(NO OF PAGES: 2)

FOR THE CANDIDATES ADMITTED DURING THE ACADEMIC YEAR 2019 ONLY)

19UCS6E4

REG.NO

NGM COLLEGE (AUTONOMOUS) POLLACHI

END-OF-SEMESTER EXAMINATIONS: JULY- 2022

B. Sc – COMPUTER SCIENCE (Aided & SF) **MAXIMUM MARKS: 75**

VI SEMESTER **TIME: 3 HOURS**

PART-III

CORE ELECTIVE II: DATA MINING AND WAREHOUSING

SECTION – A **(10 X1 = 10 MARKS)**

ANSWER THE FOLLOWING QUESTIONS
MULTIPLE CHOICE QUESTIONS **(K1)**

1. _____ is a process of automatically discovering useful information in large data repositories
   a) Data Mining b) Data Warehousing
   c) Metadata d) Knowledge discovery database

2. _____ is a complex chemical process in neural networks
   a) Receiving process b) Sending process.
   c) Transmission process d) Switching process

3. Association rules are always defined on_____.
   a) Single attribute b) Binary attribute
   c) Relational database d) Multidimensional attribute

4. OLAP is stands for _____.
   a) Online Application Processing b) Online Analytical Processing
   c) Online Aptitude Processing d) Online Administration and Processing

5. _____ is concerned with discovering the model underlying the link structures of the web.
   a) Data structure mining b) Text mining
   c) Web structure mining d) Image mining

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

6. Define Data cleaning
7. What is decision tree?
8. Write the two measures which are used to find the interestingness of data.
9. What is meant by fact table?
10. What is temporal mining?

(CONTD…. 2)

SECTION – B (5 X 5 = 25 MARKS)

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

11. (a) Explain the various data mining issues
(OR)
(b) Summarize the social implications of data mining

12. (a) Describe statistical perspective on data mining
(OR)
(b) Illustrate the distance based algorithm

13. (a) Write short notes on similarity and distance measures
(OR)
(b) What is meant by parallel & distributed algorithms? Explain

14. (a) Discuss the differences between operational database system and data warehouses
(OR)
(b) Explain the characteristics of OLAP system

15. (a) Describe briefly the spatial data mining techniques
(OR)
(b) What is a time series database? Explain the similarity search in time series analysis

SECTION - C (4 X 10 = 40 MARKS)

ANSWER ANY FOUR OUT OF SIX QUESTIONS.
(16TH QUESTION IS COMPULSORY AND ANSWER ANY THREE QUESTIONS FROM Q.NO: 17 TO 21) (K4) OR (K5)

16. Explain how the association rules mining technique is used to mine the frequent patterns. give example.

17. Discuss the data preprocessing stages for discovery of new knowledge from large databases.

18. Explain the classification by decision tree based algorithm with an example.

19. What are the desirable features of a cluster analysis? Briefly describe the Hierarchical approach of clustering methods.

20. Explain the stars and snowflakes schemas for multi-dimensional databases with example

21. Explain the following.
(i). Web Structured Mining
(ii). Web Content mining