SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF INFORMATION TECHNOLOGY

(COMMON TO COMPUTER SCIENCE AND ENGINEERING)

QUESTION BANK

V SEMESTER

1908507-DATA WAREHOUSING AND DATA MINING

Regulation – 2019

Academic Year 2022– 2023(Odd Semester)

Prepared by

Ms.S. Shenbagavadivu , Assistant Professor(Sr.G)/IT

Ms.S.Shanthi, Assistant Professor (Sr.G)/CSE

Ms.S.Benila , Assistant Professor(Sr.G)/CSE

Ms.V.Prema, Assistant Professor (Sr.G)/CSE

SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar, Kattankulathur – 603 203.

DEPARTMENT OF INFORMATION TECHNOLOGY

QUESTION BANK

SUBJECT : 1908507—DATA WAREHOUSING AND DATAMINING

SEM / YEAR: V -Third year

UNIT -1- DATA WAREHOUSING			
Data warehousing Components –Building a Data warehouse –- Mapping the Data Warehouse to a Multiprocessor Architecture – DBMS Schemas for Decision Support – Data Extraction, Cleanup, and Transformation Tools –Metadata.			
PART A			
Q.NO	**QUESTIONS**	**COMPETENCE**	**LEVEL**
1.	How is data warehouse different from a database? **Identify** the similarity.	Remember	BTL-1
2.	**Differentiate** metadata and data mart.	Understand	BTL-2
3.	**Analyze** why one of the biggest challenges when designing a Datawarehouse is the data placement and distribution strategy.	Analyze	BTL-4
4.	How would you **evaluate** the goals of data mining?	Evaluate	BTL-5
5.	**List** the two ways the parallel execution of the tasks within SQL statements can be done.	Remember	BTL-1
6.	What elements would you use to **relate** the design of data warehouse?	Apply	BTL-3
7.	**Reframe** about Data mart.	Evaluate	BTL-5
8.	**Define** star schema.	Remember	BTL-1
9.	What is Data warehousing? **Explain** the benefits of Data warehousing.	Evaluate	BTL-5
10.	Why data transformation is essential in the process of Knowledge discovery? **Describe** it.	Remember	BTL-1
11.	**Describe** the alternate technologies used to improve the performance in data warehouse environment	Understand	BTL-2
12.	**Distinguish** STAR join and STAR index.	Understand	BTL-2
13.	**Analyze** the types of data mart.	Apply	BTL-3
14.	**Formulate** what is data discretization.	Create	BTL-6
15.	**Point out** the major differences between the star schema and the snowflake schema.	Analyze	BTL-4

16.	**Point out** the features of Metadata repository in data warehousing.	Analyze	BTL-4
17.	**Define** Metadata repository.	Remember	BTL-1
18.	**Discuss** metadata with an example.	Understand	BTL-2
19.	**Illustrate** the benefits of metadata repository.	Apply	BTL-3
20.	**Design** the datawarehouse architecture.	Create	BTL-6
21.	**List** the characteristic of Data warehousing.	Remember	BTL-1
22.	**Differentiate** the Row wise and Column wise storage.	Understand	BTL-2
23.	**Illustrate** the two goals needed in searching better performance and scalability.	Apply	BTL-3
24.	**Point out** the features of Parallel RDBMS.	Analyze	BTL-4
	PART B		
1.	What is datawarehouse? **Give** the Steps for design and construction of Data warehouses and explain with three tier architecture diagram.(13)	Understand	BTL-2
2.	Diagrammatically **illustrate** and discuss the following preprocessing techniques: (i) Data cleaning (3) (ii) Data Integration (3) (iii) Data transformation (3) (iv) Data reduction (4)	Apply	BTL-3
3.	(i).Draw the data warehouse architecture and **explain** its components.(7) (ii).**Explain** the different types of OLAP tools. (6)	Analyze	BTL-4
4.	**(i).Describe** in detail about Mapping the Data warehouse to a multiprocessor architecture. (8) (ii).**Describe** in detail on datawarehouse Metadata. (5)	Remember	BTL-1
5.	(i).**Explain** the steps in building a datawarehouse. (8) (ii).**Analyze** the information needed to support DBMS schemas for Decision support. (5)	Analyze	BTL-4
6.	(i).**Discuss** in detail about access tools types? (6) ii**) Describe** the overall architecture of data warehouse? (7)	understand	BTL-2
7.	**(i) Discuss** the different types of data repositories on which mining can be performed. (7) **(ii) Differentiate t**angible and intangible benefits of data warehouse. (6)	Understand	BTL-2
8.	i) **Describe** in detail about data extraction. (6) ii) **Describe** in detail about transformation tools. (7)	Remember	BTL-1
9.	(i).Suppose that a data warehouse consists of four dimensions customer, product, salesperson and sales time, and the three measure sales Amount (in rupees), VAT (in rupees) and payment type(in rupees). Draw the different classes of schemas that are popularly used for modeling data warehouses and **explain** it. (7) (ii).How would you **explain** Metadata implementation with examples? (6)	Evaluate	BTL-5

