NARASIMHA REDDY ENGINEERING COLLEGE

(Autonomous)

Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad

Accredited by NAAC with A Grade, Accredited by NBA

COMPUTER SCIENCE AND ENGINEERING

(DATA SCIENCE)

QUESTION BANK

Course Title : INTRODUCTION TO DATA SCIENCE

Course Code : DS3104PC

Regulation : NR 20

Course Objectives

- Learn data science project concepts
- Learn to collect data and process
- Learn to visualize data

Course Outcomes (CO's)

- Able to collect data from various resources and process data
- Able to collect data and apply different techniques on data
- Able to plot data using various methods
- Able to Model and develop the dataset
- Able to visualize data from various data sources

UNIT–I

INTRODUCTION TO DATA SCIENCE

S.No		Questions	BT	CO	PO
Part – A (Short Answer Questions)					
1		Define data science.	L2	CO1	PO1
2		Explain the roles and responsibilities of data scientist in data science	L1	CO1	PO1
3		Explain the data preprocessing stage in the data science project.	L1	CO1	PO1
4		Explain how the data science work in the health-care application.	L1	CO1	PO1
5		Give the neat diagram of stages in a data science project.	L1	CO1	PO1
6		Briefly explain any two major applications of data science.	L1	CO1	PO1
7		What is data security in data science?	L4	CO1	PO1
8		Define data protection in data security.	L2	CO1	PO1
9		Explain the insider threats in data security.	L1	CO1	PO1
10		Define authentication and authorization.	L2	CO1	PO1
Part – B (Long Answer Questions)					
11	a)	What is data science? Briefly explain about data science life cycle.	L4	CO1	PO1
	b)	Briefly explain about evaluation of data science.	L1	CO1	PO1
12	a)	Explain the roles and responsibilities of data analyst and data engineer and ML engineer in data science	L1	CO1	PO1
	b)	Briefly explain the different stages in a data science project.	L1	CO1	PO1
13	a)	Explain the roles and responsibilities of data architect and statistician and business analytics in data science	L1	CO1	PO1
	b)	Give the major applications of data science in various fields and explain.	L1	CO1	PO1
14	a)	Briefly explain any five major applications of data science.	L1	CO1	PO1

	b)	Define data security. What are different data security risks? Explain them	L2	CO1	PO1
15	a)	Explain briefly data security issues in data science.	L1	CO1	PO1
	b)	Briefly explain the common data security solutions and techniques in data science security.	L1	CO1	PO1

UNIT–II

DATA COLLECTION AND DATA PREPROCESSING

S.No		Questions	BT	CO	PO
Part – A (Short Answer Questions)					
1		What is data collection?	L4	CO2	PO1
2		Explain the primary data collection methods.	L1	CO2	PO1
3		Explain the secondary data collection methods.	L1	CO2	PO1
4		What is data preprocessing?	L4	CO2	PO1, PO2
5		Define data profiling and data cleansing.	L2	CO2	PO2
6		What are the different steps of data preprocessing in machine learning.	L4	CO2	PO2
7		Define data cleaning in data preprocessing.	L2	CO2	PO2, PO3
8		Define tight coupling and loose coupling in data integration.	L2	CO2	PO2, PO3
9		What are the different techniques in the data transformation.	L4	CO2	PO5
10		What is the data reduction? What are the different techniques in thedata reduction?	L2	CO2	PO5
Part – B (Long Answer Questions)					
11	a)	What is data collection? Briefly explain the data collection methods.	L4	CO2	PO1
	b)	What is data pre processing? Explain the different methods and key steps in the data preprocessing.	L4	CO2	PO1
12	a)	Explain the data preprocessing steps in machine learning	L1	CO2	PO1, PO2
	b)	Define data cleaning? Explain the different data cleaning steps and methods in data mining.	L2	CO2	PO2, PO3,
13	a)	What is data integration? Briefly explain with neat diagram.	L4	CO2	PO2
	b)	What is data transformation? What are the different techniques in the data transformation and explain them.	L4	CO2	PO2, PO5
14	a)	Briefly explain about data normalization, data discretization and data generalization.	L1	CO2	PO2
	b)	What is data reduction? What are the techniques of data reduction? Explain them.	L4	CO2	PO2, PO5
15	a)	Give the complete description about dimensionality reduction, numerosity reduction and data cube aggregation in data reduction.	L1	CO2	PO2, PO3
	b)	What is data discretization? Briefly explain the techniques of data discretization.	L4	CO2	PO2, PO5

UNIT–III

EXPLORATORY DATA ANALYTICS

S.No		Questions	BT	CO	PO
Part – A (Short Answer Questions)					
1		Explain central tendency of data in descriptive statistics.	L1	CO3	PO3, PO4

