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:ADS	Pattern:2023
Name of Course:Data Science and Big Data	Course Code:2311303
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)	Define dimensionality reduction? Why is it important?	(6)	CO1
Question No. 2			
2a)	List out different performance metrics for evaluating the regression. Write down the formula for calculating it.	(6)	CO2
Question No. 3			
3a)	Given the following training instances ,each having two attributes (x1 and x2). Compute the class label for test instance l5=(3,7) using three nearest neighbors (k=3).	(8)	CO3

Training Instance	X1	X2	Output
I1	7	7	0
I2	7	4	0
I3	3	4	1
I4	1	4	1

OR

3b)	Demonstrate entropy and information gain in building a decision tree?	(8)	CO3
3c)	Explain One-vs-One and One-vs-All strategies used in multiclass classification. Discuss their advantages and disadvantages.	(8)	CO3

OR

3d) Consider the confusion matrix given below for a binary classifier predicting the presence of a disease. The classifier made a total of 150 predictions. Out of those 150 cases,the classifier predicted yes 100 times and no 50 times. Calculate precision, recall,F1-score and accuracy. (8) CO3

	Predicted Yes	Predicted No
Actual No	45	5
Actual Yes	5	95

Question No. 4

4a) Compare Agglomerative and Divisive Hierarchical Clustering methods. (8) CO4

OR

4b) Compare K-Means and K-Medoids in terms of working, robustness, and applications. (8) CO4

4c) Define the term intra-cluster similarity and inter-cluster dissimilarity (8) CO4

OR

4d) Define Multi-view Clustering and state one application. How is it used in clustering? (8) CO4

Question No. 5

5a) List and briefly explain the six phases of Data Analytics Lifecycle. (8) CO5

OR

5b) Define Big Data and explain its characteristics in detail. (8) CO5

5c) Explain Model Planning and Model Building phases of the Data Analytics Lifecycle (8) CO5

OR

5d) What is Data Explosion? Give examples. Explain the 5 V's of Big Data. (8) CO5

····· End of question paper·····