Register Number:……..

KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2023

(Regulation 2018)

Sixth Semester

COMPUTER SCIENCE AND ENGINEERING

U18CSI6203**:** Data warehousing and Data Mining

COURSE OUTCOMES

CO1: Demonstrate data warehouse schema and process of data retrieval for real time applications. [K3]

CO2: Identify necessity of data pre-processing and apply the appropriate procedure. [K4, S2]

CO3: Design and deploy appropriate Classification/ Clustering techniques for various problems with high dimensional data using modern tools. [K5, S2]

CO4: Apply the association rules for real life mining applications. [K4, S2]

CO5: Synthesize various mining techniques and work in teams to develop project on complex data objects. [K5, S3]

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

1.	Classify OLAP tools	CO1	[K_1]
2.	What is an apex cuboid? Give example.	CO1	[K_1]
3.	State why data preprocessing an important issue for data warehousing and data mining.	CO2	[K_1]
4.	What is data transformation? Give example.	CO2	[K_1]
5.	An airport security screening station wants to determine if passengers are criminals or not. To do this, the faces of passengers are scanned and kept in a database. Is this a classification or prediction task? Justify	CO3	[K_3]
6.	Discuss the significance of tree pruning in decision tree algorithms?	CO3	[K_1]
7.	Given two objects represented by the tuples (22, 1, 42, 10) and (20,0,36,8) .Compute the Manhattan distance between the two objects.	CO4	[K_3]
8.	Differentiate Agglomerative and Divisive hierarchical clustering method.	CO4	[K_2]
9.	Enumerate the components of time series data.	CO5	[K_1]
10.	Define Spatiotemporal data mining	CO5	[K_1]

Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11.	a)	What is data warehouse? Illustrate and discuss the data warehousing architecture.	8	CO1	[K_1]
	b)	Suppose that a data warehouse for Big University consists of the following four dimensions: student, course, semester, and instructor and two measures count and average grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the average grade measure stores the actual course grade of the student. At higher conceptual levels, average grade stores the average grade for the given combination. Draw a snowflake schema diagram for the data warehouse. Starting with the base cuboid [student, course, semester, instructor], what specific OLAP operations (e.g., roll-up from semester to year) should one perform in order to list the average grade of CS courses for each Big University student.	8	CO1	[K_3]
12.	a)	Paraphrase on data mining task primitives with the syntax of DMQL specification?	8	CO2	[K_2]
	b)	How data mining systems are classified? Discuss each classification with an example.	8	CO2	[K_1]
13.	a)	Discuss the Apriori algorithm for discovering frequent item sets. Apply the Apriori algorithm to the following data set.	16	CO4	[K_3]

101	Strawberry,Litchi,Oranges
102	Strawberry, Butter Fruit
103	Butter Fruit, Vanilla
104	Strawberry,Litchi,Oranges
105	Banana, Oranges
106	Banana
107	Banana, Butter Fruit
108	Strawberry,Litchi,Apple,Oranges
109	Apple, Vanilla
110	Strawberry, Litchi

The set of items is {Strawberry, Litchi, Apple, Oranges, Vanilla, Banana, Butter Fruit}.Use 0.3 as minimum support value. Generate all the association rules from the frequent item set mined.

14. a) Develop an algorithm for classification using Decision Tree building. 8 CO3 [K_2]

b) For the given data set, apply Naive – Bayes algorithm and predict the outcome for **Car = {Red, Domestic, SUV}** 8 CO3 [K_3]

Color	Type	Origin	Stolen
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Yellow	Sports	Domestic	No
Yellow	Sports	Imported	Yes
Yellow	SUV	Imported	No
Yellow	SUV	Imported	Yes
Yellow	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

15. a) Consider the Eight points {X_1,X_2,X_3,X_4,X_5,X_6,X_7,X_8} with the following coordinates at a two dimensional plane for clustering: 8 CO3 [K_3]

X_1-(2, 10), X_2-(2, 5), X_3-(8, 4), X_4-(5, 8), X_5-(7, 5), X_6-(6, 4), X_7-(1, 2), X_8-(4, 9)

Illustrate the use of k-means algorithm and Euclidean distance to cluster the above data points into 3 clusters.

b) Identify the clusters using Single Link Technique. Use Euclidean distance and draw the dendrogram. 8 CO3 [K_3]

	X	Y
P1	0.40	0.53
P2	0.22	0.38
P3	0.35	0.32
P4	0.26	0.19
P5	0.08	0.41
P6	0.45	0.30

16. a) What is Graph pattern mining? List the methods of graph pattern mining and explain them in detail with suitable examples. 16 CO5 [K_1]

****