10.	**Describe** in detail about i) Bitmapped indexing. (6) ii)STAR join and index. (7)	Remember	BTL-1
11.	(i).What is data Preprocessing**? Explain** the various data preprocessing techniques. (7) (ii).**Explain** the basic methods for data cleaning. (6)	Analyze	BTL-4
12.	**Explain** the diagrammatic representation for relationship between operational data, a data warehouse and data marts. (13)	Evaluate	BTL-5
13.	**i)Demonstrate** in detail about Data marts. (6) **ii)Demonstrate** data warehouse administration and management. (7)	Apply	BTL-3
14.	(i) **Generalize** the potential performance problems with star schema. (6) (ii) **Design** and discuss about the star and snowflake schema models of a Data warehouse. (7)	Create	BTL-6
15.	**Describe** the various issues to be considered when designing and implementing a data warehousing environment. (13)	Remember	BTL-1
16.	**Discuss** the following: (i) Column Local Storage. (7) (ii) Complex Data type. (6)	Understand	BTL-2
17.	What is Meta data? **Illustrate** the various classification of Meta data with example and explain the same.(13)	Apply	BTL-3
	PART C		
1.	**Explain** mapping data warehouse with multiprocessor architecture with the concept of parallelism and data partitioning. (15)	Evaluate	BTL-5
2.	**Design** a star-schema , snow-flake schema and fact- constellation schema for the data warehouse that consists of the following four dimensions (Time, Item, Branch And Location) . Include the appropriate measures required for the schema.(15)	Create	BTL-6
3.	**i) Generalize** why we need data preprocessing step in data warehousing.(8) **ii)Explain t**he various methods of data cleaning and data reduction Technique. (7)	Create	BTL-6
4.	i)**Compare** the similarities and differences between the database and datawarehouse. (8) ii) **Explain** data visualization. How it helps datawarehousing.(7)	Evaluate	BTL-5
5.	**Explain** Star Schema model for e-wallet and Explain the same with diagram.(15)	Evaluate	BTL-5
	UNIT 2- BUSINESS ANALYSIS		
	Reporting and Query tools and Applications – Tool Categories – The Need for Applications – Cognos Impromptu – Online Analytical Processing (OLAP) – Need – Multidimensional Data Model – OLAP Guidelines – Multidimensional versus Multi-relational OLAP – Categories of Tools – OLAP Tools and the Internet.		
	PART A		
1.	**List** the categories of tools in business analysis.	Remember	BTL-1

