

SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK



VI SEMESTER

CS8075-DATAWARE HOUSING AND DATA MINING

Regulation – 2017

Academic Year 2019 – 20

Prepared by

Ms.S.SUMA, Assistant Professor/CSE

Ms.C.PABITHA, Assistant Professor/CSE



SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar, Kattankulathur – 603 203.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



QUESTION BANK

SUBJECT : CS8075 -DATAWAREHOUSING AND DATAMINING

SEM / YEAR : VI- Third year

UNIT I DATA WAREHOUSING, BUSINESS ANALYSIS AND ON-LINE ANALYTICAL PROCESSING (OLAP)			
Basic Concepts - Data Warehousing Components – Building a Data Warehouse – Database Architectures for Parallel Processing – Parallel DBMS Vendors - Multidimensional Data Model – Data Warehouse Schemas for Decision Support, Concept Hierarchies -Characteristics of OLAP Systems – Typical OLAP Operations, OLAP and OLTP.			
PART A			
Q.NO	QUESTIONS	COMPETENCE	LEVEL
1.	How is data ware house different from a database? Also 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 data warehouse 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.	Can you list the categories of tools in business analysis?	Remember	BTL-1
8.	How would you describe the basic functions used in reporting tools?	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.	Give the need for OLAP.	Understand	BTL-2
12.	Distinguish STAR join and STAR index.	Understand	BTL-2
13.	How would you show your understanding of Multidimensional data model?	Apply	BTL-3
14.	Generalize the function of OLAP tools in the internet.	Create	BTL-6

15.	Compare OLTP and OLAP system.	Analyze	BTL-4
16.	Compare drill down with roll up approach.	Analyze	BTL-4
17.	List out the key features in business applications using OLAP.	Remember	BTL-1
18.	Differentiate MOLAP and ROLAP.	Understand	BTL-2
19.	How would you classify the ideas of multidimensional data model with multirelational OLAP?	Apply	BTL-3
20.	Design the data warehouse architecture.	Create	BTL-6
PART B			
1.	What is data warehouse? Give the steps for design and construction of Data Warehouses and explain with three tier architecture diagram. (13)	Understand	BTL-2
2.	Examine the relevant examples discuss multidimensional online analytical processing and multi relational online analytical processing. (13)	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 the Database architecture for parallel processing. (8) (ii) Describe the importance of Metadata in data warehouse. (5)	Remember	BTL-1
5.	(i) Explain the steps in building a data warehouse. (8) (ii) Analyze the information needed to support DBMS schemas for Decision support. (5)	Analyze	BTL-4
6.	(i) Discuss in detail about components of data warehousing. (7) (ii) Describe the overall architecture of data warehouse? (6)	understand	BTL-2
7.	(i) Describe the need for OLAP. (7) (ii) Discuss in detail about the OLAP guidelines. (6)	Understand	BTL-2
8.	(i)Diagrammatically illustrate and describe the architecture of MOLAP and ROLAP. (7) (ii) Identify the major differences between MOLAP and ROLAP. (6)	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 Amt(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 (7) ii) STAR join and index. (6)	Remember	BTL-1
11.	(i) Compare OLTP and OLAP systems. (7) (ii) Classify the various OLAP operations in the Multidimensional Data Model. (6)	Analyze	BTL-4
12.	Describe Database architecture for parallel processing and list out the various parallel vendors. (13)	Remember	BTL-1
13.	i) Demonstrate in detail about Data marts. (7) ii) Examine data warehouse administration and management. (6)	Apply	BTL-3
14.	(i) Generalize the potential performance problems with star schema. (7) (ii) Design and discuss about the star and snowflake schema models of a Data	Create	BTL-6

	warehouse. (6)		
PART C			
1.	Explain the typical OLAP operations and compare OLAP with OLTP. (15)	Analyze	BTL-4
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) Depict the 3 tier data warehousing architecture and explain its features in Detail. (8) ii) Explain the different types of OLAP servers (7)	Create	BTL-6
4.	i) Compare the similarities and differences between the database and data warehouse. (8) ii) Explain what data visualization is. How it helps in data warehousing. (7)	Evaluate	BTL-5
UNIT II DATA MINING – INTRODUCTION			
Introduction to Data Mining Systems – Knowledge Discovery Process – Data Mining Techniques – Issues – applications- Data Objects and attribute types, Statistical description of data, Data Preprocessing – Cleaning, Integration, Reduction, Transformation and discretization, Data Visualization, Data similarity and dissimilarity measures.			
PART A			
1.	Define Data mining. List out the steps in data mining.	Remember	BTL-1
2.	List the steps involved in the process of KDD. How does it relate to data mining?	Remember	BTL-1
3.	Compare Discrete versus Continuous Attributes.	Analyze	BTL-4
4.	Identify the various kinds of attributes and brief them.	Remember	BTL-1
5.	How will you Apply the statistical description of Data?	Apply	BTL-3
6.	Give the applications of Data Mining.	Understand	BTL-2
7.	Analyze the primitives for specification of a data mining task.	Analyze	BTL-4
8.	Show the Displays of Basic Statistical Descriptions of Data.	Analyze	BTL-4
9.	Analyze the issues in Data Mining Techniques.	Apply	BTL-3
10.	Express why we need data transformation. Mention the ways by which data can be transformed?	Understand	BTL-2
11.	Analyze data characterization related to data discrimination.	Apply	BTL-3
12.	Generalize in detail about Numeric Attributes.	Create	BTL-6
13.	Express how the attribute selection set is important in data reduction.	Understand	BTL-2

