



QP CODE: 22002582

Reg No :

Name :

MSc DEGREE (CSS) EXAMINATION , NOVEMBER 2022

Second Semester

M.Sc. ARTIFICIAL INTELLIGENCE

CORE - AI010203 - DATA MINING TECHNIQUES

2019 Admission Onwards

EFF63C6F

Time: 3 Hours

Weightage: 30

Part A (Short Answer Questions)

*Answer any **eight** questions.*

Weight 1 each.

1. Explain a scheme for consolidating data in multidimensional space.
2. Differentiate between Operational Database Systems and Data Warehouse.
3. What is the motivation behind Data Mining ?
4. Explain log-linear models.
5. Which property is used to reduce the search space in frequent itemset generation?
6. What is Item merging?
7. What are prediction problems?
8. Define rule-based classification.
9. What is the criteria for comparing clustering methods?
10. Explain the k-means method for partitioning.

(8×1=8 weightage)





Part B (Short Essay/Problems)

*Answer any **six** questions.*

Weight 2 each.

11. What are Snowflakes?
12. Briefly explain about syntax for specifying the kind of knowledge to be mined.
13. Explain min-max normalization. Suppose that the minimum and maximum values for the attribute income are \$12,000 and \$98,000, respectively. We would like to map income to the range [0.0, 1.0]. Normalize it using min-max normalization.
14. Explain in detail about support and confidence measures with an example.
15. Explain sampling method for improving apriori efficiency with an example.
16. How bagging improves the accuracy of classification?
17. Discuss the measures used to calculate distance between clusters and various algorithms where they are employed.
18. Discuss any one grid-based multiresolution clustering technique.

(6×2=12 weightage)

Part C (Essay Type Questions)

*Answer any **two** questions.*

Weight 5 each.

19. Explain OLAP operations with the help of diagrams.
20. Explain data integration with proper examples.
21. Explain apriori algorithm for the following example. Support= 60 %, Confidence = 80 % Transaction ID Items 1 {A, C, D} 2 {B, C, E} 3 {A, B, C, E} 4 {B, E} 5 {A, B, C, E}





22. Use Bayesian classification to predict class label using the training data given below. The tuple to be classified is $X=(\text{Age}=\text{Youth}, \text{Income}=\text{Medium}, \text{Student}=\text{Yes}, \text{Credit-rating}=\text{fair})$

Class-Labeled Training Tuples from the AllElectronics Customer Database

ID	Age	Income	Student	Credit-rating	Class: buys _computer
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle_aged	High	No	Fair	yes
4	Senior	Medium	No	Fair	yes
5	Senior	low	yes	Fair	yes
6	Senior	low	yes	Excellent	no
7	Middle_aged	low	yes	Excellent	yes
8	Youth	Medium	No	Fair	no
9	Youth	low	yes	Fair	yes
10	Senior	Medium	yes	Fair	yes
11	Youth	Medium	yes	Excellent	yes
12	Middle_aged	Medium	No	Excellent	yes
13	Middle_aged	high	yes	Fair	yes
14	Senior	Medium	No	Excellent	no

(2×5=10 weightage)

