



B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 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. Identify the steps involved in the process of KDD. How does it relate to data mining. | CO1 | [K ₁] |
| 2. Distinguish between data cleaning and noisy data. | CO1 | [K ₁] |
| 3. Discuss the advantages of dimensional modeling. | CO2 | [K ₂] |
| 4. List out the key features in business applications using OLAP. | CO2 | [K ₂] |
| 5. Formulate the principle frequent item set and closed item set | CO3 | [K ₄] |
| 6. How would you evaluate accuracy of a classifier? | CO3 | [K ₃] |
| 7. Identify what changes you would make to solve the problem in cluster analysis. | CO4 | [K ₂] |
| 8. Evaluate what information is used by outlier detection method. | CO4 | [K ₄] |
| 9. Name some of the data mining applications. | CO5 | [K ₁] |
| 10. List the organization of Text Data. | CO5 | [K ₁] |

Answer any FIVE Questions: -

PART B (5 x 16 = 80 Marks)

(Answer not more than 400 words)

- | | | | |
|------------------------------------------------------------------------------------------|---|-----|-------------------|
| 11. a) Infer and Explain the various classification of data mining systems with example. | 8 | CO1 | [K ₁] |
| b) Assume the various phases of data mining and explain the different steps | 8 | CO1 | [K ₁] |

involved in preprocessing with their significance before mining, Give an example for each process

- | | | | | | |
|-----|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|-----|-------------------|
| 12. | a) | List and discuss the basic features that are provided by reporting and query tools used for business analysis. | 8 | CO2 | [K ₃] |
| | b) | Tell about the multidimensional data model. Describe how it is used in data warehousing. | 8 | CO2 | [K ₃] |
| 13. | a) | Outline an example explain how support vector machines can be used for classification | 8 | CO3 | [K ₄] |
| | b) | What is the classification of association rules based on various criteria? | 8 | CO3 | [K ₄] |
| 14. | a) | Examine five points {X1, X2, X3, X4, X5} with the following coordinates as a two dimensional sample for clustering: X1 = (0,2.5); X2 = (0,0); X3= (1.5,0); X4 = (5,0); X5 = (5,2) Illustrate the K-means partitioning algorithm using the above data set. | 8 | CO4 | [K ₅] |
| | b) | What is outlier mining important? Briefly describe the different approaches behind statistical –based outlier detection, distance based outlier detection and deviation based outlier detection. | 8 | CO4 | [K ₄] |
| 15. | a) | Name some specific application oriented databases, explain with examples. | 8 | CO5 | [K ₂] |
| | b) | What is Time Series and Sequence Mining in the context of data mining? | 8 | CO5 | [K ₅] |
| 16. | a) | Find all frequent item sets for the given training set using Apriori and FP growth respectively. Compare the efficiency of the two mining processes. | 8 | CO3 | [K ₄] |
| | | TID ITEMS BOUGHT | | | |
| | | 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) | Discuss and generalize about constraint based association rule mining with examples and state how association mining to correlation analysis is dealt with. | 8 | CO3 | [K ₄] |