2.	Illustrate the basic functions used in reporting tools?	Analyze	BTL-4
3.	**Explain** the tool for corporate data?	Analyze	BTL-4
4.	**Define** the various access types to the data stored in a data warehouse.	Remember	BTL-1
5.	What would **demonstrate** the result if an organization uses a familiar approach to build a query and reporting environment for the data warehouse?	Apply	BTL-3
6.	What is a reporting tool? **Give** examples for managed query tools.	Understand	BTL-2
7.	**Explain** the need of Cognos Impromptu.	Evaluate	BTL-5
8.	**Compare** OLTP and OLAP system.	Analyze	BTL-4
9.	**Illustrate** about reporting tools.	Apply	BTL-3
10.	**Discuss** the need for OLAP.	Understand	BTL-2
11.	How would you **classify** the ideas of multidimensional data model with multi-relational OLAP?	Apply	BTL-3
12.	Which parameter and **design** choices determine the MOLAP?	Create	BTL-6
13.	**Summarize** the distinct features of OLTP with OLAP.	Understand	BTL-2
14.	**List** any four tools for performing OLAP.	Remember	BTL-1
15.	**Explain** about query tools with example.	Evaluate	BTL-5
16.	**Generalize** the function of OLAP tools in the internet.	Create	BTL-6
17.	**Describe** MOLAP and ROLAP.	Understand	BTL-2
18.	How would you **explain** the key features of OLAP components?	Evaluate	BTL-5
19.	**Describe** about the Internet tools.	Remember	BTL-1
20.	**List** out the key features in business applications using OLAP.	Remember	BTL-1
21.	**Define** data cube?	Remember	BTL-1
22.	**Discuss** the advantages of dimensional modelling?	Understand	BTL-2
23.	**Illustrate** the activities and processes that governors can control.	Apply	BTL-3
24.	**Draw** the MOLAP architecture.	Analyze	BTL-4
	PART B		
1.	What is business analysis? **List** and discuss the basic features that are provided by reporting and query tools used for business analysis. (13)	Remember	BTL-1
2.	Write the **Value features** of Cognos Impromptu business analysis tool. (13)	Evaluate	BTL-5

3.	(i).**Compare** OLTP and OLAP systems. (7) (ii).**Summarize** the various OLAP operations in the Multidimensional Data Model. (6)	Evaluate	BTL-5
4.	i)How would you **describe** in detail about reporting query classification? (7) ii) **Describe** in detail about application of OLAP tools. (6)	Understand	BTL-2
5.	**Describe** in detail about tool categories in business analysis. (13)	Remember	BTL-1
6.	**Examine** with an example the reporting and query tools in business analysis. (13).	Apply	BTL-3
7.	**(i). Describe** the need for OLAP. (6) **(ii).Discuss** in detail about the OLAP guidelines. (7)	Understand	BTL-2
8.	(i). Diagrammatically illustrate and **describe** the architecture of MOLAP and ROLAP. (7) (ii). **Identify** the Major difference between MOLAP and ROLAP. (6)	Remember	BTL-1
9.	**(i).** Write about the multidimensional data model. **Describe** how it is used in data warehousing. (7) **(ii). Describe** in detail about tool categories in datawarehouse. (6)	Remember	BTL-1
10.	**i). Discuss** in detail about the OLAP tools. (6) ii) **Discuss** in detail about application in the internet (7)	Understand	BTL-2
11.	**Explain** in detail about the characteristics of OLAP cube. (13)	Analyze	BTL-4
12.	**Examine** the relevant examples discuss multidimensional online analytical processing and multi relational online analytical processing. (13)	Apply	BTL-3
13.	**Generalize** the topic on (i). Operations in Multidimensional Data Model. (7) (ii). OLAP Tools and the Internet. (6)	Create	BTL-6
14.	**Explain** in detail about the features of OLTP and OLAP. (13)	Analyze	BTL-4
15.	**Discuss** why organizations use a familiar application development approach to build a query and reporting environment.(13)	Understand	BTL-2
16.	**Examine** how a client server applications can be constructed using Power Builder Painters.(13)	Apply	BTL-3
17.	**Explain** how Forte application provide facility to develop and partition application in three-tier architecture. (13)	Analyze	BTL-4
	PART C		
1.	**Illustrate and criticize** the various multidimensional data modeling to design a data warehouse . (15)	Evaluate	BTL-5
2.	**Summarize** different tool categories in data warehouse business analysis. (15)	Create	BTL-6

