

Code: 20IT2701A

**IV B.Tech - I Semester – Regular Examinations - DECEMBER 2023**

**FUNDAMENTALS OF DATA SCIENCE**  
(Common for ALL BRANCHES)

Duration: 3 hours

Max. Marks: 70

Note: 1. This paper contains questions from 5 units of Syllabus. Each unit carries 14 marks and have an internal choice of Questions.

2. All parts of Question must be answered in one place.

BL – Blooms Level

CO – Course Outcome

			BL	CO	Max. Marks
<b>UNIT-I</b>					
1		Explain the steps involved in the process of data science project.	L2	CO1	14 M
<b>OR</b>					
2		Explain with example on the sampling of data for modeling and validation.	L2	CO1	14 M
<b>UNIT-II</b>					
3		Explain about any 4 tasks that could be performed by machine learning algorithms. Give suitable example for each.	L2	CO2	14 M
<b>OR</b>					
4	a)	Explain about classification and any one algorithm to perform the task of classification.	L2	CO2	7 M
	b)	Apply the classification algorithm to classify the image dataset into binary classes of cats and not cats.	L2	CO2	7 M

<b>UNIT-III</b>																					
5	Explain Linear regression with a suitable example application.			L2 L3	CO2	14 M															
<b>OR</b>																					
6	Explain Logistic regression with a suitable example application.			L2 L3	CO2	14 M															
<b>UNIT-IV</b>																					
7	a)	Explain K-Means clustering algorithm.		L2	CO2	7 M															
	b)	Cluster the following eight points (with (x, y) representing locations) into three clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9).		L3	CO2	7 M															
<b>OR</b>																					
8	a)	Describe the algorithm for Association rule mining.		L2	CO2	7 M															
	b)	Trace the results of using the Apriori algorithm on the grocery store example with support threshold $s=33.34\%$ and confidence threshold $c=60\%$ . Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence.		L3	CO2	7 M															
		<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center;">Transaction ID</th> <th style="text-align: center;">Items</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">T1</td> <td>HotDogs, Buns, Ketchup</td> </tr> <tr> <td style="text-align: center;">T2</td> <td>HotDogs, Buns</td> </tr> <tr> <td style="text-align: center;">T3</td> <td>HotDogs, Coke, Chips</td> </tr> <tr> <td style="text-align: center;">T4</td> <td>Chips, Coke</td> </tr> <tr> <td style="text-align: center;">T5</td> <td>Chips, Ketchup</td> </tr> <tr> <td style="text-align: center;">T6</td> <td>HotDogs, Coke, Chips</td> </tr> </tbody> </table>		Transaction ID	Items	T1	HotDogs, Buns, Ketchup	T2	HotDogs, Buns	T3	HotDogs, Coke, Chips	T4	Chips, Coke	T5	Chips, Ketchup	T6	HotDogs, Coke, Chips				
Transaction ID	Items																				
T1	HotDogs, Buns, Ketchup																				
T2	HotDogs, Buns																				
T3	HotDogs, Coke, Chips																				
T4	Chips, Coke																				
T5	Chips, Ketchup																				
T6	HotDogs, Coke, Chips																				

**UNIT-V**

9	Explain about Web Mining and Web Structure Mining.	L2	CO3	14 M
---	--	----	-----	------

**OR**

10	Explain about Text Mining algorithm and its advantages and applications.	L2	CO3	14 M
----	--	----	-----	------