

Gandhi Institute of Engineering and Technology University, Odisha, Gunupur (GIET University)

M.C.A. (Second Semester - Regular) Examinations, May - 2026

MCA 252004 – Introduction to Data Science



Time: 3 hrs

Maximum: 60 Marks

(The figures in the right hand margin indicate marks)

PART – A

(2 x 5 = 10 Marks)

Q.1. Answer **ALL** questions

- | | CO # | Blooms
Level |
|--------------------------------------|------|-----------------|
| a. Recall Null Hypothesis (H_0). | CO1 | K1 |
| b. Define Eigenvalue and Eigenvector | CO2 | K2 |
| c. Define Multicollinearity. | CO3 | K1 |
| d. Explain Naive Bayes. | CO4 | K2 |
| e. Define K-Means clustering. | CO5 | K2 |

PART – B

(10 x 5 = 50 Marks)

Answer **ALL** questions

- | | Marks | CO # | Blooms
Level | | | | | | | | | | | | |
|---|-------|------|-----------------|---|---|---|---|---|---|---|---|---|--|--|--|
| 2. a. Explain Descriptive Analytics with examples. | 5 | CO1 | K1 | | | | | | | | | | | | |
| b. Explain Probability Theory and important distributions in detail? | 5 | CO1 | K2 | | | | | | | | | | | | |
| (OR) | | | | | | | | | | | | | | | |
| c. Find the Median of the dataset
Data: 7, 3, 15, 10, 9, 16, 19, 12 | 5 | CO1 | K2 | | | | | | | | | | | | |
| d. Differentiate between Data Mining and Data Warehousing. | 5 | CO1 | K3 | | | | | | | | | | | | |
| 3.a. Explain Dimensionality Reduction techniques in detail? | 5 | CO2 | K2 | | | | | | | | | | | | |
| b. Calculate Pearson Correlation | 5 | CO2 | K3 | | | | | | | | | | | | |
| <table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="padding: 2px 5px;">X</td> <td style="padding: 2px 5px;">1</td> <td style="padding: 2px 5px;">2</td> <td style="padding: 2px 5px;">3</td> <td style="padding: 2px 5px;">4</td> <td style="padding: 2px 5px;">5</td> </tr> <tr> <td style="padding: 2px 5px;">Y</td> <td style="padding: 2px 5px;">2</td> <td style="padding: 2px 5px;">4</td> <td style="padding: 2px 5px;">5</td> <td style="padding: 2px 5px;">4</td> <td style="padding: 2px 5px;">5</td> </tr> </table> | X | 1 | 2 | 3 | 4 | 5 | Y | 2 | 4 | 5 | 4 | 5 | | | |
| X | 1 | 2 | 3 | 4 | 5 | | | | | | | | | | |
| Y | 2 | 4 | 5 | 4 | 5 | | | | | | | | | | |
| (OR) | | | | | | | | | | | | | | | |
| c. Describe PCA with steps and mathematical explanation. | 5 | CO2 | K1 | | | | | | | | | | | | |
| d. Explain Feature Selection and its methods. | 5 | CO2 | K3 | | | | | | | | | | | | |
| 4.a. Describe Residual Analysis with examples. | 5 | CO3 | K2 | | | | | | | | | | | | |
| b. Explain Multicollinearity and its effects. | 5 | CO3 | K2 | | | | | | | | | | | | |
| (OR) | | | | | | | | | | | | | | | |
| c. Explain residual analysis and detection of influencers with methods. | 10 | 10 | CO3 | | | | | | | | | | | | |
| 5.a. Explain Logistic Regression with example. | 5 | 5 | CO4 | | | | | | | | | | | | |
| b. Explain Decision Tree and splitting criteria. | 5 | 5 | CO4 | | | | | | | | | | | | |
| (OR) | | | | | | | | | | | | | | | |
| c. Explain Optimization techniques in Data Science with examples. | 10 | 10 | CO5 | | | | | | | | | | | | |
| 6.a. Explain Naive Bayes classifier and assumptions? | 5 | 5 | CO5 | | | | | | | | | | | | |
| b. Explain Random Forest algorithm. | 5 | 5 | CO5 | | | | | | | | | | | | |
| (OR) | | | | | | | | | | | | | | | |
| c. Explain DBSCAN algorithm with examples. | 5 | 5 | CO4 | | | | | | | | | | | | |
| d. Explain K-Means algorithm with steps.
Perform 1 iteration of K-Means
Data: (2,2), (4,4), (6,6), (8,8)
Initial centroids: (2,2) and (8,8) | 5 | 5 | CO4 | | | | | | | | | | | | |

--- End of Paper ---