate frequent item [14M] sets and to mine association rules using Apriori algorithm. 7. Explain density based algorithm for clustering data. Discuss its merits and [7M] a) demerits. Discuss the working procedure of bisecting K-means algorithm. How it improves **b**) [7M] the performance of K-means? Explain. \*\*\*\*

## [["["]["]["]["][Www.manaresults.co.in