



10IS74

Seventh Semester B.E. Degree Examination, Jan./Feb. 2023

## Data Warehousing and Data Mining

Time: 3 hrs.

Max. Marks:100

*Note: Answer any FIVE full questions, selecting at least TWO questions from each part.*

### PART - A

- 1 a. Define data mining. Why is it called knowledge discovery database? Explain five steps involved in data mining. (10 Marks)
- b. Explain the structure of data warehouse with Operational Data Source (ODS). (10 Marks)
- 2 a. Explain the characteristics of OLAP (Online Analytical Processing). (10 Marks)
- b. Explain data cube operations with suitable example. (10 Marks)
- 3 a. What is data preprocessing? Explain the following pre-processing techniques, aggregation, sampling. (10 Marks)
- b. Explain four core data mining analysis. (10 Marks)
- 4 a. Write and explain the Apriori algorithm for frequent item set generation. (10 Marks)
- b. Define maximal frequent and close frequent item sets. Show how closest frequent item set is advantages. (10 Marks)

### PART - B

- 5 a. Explain what node impurity is? How to identify nodes impurity by Gini Method. (10 Marks)
- b. Explain Hunt's algorithm with an example. (10 Marks)
- 6 a. Explain Bayesial classifiers. Show how Naïve bayes classifiers are used to calculate probability to identify class among attributes which are independent. (10 Marks)
- b. Explain DBSCAN algorithm with an example. (10 Marks)
- 7 a. What is cluster analysis? Illustrate the features of cluster analysis. (10 Marks)
- b. Explain hierarchical method of cluster analysis with an example. (10 Marks)
- 8 Write short notes on:
  - a. Spatial data mining
  - b. Text data mining
  - c. World Wide Data Mining
  - d. Social Impacts of data mining (20 Marks)

\* \* \* \* \*

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.