USN | | | | | | | | | | |

202MC009

# Ph.D./M.S. (Engg.) by Research Degree Examination, Nov./Dec.2024
# Data Mining and Data Warehousing

Time: 3 hrs. Max. Marks: 100

***Note: Answer any FIVE full questions, choosing ONE full question from each module.***

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

VTU Library - Nov/Dec.2024

## Module-1

**1** a. What is data mining? What is the need for data mining? **(04 Marks)**
b. What kinds of data can be mined? **(06 Marks)**
c. What kinds of patterns can be mined? **(10 Marks)**

**OR**

**2** a. Which are the different technologies used in data mining? **(08 Marks)**
b. What is business intelligence? **(02 Marks)**
c. Explain the major issues in data mining. **(10 Marks)**

## Module-2

**3** a. Explain the four characteristics of data ware house which distinguish them from other data repository systems. **(04 Marks)**
b. Explain the multitiered architecture of data warehouse with a neat diagram. **(10 Marks)**
c. What is a datacube? Explain a star scheme. **(06 Marks)**

**OR**

**4** a. What is OLAP? What is a concept hierarchy? **(04 Marks)**
b. Explain the typical operations of OLAP. **(10 Marks)**
c. Discuss the process of data warehouse design. **(06 Marks)**

## Module-3

**5** a. What is data classification? Differentiate between supervised and unsupervised learning. **(04 Marks)**
b. Write the basic algorithm for inducing a decision tree from training tuples. **(10 Marks)**
c. What is information gain? Discuss information gain. **(06 Marks)**

**OR**

**6** a. Discuss the Bayes theorem. **(06 Marks)**
b. Explain the metrics for evaluating classifier performance. **(14 Marks)**

## Module-4

**7** a. What is clustering? Discuss the requirements of clustering in data mining. **(12 Marks)**
b. Which are the different measures used to measure the distance between two cluster? **(08 Marks)**

**OR**

**8** a. Discuss briefly the basic clustering methods of partitioning method, hierarchical methods, density based methods and grid based methods. **(12 Marks)**
b. Explain the three major tasks of clustering evaluation. **(08 Marks)**

## Module-5

| | | | |
|---|---|---|---|
| **9** | a. | Discuss mining of time series data. | **(08 Marks)** |
| | b. | Discuss sequential pattern mining. | **(08 Marks)** |
| | c. | Discuss any two methods or techniques and in statistical data mining. | **(04 Marks)** |

**OR**

| | | | |
|---|---|---|---|
| **10** | a. | Discuss any two applications of data mining with suitable examples. | **(10 Marks)** |
| | b. | Discuss the classification of privacy preserving data mining methods. | **(10 Marks)** |

* * * * *

VTU Library - Nov/Dec.2024