
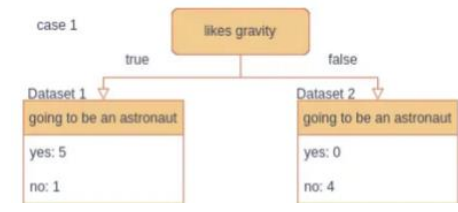
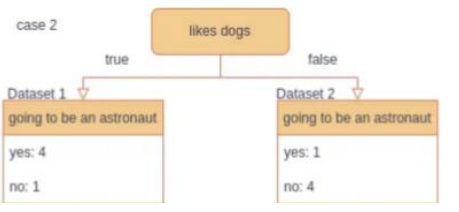


Name:			
Enrolment No:			
<b>UPES</b> <b>End Semester Examination, December 2024</b>			
<b>Course:</b> Predictive Analytics <b>Program:</b> BTech CSE Data Science and AIML <b>Course Code:</b> CSBA3017		<b>Semester :</b> 5 <b>Time :</b> 03 hrs. <b>Max. Marks:</b> 100	
<b>Instructions:</b> i. Please attempt according to the provided time and given weightage. ii. Use of calculator is allowed.			
<b>SECTION A</b> <b>(5Qx4M=20Marks)</b>			
S. No.		Marks	CO
Q 1	How can you utilize clustering as a preprocessing tool? Explain two measures.	4	CO2
Q2	Differentiate between i. Nominal data and Ordinal data ii. Standardization and normalization	4	CO1
Q3	Define k-fold cross validation. A dataset has 1000 instances. You perform 10-fold cross-validation. How many instances are in each fold?	4	CO3
Q4	Discuss four issues to consider during data integration as a pre-processing step.	4	CO2
Q5	You are given the following data points representing two documents in a text classification task, with each value representing the frequency of a certain term across six terms: Document 1: (4, 2, 0, 3, 6, 1) Document 2: (3, 1, 2, 4, 5, 0) Calculate the cosine similarity between the two documents across all six terms.	4	CO1
<b>SECTION B</b> <b>(4Qx10M= 40 Marks)</b>			
Q6	Outline the major five research challenges of KDD process when applied in one specific application domain stream/sensor data analysis. Also, list five benefits of adopting data mining for the specified application.	5+5 =10	CO1
Q7	i. Define two evaluation metrics used for regression analysis. ii. Given a confusion matrix with the following values: TP = 80, FP = 20, FN = 10, TN = 90 Calculate the accuracy, precision, recall, and F1-score.	5+5 =10	CO3
Q8	Compare and contrast web mining and traditional data mining in terms of their definition, data sources, techniques, goals and applications.	5+5 =10	CO4
Q9	i. How does SVM handle high-dimensional and non-linear data. ii. Discuss three major clustering approaches, detailing which approach is preferred for which type of data/application.	5+5 =10	CO2

	<p>Or</p> <p>i. List three strengths and three weaknesses of artificial neural network.</p> <p>ii. Demonstrate how a linear regression model can be used to predict future values of a target variable?</p>		
--	---	--	--

**SECTION-C**  
**(2Qx20M=40 Marks)**

Q10	<p>i. Using a block diagram, demonstrate the process of data preparation in Web Usage mining.</p> <p>ii. Explain the concepts of Pageview Identification, User Identification and Session Identification with help of diagrams.</p>	<b>10+10</b> <b>=20</b>	<b>CO4</b>
-----	---	----------------------------	------------

Q11	<p>Given a dataset for predicting whether a person is going to be an astronaut, depending on whether they like dogs and whether they like gravity.</p> <p>i. Using Gini Index as the attribute selection measure, compute and decide which of cases shown below will be considered for creating the decision tree.</p> <p>ii. Construct the final decision tree obtained. Using If/Else AND/OR formulate the prediction rules for cases where a person is going to be an astronaut.</p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="background-color: #fff9c4;">likes dogs</th> <th style="background-color: #fff9c4;">likes gravity</th> <th style="background-color: #fff9c4;">going to be an astronaut</th> </tr> </thead> <tbody> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </tbody> </table> <div style="display: flex; justify-content: space-around; margin-top: 20px;"> <div style="text-align: center;"> <p>case 1</p>  </div> <div style="text-align: center;"> <p>case 2</p>  </div> </div>	likes dogs	likes gravity	going to be an astronaut	0	0	0	1	1	1	0	1	1	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	0	1	0	1	1	1	<b>10+10</b> <b>=20</b>	<b>CO3</b>
likes dogs	likes gravity	going to be an astronaut																																			
0	0	0																																			
1	1	1																																			
0	1	1																																			
0	0	0																																			
0	0	0																																			
1	1	1																																			
1	0	0																																			
1	1	1																																			
0	1	0																																			
1	1	1																																			