



B.E/B.TECH DEGREE EXAMINATIONS: DECEMBER 2022

(Regulation 2018)

Fifth Semester

INFORMATION TECHNOLOGY

U18ITI5201: DATA MINING TECHNIQUES

COURSE OUTCOMES

- CO1:** Summarize the data pre - processing process
CO2: Explain the association rule Mining algorithm for correlation analysis
CO3: Apply decision tree algorithm for classification
CO4: Apply and analyze Bayesian networks algorithm for classification
CO5: Apply various clustering algorithms for different datasets

Time: Three Hours

Maximum Marks: 100

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

- | | | | |
|-----|---|-----|-------------------|
| 1. | If a nationalized bank wants to set up ATMs across a city, it would use the concept of Classification, Clustering or Association. Give your reason. | CO1 | [K ₃] |
| 2. | Define Data mining. Why is it considered to be a misnomer? | CO1 | [K ₂] |
| 3. | How is a concept hierarchy useful in solving real world problems? | CO2 | [K ₂] |
| 4. | How is FP growth algorithm advantageous over apriori algorithm? | CO2 | [K ₃] |
| 5. | Why decision trees are not the best option for large datasets? | CO3 | [K ₃] |
| 6. | Differentiate between lazy learner and eager learner. | CO3 | [K ₂] |
| 7. | In the context of Naïve Bayes classifier, what does the word “naïve” mean? | CO4 | [K ₃] |
| 8. | What is an outlier analysis? | CO5 | [K ₂] |
| 9. | What is good clustering? | CO5 | [K ₂] |
| 10. | Summarize the importance of web usage mining. | CO5 | [K ₂] |

Answer any FIVE Questions:-

PART B (5 x 16 = 80 Marks)

(Answer not more than 400 words)

- | | | | | | |
|-----|----|--|----|-----|-------------------|
| 11. | a) | Explain the seven stages of data mining process in detail | 10 | CO1 | [K ₂] |
| | b) | Give a brief account of how data mining can be used in any two applications. | 6 | CO1 | [K ₃] |

| 12. | Transaction ID | List of items in the transaction | | CO2 | [K ₃] |
|-----|----------------|----------------------------------|--|-----|-------------------|
| | T1 | B , A , T | | | |
| | T2 | A , C | | | |
| | T3 | A , S | | | |
| | T4 | B , A , C | | | |
| | T5 | B , S | | | |
| | T6 | A , S | | | |
| | T7 | B , S | | | |
| | T8 | B , A , S , T | | | |
| | T9 | B , A , S | | | |

Find the frequent itemsets and generate strong association rules using apriori and FP growth algorithm. Minimum support =2 and minimum confidence = 70%. Compare and comment on the performance of both algorithms.

- | | | | | | |
|-----|----|---|----|-----|-------------------|
| 13. | a) | Explain in detail the working of decision tree algorithm. | 10 | CO3 | [K ₂] |
| | b) | Bring out the important aspects of prediction. | 6 | CO3 | [K ₂] |
| 14. | | Compare and relate in detail between spatial data mining and time series mining. | | CO4 | [K ₃] |
| 15. | | Explain in detail about web content mining, web usage mining and web structure mining. Comment on applications pertaining to each of the above. | | CO4 | [K ₂] |
| 16. | | List and bring out the key aspects of major clustering approaches. | | CO5 | [K ₃] |
