PVP 20

Code: 20CS4501A

III B.Tech - I Semester – Regular Examinations - DECEMBER 2022

DATA SCIENCE
(COMPUTER SCIENCE & ENGINEERING)

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
UNIT-I					
1	a)	Summarize the steps and benefits of Data Preparation process with an example.	L2	CO1	10 M
	b)	Discuss Exploratory Data Analysis.	L2	CO1	4 M
OR					
2	a)	Illustrate the examples to improve your Business Understanding in Data science.	L2	CO1	7 M
	b)	Explain the popular metrics of model evaluation in Data Science.	L2	CO1	7 M
UNIT-II					
3	a)	Relate the need of Data Integration in preprocessing.	L3	CO2	6 M
	b)	Differentiate Normalization, Aggregation and Discretization.	L2	CO1	8 M
OR					

4	a)	Show the different methods of Attribute subset selection process.	L3	CO2	10 M
	b)	Explain Attribute Selection measure.	L2	CO1	4 M
		UNIT-III			
5	a)	Compare the following different discrete distribution models. Uniform, Binomial, Bernoulli and Poisson.	L3	CO3	10 M
	b)	Explain Stratified sampling.	L2	CO1	4 M
		OR			
6	a)	Illustrate the process of Student's T distribution with example.	L3	CO3	7 M
	b)	"Cluster sampling is a sampling plan used when mutually homogeneous yet internally heterogeneous groupings are evident in a statistical population". Infer the above statement with examples.	L3	CO3	7 M
		UNIT-IV			
7	a)	Implement Linear Regression model with an example.	L3	CO4	8 M
	b)	Give a key note on Regression Coefficient and properties and assumptions of Linear Regression.	L2	CO1	6 M
		OR			
8	a)	Discuss the need for fitting the model in multiple regression.	L2	CO1	6 M
	b)	Illustrate Linear Discriminative analysis with an example.	L3	CO4	8 M

UNIT-V					
9	a)	Explain the bias/ variance dilemma about the model complexity.	L2	CO1	7 M
	b)	Explain k-fold cross validation and how it can be implemented for building a model.	L3	CO4	7 M
OR					
10	a)	How to calculate prediction error? Justify the answer by describing which is better prediction or estimation?	L3	CO4	7 M
	b)	What is Minimum Description Length (MDL)? How MDL is used in Model Class Selection?	L2	CO1	7 M