

(FOR THE CANDIDATES ADMITTED

22UCS5E1

DURING THE ACADEMIC YEAR 2022

ONLY)

REG.NO.

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

END-OF-SEMESTER EXAMINATIONS : NOVEMBER 2024

BSC COMPUTER SCIENCE(AIDED & SF)

MAXIMUM MARKS: 50

SEMESTER-V

TIME : 3 HOURS

**PART – III**

**22UCS5E1– DATAMINING AND WAREHOUSING**

**SECTION – A**

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

**ANSWER THE FOLLOWING QUESTIONS.**

**(K1)**

1. Which one of the following is a main task of Data Mining?  
a) Data Entry   b) Predictive Modeling   c) Data Storage   d) Network Configuration
2. What is the first step in the Data Mining Methodology?  
a) Select Appropriate Data   b) Build Models  
c) Translate the Business Problem into a Data Mining Problem  
d) Assess Models
3. What does OLAP stand for?  
a) Online Linear Analytics Processing  
b) Online Logical Analytics Processing  
c) Online Analytical Processing  
d) Online Analytical Pipeline
4. Which of the following is a key aspect of preparing data for Data Mining?  
a) Selecting data from a single source  
b) Aggregating data from multiple sources  
c) Ignoring missing values  
d) None of the above
5. In the K-Means algorithm, the value of 'k' refers to: \_\_\_\_\_  
a) Number of data points   b) Number of clusters   c) Number of attributes   d) Number of iterations

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

**(K2)**

6. Why has Data Mining become more prominent now why?
7. Why "Build Models," step important in the Data Mining process?
8. What is a Data Cube in the context of OLAP?
9. Why Handling Missing data important in data mining?
10. What is a stopping criterion in the context of Decision Trees?

**SECTION – B****(5 X 3 = 15 MARKS)****ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS. (K3)**

11. a) Describe classification and clustering in Data Mining.  
(OR)
- b) Explain how Data Mining can be used in Analytic Customer Relationship Management (CRM).
12. a) Describe the importance of translating the Business Problem into a Data Mining problem.  
(OR)
- b) Explain the role of Data Transformation in the Data Mining methodology.
13. a) Explain the relationship between Data Warehousing and OLAP.  
(OR)
- b) Describe the role of Data Mining in a Data warehousing environment.
14. a) Explain the concept of 'Model Roles in Modeling' and its significance in data mining.  
(OR)
- b) What are the computational issues that arise in data mining, and how can they be mitigated?
15. a) Explain the concept of Support And Confidence In Association Rule Mining.  
(OR)
- b) What is Pruning in the context of Decision Trees, and why is it important?

**SECTION – C****(5 X 5 = 25 MARKS)****ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.(K4 (Or) K5)**

16. a) Discuss the reasons why Data Mining has gained popularity in recent years and provide examples of its application in businesses today.  
(OR)
- b) What are the ethical considerations and challenges associated with the use of Data Mining in business applications? Explain in detail.
17. a) Discuss the steps involved in preparing data for Model Building.  
(OR)
- b) What are the key considerations when assessing and deploying models in Data Mining, and why is it important to begin the process again?
18. a) Discuss the general Architecture of Data Warehousing and explain how OLAP integrates with this architecture.  
(OR)
- b) Compare and contrast OLAP and Data Mining in terms of their purpose, approach, and outcomes within the context of data warehousing.
19. a) Discuss the importance of 'The Dark Side of Data' in data mining and how it affects the overall data mining process.  
(OR)
- b) Explain the process of creating 'The Customer Signature' and its role in data mining.

20. a) Given a transaction dataset as follows:

Transaction ID	Items Purchased
T1	{Bread, Milk}
T2	{Bread, Diapers, Beer, Eggs}
T3	{Milk, Diapers, Beer, Coke}
T4	{Bread, Milk, Diapers, Beer}
T5	{Bread, Milk, Coke}

Using the Apriori algorithm, find all the frequent itemsets with a minimum support threshold of 2. Then, generate Association Rules with a minimum confidence of 50%.

(OR)

- b) You have the following 2-dimensional data points:

Point	X	Y
P1	2	3
P2	3	3
P3	6	7
P4	8	7
P5	1	5

Apply the K-Means algorithm with  $k=2$  to cluster these points. Start with initial Centroids at (2, 3) and (6, 7). Perform one iteration and assign points to the nearest Centroid, then recalculate the centroids.

\*\*\*\*\*