


Name:			
Enrolment No:			
<b>UPES</b> <b>End Semester Examination, December 2024</b>			
<b>Course: Foundation of Data Science</b> <b>Program: MCA</b> <b>Course Code: CSDS7002_4</b>		<b>Semester : I Semester</b> <b>Time : 03 hrs.</b> <b>Max. Marks: 100</b>	
<b>Instructions:</b>			
<b>SECTION A</b> Attempt <b>FIVE</b> questions			
S. No.		Marks	CO
Q 1	Explain the differences between Business Analytics, Data Analytics, and Data Science.	4	CO1
Q 2	What is data wrangling, and why is it essential in Data Science?	4	CO2
Q 3	Differentiate between structured, unstructured, and semi-structured data.	4	CO3
Q 4	Explain the concept of conditional probability with an example.	4	CO1
Q 5	Differentiate between IaaS, PaaS, and SaaS in cloud service models.	4	CO3
Q 6	Discuss the role of Hadoop in Big Data processing.	4	CO3
<b>SECTION B</b> Attempt <b>FOUR</b> questions			
Q 7	What is the purpose of a confusion matrix in model evaluation?	10	CO4
Q 8	Explain the k-means clustering algorithm.	10	CO5
Q 9	Explain how data normalization impacts model accuracy and training.	10	CO4
Q 10	Describe the multiple linear regression model and its assumptions.	10	CO4
Q 11	Discuss the role of decision trees in classification and provide an example.	10	CO4
<b>SECTION-C</b> Attempt <b>TWO</b> questions			
Q 12	Analyze how a healthcare organization could use predictive modeling to manage patient care and resources. Outline the steps from data collection to model deployment.	20	CO5
Q 13	Design a data pipeline for real-time fraud detection using Big Data technologies. Discuss the tools and frameworks you would utilize.	20	CO5

Q 14	a. A dataset contains two variables, income and expenditure. Using the provided dataset, calculate the correlation coefficient and interpret the results. b. Explain principal component analysis (PCA) and its applications in dimensionality reduction.	<b>20</b>	<b>CO5</b>
------	--	-----------	------------