3.	1. **Design** a multidimensional cube with your own example. (5) 2. Suppose that a data warehouse consists of the four dimensions date, spectator, location, and game, and the two measures count and charge, where charge is the fare that a spectator pays when watching a game on a given date spectatorsmay be students, adults, or seniors, with each category having its own charge rate. **i.** **Draw** a star schema diagram for the data warehouse. (5) **ii.** Starting with a base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at GM_Place in 2000? (5).	Create	BTL-6
4.	i) Depict the 3 tier data warehousing architecture and **explain** its features in detail. (8) ii).**Explain** the different types of OLAP servers. (7)	Evaluate	BTL-5
5.	**Investigate** in detail about Cactus and FOCUS Fusion products from Information Builders. (15)	Create	BTL-6
	UNIT 3- DATA MINING		
	Introduction – Data – Types of Data – Data Mining Functionalities – Interestingness of Patterns – Classification of Data Mining Systems – Data Mining Task Primitives – Integration of a Data Mining System with a Data Warehouse – Issues –Data Preprocessing.		
	PART A		
1.	**Define** Data mining. **List** out the steps in data mining?	Remember	BTL-1
2.	**Identify the** steps involved in the process of KDD. How does it relate to data mining?.	Understand	BTL-2
3.	**List** the ways in which interesting patterns should be mined.	Remember	BTL-1
4.	**Compare** drill down with roll up approach.	Analyze	BTL-4
5.	**Describe** the other kinds of data in data mining.	Understand	BTL-2
6.	How would you **illustrate** Handling outlier or incomplete data?	Apply	BTL-3
7.	**Analyze** data characterization related to data discrimination.	Analyze	BTL-4
8.	**Define** association and correlations.	Remember	BTL-1
9.	**List** the five primitives for specification a data mining task.	Remember	BTL-1
10.	**Evaluate** the major tasks of data preprocessing.	Evaluate	BTL-5
11.	Are all patterns generated are interesting and useful? **Give** reasons to justify.	Understand	BTL-2
12.	**Classify** different types of reductions.	Apply	BTL-3
13.	**Distinguish** between data cleaning and noisy data.	Understand	BTL-2
14.	**Explain** the principle elements of missing values in data cleaning.	Analyze	BTL-4

15.	**Discuss** the roles of noisy data in data preprocessing.	Understand	BTL-2
16.	**Consider** that the minimum and maximum values for the attribute "salary" are 12,000 and 98,000 respectively and the mapping range of salary is [0.0 , 1.0]. Find the transformation for the salary 73,600 using min-max normalization.	Create	BTL-6
17.	**Generalize how** the attribute selection set is important in data reduction.	Apply	BTL-6
18.	Consider the following set of data X = {15,27,62,35,39,50,44,44,22,98} Do preprocessing using smoothing by bin means and bin boundary to smooth the data, using a bin of depth 3.**Evaluate** it.	Evaluate	BTL-5
19.	**Demonstrate why** we need data transformation. Mention the ways by which data can be transformed.	Apply	BTL-3
20.	**Define** an efficient procedure for cleaning the noisy data.	Remember	BTL-1
21.	**Illustrate** with example pattern and pattern evaluation.	Apply	BTL-3
22.	**List out** the data mining functionalities.	Remember	BTL-1
23.	**Explain** the various types of data?	Analyze	BTL-4
24.	**Explain** interestingness of pattern	Evaluate	BTL-5
	PART B		
1.	**i) Demonstrate** in detail about data mining steps in the process of knowledge discovery? (8) **ii)** List the application area of data mining? (5)	Apply	BTL-3
2.	**Explain** in detail about the Evolution of Database Technology. (13)	Evaluate	BTL-5
3.	(i).What is data? How different type of data and attributes can be **designed**? (6) (ii). **Design** and discuss in detail about Primitives for specifying a data mining task (7)	Create	BTL-6
4.	(i).**Discuss** whether or not each of the following activities is a data mining task. (5) 1. Credit card fraud detection using transaction records. 2. Dividing the customers of a company according to their gender. 3. Computing the total sales of a company 4. Predicting the future stock price of a company using historical records. 5. Monitoring seismic waves for earthquake activities. (ii). **Discuss** on descriptive and predictive data mining tasks with illustrations.(8)	Understand	BTL-2
5.	**(i)State and Explain** the various classification of data mining systems with example. (7) **(ii)Explain** the various data mining functionalities in detail. (6)	Analyze	BTL-4

