

(FOR THE CANDIDATES ADMITTED  
DURING THE ACADEMIC YEAR 2021 ONLY)

21UBC6E4

REG.NO. :

**N.G.M.COLLEGE (AUTONOMOUS) : POLLACHI**

**END-OF-SEMESTER EXAMINATIONS: MAY-2024**

**COURSE NAME: B.C.A**

**MAXIMUM MARKS: 70**

**SEMESTER: VI**

**TIME : 3 HOURS**

**PART - III**

**DATAMINING AND WAREHOUSING**

**SECTION - A**

**(10 X 1 = 10 MARKS)**

**ANSWER THE FOLLOWING QUESTIONS.**

**MULTIPLE CHOICE QUESTIONS.**

**(K1)**

1. In which application domain might data mining be used for fraud detection?
  - a) Social Media
  - b) Healthcare
  - c) Finance
  - d) Business and Marketing
2. What is Tuple Duplication in the context of data integration?
  - a) Duplicating entire datasets
  - b) Duplicating attribute values within a tuple
  - c) Removing duplicate tuples
  - d) Ignoring duplicate tuples
3. Which technique is used for finding patterns based on likelihood and distance?
  - a) Decision Trees
  - b) K-Nearest Neighbour
  - c) Association Rules
  - d) Neural Networks
4. Which of the following is NOT a part of the 10 Golden Rules?
  - a) Understand the Business
  - b) Keep Data Simple
  - c) Ignore Ethical Considerations
  - d) Ensure Transparency
5. What does Attribute-Oriented Induction focus on?
  - a) Data complexity
  - b) Identifying key attributes in a dataset
  - c) Increasing data redundancy
  - d) Ignoring data warehousing design

**ANSWER THE FOLLOWING IN ONE (OR) TWO SENTENCES.**

**(K2)**

6. Name the data visualization technique used in data mining.
7. What is tuple duplication in the context of data cleaning?
8. What is the primary objective of data selection in the Knowledge Discovery Process?
9. Identify the primary goal of Knowledge Discovery in Databases.
10. List the characteristics of a data warehouse.

**(CONTD ... 2)**

**SECTION – B****(5 X 4 = 20 MARKS)****ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS. (K3)**

11. a) Differentiate between structured and unstructured data.

**(OR)**

b) Examine the applications and developments in data mining.

12. a) Outline the steps involved in a systematic data cleaning process.

**(OR)**

b) How do dimensionality reduction techniques contribute to data reduction? Analyze.

13. a) Describe various OLAP operations performed on Multidimensional Data Model.

**(OR)**

b) Elucidate the basic concept behind the K-Nearest Neighbour algorithm.

14. a) Outline the steps involved in identifying and selecting relevant data sources for a KDD project.

**(OR)**

b) Illustrate the key steps involved in implementing machine learning models in the coding phase.

15. a) Write short notes on Meta Data in Warehouse.

**(OR)**

b) Differentiate between logical data warehouse and physical data warehouse.

**SECTION - C****(4 X 10 = 40 MARKS)****ANSWER ANY FOUR OUT OF SIX QUESTIONS****(16<sup>th</sup> QUESTION IS COMPULSORY AND ANSWER ANY THREE QUESTIONS****(FROM Qn. No : 17 to 21)****(K4 (Or) K5)**

16. Differentiate between association, clustering and classification patterns in Data Mining. Provide real-world examples for each type.

17. Discuss the various major issues in data mining.

18. Discuss the methods for handling redundancy in a dataset during the data integration process.

19. Illustrate the steps involved in Knowledge Discovery Process. Draw a neat diagram.

20. Describe the importance of selecting appropriate data subsets based on analysis objectives.

21. Illustrate the three-tier data warehousing architecture with neat diagram.

\*\*\*\*\*