

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
MID TERM EXAM
SUMMER SEMESTER 2024-2025


SLOT: F1+TF1+F2+TF2

Programme Name & Branch : B.Tech (Common to All)
Course Code and Course Name : BCSE334L & Predictive Analytics
Faculty Name(s) : Dr PADMAVATHY T
Class Number(s) : VL2024250700182
Date of Examination : 10.06.2025
Exam Duration : 90 minutes **Maximum Marks: 50**

General instruction(s):

- Answer All Questions
 - M - Max mark; CO - Course Outcome; BL - Blooms Taxonomy Level (1 - Remember, 2 - Understand, 3 - Apply, 4 - Analyse, 5 - Evaluate, 6 - Create)
 - Course Outcomes (Type the CO statements covered in this question paper. Use the CO number as per the syllabus copy)
1. Understand the importance of predictive analytics and processing of data for analysis.
 2. Describe different types of predictive models.
 3. Apply regression and classification model on applications for decision making and evaluate the performance.

Q. No	Question	M	CO	BL																									
1.	Compare and contrast predictive analysis and prescriptive analysis.	5																											
2.	Explain in detail about the various visualization techniques for categorical data.	5	1	2																									
	<p>You're analysing customer purchase behaviour for an e-commerce platform where the amount spent is given as below:</p> <table border="1" style="margin: 10px auto; border-collapse: collapse;"> <thead> <tr> <th>Customer ID</th> <th>CO01</th> <th>CO02</th> <th>CO03</th> <th>CO04</th> <th>CO05</th> <th>CO06</th> <th>CO07</th> <th>CO08</th> <th>CO09</th> <th>CO010</th> </tr> </thead> <tbody> <tr> <td>Amount Spent</td> <td>50</td> <td>75</td> <td>100</td> <td>200</td> <td>500</td> <td>600</td> <td>450</td> <td>1200</td> <td>1500</td> <td>300</td> </tr> </tbody> </table> <p>As a data analyst, identify the problem with the data and apply necessary transformation to make the data suitable for modelling. Also, explain the alternative data transformation methods.</p>	Customer ID	CO01	CO02	CO03	CO04	CO05	CO06	CO07	CO08	CO09	CO010	Amount Spent	50	75	100	200	500	600	450	1200	1500	300	10	2	4			
Customer ID	CO01	CO02	CO03	CO04	CO05	CO06	CO07	CO08	CO09	CO010																			
Amount Spent	50	75	100	200	500	600	450	1200	1500	300																			
3.	<p>As a data scientist, you are analysing the patient vitals to predict a cardiac event.</p> <table border="1" style="margin: 10px auto; border-collapse: collapse;"> <thead> <tr> <th>Patient ID</th> <th>HRV</th> <th>Blood Pressure</th> <th>Oxygen</th> <th>Respiration</th> </tr> </thead> <tbody> <tr> <td>P001</td> <td>80</td> <td>120</td> <td>98</td> <td>16</td> </tr> <tr> <td>P002</td> <td>78</td> <td>118</td> <td>97</td> <td>15</td> </tr> <tr> <td>P003</td> <td>500</td> <td>300</td> <td>90</td> <td>30</td> </tr> <tr> <td>P004</td> <td>82</td> <td>122</td> <td>98</td> <td>17</td> </tr> </tbody> </table> <p>Apply necessary transformations to the predictors so that the model's decision boundary is not distorted and the model's performance is not affected.</p>	Patient ID	HRV	Blood Pressure	Oxygen	Respiration	P001	80	120	98	16	P002	78	118	97	15	P003	500	300	90	30	P004	82	122	98	17	10	2	4
Patient ID	HRV	Blood Pressure	Oxygen	Respiration																									
P001	80	120	98	16																									
P002	78	118	97	15																									
P003	500	300	90	30																									
P004	82	122	98	17																									

 (Deemed to be University under section 3 of UGC Act, 1956)
SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
MID TERM EXAM
SUMMER SEMESTER 2024-2025

SLOT: F1+TF1+F2+TF2

4.	Discuss in detail the various clustering models used in predictive analytics. Highlight their underlying principles, key characteristics, and typical applications.	10	2	2																		
5.	Explain the application of linear regression A real estate company wants to estimate house prices based only on the area (in square feet) of each house. They have collected the following data:	10	3	3																		
<table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th style="width: 30%;">HouseID</th> <th style="width: 30%;">Area (sqft)</th> <th style="width: 30%;">Price (\$1000s)</th> </tr> </thead> <tbody> <tr> <td>H001</td> <td>1000</td> <td>200</td> </tr> <tr> <td>H002</td> <td>1200</td> <td>240</td> </tr> <tr> <td>H003</td> <td>1500</td> <td>300</td> </tr> <tr> <td>H004</td> <td>1800</td> <td>360</td> </tr> <tr> <td>H005</td> <td>2000</td> <td>400</td> </tr> </tbody> </table>					HouseID	Area (sqft)	Price (\$1000s)	H001	1000	200	H002	1200	240	H003	1500	300	H004	1800	360	H005	2000	400
HouseID	Area (sqft)				Price (\$1000s)																	
H001	1000	200																				
H002	1200	240																				
H003	1500	300																				
H004	1800	360																				
H005	2000	400																				
a) Fit a simple linear regression model using Area as the independent variable. b) Interpret the slope and intercept of the regression line. c) Use the model to predict the price of a 1600 sq. ft. house. d) Explain the assumptions of simple linear regression in this context.																						
