VASAVI COLLEGE OF ENGINEERING *(Autonomous)*, HYDERABAD
M.C.A. (CBCS) IV-Semester Main Examinations, May/June-2018

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. Describe the binning process with an example.

2. Differentiate between similarity and dissimilarity.

3. Point out the difference between snowflake schema and fact constellation schema.

4. Name and explain any two cube operations.

5. Define confidence measure. Explain with formula and example.

6. Write Apriori property.

7. Which attribute selection measure is best? Explain with reason.

8. How does classification work?

9. Minimum intra cluster distance is the stopping condition of iteration in clustering algorithm. True or false? Justify.

10. Differentiate between k-means and k-medoid algorithm.

Part-B (5 × 10 = 50 Marks)

11. a) Explain any three type of data discretization techniques with proper example.   [6]

    b) Differentiate between OLAP system and OLTP system.   [4]

12. a) Construct a data warehouse of a retail market sales using Snowflake schema. Table sales   [5]
       (time, TID, CID, list of Items, Total price, payment mode) and table customer (CID,
       age, gender, Mobile, city)

    b) Write about starnet query model for querying multidimensional data basis.   [5]

13. a) Explain the correlation measure LIFT for mining large data set.   [3]

    b) Explain the association rule mining problem. Consider an example with the following   [7]
       set of transactions. There are 10 items in the shop. Find all the frequent item sets with
       30% support using Fp-Growth algorithm.

TID	Items bought	TID	Items bought
001	B,M,T,Y	006	T,Y,E,M
002	B,M	007	A,B,M
003	A,T,S,P	008	B,C,D,T,P
004	A,B,C,D	009	D,T,S
005	A,B	010	A,B,M

Contd... 2

14. a) Explain naïve Bayesian classification technique. [5]

b) Build a decision tree classification model using gain ratio attribute selection measure to classify bank loan application by assigning applications to one of three risk classes. [5]

Owns home	Married	Gender	Employed	Credit rating	Risk class (Class Label)
Yes	Yes	Male	Yes	A	B
No	No	Female	Yes	A	A
Yes	Yes	Female	Yes	B	C
Yes	No	Male	No	B	B
No	Yes	Female	Yes	B	C
No	No	Female	Yes	B	A
No	No	Male	No	B	B
Yes	No	Female	Yes	A	A
No	Yes	Female	Yes	A	C
Yes	Yes	Female	Yes	A	C

15. a) Explain K-means partitioning method for clustering. [5]

b) Differentiate between STING and CLIQUE of Grid based methods [5]

16. a) Write about the basic methods for data cleaning. [5]

b) Write the importance of Extraction and Transformation in the process of data warehouse creation. [5]

17. Answer any *two* of the following:

a) Procedure of association rule generation from frequent item set. [5]

b) Back propagation algorithm. [5]

c) Name and write about the techniques of density based clustering. [5]

ങ്കരുങ്കരുങ്കരുങ്കരുങ്ക