KUMARAGURU
college of technology
character is life

M.E DEGREE EXAMINATIONS: JUNE 2017

(Regulation 2015)

Second Semester

COMPUTER SCIENCE AND ENGINEERING

P15CST201: Data Warehousing and Data Mining

COURSE OUTCOMES

CO1: Explain the concepts of Data Warehousing architecture and implementation.

CO2: Apply the association rules for mining applications.

CO3: Discuss on appropriate Classification/ Clustering techniques for various problems with high dimensional data.

CO4: Discover the knowledge imbibed in the high dimensional system.

CO5: Illustrate various data mining techniques on complex data objects.

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

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. Assertion (A): Aggregation is the technique wherein, as the name suggest, summary or   CO1   [K2]

    aggregation operation are applied to the data

    Reason (R): The data basically deals with the concept wherein low level or primitive

    data is replaced by high level concept.

    a) Both A and R are individually true and   b) Both A and R are Individually true but
        R is the correct explanation of A   R is not the correct explanation of A

    c) A is true but R is false   d) A is false but R is true

2. A _____ for an attribute A partitions the data distribution of A into   CO1   [K1]

    disjoint subsets or buckets.

    a) Sampling   b) Clustering

    c) Histogram   d) Classification

3. _____ is a pragmatic collection of related facts, but does not have to be   CO1   [K1]

    exhaustive or exclusive

    a) Data mart   b) Knowledge Database

    c) Data set   d) Database

4. Match the following   CO3   [K_2]

List I	List II
A. OLAP	i. Registration
B. OLTP	ii. Data Warehouse
C. Decision Tree	iii. RDBMS
D. Neural Network	iv. Classification

	A	B	C	D
a)	ii	iii	i	iv
b)	ii	iii	iv	i
c)	iii	ii	i	iv
d)	iii	ii	iv	i

5. Assertion (A): Data warehouse is constructed by integration of data from CO4   [K_2] heterogeneous sources such as relational databases, flat files etc.

Reason (R): Integration enhances the effective analysis of data.

a) Both A and R are Individually true and R is the correct explanation of A   b) Both A and R are Individually true but R is not the correct explanation of A

c) A is true but R is false   d) A is false but R is true

6. FP – growth which adopts a   CO3   [K_1]

a) Divide and continue strategy   b) Divide and conquer strategy

c) Dimensionality reduction   d) Generalization

7. Which of the following statements are false?   CO1   [K_2]

    1. Joining is the process of partitioning data according to predefined criteria.

    2. Data in the data warehouse are loaded and refreshed from operational systems.

    3. Successful data warehousing requires that a formal program in total quality management (TQM) be implemented.

    4. Periodic data are data that are physically altered once added to the store.

a) 1,3   b) 1,4

c) 1,2   d) 2,3

8. _____ is a special case of k-fold cross- validation where k is set to the number CO1   [K_1] of initial tuples.

a) Bootstrap   b) Bagging

c) Leave-one-out   d) Boosting

9. Document clustering analysis is one of the _____ approaches   CO3   [K₁]

    a)   Text Mining   b)   Multimedia data Mining

    c)   WWW Mining   d)   Spatial data Mining

10. Sequence the steps in Process of Query-Driven Approach   CO2   [K₂]

    1. When a query is issued to a client side, a metadata dictionary translates the query into an appropriate from for individual heterogeneous sites involved.

    2. The results from heterogeneous sites are integrated into a global answer set.

    3. Now these queries are mapped and sent to the local query processor

    a)   2-3-1   b)   1-3-2

    c)   3-2-1   d)   3-1-2

PART B (10 x 2 = 20 Marks)

11. Define the term Knowledge Discovery in Databases (KDD).   CO1   [K₁]

12. Differentiate between quantitative and qualitative attributes   CO1   [K₁]

13. What do you mean by discretization?   CO2   [K₁]

14. Differentiate clustering and classification techniques.   CO3   [K₁]

15. Write the two measures of association rule.   CO2   [K₂]

16. Define conditional pattern base.   CO2   [K₁]

17. List the factors that affecting the complexity of Apriori algorithm.   CO3   [K₂]

18. What is worst case time complexity of K-Means cluster algorithm with m points and n attributes?   CO3   [K₃]

19. Write short notes on text mining.   CO5   [K₁]

20. What kind of association can be mined from multimedia data?   CO5   [K₂]

PART C (6 x 5 = 30 Marks)

21. Discuss the data warehouse architecture with suitable diagram.   CO1   [K₂]

22. Differentiate OLTP with OLAP systems.   CO1   [K₂]

23. Discuss the working of Partitioning Around Medoids algorithm.   CO3   [K₂]

24. Explain how data mining system can be integrated with database/ data warehouse system?   CO4   [K₂]

25. Discuss in detail the application of data mining for financial data analysis.   CO5   [K₂]

26. Explain in detail about data mining tools.   CO5   [K₂]

Answer any FOUR Questions

PART D (4 x 10 = 40 Marks)

27. Explain data preprocessing techniques in detail with suitable examples.   CO2   [K₂]

28. Explain the KDD process with a neat diagram. Explain each step of KDD process with their significance.   CO1   [K₂]

29. Suppose that the university course database for ABC-University includes the following attributes describing students name, address, status (under graduate or post graduate), department and CGPA (Cumulative Grade Point Average).   CO4   [K₃]

    i)   Propose a concept hierarchy for the attributes address, status, department and CGPA.

    ii)   For each concept hierarchies that have proposed above, what type of concept hierarchy is it?

    iii)   Define each hierarchy using DMQL syntax.

    iv)   Write a DMQL- query to find the characteristics of students who have an excellent CGPA.

30. Briefly outline the major steps of decision tree classification.   CO3   [K₂]

31. Summarize the role of data mining in Web.   CO5   [K₂]

****