2		Define mean, median, mode with example.	L2	CO3	PO3, PO4
3		Briefly explain standard deviation with example.	L1	CO3	PO3, PO4
4		What is the relationship between standard deviation and variance.	L4	CO3	PO3, PO4
5		What is skewness? Give difference between positive skweness and negative skewness.	L4	CO3	PO3, PO4
6		What is kurtosis? Give different types of kurthosis.	L4	CO3	PO3, PO4
7		Describe how to calculate lower quartile and upper quartile in box plot.	L3	CO3	PO3, PO4
8		What is the pivot table? Give the syntax for pivot table.	L4	CO3	PO3, PO4
9		What is heat map? Give any two uses of heat map.	L4	CO3	PO3, PO4
10		What is correlation? Give the relationship between two variables of correlation coefficient.	L4	CO3	PO3, PO4
Part – B (Long Answer Questions)					
11	a)	What are different types of descriptive statistics? And explain them with example.	L4	CO3	PO3, PO4
	b)	What is standard deviation(SD)? Give the steps to calculate SD with example.	L4	CO3	PO3, PO4
12	a)	What is skewness? Give the importance of skewness and distinguish the skewness with normal distribution with diagrams.	L4	CO3	PO3, PO4
	b)	Describe different types of skewness. Give the relationship between mean, median and mode of skewness in curve graph.	L3	CO3	PO3, PO4
13	a)	What is kurtosis? Give the types of kurtosis. Give steps to calculate kurtosis with example.	L4	CO3	PO3, PO4
	b)	What is box plot? Give the five number summary calculation of box plot with example.	L4	CO3	PO3, PO4
14	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	L4	CO3	PO3, PO4
	b)	Give the complete description about pivot table with example in statistics.	L1	CO3	PO3, PO4
15	a)	What is heat map? Give the different types and uses of heat map with diagrams.	L4	CO3	PO3, PO4
	b)	Define correlation coefficient. Give the brief description about correlation with example.	L2	CO3	PO3, PO4

UNIT–IV

MODEL DEVELOPMENT

S.No	Questions	BT	CO	PO
Part – A (Short Answer Questions)				
1	Distinguish between simple and multiple regression.	L4	CO4	PO2
2	What is multiple regression in machine learning?	L4	CO4	PO3, PO5
3	What is model evaluation?	L4	CO4	PO3, PO5
4	Define data visualization.	L2	CO4	PO3, PO5
5	What is residual plot with example?	L4	CO4	PO3, PO5

6		Give the two characteristics of good residual plots.	L1	CO4	PO3, PO5
7		Define polynomial regression. Give the types of polynomial regression.	L2	CO4	PO3
8		Explain briefly data science pipeline.	L1	CO4	PO3, PO12
9		What are the stages in data science pipeline?	L4	CO4	PO3, PO12
10		What is prediction and decision making in development?	L4	CO4	PO3, PO12
Part – B (Long Answer Questions)					
11	a)	What is regression analysis? Briefly describe about simple and multiple regression.	L4	CO4	PO1, PO3
	b)	Describe briefly about model evaluation using visualization.	L3	CO4	PO3
12	a)	Describe how to use residual plots for regression model validation.	L4	CO4	PO3, PO5
	b)	What is polynomial regression? Give the types and assumptions of polynomial regression with diagrams.	L4	CO4	PO3, PO5
13	a)	What is pipeline in data science? Explain how the working of data science pipeline.	L4	CO4	PO3, PO5
	b)	What are the different stages of pipeline? Describe benefits and features of data science pipeline.	L4	CO4	PO3, PO5
14	a)	Briefly explain about how to measure in-sample evaluation in data science?	L2	CO4	PO3, PO12
	b)	Explain about prediction and decision making in data science with diagrams.	L2	CO4	PO3, PO12
15	a)	Briefly explain about distribution plot with diagram.	L2	CO4	PO3, PO12
	b)	Give complete description about model evaluation with a dataset example.	L1	CO4	PO3, PO12

UNIT–V

MODEL EVALUATION GENERALIZATION ERROR

S.No		Questions	BT	CO	PO
Part – A (Short Answer Questions)					
1		Define generalization error.	L2	CO5	PO2
2		What is cross validation?	L4	CO5	PO2
3		Explain any two common techniques used for cross validation.	L2	CO5	PO5
4		Define over fitting.	L2	CO5	PO2, PO4
5		Differentiate over fitting and under fitting.	L4	CO5	PO2, PO4
6		What is ridge regression?	L4	CO5	PO4, PO5
7		What is model selection? Explain the types of model selection.	L4	CO5	PO4, PO5
8		Explain any two over fitting avoid techniques.	L2	CO5	PO4, PO5
9		What is grid search?	L4	CO5	PO4, PO5
10		Explain cross validation with diagram.	L2	CO5	PO4, PO5
Part – B (Long Answer Questions)					
11	a)	Briefly explain about generalization error with diagram.	L2	CO5	PO2, PO3
	b)	Give and explain the all evaluation metrics in data science.	L2	CO5	PO2, PO3

12	a)	Define cross validation. Briefly explain about cross validation techniques.	L2	CO5	PO2, PO4
	b)	Explain about holdout cross validation and k-fold cross validation.	L2	CO5	PO2, PO4
13	a)	Briefly explain about over fitting and under fitting with diagrams.	L2	CO5	PO2, PO4
	b)	Explain how to detect the over fitting and under fitting.	L2	CO5	PO2, PO4
14	a)	Briefly explain about ridge regression.	L2	CO5	PO4, PO5
	b)	Explain bias vs variance in the ridge regression.	L2	CO5	PO4, PO5
15	a)	Briefly explain grid search parameters.	L2	CO5	PO4, PO5
	b)	Explain how to implement grid search by using a dataset.	L2	CO5	PO4, PO5

* **Blooms Taxonomy Level (BT)**(L1 – Remembering; L2 – Understanding; L3 – Applying; L4 – Analyzing; L5 – Evaluating; L6 – Creating)

Course Outcomes (CO)

Program Outcomes (PO)

Prepared By: P.LAKSHMI PRSANNA **HOD,**