Hall Ticket Number:

Code No.: 32512 AS

VASAVI COLLEGE OF ENGINEERING *(Autonomous)*, HYDERABAD
B.E. (I.T.) III Year II-Semester Advanced Supplementary Examinations, June/July-2017

Data Warehousing and Data Mining

Time: 3 hours

Max. Marks: 70

Note: Answer ALL questions in Part-A and any FIVE from Part-B

Part-A (10 × 2 = 20 Marks)

1. Define Data Characterization.
2. What is the need of Data Reduction?
3. Define Data Warehouse.
4. Describe structured pattern mining with example.
5. How rule based classifier works for classification?
6. Why is tree pruning useful in decision tree induction?
7. What are Interval-scaled variables?
8. List the applications of Outlier detection and analysis.
9. Define Multimedia Data Mining.
10. What a time-series database consists of?

Part-B (5 × 10 = 50 Marks)

11. a) Illustrate various steps in data mining for knowledge discovery. [6]
    b) How the Missing values are handled in Data cleaning? Explain. [4]

12. a) With a suitable example explain Apriori algorithm for mining frequent item sets for Boolean association rules. [6]
    b) Explain FP-Growth algorithm with suitable example. [4]

13. a) Describe the preprocessing steps that applied to the data to improve the accuracy, efficiency, and scalability of the classification or prediction process. [5]
    b) Explain how to evaluate the accuracy of classifier and predictor. [5]

14. a) What is Cluster analysis? Explain different types of data that often occur in cluster analysis. [4]
    b) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8) [6]
       i) Compute the Euclidean distance between the two objects.
       ii) Compute the Manhattan distance between the two objects.
       iii) Compute the Minkowski distance between the two objects, using $q = 3$.

15. a) What are the three types of dimensions in a spatial data cube? Explain. [4]
    b) Explain Hidden Markov Model for Biological Sequence Analysis. [6]

16. a) What are the major issues in data mining regarding mining methodologies? [5]
    b) List and explain the three kinds of Data Warehouse applications. [5]

17. Answer any *two* of the following:
    a) What is the purpose of "Attribute selection measures" in classification by decision tree induction? How we can use the "Tree pruning" in classification? [5]
    b) Explain agglomerative Hierarchical Clustering. [5]
    c) Describe Mining the Web Page Layout Structure in detail. [5]

ಜ಼ಜ಼ಜ಼ಜ಼ಜ಼ಜ಼ಜ಼