VASAVI COLLEGE OF ENGINEERING *(Autonomous)*, HYDERABAD
M.C.A. II Year II-Sem. Advanced Supplementary Examinations, June/July-2017

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. List the different types of Data Mining functionalities.
2. Mention the most commonly used Data Reduction techniques.
3. List the key features of a Data Warehouse.
4. What is the role of Meta data repository in a Data Warehouse?
5. Define Support and Confidence.
6. Differentiate between Closed Frequent Itemset and Maximal frequent Itemset.
7. Differentiate Eager Learners from Lazy Learners.
8. Write the two step process of Data Classification.
9. Illustrate the Strengths and Weakness of k-Means in comparison with k-Medoids.
10. What is outlier detection?

Part-B *(5 × 10 = 50 Marks)*
(All bits carry equal marks)

11. a) Describe briefly about the different types of data that can be mined.

    b) Discuss the different measures used for assessing the spread or dispersion of Numeric data.

12. a) Differentiate between Operational Database Systems and Data Warehouses.

    b) Briefly compare Snowflake Schema, Fact Constellation and Starnet Query Model.

13. a) Describe Apriori Frequent Pattern Mining Algorithm with an example.

    b) Discuss the several variations proposed to increase the efficiency and scalability of the Apriori algorithm.

14. a) Why is tree pruning useful in decision tree induction? Discuss the common approaches to tree pruning.

    b) Discuss about Classification by Backpropagation.

15. a) Explain in detail 'DBSCAN' Clustering Method.

    b) Discuss briefly about the Basic Clustering methods.

16. a) Discuss in detail the different strategies used for Data Transformation.

    b) Compare MOLAP, HOLAP and ROLAP servers.

17. Answer any *two* of the following:
    a) Explain in detail the procedure of Mining Multidimensional Association rules.
    b) Illustrate how k-nearest Neighbor Classifier classifies the samples.
    c) How is Similarity Search performed on Time Series Data? Explain.

ᘓᘓᘓᘓᘓ