



QP CODE: 25802549



Reg No :

Name :

INTEGRATED MSC DEGREE EXAMINATION, MAY 2025

Fourth Semester

CORE - ICSC4CR3 - DATA MINING

INTEGRATED MSC COMPUTER SCIENCE-ARTIFICIAL INTELLIGENCE AND MACHINE
LEARNING, INTEGRATED MSC COMPUTER SCIENCE-DATA SCIENCE

2020 Admission Onwards

44C760F5

Time: 3 Hours

Weightage: 30

Part A (Short Answer Questions)

*Answer any **eight** questions.*

Weight 1 each.

1. What is Over-fitting?
2. Explain Quantitative attribute.
3. List out the factors comprising of data quality.
4. What do you mean by entity identification problem?
5. Suppose the data contain the frequent itemset $I = \{I_1, I_2, I_5\}$. What are the association rules that can be generated from I ?
6. What are the limitations of KNN?
7. What is Holdout method for evaluating classifier?
8. What do you mean by partitioning methods of clustering?
9. What is AGNES in clustering?
10. What is CLIQUE?

(8×1=8 weightage)

Part B (Short Essay/Problems)

*Answer any **six** questions.*

Weight 2 each.

11. Differentiate between qualitative and quantitative data . Which measure of central tendency can be used for quantitative data?
12. Explain the steps in PCA.





13. i)What are the advantages of z-score normalization?
ii)Consider the marks data: 8, 10, 15, 20. Normalize this data set using z-score normalization
14. Explain the terms frequency,support count and confidence with suitable example.
15. Explain ROC Curves in detail.
16. What is meant by overfitting and underfitting of classifier? How to avoid overfitting?
17. Briefly explain the characteristics of different categories of clustering methods.
18. What are outliers? Explain various types of outliers.

(6×2=12 weightage)

Part C (Essay Type Questions)

Answer any **two** questions.

Weight 5 each.

19. With neat block diagram, explain in detail the knowledge discovery process.
20. Consider the below the transaction database and generate the frequent itemsets using FP-growth algorithm where the minimum support count is 2.

TID	Items
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

21. Explain classification by decision tree induction in detail.
22. Explain DBSCAN in detail.Describe its advantages and limitations.

(2×5=10 weightage)

