

QP CODE: 24803211



Reg No : .....

Name : .....

**M.C.A DEGREE EXAMINATION, JUNE 2024**  
**Second Semester**  
MASTER OF COMPUTER APPLICATION  
**CORE - MCACT204 - DATA SCIENCE & BIG DATA ANALYSIS**

2020 Admission Onwards

35330749

Time: 3 Hours

Maximum: 75 Marks

**Part A**

*Answer any **ten** questions*

*Each question carries **3** marks*

1. What are the major components in datamining architecture?
2. Distinguish between the relational databases and transactional databases.
3. Explain how we handle noisy Data.
4. Discuss the disadvantages of Apriori Algorithm.
5. How can you estimate the accuracy of the classifier built?
6. Why did you say that clustering is an unsupervised learning process?
7. Explain with diagram credit risk management.
8. How is Big Data useful in healthcare?
9. Explain the Types of Big data.
10. What is Hadoop?
11. What is YARN?
12. How Hive can be accessed?

(10×3=30 marks)

**Part B**

*Answer **all** questions*

*Each question carries **9** marks*

13. a) Explain the different kinds of Databases used in DataMining.





OR

b) Illustrate the discretization and concept hierarchy generation methods.

14. a) Explain Apriori Algorithm in detail.

OR

b) Briefly explain the various Clustering methods.

15. a) Which are the facets of data science?

OR

b) Explain Big Data in healthcare.

16. a) Briefly explain the technologies for handling Big Data.

a. Distributed and Parallel computing

b. In-Memory Computing for Big data.(IMC )

OR

b) What is Hadoop? Explain the relevance of HBase in Big data.

17. a) Write notes on a) Polyglot persistence b) CAP theorem c)Advantages of YARN

OR

b) What are the data types in Hive ? Explain Hive services

(5×9=45 marks)

