



43456

Library
Brainware University
398, Ramkrishnapur Road, Barasat
Kolkata, West Bengal-700125



BRAINWARE UNIVERSITY

Term End Examination 2025-2026

Programme – MCA-2024

Course Name – Big Data Analytics

Course Code - MCA40115B

(Semester IV)

Full Marks : 60

Time : 2:30 Hours

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group-A

(Multiple Choice Type Question)

1 x 15=15

1. Choose the correct alternative from the following :

(i) What is Big Data?

- | | |
|-------------------------|-------------------------------|
| a) Small datasets | b) Large and complex datasets |
| c) Only relational data | d) Backup data |

(ii) Which of the following is NOT a core characteristic of Big Data?

- | | |
|------------|-------------|
| a) Volume | b) Velocity |
| c) Variety | d) Accuracy |

(iii) Which industry heavily uses Big Data for fraud detection?

- | | |
|--------------|----------------|
| a) Banking | b) Agriculture |
| c) Education | d) Tourism |

(iv) Why are traditional databases unsuitable for Big Data?

- | | |
|---------------------|------------------------|
| a) Lack of security | b) Limited scalability |
| c) Too expensive | d) Too simple |

(v) What does data variety refer to?

- | | |
|---------------------------|------------------|
| a) Data size | b) Data speed |
| c) Different data formats | d) Data accuracy |

(vi) Which Big Data use case supports algorithmic trading?

- | | |
|------------------------------|---------------------|
| a) Real-time market analysis | b) Static reporting |
| c) Manual auditing | d) File archiving |

(vii) Which factor mainly affects K-Means convergence?

- | | |
|----------------------|--------------------|
| a) File format | b) Storage size |
| c) Initial centroids | d) Data encryption |

(viii) In which type of situation is hierarchical clustering most appropriate?

- | | |
|------------------------|-------------------------------|
| a) Real-time streams | b) Hierarchical relationships |
| c) Very large datasets | d) Labeled data |

- (ix) Overfitting in Decision Trees occurs when:
- a) Tree is too deep
 - b) Dataset is small
 - c) Pruning is applied
 - d) Entropy is low
- (x) Pruning in Decision Trees is used to:
- a) Increase depth
 - b) Reduce overfitting
 - c) Increase bias
 - d) Add noise
- (xi) Which factor should be analyzed before choosing K in K-Means?
- a) File system
 - b) Data distribution
 - c) Programming language
 - d) Storage device
- (xii) Naïve Bayes works best when features are:
- a) Highly correlated
 - b) Random
 - c) Conditionally independent
 - d) Sparse
- (xiii) Graph-based models are preferred in fraud detection because they:
- a) Reduce storage
 - b) Identify hidden relationships
 - c) Encrypt data
 - d) Simplify schema
- (xiv) Recommendation engines add business value by:
- a) Reducing users
 - b) Increasing engagement
 - c) Removing data
 - d) Manual analysis
- (xv) Which real-time platform is used for Twitter data analysis?
- a) Hadoop
 - b) Spark
 - c) Hive
 - d) Pig

Library
Brainware University
398, Ramkrishnapur Road, Barasat
Kolkata, West Bengal-700125

Group-B

(Short Answer Type Questions)

3 x 5=15

2. Define Big Data and list the core 3Vs. (3)
3. Explain the role of Big Data analytics in healthcare. (3)
4. Apply the concept of variety to social media data. (3)
5. Analyze how graph analytics improves Big Data decisions. (3)
6. Evaluate suitability of Apriori algorithm for large datasets. (3)

OR

- Assess reliability of sentiment analysis models. (3)

Group-C

(Long Answer Type Questions)

5 x 6=30

7. Apply Big Data concepts to explain how web analytics improves user experience. (5)
8. Describe the role of entropy in decision tree construction. (5)
9. Examine the effect of initial centroid selection on clustering results. (5)
10. Analyze limitations of relational models for unstructured data. (5)
11. Examine data modeling challenges in document-oriented databases. (5)
12. Justify the choice of graph-based models for recommendation engines. (5)

OR

- Evaluate predictive accuracy limitations in stock market analytics. (5)
