

--	--	--	--	--	--	--	--

MCA DEGREE FIFTH SEMESTER EXAMINATION, JANUARY 2022

19-381-05047 DATA SCIENCE AD ANALYTICS

(Regular)

Q.No		QUESTIONS	Marks	CO	BL	PI
1.	(a)	Differentiate the role of data scientist and data analyst.	6	CO1	L2	1.7.1
	(b)	What are the techniques used to handle the missing data in a data science process.	4			
2.	(a)	Explain the parameters that are used to represent the amount of dispersion in a dataset.	7	CO2	L2	1.2.2
	(b)	Consider the following set of data: 14, 16, 7, 9, 11, 13, 8, 10. Calculate the variance and standard deviation.	3		L1	
3	(a)	Explain how support vector machine classify data?	3	CO2	L2	1.7.1
	(b)	Write two applications to classify data using SVM.	2			
	(c)	Explain how SVM handles non- linearly separable data.	5			
4.	(a)	Draw the Box plot in python for the given data. Data 1 = {100,10,200} Data 2 = {90,20,200} Data 3 = {80,30,200} Data 4 = { 70,40,200} Also draw the box plot and mark the outliers if any.	10	CO4	L3	1.7.1
	(b)	Explain in detail about retinal variables.				
5.	(a)	Write python script to create a line plot ,bar plot, for the following data x={1,2,3,4,5} y={2,4,6,8,10} Also use the random library in python to create scatter plot.	6	CO3	L3	1.7.1
			2			2.6.3
	(b)	List the difference in python libraries math plotlib and bokeh.	2			2.6.3

6.		Use K means algorithm and Eudidean distance to cluster the following points (2,10) (2,5) (8,4) (5,8) (7,5) (6,4)(1,2)(4,9) into three clusters . Initial centers are (2,10) (5,8) & (1,2) and write python code to cluster the above data points.	10	CO5	L3	1.7.1
7	(a)	Discuss the advantages of deep learning over other machine learning algorithms.	3	CO5	L2	1.7.1
	(b)	Explain RNN and its variant based models in differential sequential model.	7			
