Shiv Nadar University Chennai

End Semester Examinations, 2023-2024 Even

Question Paper

Name of the Program: B.Tech. Artificial Intelligence and Data Science	Semester: II
Course Code & Name: CS1704 FOUNDATIONS OF DATA SCIENCE + LAB	
Regulation 2021	
Time: 3 Hours	Maximum: 100 Marks

Q.No.		Questions	Marks	CO#	KL#
1	a	What does the K in K-means and KNN mean? What are the machine learning scenarios you pick them?	5	CO3	KL2
2	a	Define skewness, represent the order of mean, mode and median for right and left skewed data compared to the normal distribution.	5	CO2	KL1
3	a	Discuss the use of multiple Integrated Development Environments (IDEs) in data science, highlighting their respective advantages and disadvantages.	5	CO1	KL2
4	a	Consider a business scenario where a credit card company wants to assess the performance of its fraud detection model. Here's a hypothetical confusion matrix based on the model's predictions: | | Predicted Non-Fraud | Predicted Fraud | | Actual Non-Fraud | 920 | 30 | | Actual Fraud | 20 | 230 | Calculate the following metrics Accuracy, Precision, Sensitivity, F1 score. And give your conclusion.	5	CO4	KL3
5	a	Write NumPy programs for the following along with your own input and output: 1. Reverse an array 2. Convert a 1-D array to 2-D array 3. Create a 2-D array with 2000 zeroes. The array should have 50 rows and 40 columns. 4. Change the sign of 4th and 5th element in a 1-D array Eg i/p: [1,2,3,4,5,6] ; o/p : [1,2,3,-4,-5,6] 5. Create a 3D array with ones on a diagonal and zeros elsewhere.	5	CO1	KL2
6	a	Imagine you're tasked with analyzing a dataset containing information about customer purchases for an e-commerce company. Describe how you would approach this project using the data science lifecycle.	15	CO4	KL2
7	a	A data scientist for a market research firm is tasked with conducting a survey to understand consumer preferences for a new product launch. Discuss about different sampling techniques available in data science and choose an appropriate sampling technique for this scenario and justify.	15	CO1	KL3
8	a	Consider a company is launching a new advertising campaign aimed at increasing brand awareness. They want to assess the effectiveness of the campaign by comparing brand recognition before and after the campaign. Design a market research study to investigate whether the new advertising campaign has a statistically significant impact on brand recognition among 3 different age groups using hypothesis testing with confidence intervals. [Note: For a 95% confidence level, the critical value is 1.96]	15	CO2	KL5

			Participant ID	Age Group	Brand Recognition (Before)	Brand Recognition (After)			
			1	18-25	15	20			
			2	26-35	12	18			
			3	36-45	14	22			
			4	18-25	16	19			
			5	26-35	13	21			
			6	36-45	17	23			
			7	18-25	11	17			
			8	26-35	14	20			
			9	36-45	12	18			
			10	18-25	15	22			
9	a	A real estate reselling company wants to predict house prices for their properties based on various factors such as square footage, number of bedrooms, location, and age of the property. Considering these factors, which machine learning algorithm you would choose and explain its assumption, working and metrics?					15	CO3	KL3
10	a	A company offers a training program to its sales employees to improve their performance. The training includes modules on product knowledge, sales techniques, and customer relationship management. The management wants to analyze the data of the number of hours spent in training and the sales performance of the employees after training. Explain the statistical measure and calculate the mean, variance, standard deviation, covariance, and correlation for the dataset.					15	CO2	KL3

Employee	Hours Trained (X)	Monthly Sales (in $1000s)
1	5	50
2	10	55
3	8	52
4	15	85
5	12	70
6	20	95
7	18	80
8	9	65
9	14	80
10	11	60

KL – Bloom's Taxonomy Levels

(KL1: Remembering, KL2: Understanding, KL3: Applying, KL4: Analyzing, KL5: Evaluating, KL6: Creating)

CO – Course Outcomess
