

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: 50051

B.E. / B.Tech. DEGREE END-SEMESTER EXAMINATIONS – JAN. 2025

Sixth Semester

Computer Science and Engineering

U19CSV32 – DATA SCIENCE AND ANALYTICS

(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.	What is Data Science? Differentiate between Data Analytics and Data Science.	2	K2	CO1
2.	The given data set is consisting of variables with more than 30 percent missing values. How will you deal with them?	2	K3	CO1
3.	List down the various data analytic challenges faced in the conventional system.	2	K1	CO2
4.	Why do we want to resample an existing Data?	2	K2	CO2
5.	How to find the hyper plane dimension given the dimension of data in Support Vector Machine classification?	2	K2	CO3
6.	State the purpose of Multilayer Feed forward Neural Networks.	2	K2	CO3
7.	Mention various issues faced in Data Stream Query Processing.	2	K2	CO4
8.	What does the K-th order moment of the data Stream Mean?	2	K2	CO4
9.	Write a python code for line chart for the following data. years = [1950, 1960, 1970, 1980, 1990, 2000, 2010] gdp = [300.2, 543.3, 1075.9, 2862.5, 5979.6, 10289.7, 14958.3].	2	K3	CO5
10.	Define Collective inference in Social Network Analysis.	2	K2	CO5

PART – B

(5 x 13 = 65 Marks)

Q.No.	Questions	Marks	KL	CO
11. a)	i. Outline the purpose of data cleansing. How are missing values, omit the values and Impute a static value handled and modified during preprocessing stage?	8	K2	CO1
	ii. Write a brief summary about the challenges faced in processing the big data in data science.	5		
(OR)				
b)	Describe the life cycle of Data Science with neat diagram and explain the objectives of Exploratory Data Analysis.	13	K2	CO1
12. a)	i. Illustrate the evolution of Analytic Scalability from conventional Database Management System.	8	K2	CO2
	ii. Summarize the most commonly used Modern data analytical tools used by Data Analyst.	5		
(OR)				
b)	i. Explain the statistical concepts used in sampling distribution in data analytics.	8	K2	CO2
	ii. Discuss Analysis vs. Reporting.	5		
13. a)	i. Write the formula for Bayes Theorem and explain Naïve Bayes classifier with necessary routine.	8	K2	CO3
	ii. Brief about Principal Component Analysis (PCA) Method.	5		
(OR)				
b)	Construct a decision tree for sample data of your own and evaluate various decisions that can be arrived based on the decision tree.	13	K3	CO3
14. a)	i. Describe the Flajolet-Martin algorithm that estimates the number of distinct elements with suitable examples.	8	K2	CO4
	ii. Write short notes on Bloom filter.	5		
(OR)				
b)	Explain in detail about Real-time Analytics Platform (RTAP) applications with suitable examples.	13	K2	CO4

15.	a)	i.	Explain different visualization techniques in detail with an example.	8	K2	CO5
		ii.	Write short notes on Egonets.	5		
(OR)						
	b)		Elucidate in detail about Social Network Analysis.	13	K2	CO5

PART – C

(1 x 15 = 15 Marks)

Q.No.	Questions	Marks	KL	CO	
16.	a)	Develop a case study of Sentiment Analysis in Twitter.	15	K4	CO4
(OR)					
	b)	Assume that we have a stream of items of large and unknown length that we can only iterate over once. Devise an effective sampling algorithm that randomly chooses an item from this stream such that each item is equally likely to be selected. Define a Decaying window and explain how finding of most popular elements work in online streaming.	15	K4	CO4