SHRI ANGALAMMAN COLLEGE OF ENGINEERING & TECHNOLOGY
(An ISO 9001:2008 Certified Institution)
SIRUGANOOR,TRICHY-621105.

FS 81504

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Year / Semester: IV/VII

CS1011-DATA WAREHOUSING AND DATA MINING

UNIT-I BASICS OF DATA WAREHOUSING

PART-A

1. Define data warehouse.
2. What are the characteristics of data warehouse?
3. Define data mart.
4. What are the difference between OLTP and OLAP?
5. Define data model.
6. Define data cube.
7. What are Fact table and Dimension table?
8. What is Facts?
9. What are the different forms of schema?
10. What is DMQL?
11. How the measures are categorized?
12. What is a concept hierarchy?
13. Name the different types of OLAP operations?
14. How roll-up differ from drill-down?
15. Define starnet model.
16. What are the four different views in data warehouse design?
17. What are the steps to be followed in data warehouse design process?
18. What are different kinds of OLAP servers exist?
19. What are the optimization techniques used In ROLAP cube computation?
20. Define bitmap indexing and join indexing.
21. What are metadata?
22. What are the functions available in back-end tools and utilities of data warehouse?
23. How the tools for data warehousing can be categorized?

PART-B

1. Explain in detail about the multidimensional data model with example.
2. Explain in detail the different types of Schema by representing in their own structure with example.
3. Elaborately explain the three tier data warehouse architecture with neat diagram.
4. How do you map OLAP to OLAM and explain the architecture of OLAM in detail with neat diagram.
5. What is OLAP? List the various OLAP operations and explain the operation in detail with diagram.
6. Elaborate the steps to be consider for the design and construction of data warehouse.
7. Examine the different methods for the efficient implementation of data warehouse systems.
8. Explain data ware house architecture and operational data stores with neat diagram.

UNIT-II DATA PREPROCESSING, LANGUAGE, ARCHITECTURES, CONCEPT DESCRIPTION

PART-A

1. What is the need of preprocessing the data?
2. What are the various forms of data preprocessing?
3. Define data cleaning?
4. What are the different methods available to fill in the missing values?
5. What is noise?
6. Write the different types of data smoothing techniques?
7. Define data integration and transformation?
8. Name the three methods used for data normalization?
9. What is data reduction?
10. What are the different strategies used for data reduction?
11. What are the two types of compression used?
12. What are the two popular methods of lossy data compression?
13. What are the techniques available for heuristic methods of attribute selection?
14. What are linear regression and multiple regression?
15. Expand SRSWOR and SRSWR.
16. What is discretization?
17. What are the methods used for numeric concept hierarchy generation?
18. What is 3-4-5 rule?
19. Name the several methods used for the generation of concept hierarchies for categorical data?
20. List the five data mining primitives used in data mining system?
21. What are the various forms available to display the discovered patterns?

22. What are functional components used in data mining GUI?
23. What are different coupling schemes used to integrate data mining with database/data ware house?
24. What is concept description?
25. What are differences between concept description in large databases and OLAP?
26. How the data mining classified into two categories?
27. Define data generalization?
28. What are the two approaches used foe efficient and flexible generalization of large data sets?
29. What is class comparision?
30. How is class comparision performed?
31. Define dispersion/variance and name the common measures of data dispersion?

PART-B

1. What is cleaning? Explain in detail the various methods used for data cleaning?
2. Define data integration and discuss the issues to consider during data integration?
3. Explain the two popular and effective methods of lossy data compression?
4. Explain in detail the discretization and concept hierarchy generation for numeric and categorical data?
5. Discuss the five primitives in specifying a data mining task.
6. Discuss the importance of establishing a standardized data mining query language. What are the potential benefits and challenges involved in such a task?
7. Elaborately explain the various schemes of coupling?
8. Elaborately explain the attribute oriented induction approach with necessary examples and tables
9. Write the algorithm of attribute oriented induction.
10. Explain the various descriptive statistical measures in large databases.

