**Code No: 126VW** **R15**

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD**
**B. Tech III Year II Semester Examinations, December - 2018**
**DATA WAREHOUSING AND DATA MINING**
**(Information Technology)**

**Time: 3 hours** **Max. Marks: 75**

**Note:** This question paper contains two parts A and B.
Part A is compulsory which carries 25 marks. Answer all questions in Part A. Part B consists of 5 Units. Answer any one full question from each unit. Each question carries 10 marks and may have a, b, c as sub questions.

## PART - A

**(25 Marks)**

| | | |
|---|---|---|
| 1.a) | What is a data warehouse? How is it differs from DBMS? | [2] |
| b) | How to index OLAP data? | [3] |
| c) | Define dimensionality reduction. | [2] |
| d) | Explain attribute subset selection. | [3] |
| e) | What are frequent patterns? Give an example. | [2] |
| f) | How are meta rules useful in data mining? | [3] |
| g) | What is associative classification? | [2] |
| h) | How is prediction different from classification? | [3] |
| i) | What is categorical variable? | [2] |
| j) | What are interval-scaled variables? | [3] |

## PART - B

**(50 Marks)**

2.a) Give examples for defining star, snowflake and fact constellation schemas.
b) Discuss about a three-tier data warehouse architecture. [5+5]

**OR**

3.a) What are the various types of OLAP servers? Explain.
b) Describe the measures of multidimensional data model. [5+5]

4. Discuss about discretization and concept hierarchy generation for numerical data. [10]

**OR**

5.a) Describe the process of data cleaning.
b) Write a brief note on relational databases. [5+5]

6. Can we design a method that mines the complete set of frequent item sets without candidate generation? Explain with example. [10]

**OR**

7. Explain in detail about multilevel association rules. [10]

8. Why is naive Bayesian classification called "naive"? Briefly outline the major ideas of naive Bayesian classification. Explain Naive-Bayes classification. [10]

**OR**

9.a) What are the methods for expressing attribute test conditions? Explain.
b) Write an algorithm for *k-nearest-neighbor classification* given *k* and *n*, the number of attributes describing each tuple. [5+5]

10.a) What are typical requirements of clustering in data mining? Explain.
b) How does the PAM algorithm work? Explain. [5+5]

**OR**

11.a) What are key issues in hierarchical clustering? Explain.
b) Explain about the basic Agglomerative Hierarchical clustering algorithm. [5+5]

**–--ooOoo----**