KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2024

(Regulation 2018)

Fourth Semester

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

U18AII4203**:** Data Mining and Modeling

COURSE OUTCOMES

CO1: Understand about data mining basics, issues and the working principle of classification technique.

CO2: Explain the basic concepts of Association Rule Mining and evaluate the working of various Association Rule Mining algorithms.

CO3: Implement classification and prediction techniques.

CO4: Analyze the working of different clustering algorithms.

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

Answer all the Questions: -
PART A (10 x 2 = 20 Marks)
(Answer not more than 40 words)

1. Define data mining and its main goals.	CO1	[K_1]
2. List three common data mining techniques.	CO1	[K_1]
3. Differentiate between OLAP and OLTP.	CO1	[K_2]
4. Name two methods used for mining frequent patterns.	CO2	[K_1]
5. Differentiate between strong and weak associations.	CO3	[K_2]
6. What is an association rule in data mining?	CO2	[K_1]
7. List one application of associative classification.	CO3	[K_1]
8. Identify the significance of Bayesian classification in data mining.	CO3	[K_2]
9. Define cluster analysis in the context of data mining.	CO4	[K_1]
10. Explain the concept of density-based clustering.	CO4	[K_2]

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

11. a) You are tasked with cleaning a dataset of customer transaction records that contain many null values and inconsistencies. Describe the data cleaning steps you would apply and explain how each step improves data quality.	8	CO1	[K_2]

b) A startup wants to understand customer purchase patterns over time. Explain how you would transform raw transaction data using data transformation and reduction techniques to help with their analysis.   8   CO1   [K₃]

12. a) You have a dataset with transaction details from a supermarket. Describe how you would apply association rule mining to identify common product associations and suggest marketing strategies based on the results.   8   CO2   [K₃]

b) A retail company wants to optimize shelf space by understanding associations among purchased products. Explain how constraint-based association mining can help the company tailor the product placement.   8   CO2   [K₂]

13. a) You are given a large dataset with labeled customer reviews and need to predict the sentiment (positive or negative) of new reviews. Describe how you would use rule-based classification for the given scenario, and explain the steps involved.   8   CO3   [K₃]

b) Given a dataset on customer behaviours, illustrate how lazy learners could be used to predict future purchasing decisions. Discuss the potential limitations of this approach.   8   CO3   [K₂]

14. a) A company wants to categorize its customers based on purchasing behavior to tailor its marketing. Describe how you would use K-means clustering for this analysis and the steps involved in optimizing the clusters.   8   CO4   [K₃]

b) You are tasked with identifying outliers in a credit card transaction dataset to detect potential fraud. Explain the outlier analysis process and discuss how clustering can be applied for this task.   8   CO4   [K₂]

15. a) A healthcare organization wants to categorize patients based on their medical history and symptoms to improve personalized treatment plans. Describe how you would apply hierarchical clustering for this purpose and discuss the advantages and challenges of using hierarchical methods in healthcare data.   8   CO1   [K₃]

b) An e-commerce company wants to implement correlation analysis on customer purchase history to boost product recommendations. Describe the steps you would follow, and the insights correlation analysis might reveal.   8   CO2   [K₂]

16. a) A bank wants to classify loan applicants based on their risk levels. Explain how decision tree induction can help and describe the steps you would take to implement this using RapidMiner or Weka.   8   CO3   [K$_L$]

b) A telecom company wants to analyze customer location data to optimize network coverage. Describe how grid-based clustering could be used for this analysis, the steps involved, and how it might differ from other clustering approaches in terms of efficiency and data handling.   8   CO4   [K$_L$]

****