



K. K. Wagh Institute of Engineering Education & Research, Nashik
(An Autonomous Institute From A.Y. 2022-23)

WINTER-2025	
Exam Seat No.:	
Academic Year:2025-2026	Semester:V
Class:TY	Program:B.Tech
Branch Code:CSD	Pattern:2023
Name of Course:Data Science and Big Data	Course Code:2313306B
Max. Marks:60	Duration:2.30 Hrs.

Instructions: Candidates should read carefully the instructions printed on the Question Paper and on the cover page of the Answer Book, which is provided for their use.

1. This question paper contains 02 page(s).
2. Answer to each new question is to be started on a new page.
3. Assume suitable data wherever required, but justify it.
4. Draw the neat labelled diagrams, wherever necessary.
5. The last columns indicates the Course Outcome and level of Blooms Taxonomy of the Question/sub-question.

Marks CO

Question No. 1

- 1a) What is dimensionality reduction. Explain any two dimensionality reduction techniques. (6) CO1

Question No. 2

- 2a) What is bias and variance in regression and what is bias-variance trade off (6) CO2

Question No. 3

- 3a) Explain the algorithm of decision tree classifier used in machine learning (8) CO3

OR

- 3b) Explain bagging and boosting ensemble learning techniques (8) CO3

- 3c) What is confusion matrix and explain different performance metrics of classifier (8) CO3

OR

- 3d) What is meant by multi-class classification. Explain one Vs one and one Vs all. (8) CO3

Question No. 4

- 4a) Given the data points and Euclidean distance between the datapoint perform the agglomerative clustering using average linkage (8) CO4

A	0	2	6	10
B	2	0	5	9
C	6	5	0	4
D	10	9	4	0

OR

- 4b) Explain K Medoid classification algorithm. (8) CO4

- 4c) Explain how clustering performance can be evaluated in extrinsic method. (8) CO4

OR

- 4d) Explain the concept of cluster analysis and differentiate it from classification with suitable examples. (8) CO4

Question No. 5

- 5a) Mother's Care Hospital is facing complaints from patients about long waiting times in the emergency department (ED). The hospital's management wants to reduce wait times and improve patient satisfaction. They approach your data analytics team to investigate and provide actionable insights. How could it be done? (8) CO5

OR

- 5b) What is data explosion. Explain main sources and characteristics of Big Data (8) CO5
- 5c) Which stage in data analytics takes much time. Give reasoning for it (8) CO5

OR

- 5d) What is model planning phase in data analytics life. Explain with suitable example. (8) CO5

..... End of question paper.....