

COLLEGE: ST. JOSEPH'S COLLEGE OF ENGINEERING AND TECHNOLOGY,PALAI																															
EIGHTH SEMESTER B.TECH DEGREE EXAMINATION, JUNE 2021																															
Course Code: CS 402																															
Course Name: DATA MINING AND WAREHOUSING																															
Max. Marks: 70			Duration: 2.15 Hours																												
PART A																															
		Answer any two full questions, each carries 10.5 marks.	Marks																												
1	a)	With a neat sketch, discriminate the processes in knowledge discovery.	(7)																												
	b)	Suppose that a data warehouse consists of the four dimensions date,spectator, location and game and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given date.Spectators may be students, adults or seniors with each category having its own charge rate.Draw a star schema diagram for the data warehouse.	(3.5)																												
2	a)	Why is sampling used as a data reduction technique? Use suitable sketches to discriminate any three sampling techniques.	(7)																												
	b)	A set of data is given: A={116,234,486,544}. Normalize the data by min-max normalization by setting min=0 and max=1	(3.5)																												
3	a)	Define OLAP and demonstrate any three OLAP operations on multidimensional data with suitable sketches.	(7)																												
	b)	Sort the given data :20,16,16,19,20,13,15,21,22,25,25,22 and use smoothing by bin means to smooth the data, using a bin depth of 3.	(3.5)																												
PART B																															
Answer any two full questions, each carries 10.5 marks.																															
4	a)	With an example illustrate the steps in the data classification process.	(7)																												
	b)	Symptoms of patients and diagnosis of flu are given in the following table. <table><tr><td>headache</td><td>runny nose</td><td>fever</td><td>flu</td></tr><tr><td>N</td><td>Y</td><td>Y</td><td>N</td></tr><tr><td>Y</td><td>N</td><td>N</td><td>N</td></tr><tr><td>N</td><td>N</td><td>N</td><td>N</td></tr><tr><td>Y</td><td>Y</td><td>Y</td><td>Y</td></tr><tr><td>Y</td><td>Y</td><td>N</td><td>Y</td></tr><tr><td>N</td><td>N</td><td>Y</td><td>Y</td></tr></table>	headache	runny nose	fever	flu	N	Y	Y	N	Y	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y	Y	(3.5)
headache	runny nose	fever	flu																												
N	Y	Y	N																												
Y	N	N	N																												
N	N	N	N																												
Y	Y	Y	Y																												
Y	Y	N	Y																												
N	N	Y	Y																												

		<p>Using Naïve Bayes classifier algorithm identify a patient with the following symptoms has flu or not.</p> <table><tr><td>headache</td><td>runny nose</td><td>fever</td><td>flu</td></tr><tr><td>Y</td><td>N</td><td>Y</td><td>?</td></tr></table> <p>use below details for the calculation.</p> <table><tr><td>P(headache=Y/flu=N)</td><td>1/3</td></tr><tr><td>P(headache=Y/flu=Y)</td><td>2/3</td></tr><tr><td>P(runny nose=N/flu=N)</td><td>2/3</td></tr><tr><td>P(runny nose=N/flu=Y)</td><td>1/3</td></tr><tr><td>P(fever=Y/flu=N)</td><td>1/3</td></tr><tr><td>P(fever=Y/flu=Y)</td><td>2/3</td></tr><tr><td>P(flu=N)</td><td>1/2</td></tr><tr><td>P(flu=Y)</td><td>1/2</td></tr></table>	headache	runny nose	fever	flu	Y	N	Y	?	P(headache=Y/flu=N)	1/3	P(headache=Y/flu=Y)	2/3	P(runny nose=N/flu=N)	2/3	P(runny nose=N/flu=Y)	1/3	P(fever=Y/flu=N)	1/3	P(fever=Y/flu=Y)	2/3	P(flu=N)	1/2	P(flu=Y)	1/2	
headache	runny nose	fever	flu																								
Y	N	Y	?																								
P(headache=Y/flu=N)	1/3																										
P(headache=Y/flu=Y)	2/3																										
P(runny nose=N/flu=N)	2/3																										
P(runny nose=N/flu=Y)	1/3																										
P(fever=Y/flu=N)	1/3																										
P(fever=Y/flu=Y)	2/3																										
P(flu=N)	1/2																										
P(flu=Y)	1/2																										
5	a)	How does back propagation work ?Explain with a suitable diagram.	(7)																								
	b)	<p>An example confusion matrix for a binary classifier has given</p> <table><tr><td>N=165</td><td>Predicted :YES</td><td>Predicted :NO</td></tr><tr><td>Actual: YES</td><td>100</td><td>5</td></tr><tr><td>Actual: NO</td><td>10</td><td>50</td></tr></table> <p>Calculate: i) Accuracy ii) Recall</p>	N=165	Predicted :YES	Predicted :NO	Actual: YES	100	5	Actual: NO	10	50	(3.5)															
N=165	Predicted :YES	Predicted :NO																									
Actual: YES	100	5																									
Actual: NO	10	50																									
6	a)	What is overfitting in the decision tree? How can it be avoided?Explain with suitable examples.	(7)																								
	b)	With an example illustrate K-Nearest Neighbor classifier.	(3.5)																								
PART C																											

Answer any two full questions, each carries 14 marks.

7	a)	<p>A database has five transactions. Let minimum support count = 3 and minimum confidence = 80%.</p> <table><tr><th><i>TID</i></th><th><i>items_bought</i></th></tr><tr><td>T100</td><td>{M, O, N, K, E, Y}</td></tr><tr><td>T200</td><td>{D, O, N, K, E, Y }</td></tr><tr><td>T300</td><td>{M, A, K, E}</td></tr><tr><td>T400</td><td>{M, U, C, K, Y}</td></tr><tr><td>T500</td><td>{C, O, O, K, I,E}</td></tr></table> <p>find frequent itemset using Apriori algorithm.</p>	<i>TID</i>	<i>items_bought</i>	T100	{M, O, N, K, E, Y}	T200	{D, O, N, K, E, Y }	T300	{M, A, K, E}	T400	{M, U, C, K, Y}	T500	{C, O, O, K, I,E}	(8)
<i>TID</i>	<i>items_bought</i>														
T100	{M, O, N, K, E, Y}														
T200	{D, O, N, K, E, Y }														
T300	{M, A, K, E}														
T400	{M, U, C, K, Y}														
T500	{C, O, O, K, I,E}														
	b)	<p>Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):</p> <p>i) Compute the Euclidean distance between the two objects.</p> <p>ii) Compute the Manhattan distance between the two objects.</p>	(6)												
8	a)	Differentiate AGNES and DIANA with neat sketches.	(8)												
	b)	Discriminate characteristics of social networks.	(6)												
9	a)	Summarize BIRCH algorithm with suitable diagram.	(8)												
	b)	Outline the procedure of k-medoids	(6)												