UNIT-III ASSOCIATION RULES

PART-A

1. What is association analysis?
2. What is association rule mining? Give an example.
3. Differentiate itemset and K-itemset?
4. What is strong association rule?
5. What is minimum support?
6. Define frequent itemset?
7. How is association rules mined from large database? Or what are the two steps of association rule Mining?
8. How the association rules classified?
9. What is Boolean association rule?
10. What is quantitative association rule?

11. Compare single dimensional association rule with multidimensional association rule?
12. Define maxpattern?
13. What is frequent closed itemset?
14. What are the interestingness measures of association rule mining?
15. What is need of using apriori property?
16. What are the two steps of apriori property?
17. What is FP-growth?
18. Define iceberg query?
19. Define multilevel association rules?
20. What are the different approaches of mining multilevel association rules?

PART-B

1. Explain in detail the apriori algorithm with an example?
2. Elaborately explain the variations that have been proposed to improve the efficiency of apriori?
3. With neat diagram and example explain the method that mines the complete set of frequent itemsets without candidate generation?
4. Write the algorithm for FP-growth?
5. With an example discuss multilevel association rule.
6. Discuss the single dimensional Boolean association rule mining for transaction database?

UNIT-IV CLASSIFICATION AND CLUSTERING

PART-A

1. What are the factors to be considered, when comparing classification methods?
2. How is prediction different from classification?
3. What is a decision tree?
4. What is information gain measure?
5. What are the two common approaches of tree pruning?
6. How can we extract classification rules form decision tree?
7. What are the three main problems faced by decision tree induction method?
8. What are Bayesian classifiers?
9. Define belief networks or Bayesian networks.
10. Expand: ARCS, CAEP.
11. Write the three methods used for association rule based mining?
12. Define prediction.
13. How can we convert the nonlinear model to linear model give a example?
14. What is classifier accuracy?
15. Draw the holdout method for estimating classifier accuracy?
16. Define clustering?

17. Differentiate conceptual clustering and conventional clustering?
18. Mention the various types of data available in data mining?
19. How can the major clustering methods are classified into categories?
20. What are the different types of algorithm available for partitioning methods?
21. How can we improve the quality and scalability of CLARA?
22. What is an outlier?
23. Define outlier mining?
24. List the applications of outlier mining?
25. What are the three approaches used in outlier detection?

PART-B

1. Briefly discuss the major steps involved in the induction of decision trees using the ID3 algorithm?
2. Elaborate the Bayesian classification with Bayes theorem in detail?
3. Explain in detail the other methods used for classification?
4. Discuss in detail the different types of data available in cluster analysis?
5. How clustering differ from classification? Briefly explain the requirements of clustering in data mining?
6. What is partitioning method? List the various partitioning methods and elaborately explain the methods with algorithms?
7. What is an outlier? Elaborately explain the computer based methods for outlier detection?

UNIT-V RECENT TRENDS

PART-A

1. List the spatial operations?
2. Define spatial merge?
3. What is class composition hierarchy?
4. Define plan, plan database & plan mining
5. Define spatial database?
6. Define spatial data mining?
7. Define spatial data warehouse?
8. What are the three types of dimension available in a spatial data cube?
9. What are the two types of measures in a spatial data cube?
10. What is multimedia database?
11. What are the two families of multimedia indexing and retrieval system?
12. What are the two kinds of queries in content-based retrieval system?
13. What is a time-series database?
14. What is a sequence database?
15. What are the four major components or movements used to characterize time-series data?
16. What is similarity search?

17. What is sequential pattern mining?
18. How the problem of mining periodic patterns can be partitioned into various categories?
19. What is information retrieval? Name the methods used for that?
20. What are the two basic measures for assessing the quality of text retrieval?
21. Name the difficulties faced by keyword based retrieval system?
22. Define web mining?
23. How the web mining tasks are classified?
24. List the some applications of data mining?
25. Name some of the trends in data mining?
26. What are the stages of life cycle technology adoption?

PART-B

1. Explain in detail how to mine the Text databases?
2. Describe the applications and trends in data mining in detail.
3. Explain the Visual and Audio data mining?
4. Elaborately explain the social impacts of data mining?
5. Explain the Scientific and Statistical data mining?
6. Explain in detail the mining of spatial databases?
7. Discuss in detail the time series and sequence data?