6.	Suppose that the data for analysis include the attributed age. The age values for the data tuples are 13,15,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35, 36,40,45,46,52,70. (i).use smoothing by bin means to smooth the above data using a bin depth of 3. **Illustrate** your steps. (6) (ii)**Classify** the various methods for data smoothing. (7)	Apply	BTL-3
7.	Sketch the various phases of data mining and explain the **different** steps involved in preprocessing with their significance before mining, Give an example for each process. (13)	Understand	BTL-2
8.	**Describe** in detail about the issues of data mining. (13)	Remember	BTL-1
9.	**Describe** in detail about data reduction in data preprocessing (13)	Remember	BTL-1
10.	**Describe** in detail about various data transformation techniques. (13)	Remember	BTL-1
11.	**List** and explain the primitives for specifying a data mining task.(13)	Remember	BTL-1
12.	**(i).** How will you handle missing value in a dataset before mining process? **(4)** **(ii).Give** the architecture of a typical data mining system. (9)	Understand	BTL-2
13.	**i) Explain** how integration is done with a database or data warehouse system.(5) **ii)Consider** the following data for the attribute AGE:4,8,21,5,21,24,34,28,25. Perform smoothing by bin means and bin boundaries using a bin depth of 3.(8)	Analyze	BTL-4
14.	**Analyze** Using Equi-depth binning method, partition the data given below into 4 bins and perform smoothing according to the following methods.(8) 1. Smoothing by bin means 2. Smoothing by bin median 3. Smoothing by bin boundaries 24,25,26,27,28,56,67,70,70,75,78,89,89,90,91,94,95,96,100,102,103,107,109,112. (ii).What motivated data mining? Why is it important? (5)	Analyze	BTL-4
16.	**Explain** various methods of data cleaning /Data Preprocessing in detail.(13)	Understand	BTL-2
17.	**Illustrate** about the data mining task Primitives.(13)	Apply	BTL-3
18.	**Define** the following about data summarization: (i)Measuring the central tendency(6) (ii)Measuring the dispersion of data(7)	Evaluate	BTL-5
	PART C		
1.	**Describe** and **judge** the Major issues in data warehousing and data mining. (15)	Evaluate	BTL-5
2.	i) What is interestingness of a pattern? (5) ii) **Summarize** the integration of data mining system with a data warehouse.(10)	Create	BTL-6
3.	**Reframe** the major data preprocessing techniques and **explain** in detail withexamples.(15)	Evaluate	BTL-5
4.	**i)Generalize** in detail how data mining system are classified (5) **ii)**Discuss each classification with an example. (10)	Create	BTL-6
5.	**Defend** the detail data mining functionalities and the different kinds of patterns can be mined.(15)	Evaluate	BTL-5

UNIT-4- ASSOCIATION RULE MINING AND CLASSIFICATION

Mining Frequent Patterns, Associations and Correlations – Mining Methods – Mining various Kinds of Association Rules – Correlation Analysis – Constraint Based Association Mining – Classification and Prediction - Basic Concepts - Decision Tree Induction - Bayesian Classification – Rule Based Classification – Classification by Back propagation – Support Vector Machines – Associative Classification – Lazy Learners – Other Classification Methods – Prediction.

PART A

1.	**Define** correlation and market basket analysis.	Remember	BTL-1
2.	**Formulate** the principle frequent itemset and closed itemset.	Create	BTL-6
3.	How would you **explain** the principle of Apriori algorithm? How can the Efficiency of an Apriori algorithm be improved?	Evaluate	BTL-5
4.	**Define** Data pruning. State the need for pruning phase in decision treeconstruction.	Remember	BTL-1
5.	**Compare** the advantages of FP growth algorithm over apriori algorithm.	Analyze	BTL-4
6.	**Explain** how you will generate association rules from frequent item sets.	Analyze	BTL-4
7.	What is naïve Bayesian classification? How is it **differing** from Bayesian Classification?	Analyze	BTL-4
8.	**Discuss** association rule mining.	Understand	BTL-2
9.	**Describe** the uses correlation.	Understand	BTL-2
10.	**Discuss** the features of Decision tree induction.	Understand	BTL-2
11.	How would you **evaluate** accuracy of a classifier?	Evaluate	BTL-5
12.	**List** the two interesting measures of an association rule.	Remember	BTL-1
13.	**Define** Back propagation.	Remember	BTL-1
14.	**Illustrate** support vector machine with example.	Apply	BTL-3
15.	How would you **show** your understanding about rule based classification?	Apply	BTL-3
16.	**Discuss why** pruning is needed in decision tree.	Understand	BTL-2
17.	What inference can you **Relate** with Bayes theorem?	Analyze	BTL-4
18.	**Demonstrate** the Bayes classification methods.	Apply	BTL-3
19.	**Define** Lazy learners with an example.	Remember	BTL-1
20.	**What** are eager learners?	Remember	BTL-1
21.	**Define** predication.	Apply	BTL-3
22.	**Write** concept about classification?	Understand	BTL-2
23.	**Prioritize the** steps involved in preparing the data for classification.	Evaluate	BTL-5

