



## Avinashilingam Institute for Home Science and Higher Education for Women

(Deemed to be University under Category 'A' by MHRD, Estd. u/s 3 of UGC Act 1956)

Re-accredited with 'A+' Grade by NAAC. Recognised by UGC Under Section 12B

Coimbatore - 641 043, Tamil Nadu, India

### Master's Degree Examination – June / July 2021

#### II Semester

Class : I PG

Major : Computer Science

Time : 3 Hours

Max. Marks: 100

#### 20MCSC08 Data Mining and Warehousing

##### Part A

10 x 1 = 10

##### Choose the Correct Answer

- The star schema is composed of \_\_\_\_\_ fact table. CO1K1
  - one
  - two
  - three
  - four
- Removing duplicate records is a process called \_\_\_\_\_. CO1K2
  - recovery
  - data cleaning
  - data cleansing
  - data pruning
- All set of items whose support is greater than the user-specified minimum support are called as \_\_\_\_\_. CO2K2
  - border set
  - maximal frequent set
  - frequent set
  - lattice
- The FP-growth algorithm has \_\_\_\_\_ phases CO2K1
  - one.
  - two
  - three.
  - four.
- \_\_\_\_\_ can be thought of as classifying an attribute value into one of a set of possible classes CO3K1
  - Estimation
  - Prediction.
  - Prediction.
  - Clarification.
- Rule based classification algorithms generate \_\_\_\_\_ rule to perform the classification CO31
  - while
  - if-then
  - do while
  - switch
- Pick out a hierarchical clustering algorithm CO2K2
  - DBSCAN
  - BIRCH
  - PAM
  - CURE
- Which of the two are closely related to each other CO4K1
  - association rules & classification
  - classification & prediction
  - association rules & clustering
  - classification & clustering
- Business Intelligence and data warehousing is used for CO4K1
  - Forecasting
  - Data Mining
  - Analysis of large volumes of product sales data
  - All the above
- Data mining application is a \_\_\_\_\_ type of system CO5K1
  - Artificial intelligence
  - Expert
  - Decision support system
  - Statistical system

**Part - B**  
**Answer ALL questions**

**5 x 6 = 30**

**Each answer should not exceed 400 words or two pages**

- |        |   |       |
|--------|---|-------|
| 11. a. | Point out the needs of preprocessing the data   | CO1K3 |
|        | (or)  |       |
| 11. b. | Analyze the three categories of measures of Data Cube                                 | CO1K3 |
| 12. a. | Compute the support and confidence measures of association rule mining with example   | CO2K3 |
|        | (or)  |       |
| 12. b. | Discuss briefly the constrain based association rule mining                           | CO2K2 |
| 13. a. | Summarize the basic principle of Naïve Bayesian Classification                        | CO3K4 |
|        | (or)  |       |
| 13. b. | Differentiate linear regression and non– linear regression                            | CO4K3 |
| 14. a. | Write short notes on categorization of major clustering methods                       | CO4K3 |
|        | (or)  |       |
| 14. b. | Assess the clustering based outlier approaches  | CO4K4 |
| 15. a. | What is a time series database? Explain the similarity search in time series analysis | CO5K2 |
|        | (or)  |       |
| 15. b. | Describe briefly about text mining process  | CO5K2 |

**Part C**  
**Answer ALL questions**

**5 x 12 = 60**

**Each answer should not exceed 800 words or four pages**

- |        |   |       |
|--------|---|-------|
| 16. a. | What is Date Warehouse? Explain the stars, snowflakes schemas for multidimensional data base with example                           | CO1K3 |
|        | (or)  |       |
| 16. b. | Discuss briefly the three tier architecture of data warehouse   | CO1K1 |
| 17. a. | Generate the Apriori algorithm to find frequent items sets using candidate generation   | CO2K5 |
|        | (or)  |       |
| 17. b. | Construct the FP Growth algorithm to find frequent items with example   | CO2K4 |
| 18. a. | Describe the basic algorithm for inducing Decision Tree from training samples   | CO3K3 |
|        | (or)  |       |
| 18. b. | What is Back-propagation? Explain briefly the Support Vector Machines   | CO3K2 |
| 19. a. | What is Clustering analysis? Describe the important approaches of Partitioning clustering methods                                   | CO4K2 |
|        | (or)  |       |
| 19. b. | Assess the hierarchical clustering methods to group the similar object in data mining   | CO4K4 |
| 20. a. | Point out the potential applications and major issues in Data mining  | CO5K3 |
|        | (or)  |       |
| 20. b. | How WEKA Tool is used to construct the experimental modules for data preprocessing association and classification methods? Explain. | CO5K4 |

