

Reg.No.:



VIVEKANANDHA COLLEGE OF ENGINEERING FOR WOMEN
[AUTONOMOUS INSTITUTION AFFILIATED TO ANNA UNIVERSITY, CHENNAI]
Elayampalayam – 637 205, Tiruchengode, Namakkal Dt., Tamil Nadu.

Question Paper Code: 13002

B.E. / B.Tech. DEGREE END-SEMESTER EXAMINATIONS – MAY / JUNE 2024

Sixth Semester

Computer Science and Technology

U19CT611 – FOUNDATIONS OF DATA SCIENCE

(Regulation 2019)

Time: Three Hours

Maximum: 100 Marks

Answer ALL the questions

Knowledge Levels (KL)	K1 – Remembering	K3 – Applying	K5 - Evaluating
	K2 – Understanding	K4 – Analyzing	K6 - Creating

PART – A

(10 x 2 = 20 Marks)

Q.No.	Questions	Marks	KL	CO
1.	Define Data Science.	2	K1	CO1
2.	List out the data security issues.	2	K1	CO1
3.	List an overview of common errors in retrieving data and the cleaning solutions to be employed.	2	K2	CO2
4.	What is data discretization?	2	K2	CO2
5.	List out the use of heat map.	2	K2	CO3
6.	What is exploratory data analytics?	2	K2	CO3
7.	Define data sampling.	2	K2	CO4
8.	Mention the validation techniques in machine learning.	2	K2	CO4
9.	List the different types of plot() functions.	2	K2	CO5
10.	Write the significance of matrix plots?	2	K2	CO5

PART – B

(5 x 13 = 65 Marks)

Q.No.	Questions	Marks	KL	CO
11. a)	Explain the various steps associated with data science process with suitable diagrams and example.	13	K2	CO1
(OR)				
b)	Explain in detail about applications of data science in various fields.	13	K1	CO1

12.	a)	Elaborate in detail about the different facets of data with the challenges in their processing.	13	K2	CO2
		(OR)			
	b)	Outline in detail about various steps in data preprocessing with suitable examples.	13	K1	CO2
13.	a)	i. What is mode? Can there be distributions with no mode (or) more than one mode?	3	K2	CO3
		ii. What is median? Outline the steps to find the median and find the median for the following scores; first, set of five scores 2, 8, 2, 7, 6 and second, set of six scores 3, 8, 9, 3, 1, 8 with steps.	10	K2	CO3
		(OR)			
	b)	Calculate the correlation co-efficient for the heights 'in inches' of father' (x) and their son's (y) with the data presented below:	13	K2	CO3
		x 66 68 68 70 71 72 72			
		y 68 70 69 72 72 72 74			
14.	a)	Explain in detail the K-means algorithm with suitable examples.	13	K1	CO4
		(OR)			
	b)	i. Summarize in detail about various methods for evaluating clustering models.	6	K1	CO4
		ii. Explain in detail about the different types of data used in cluster analysis.	7	K1	CO4
15.	a)	Illustrate in detail about the methods in graphical analysis with suitable examples.	13	K1	CO5
		(OR)			
	b)	Discuss in brief about the methods of displaying multivariate data.	13	K1	CO5

PART – C

(1 x 15 = 15 Marks)

Q.No.	Questions	Marks	KL	CO
16.	a) Consider the following dataset with one response variable y and two predictor variables x_1 and x_2 .	15	K2	CO4
	Y 140 155 159 179 192 200 212 215			
	x_1 60 62 67 70 71 72 75 78			
	x_2 22 25 24 20 15 14 14 11			

Fit a multiple linear regression model to this dataset.

(OR)

- b) Perform an exploratory data analysis for the following data with different types of plots. The dataset contains cases from a study that was conducted between 1958 and 1970 at the University of Chicago's billings hospital on the survival of patients who had undergone surgery for breast cancer. .15 K3 CO3

Data Attributes:

Age of patient at the time of operation (numerical)

Patients' year of operation (year-1900, numerical)

Number of positive axillary nodes detected (numerical)

Survival status (class attribute) 1 = the patient survived 5 years (or) longer

2 = the patient died within 5 year.

The first part of the report is a general introduction to the subject of the study. It discusses the importance of the study and the objectives of the research. The second part of the report is a detailed description of the methodology used in the study. This includes a description of the data sources, the sampling method, and the statistical methods used to analyze the data. The third part of the report is a discussion of the results of the study. This includes a description of the findings and a discussion of their implications. The final part of the report is a conclusion and a list of references.

The second part of the report is a detailed description of the methodology used in the study. This includes a description of the data sources, the sampling method, and the statistical methods used to analyze the data.