(NO. OF PAGES: 2)

(FOR THE CANDIDATES ADMITTED DURING THE ACADEMIC YEAR 2023 ONLY) **23UDA412**

REG.NO. :

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

END-OF-SEMESTER EXAMINATIONS: MAY - 2025

B.Sc (CS with Data Analytics)-SF **MAXIMUM MARKS: 75**

SEMESTER: IV **TIME : 3 HOURS**

PART - III

23UDA412 – DATA MINING AND WAREHOUSING

SECTION – A (10 X 1 = 10 MARKS)

ANSWER THE FOLLOWING QUESTIONS. (K1)

1. ___ predicts future trends & behaviors, allowing business managers to make proactive, knowledge driven decisions.

   a) Data ware house b) Data mining c) Data marts d) Meta data

2. _____is the input to KDD.

   a) Data b) Information c) Query d) Process

3. The data is stored, retrieved & updated in ___.

   a) OLAP b) OLTP c) SMTP d) FTP

4. Removing duplicate records is a process called _____

   a) Recovery b) Data cleaning c) Data cleansing d) Data pruning

5. In ___ each cluster is represented by one of the objects of the cluster located near the center.

   a) K-medoid b) k-means c) STIRR d) ROCK

ANSWER THE FOLLOWING IN ONE (OR) TWO SENTENCES (K2)

.6. What is Datamining?

7. Define the term data preprocessing.

8. What is data warehousing?

9. Define Association Ruile.

10. What is cluster?

SECTION – B (5 X 5 = 25 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS. (K3)

11. a) Compare the data ware house and database with suitable example.

(OR)

b)Discuss Data warehouse model.

12. a) What are Data objects and attributes? Explain in detail.

(OR)

b) Discuss the various steps involved in Data preprocessing.

13. a) Sketch the data warehouse architecture and explain it.

(OR)

b) Explain various types of Data warehouse models.

14. a) A data set for analysis includes only one attribute X:
X={7,12,5,18,5,9,13,12,19,7,12,12,13,3,4,5,13,8,7,6}
a) What is the mean of the data set X? b) What is the median?
c) Find the standard deviation for X.

(OR)

b) Describe in detail Decision tree Induction.

15. a) Analyze the partitioning methods available in cluster analysis.

(OR)

b) Explain the Distance Measures in Algorithmic Methods.

SECTION – C (5 X 8 = 40 MARKS)

ANSWER EITHER (a) OR (b) IN EACH OF THE FOLLOWING QUESTIONS.(K4 (Or) K5)

16. a) Elaborate the data mining primitives. Discuss in detail the issues in Discretization and concept hierarchy generation with an example.

(OR)

b) Explain the major application of data mining and the trends followed. Discuss a domain from medical diagnosis and social network analysis.

17. a) Summarize the Data cleaning process in detail. **(OR)**
b) Discuss Segmentation in Data analysis.

18. a) List the differences between OLTP and OLAP and Explain it.**(OR)**
b) Explain the usage of data cube.

19. a) Consider the Data set D. Given the minimum support2, apply aprior algorithm on this dataset.

Transaction ID	Items
100	A,C,D
200	B,C,E
300	A,B,C,E
400	B,E

(OR)

b) Find the frequent Itemsets and strong association rules for the following transactional database table using Aprior algorithm. Consider the threshold as support = 30% and confidence = 40 %.

TID	**ITEM IDs**
1	I1, i2, i3, i5
2	I2, i5, i7, i9
3	I1,i3,i5,i7
4	I2,i4,i6,i8
5	I1,i2,i3,i4
6	I2,i3,i4,i5
7	I3,i4,i5,i6
8	I4,i5,i6,i7
9	I5,i6,i7.i8.i9
10	I9.i1.i2.i5
11	I8,i2,i9,i7
12	I5,i6,i3,i2

20. a) Discuss the application of Data mining in Retail and Telecommunication industry.

(OR)

b) Analyze the applications of Datamining in Intrusion Detection and prevention.

****

Ethical paper