

III-B.Tech II-Semester Supplementary Examinations (BR20), SEP - 2023
DATA MINING AND DATA WAREHOUSING (CSE& CSE(AI&DS))
Max. Marks: 70

Time: 3 hours

Question Paper consists of FIVE units, each carrying 14 marks
Each unit has TWO questions; either of them should be answered
All parts of a question must be answered at one place

- 1.a) What are the goals of data mining, what are the difficulties and advantages, and why is data mining important? (7M)
- 1.b) What are the main problems with data mining, why are fascinating patterns crucial for data mining, and how does data mining function? (7M)
- (OR)
- 1.c) What types of patterns can be mined? How do data mining task primitives work? (7M)
- 1.d) What kind of patterns can be mined, data mining functionalities, and are all patterns interesting? (7M)
- 2.a) Describe data preprocessing using an example. What are the five main steps in the preparation of data? (7M)
- 2.b) What are the methods of discretization and concept hierarchy generation, and why is data summarization crucial? (7M)
- (OR)
- 2.c) What does data mining is multidimensional data model entail? What are the components of a data warehouse's architecture and what is OLAP technology? (7M)
- 2.d) What purpose does data reduction serve? Give an example of data integration and transformation. (7M)
- 3.a) What challenges are present throughout each of the two stages of association rule mining? How may mining frequent item sets be applied to classification? (7M)
- 3.b) How does the Apriori algorithm produce frequent itemsets? What is the algorithm for creating frequent itemsets? (7M)
- (OR)
- 3.c) Which algorithm is capable of producing an association rule? In Apriori, how do I create candidates? (7M)
- 3.d) Describe the Compact Representation of Frequent Itemsets and the FP growth algorithm's phases. (7M)
- 4.a) What does data mining is general approach to classification entail? How can classification issues be resolved in data mining? (7M)
- 4.b) What are the several ways to define an attribute test condition and how do you choose the optimum split? (7M)
- (OR)
- 4.c) What is decision tree induction in data mining, what are its characteristics, and what is the algorithm? (7M)
- 4.d) What is naive Bayesian classification and how is it defined in data mining? What is a practical illustration of Naive Bayes? (7M)
- 5.a) What makes bisecting K-means different from K-means? In addition, its justification of the techniques for bisecting K-means from K-means. (7M)
- 5.b) What is the purpose of hierarchical clustering? The advantages and disadvantages of hierarchical clustering. (7M)
- (OR)
- 5.c) Tell us more about DBSCAN. What is the DBSCAN Clustering Process? What does the standard density-based method entail? (7M)
- 5.d) Describe K-means as an optimization issue. What is the agglomerative hierarchical clustering algorithm? what are the advantages and disadvantages? (7M)
- *****

Received
Final Copy
U.S. For
05/09/21.