BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: BCA **SEMESTER : VI**
BRANCH: BCA **SESSION : SP/2024**

SUBJECT: CA355 DATA MINING

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO	BL
Q.1(a)	What are the steps involved in the data mining process?	[5]	1	1
Q.1(b)	What are the various issues and challenges to be considered in implementing data mining?	[5]	2	2
Q.2(a)	Calculate the similarity and distance measures for the following vectors x and y : i) x=(1,1,1,1), y= (2,2,2,2), (cosine, correlation, Euclidean). ii) X=(1,1,0,1,0,1), y= (1,1,1,0,0,1), (cosine, correlation, Jaccard).	[5]	3	3
Q.2(b)	Discuss Similarity and Dissimilarity between simple attributes.	[5]	2	2
Q.3(a)	What is association rule in data mining? Describe Support and Confidence.	[5]	1	1
Q.3(b)	What is Large itemsets? Generate association rules using Apriori Algorithm on the following datasets for which minimum support is 2 and minimum confidence is 70%.	[5]	3	3
Q.4(a)	Using K-means clustering algorithm, find two clusters for the following data: { 2,4,5,6,8,12,13,15} .	[5]	3	3
Q.4(b)	Differentiate between clustering and classification.	[5]	2	2
Q.5(a)	What is a Decision Tree? Explain with the help of an example.	[5]	3	3
Q.5(b)	Discuss Information Gain as the attribute selection measure.	[5]	2	2

TID	Items Purchased
T1	apple, orange, bread
T2	butter, orange, jelly
T3	apple, butter, orange, jelly
T4	butter, jelly

::::::23/04/2024::::::M