Q.P Code: DS3104PC **Hall Ticket No.:** | | | | | | | | | | |

NARSIMHA REDDY ENGINEERING

MODEL QUESTION PAPER

COLLEGE(UGC AUTONOMOUS)

III B.Tech I Semester (NR20) Regular Examination, January 2023

INTRODUCTION TO DATA SCIENCE

(CSE / Common to Branch Names – DS)

Time :3 hours **Maximum marks: 75**

Note:
- This question paper contains two parts A and B
- Part A is compulsory which carries 25 marks (1st 5 sub questions are one from each unit carry 2 Marks each & Next 5 sub questions are one from each unit carry 3 Marks). Answer all questions in Part A
- Part B Consists of 5 Units. Answer any one full question from each unit. Each question carries 10 Marks and may have a, b sub questions

Part-A **(25 Marks)**

Answer all questions

Q.No		Question	M	BL	CO	PO
1)	a.	Explain the roles and responsibilities of data scientist in data science	2	L1	CO1	PO1
	b.	Give the neat diagram of stages in a data science project.	2	L1	CO1	PO1
	c.	What is data security in data science?	2	L4	CO1	PO1
	d.	Explain the primary data collection methods.	2	L1	CO2	PO1
	e.	Define tight coupling and loose coupling in data integration.	2	L2	CO2	PO2, PO3
	f.	What is the data reduction? What are the different techniques in the data reduction?	3	L2	CO2	PO5
	g.	Briefly explain standard deviation with example.	3	L1	CO3	PO3, PO4
	h.	Describe how to calculate lower quartile and upper quartile in box plot.	3	L3	CO3	PO3, PO4
	i.	Define polynomial regression. Give the types of polynomial regression.	3	L2	CO4	PO3
	j.	What is model selection? Explain the types of model selection.	3	L4	CO5	PO4, PO5

Part-B **(50 Marks)**

Answer any five questions

All Questions carry equal Marks

Q.No		Question	M	BL	CO	PO
		UNIT–I				
2)	a.	Explain the roles and responsibilities of data analyst and data engineer and ML engineer in data science	5	L1	CO1	PO1
	b.	Briefly explain the different stages in a data science project.	5	L1	CO1	PO1

		OR				
3)	a.	Briefly explain any five major applications of data science.	5	L1	CO1	PO1
	b.	Briefly explain the common data security solutions and techniques in data science security.	5	L1	CO1	PO1
		UNIT–II				
4)	a.	What is data collection? Briefly explain the data collection methods.	5	L4	CO2	PO1
	b.	Explain the data preprocessing steps in machine learning	5	L1	CO2	PO1, PO2
		OR				
5)	a.	Briefly explain about data normalization, data discretization and data generalization.	5	L1	CO2	PO2
	b.	Give the complete description about dimensionality reduction, numerosity reduction and data cube aggregation in data reduction.	5	L1	CO2	PO2, PO3
		UNIT–III				
6)	a.	What are different types of descriptive statistics? And explain them with example.	5	L4	CO3	PO3, PO4
	b.	Describe different types of skewness. Give the relationship between mean, median and mode of skewness in curve graph.	5	L3	CO3	PO3, PO4
		OR				
7)	a.	What is five number summary calculation of box plot? Calculate the five number summary with range in the below dataset: 14, 19, 100, 27, 54, 52, 93, 50, 61, 87,68, 85, 75, 82, 95	5	L4	CO3	PO3, PO4
	b.	What is heat map? Give the different types and uses of heat map with diagrams.	5	L4	CO3	PO3, PO4
		UNIT–IV				
8)	a.	Describe briefly about model evaluation using visualization.	5	L3	CO4	PO3
	b.	What is polynomial regression? Give the types and assumptions of polynomial regression with diagrams.	5	L4	CO4	PO3, PO5
		OR				
9)	a.	Briefly explain about how to measure in-sample evaluation in data science?	5	L2	CO4	PO3, PO12
	b.	Briefly explain about distribution plot with diagram.	5	L2	CO4	PO3, PO12
		UNIT–V				
10)	a.	Give and explain the all evaluation metrics in data science.	5	L2	CO5	PO2, PO3
	b.	Explain about holdout cross validation and k-fold cross validation.	5	L2	CO5	PO2, PO4
		OR				
11)	a.	Explain how to detect the over fitting and under fitting.	5	L2	CO5	PO2, PO4
	b.	Briefly explain grid search parameters.	5	L2	CO5	PO4, PO5

M – Marks **CO –** Course Outcomes **PO –** Program Outcomes

BL – Bloom's Taxonomy Levels (**L1–**Remembering, **L2–**Understanding, **L3–** Applying,**L4–**Analyzing, **L5–**Evaluating, **L6–**Creating)