

BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION MO/SP20**)

CLASS: IMSC
BRANCH: MATHS & COMP.

SEMESTER : VII
SESSION : MO 2022

SUBJECT: CA532 DATA MINING AND WAREHOUSING
TIME: 03 Hours

FULL MARKS: 50

INSTRUCTIONS:

1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
4. Tables/Data handbook/Graph paper etc., if applicable, will be supplied to the candidates

-
- Q.1(a) Show the major problem areas in Data mining in the context of Efficiency and Scalability? (CO1) (BT-2) [2]
- Q.1(b) Distinguish between Classification and Regression using suitable example. (CO1) (BT-4) [3]
- Q.1(c) Examine through a numerical example that how the proximity for the binary attributes is measured. (CO1)(BT-4) [5]
- Q.2(a) "Partial materialization is preferred over no materialization". Justify this statement. (CO2) (BT-5) [2]
- Q.2(b) Distinguish between operational database systems and data warehouses. (CO2) (BT-4) [3]
- Q.2(c) Discuss and compare different data models used for designing Datawarehouse with suitable examples. [5]
- Q.3(a) List the two nontrivial costs an Apriori approach suffers from. (CO3) (BT-4) [2]
- Q.3(b) Identify and explain two techniques for improving the efficiency of Apriori Algorithm. (CO3) (BT-3) [3]
- Q.3(c) In the context of applying Apriori Property, illustrate the following: [5]
- i) The Join Step ii) The Prune Step (CO3) (BT-5)
- Q.4(a) Explain the case where 'Gain Ratio' is preferred over 'Information gain' as an attribute selection measure. (CO5) (BT-2) [2]
- Q.4(b) Examine the Naïve assumption of class-conditional independence. (CO4) (BT-4) [3]
- Q.4(c) Construct an algorithm for rule induction using Sequential Covering in rule-based classification. (CO4)(BT-3) [5]
- Q.5(a) Analyze the K-mean algorithm for its sensitivity for the outliers (CO4) (BT-4) [2]
- Q.5(b) List the major tasks of clustering evaluation. (CO5) (BT-4) [3]
- Q.5(c) Select and explain a partitioning algorithm for clustering. Also analyze its complexity. (CO5) (BT-3) [5]

:::::23/11/2022:::::E