24.	**Generalise t**he concept of Classification.	Apply	BTL-6
	PART –B		
1.	(i).**Compare** Classification and Prediction.(3) (ii). **Explain** the issues regarding classification and prediction (3). (iii).Write and **explain** the algorithm for mining frequent item sets without candidate generation. (7)	Analyze	BTL-4
2.	i) How would you **summarize** in detail about mining methods? (6) ii) Summarize in detail about various kinds of association rules. (7)	Understand	BTL-2
3.	**Describe** in detail about constraint and correlation based association mining. (13)	Remember	BTL-1
4.	(i).**Develop** an algorithm for classification using decision trees. Illustrate the algorithm with a relevant example. (7) (ii).**What** approach would you use to apply decision tree induction? (6)	Apply	BTL-3
5.	**What** is Classification? What are the features of Bayesian classification? Explain in detail with an example. (13)	Evaluate	BTL-5
6.	(i). Giving concrete example , **explain** a method that performs frequent item set mining by using the prior knowledge of frequent item set properties. (7) (ii). **Discuss** in detail the constraint based association mining. (6).	Understand	BTL-2
7.	(i).**Examine** in detail about Lazy learners with examples. (4) (ii). **Describe** about the process of multi-layer feed-forward neural network classification using back propagation learning.(9)	Remember	BTL-1
8.	(i).**Describe** in detail about frequent pattern classification. (7) (ii).**Write** an algorithm for FP-Tree Construction and discuss how frequent itemsets are generated from FP-Tree.(6)	Understand	BTL-2
9.	**Consider** a home finance loan to predict the housing loan payment. Design a general hierarchical a structure and analyze the factors using rule discovery techniques to accurately predict the number of loan payments in a given quarter/year. Loan is availed for a period of 20 to 25 years, but an average life span of the loan exists for only 7 to 10 years due to payment. Make necessary assumptions: Maintenance record of the customer details and details of the prevailing interest rates, borrower characteristics, account dare, fine tune loan prepayment such as interest rates and fees in order to maximize the profits of the company. Elaborately discuss the association rule mining issues.Also Examine on the multi-level association rules and find if you could relate any relation on from the above application. (13)	Apply	BTL-3
10.	**Generalize** the Bayes theorem of posterior probability and explain the working of a Bayesian classifier with an example.(13)	Create	BTL-6
11.	**Explain** and Apply the Apriori algorithm for discovering frequent item sets of the table. **(13)**	Analyze	BTL-4

	Trans ID / Items Purchased 101 Milk,bread,eggs 102 Milk,juice 103 Juice,butter 104 Milk,bread,eggs 105 Coffee,eggs 106 Coffee 107 Coffee,Juice 108 Milk,bread,cookies,eggs 109 Cookies,butter 110 Milk,bread Use 0.3 for the minimum support value. Illustrate each step of the Apriori Algorithm.		
12.	(i).**Define** classification? With an example explain how support vector machines can be used for classification.(7) (ii). **What** are the prediction techniques supported by a data mining systems? (6)	Remember	BTL-1
13.	(i). **Write** Bayes theorem.(4) (ii) **Explain** how the Bayesian Belief Networks are trained to perform classification.(9)	Analyze	BTL-4
14.	.**Describe** in detail about the following Classification methods. (i). Bayesian classification.(6) (ii) Classification by Back propagation.(7)	Remember	BTL-1
15.	**Illustrate** the process about preparing data for classification in detail.(13)	Apply	BTL-3
16.	(i). **List** out the density methods with suitable example.(6) (ii).**Describe** about attribute selection measure.(7)	Evaluate	BTL-5
17.	**(i) Explain** the back propagation technique.(6) **(ii) Discuss** classifier accuracy with example.(7)	Understand	BTL-2
	PART C		
1.	Find all frequent item sets for the given training set using Apriori and FP growth respectively. **Compare** the efficiency of the two mining processes. (15) 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 }	Create	BTL-6
2.	**Generalize** and Discuss about constraint based association rule mining with examples and state how association mining to correlation analysis is dealt with.(15)	Create	BTL-6

