



B.E/B.TECH DEGREE EXAMINATIONS: APRIL /MAY 2024

(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.

CO2: Identify necessity of data pre-processing and apply the appropriate procedure.

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

CO4: Apply the association rules for real life mining applications.

CO5: Synthesize various mining techniques and work in teams to develop project on complex data objects.

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

- | | | |
|--|-----|-------------------|
| 1. Enumerate different types of data that can be mined. | CO1 | [K ₁] |
| 2. List the key steps involved in KDD process. | CO1 | [K ₁] |
| 3. Differentiate between operational databases and data warehouses | CO2 | [K ₁] |
| 4. State why data preprocessing is an important issue for data warehousing and mining. | CO2 | [K ₂] |
| 5. Define support and confidence in the context of Frequent Pattern mining. | CO3 | [K ₁] |
| 6. Discuss the significance of tree pruning in decision tree algorithms. | CO3 | [K ₁] |
| 7. Recall the advantages of using grid-based methods for clustering large datasets. | CO4 | [K ₁] |
| 8. Differentiate between Agglomerative and Divisive hierarchical clustering method. | CO4 | [K ₁] |
| 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) Describe the process of data transformation and discretization, including normalization, standardization, and binning. Discuss the impact of data transformation on data mining results and provide examples of how different | 10 | CO1 | [K ₂] |
|--|----|-----|-------------------|

transformation techniques are used in practice.

- b) Describe the need of preprocessing of data with some real time examples. 6 CO2 [K₁]
12. Discuss the typical components of a data warehouse architecture and how they support various data warehouse functionalities, such as data storage, data access, and data transformation 16 CO2 [K₂]
13. a) Discuss the concept of conditional independence in the context of Naive Bayes. How does it simplify the computation of probabilities in the classification process? 8 CO3 [K₂]
- b) For the given data set, apply Naive – Bayes algorithm and predict the outcome for **Car = {Red, Domestic, SUV}** 8 CO3 [K₃]

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

14. a) Explain the working of DBSCAN algorithm. 8 CO4 [K₂]
- b) Identify the clusters using Single Link Technique. Use Euclidean distance and draw the dendrogram. 8 CO4 [K₃]

	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

15. a) Describe the Apriori algorithm for association rule mining. Outline the key steps involved and how it handles candidate generation and pruning 8 CO3 [K₁]
- b) Generate FP tree for the following transaction data set. 8 CO3 [K₃]

Transaction Id	Items
1	E,A,D,B
2	D,A,C,E,B
3	C,A,B,E
4	B,A,D
5	D
6	D,B
7	A,D,E
8	B,C

Minimum Support:30%

16. a) Define text mining and explain its role in extracting knowledge from unstructured textual data. 8 CO5 [K₂]
- b) Discuss the techniques and challenges involved in text preprocessing and sentiment analysis. 8 CO5 [K₁]
