Registration No: -

9	2	0	1	2	0	2	4	5	5

Total Number of Pages: 02 B.Tech / 22IT4PE02T

4th Semester Regular Examination: 2023-24
DATA SCIENCE FOR ENGINEERS
BRANCH: IT
Time: 3 Hours
Max Marks: 100
Q Code: Q136

Answer Question No.1 (Part-1) which is compulsory, any EIGHT from Part-II and any TWO from Part-III.
The figures in the right hand margin indicate marks.

Part-I

Q No.		CO	Level		
Q1				**Short Answer Type Questions (Answer All-10)**	**(02x10)**
	a)	1	3	Differentiate between discrete and continuous variable.	2
	b)	1	3	Differentiate between Business Intelligence and Data Science.	2
	c)	2	2	How statistics plays important role in data analysis?	2
	d)	2	1	Define quartile? How it is measured?	2
	e)	3	2	How k is decided in a k-NN technique?	2
	f)	3	3	Below you are given a summary of the output from a simple linear regression analysis from a sample of size 10, SSR=80, SST = 120. Find the coefficient of determination.	2
	g)	4	1	Write one condition to select the number of principal components in PCA while it is used for feature reduction	2
	h)	4	1	What could be a structure of a neural network to solve a decimal digit classification problem?	2
	i)	5	1	What is graphical data analysis with R/ Python?	2
	j)	2	2	What is difference between overfitting and underfitting.	2

Part-II

Q No.		CO	Level		
Q2				**Focused-Short Answer Type Questions- (Answer Any Eight out of Twelve)**	**(06x08)**
	a)	1	1	Explain different data visualization techniques.	6
	b)	2	3	How do you find a statistical test of a model? Explain with an example.	6
	c)	2	2	Explain what is sampling and discuss different types of data sampling techniques.	6
	d)	2	2	Explain Simpson's Paradox with suitable example.	6
	e)	2	2	Discuss about modeling, model evaluation and critique.	6

	f)	3	2	Explain supervised and unsupervised machine learning with suitable examples.	6
	g)	3	3	Define precision, recall, f1-score and explain how these are used for model evaluation.	6
	h)	3	3	How Bayesian techniques can be used for data classification problem? Explain with an example.	6
	i)	4	4	How do you compare Random Forest with decision Trees? Does always Random Forest give better performance than Decision Tress? Discuss their pros and cons.	6
	j)	4	2	Write F-fold cross validation.	6
	k)	5	4	In what scenarios would you use PCA, and what are the benefits of applying it to a dataset?	6
	l)	5	2	Write an R/Python code for box plot and scatter plot of a dataset.	6

Part-III

Q No.		CO	Level		
				Long Answer Type Questions (Answer Any Two out of Four)	**(02x16)**
Q3	a)	1	3	Explain the different stages of a data science project with the help of a suitable example.	8
	b)	1	1	Explain cleaning and mapping of data with an example.	8
Q4	a)	2	2	Discuss the different types of central tendency of data. Calculate the mode of the following data.	8
	b)	2	3	What is K-Fold cross validation? How Validation error is different from testing error?	8

Marks	No. of Students
30-40	12
40-50	18
50-60	32
60-70	20
70-80	13
80-90	07

Q5	a)	3	2	Discuss Bias-Variance Tradeoff in detail with suitable diagrams.	8
	b)	3	3	Perform at least two iteration of K-Means clustering algorithm. Choose K = 2. Choose first and last data points as initial centers.	8

X	Y
10	8
6	4
13	8
6	10
20	7

Q6	a)	4	2	What is the k-Nearest Neighbors (k-NN) algorithm, and how does it classify a data point?	8
	b)	4	3	Write a program using R/Python to solve a multiple linear regression problem.	8