3.	Discuss the single dimensional Boolean association rule mining for transactiondatabase. **Evaluate** the below transaction database. (15) Transaction ID | Items Bought 2000 | A,B,C 1000 | A,C 4000 | A,D 5000 | B,E,F Let minimum support 50% and minimum confidence 50% We have A=>C (50% , 66.6%) C=>A (50%, 100 %)	Evaluate	BTL-5
4.	Explain in detail about hierarchical methods of classification.(15)	Evaluate	BTL-5
5.	**Construct** the decision tree for the following training dataset using decision tree algorithm. (15) Age | Income | Student | Credit_rating | Buys_Computer <=30 | High | No | Fair | No <=30 | High | No | excellent | No 31..40 | High | No | Fair | Yes >40 | Medium | No | Fair | Yes >40 | Low | Yes | Fair | Yes >40 | Low | Yes | excellent | No 31..40 | Low | Yes | excellent | Yes <=30 | Medium | No | Fair | No <=30 | Low | Yes | Fair | Yes >40 | Medium | Yes | Fair | Yes <=30 | Medium | Yes | excellent | Yes 31..40 | Medium | No | excellent | Yes 31..40 | High | Yes | Fair | Yes >40 | Medium | No | excellent | No	Create	BTL-6

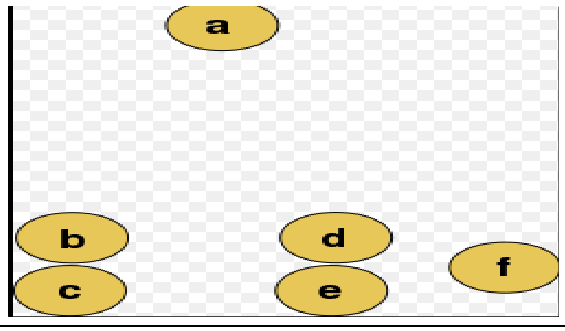
UNIT 5- CLUSTERING AND TRENDS IN DATA MINING

Cluster Analysis - Types of Data – Categorization of Major Clustering Methods – K-means– Partitioning Methods – Hierarchical Methods - Density-Based Methods –Grid Based Methods – Model-Based Clustering Methods – Clustering High Dimensional Data - Constraint – Based Cluster Analysis – Outlier Analysis – Data Mining Applications.

PART A

1.	**Identify** what changes you would make to solve the problem in cluster analysis.	Remember	BTL-1
2.	**List** the major clustering methods.	Remember	BTL-1
3.	**Distinguish** between Partition methods and Hierarchical methods.	Understand	BTL-2
4.	**Explain** why a cluster has to be evaluated.	Analyze	BTL-4
5.	**Illustrate** the intrinsic methods in cluster analysis.	Apply	BTL-3
6.	How do you **explain** the similarity in clustering?	Evaluate	BTL-5
7.	**Define** what is meant by K nearest neighbor algorithm.	Remember	BTL-1

8.	**Illustrate** some applications of clustering.	Apply	BTL-3
9.	What the services are provided by grid based clustering.	Apply	BTL-3
10.	**Formulate** challenges in clustering.	Create	BTL-6
11.	Organize the **design** goals of constraint based clustering methods.	Create	BTL-6
12.	**Classify** the hierarchical clustering methods.	Analyze	BTL-4
13.	**Distinguish** between density based clustering and grid based clustering.	Understand	BTL-2
14.	**Define** outlier. How will you determine outliers in the data?	Remember	BTL-1
15.	**Discuss** the challenges of outlier detection.	Understand	BTL-2
16.	**Distinguish** between Classification and clustering.	Understand	BTL-2
17.	**Evaluate** what information is used by outlier detection method.	Evaluate	BTL-5
18.	**Differentiate the** methods of clustering high dimensional data.	Analyze	BTL-4
19.	**List** out the difference between characterization and clustering.	Remember	BTL-1
20.	**Explain** the typical phases of outlier detection methods.	Analyze	BTL-4
21.	**Write** out the density based methods.	Understand	BTL-2
22.	**What** is a DBSCAN?	Remember	BTL-1
23.	**What is** K-means algorithm.	Understand	BTL-3
24.	**List** out the data mining applications.	Evaluate	BTL-5
	PART B		
1.	(i)**Write** and **Evaluate** the Requirements of clustering in Data Mining (8). ii) **Write** and **Evaluate** the desirable properties of Clustering algorithm.(5)	Evaluate	BTL-5
2.	**(i).Describe** in detail about categorization of major clustering methods.(8) (ii).List out the General applications of Clustering. (5)	Remember	BTL-1
3.	**What is clustering? Describe** in detail about the features of K-means partitioning method. (13)	Remember	BTL-1
4.	**i) Explain** in detail about hierarchical based method. (7) **ii) Explain** in detail about density based methods. (6)	Analyze	BTL-4
5.	What is grid based clustering? With an example explain an algorithm for grid based clustering. (13)	Remember	BTL-1
6.	**(i) Demonstrate** in detail about model based clustering methods. (7) (ii).Illustrate the following (6) 1. CLIQUE 2. DBSCAN	Apply	BTL-3

