

**Avinashilingam Institute for Home Science and Higher Education for Women
(Deemed to be University)Coimbatore-641043.**

**Master's Degree Examination – November 2018
III Semester**

**Class : II PG
Major : Computer Application**

**Time: 3 hours
Max. Marks: 60**

17MCAC16 Data Mining and Warehousing

Part A

10 x 1/2 = 5

Choose the correct answer

1. _____ mining tasks characterize the general properties of the data in the database.
a. Descriptive b. Predictive c. Web usage d. Multimedia
2. Data _____ is a form of data reduction that is very useful for the automatic generation of concept hierarchies from numerical data.
a. Redundant b. Abstraction c. Discretization d. Extraction
3. _____ is a method of mining frequent item sets without candidate generation.
a. Apriori b. ECLAT c. FP-growth d. TID
4. A measure is _____ if its value is free from the influence of null-transactions
a. Convertible b. Monotonic c. Null invariant d. Inconcertible
5. A _____ can be used to identify whether any two given attributes are statistically related.
a. Data Cleaning b. Relevance Analysis
c. Correlation Analysis d. Data Transformation
6. The ability to construct the classifier or predictor efficiently given large amounts of data are called _____
a. Speed b. Robustness c. Scalability d. Interpretability
7. In machine learning, clustering is an example of _____
a. Supervised Learning b. Unsupervised Learning c. Observation d. Data Segmentation
8. A cluster that satisfy user-specified preferences are called _____
a. Model Based Cluster b. Density Based c. Constraint Based d. Hierarchical Based
9. A _____ database consists of sequences of values or events obtained over repeated measurements of time
a. Sequence b. Sequence data c. Time-series d. Time-Parallel
10. It is the percentage of retrieval documents that are in fact relevant to the query is called _____
a. Recall b. Precision c. Relevant d. Data Analysis

Part B

5 x 4 = 20

Answer ALL questions

Each answer should not exceed 200 words or one page

11.a. Explain the missing values and noisy data in data cleaning.

Or

11.b. Expound the term Data Warehousing.

12.a. Describe mining multilevel Association rule.

Or

12.b. Write a note on Metarule-guided mining of Association rules.

13.a. Compare the classification method and Prediction method

Or

13.b. Discuss Linear Regression with example

14.a. Write about the typical requirements of clustering in data mining

Or

14.b. What is Distance-Based outlier defection? Explain

15.a. Illustrate the Hoeffding tree algorithm with example

Or

15.b. Mention some of the clustering evolving data stream

Part C

5 x 7 = 35

Answer ALL questions

Each answer should not exceed 600 words or three pages

16.a. Describe Data Integration and Transformation with example

Or

16.b. Differentiate between operational database systems and data warehouse

17.a. Explain in brief about Frequent pattern mining

Or

17.b. Illustrate the Apriori Algorithm with example

18.a. Explain the following: i) Using IF-THEN rules for classification

ii) Rule Extraction from a decision tree

Or

18.b. Explain the two techniques of ensemble methods

19.a. How Agglomerative and Divisive hierarchical method works in clustering?

Or

19.b. Explain about Expectation-Maximization in model based clustering

20.a. Write in detail about Trend Analysis.

Or

20.b. Explain the basic measures Precision and Recall for Text retrieval with example.