

Name:	
Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, May 2022

Course: Data Mining & Prediction Modeling
Program: B.Tech CSE (G&G+IOT&SC+CSF+BT+OGI)
Course Code: CSBA3001P

Semester: VI
Time : 03 hrs.
Max. Marks: 100

Instructions: All the questions are compulsory.

SECTION A
(5Qx4M=20Marks)

S. No.	Question	Marks	CO
Q 1	Describe the term segmentation and explain how to deal with missing values.	4	CO1
Q 2	Is outlier detection same as anomaly detection? Justify.	4	CO2
Q 3	How are decision trees used for classification and why decision tree classifiers are so popular?	4	CO3
Q 4	Define the following terms (a) Data aggregation (b) Data generalization	4	CO4
Q 5	Explain the term “Data Mining”. How can you differentiate data mining from data warehouse?	4	CO5

SECTION B
(4Qx10M= 40 Marks)

Q 6	What is machine learning approach to data mining? Explain KDD life cycles with diagram.	10	CO2												
Q 7	Illustrate the difference between OLAP and OLTP? Explain Multi-tiered architecture in data mining with diagram.	10	CO3												
Q 8	Draw and explain the structure of ANN and state in which manner biological neurons are relevant to ANN.	10	CO4												
Q 9	For the following transaction dataset, generate rules using Apriori algorithm. Consider the values as Support=30% and Confidence=50%	10	CO5												
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 40%;">Transaction Id</th> <th style="width: 60%;">Items Purchased</th> </tr> </thead> <tbody> <tr> <td>11</td> <td>I₁, I₂, I₃</td> </tr> <tr> <td>12</td> <td>I₂, I₄</td> </tr> <tr> <td>13</td> <td>I₂, I₃</td> </tr> <tr> <td>14</td> <td>I₁, I₃, I₄</td> </tr> <tr> <td>15</td> <td>I₁, I₃</td> </tr> </tbody> </table>				Transaction Id	Items Purchased	11	I ₁ , I ₂ , I ₃	12	I ₂ , I ₄	13	I ₂ , I ₃	14	I ₁ , I ₃ , I ₄	15	I ₁ , I ₃
Transaction Id	Items Purchased														
11	I ₁ , I ₂ , I ₃														
12	I ₂ , I ₄														
13	I ₂ , I ₃														
14	I ₁ , I ₃ , I ₄														
15	I ₁ , I ₃														

	<p style="text-align: center;">OR</p> <p>In an Anti-malarial campaign in India, Quinine was administered to 500 persons out of a total population of 2000. The number of fever cases is shown below:</p> <table border="1" data-bbox="240 411 1162 564"> <thead> <tr> <th>Treatment</th> <th>Fever</th> <th>No Fever</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Quinine</td> <td>20</td> <td>480</td> <td>500</td> </tr> <tr> <td>No Quinine</td> <td>100</td> <td>1400</td> <td>1500</td> </tr> <tr> <td>Total</td> <td>120</td> <td>1880</td> <td>2000</td> </tr> </tbody> </table> <p>Discuss the usefulness of Quinine in checking malaria by apply Chi-Square Test.</p>	Treatment	Fever	No Fever	Total	Quinine	20	480	500	No Quinine	100	1400	1500	Total	120	1880	2000		
Treatment	Fever	No Fever	Total																
Quinine	20	480	500																
No Quinine	100	1400	1500																
Total	120	1880	2000																
<p>SECTION-C (2Qx20M=40 Marks)</p>																			
Q 10	<ul style="list-style-type: none"> • Brief the key issues in data mining? (6) • What is the difference between univariate, bivariate, and multivariate analysis? (7) • Is it possible to capture the correlation between continuous and categorical variables? If yes, justify the statement. (7) 	20	CO3																
Q 11	<p>Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 36, 40.</p> <p>a) Find mean & median (5)</p> <p>b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.). (5)</p> <p>c) Explain the following terms (10)</p> <ul style="list-style-type: none"> ○ midrange of the data ○ Interquartile range (IQR) <p style="text-align: center;">OR</p> <ul style="list-style-type: none"> • How Dimensionality Reduction plays an important role in Data preprocessing? (5) • Briefly explain the advantages and disadvantages of MOLAP storage model? (5) • Define linear Regression? Discuss different situations where linear regression can work better? (10) 	20	CO5																