7.	(i).**Demonstrat**e on clustering high dimensional data. (6) (ii).Consider 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. (7)	Apply	BTL-3
8.	i) How would you **discuss** the outlier analysis in detail? (7) ii) **Discuss** in detail about the various detection techniques in outlier. (6)	Understand	BTL-2
9.	(i).**Explain** in detail about data mining applications (5). (ii). With relevant examples **summarize** in detail about constraint based cluster analysis. (8)	Evaluate	BTL-5
10.	**Design** statistical approaches in outlier detection with neat design and with examples. (13)	Create	BTL-6
11.	**Discuss** the various clustering method in Datamining. With an example (13)	Understand	BTL-2
12.	(i). **Discuss** in detail about the different types of data in cluster analysis.(5) (ii). **Discuss** the following clustering algorithm using examples. (8) 1. K.means 2. K-medoid.	Understand	BTL-2
13.	**Describe** the applications and trends in data mining in detail (13)	Analyze	BTL-4
14.	What is outlier mining important? **Briefly** describe the different approaches behind statistical –based outlier detection, distance based outlier detection and deviation based outlier detection. (13)	Remember	BTL-1
15.	**Explain** in detail about density based clustering and grid based clustering. (13)	Understand	BTL-2
16.	**Demonstrate** the similarities between the clustering algorithm and K-means algorithm in detail (13)	Apply	BTL-3
17.	(i). **Explain** outlier analysis. (6) (ii). **Discuss** about the grid based clustering methods in detail. (7)	Analyze	BTL-4
	PART C		
1.	Explain hierarchical clustering in detail. **Evaluate** the below diagram and draw thedendrogram using hierarchical clustering algorithm. (15) 	Evaluate	BTL-5

<table>
<tr><td>2.</td><td>Consider that the data mining task is to cluster the following eight points A1,A2,A3,B1,B2,B3,C1AND C2(with (X,Y) representing location) into three
clusters A1(2,10) , A2(2,5) , A3(8,4) , B1(5,8) , B2(7,5) , B3(6,4) , C1(1,2) , C2(4,9).
The distance function is Euclidean distance. Suppose initially we assign A1, B1 and C1 as the center of each cluster, respectively. Use the K-means algorithm to show the three cluster centers after the first round of execution and the final tree clusters.
(15)</td><td>Create</td><td>BTL-6</td></tr>
<tr><td>3.</td><td>Discuss the steps in K-means algorithm and evaluate the following table using K- means. (15)
<table><tr><th>Subject</th><th>A</th><th>B</th></tr><tr><td>1</td><td>1.0</td><td>1.0</td></tr><tr><td>2</td><td>1.5</td><td>2.0</td></tr><tr><td>3</td><td>3.0</td><td>4.0</td></tr><tr><td>4</td><td>5.0</td><td>7.0</td></tr><tr><td>5</td><td>3.5</td><td>5.0</td></tr><tr><td>6</td><td>4.5</td><td>5.0</td></tr><tr><td>7</td><td>3.5</td><td>4.5</td></tr></table></td><td>Evaluate</td><td>BTL-5</td></tr>
<tr><td>4.</td><td>Develop the current trends in data mining in any three fields (3x5=15).
1. Financial data analysis
2. Biological data analysis
3. Telecommunication industry
4. Intrusion detection
5. Retail industry</td><td>Create</td><td>BTL-6</td></tr>
<tr><td>5.</td><td>(i). Write the five difference between cluster analysis and outlier analysis. (5)
(ii). Investigate the case study for Data Mining in real time applications. (10)</td><td>Create</td><td>BTL-6</td></tr>
</table>