14.	Define Data integration.	Remember	BTL-1
15.	Evaluate the major tasks of data preprocessing.	Evaluate	BTL-5
16.	Formulate what is data discretization.	Create	BTL-6
17.	Distinguish between data similarity and dissimilarity.	Understand	BTL-2
18.	Explain the principle elements of missing values in data cleaning.	Evaluate	BTL-5
19.	Define an efficient procedure for cleaning the noisy data.	Remember	BTL-1
20.	Define data Visualization.	Remember	BTL-1
PART B			
1.	i) Describe the issues of data mining. (7) ii) Describe in detail about the applications of data mining (6)	Remember	BTL-1
2.	(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
3.	(i) Explain the various method to measure central tendency of data, give necessary examples. (7) (ii) Summarize the different ways to measure the dispersion of data. (6)	Evaluate	BTL-5
4.	Discuss in detail the Graphic Displays of basic Statistical Descriptions of Data with example. (13)	Understand	BTL-2
5.	(i) Describe the steps involved in Knowledge discovery in databases (KDD). (7) (ii)Draw the diagram and Describe the architecture of data mining system. (6)	Remember	BTL-1
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 depth of 3. Illustrate your steps. (6) (ii) Classify the various methods for data smoothing. (7)	Apply	BTL-3
7.	(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
8.	List and explain the primitives for specifying a data mining task. (13)	Remember	BTL-1
9.	i) Describe in detail about data extraction (7) ii) Describe in detail about transformation tools (6)	Remember	BTL-1
10.	(i) Distinguish between data generalization and characterizations. (6) (ii)Sketch the various phases of data mining and explain the different steps involved in preprocessing with their significance before mining, Give an example	Understand	BTL-2

	for each process. (7)		
11.	(i)What is Data Preprocessing? Explain the various data pre-processing techniques. (7) (ii) Explain the basic methods for data cleaning. (6)	Analyze	BTL-4
12.	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
13.	i) Compose in detail the various data transformation techniques (7) ii) Develop a short note on discretization techniques (6)	Create	BTL-6
14.	i) Discuss in detail about data integrity preprocessing (7) ii)Formulate in detail about reduction in data preprocessing (6)	Analyze	BTL-4

PART C

1.	i)Explain the major issues in data mining. (5) ii)List the major data preprocessing techniques and explain in detail with examples . (10)	Analyze	BTL-4																																								
2.	Suppose that a hospital tested the age and body fat data for 18 randomly selected adults. Develop the following results: <table border="1"><tr><td>age</td><td>23</td><td>23</td><td>27</td><td>27</td><td>39</td><td>41</td><td>47</td><td>49</td><td>50</td></tr><tr><td>%fat</td><td>9.5</td><td>26.5</td><td>7.8</td><td>17.8</td><td>31.4</td><td>25.9</td><td>27.4</td><td>27.2</td><td>31.2</td></tr><tr><td>age</td><td>52</td><td>54</td><td>54</td><td>56</td><td>57</td><td>58</td><td>58</td><td>60</td><td>61</td></tr><tr><td>%fat</td><td>34.6</td><td>42.5</td><td>28.8</td><td>33.4</td><td>30.2</td><td>34.1</td><td>32.9</td><td>41.2</td><td>35.7</td></tr></table> (a) Calculate the mean, median, and standard deviation of age and %fat. (5) (b) Draw the boxplots for age and %fat. (5) (c) Draw a scatter plot and a q-q plot based on these two variables. (5)	age	23	23	27	27	39	41	47	49	50	%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2	age	52	54	54	56	57	58	58	60	61	%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7	Create	BTL-6
age	23	23	27	27	39	41	47	49	50																																		
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2																																		
age	52	54	54	56	57	58	58	60	61																																		
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7																																		
3.	Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. (a) What is the mean of the data? What is the median? (2) (b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.). (2) (c) What is the midrange of the data? (2) (d) Can you Design (roughly) the first quartile (Q1) and the third quartile (Q3) of the data? (3) (e) Give the five-number summary of the data. (2) (f) Prepare a boxplot of the data. (2) (g) How is a quantile–quantile plot different from a quantile plot? (2)	Create	BTL-6																																								
4.	i) Generalize why do we need data preprocessing step in data mining (8) ii)Explain the various methods of data cleaning and data reduction techniques (7)	Evaluate	BTL-5																																								

UNIT III DATA MINING - FREQUENT PATTERN ANALYSIS

Mining Frequent Patterns, Associations and Correlations – Mining Methods- Pattern Evaluation Method – Pattern Mining in Multilevel, Multi Dimensional Space – Constraint Based Frequent Pattern Mining, Classification using Frequent Patterns.

PART A			
1.	Define association and correlations.	Remember	BTL-1
2.	Describe the ways to mine pattern in Multi-Dimensional space.	Remember	BTL-1
3.	List the ways in which interesting patterns should be mined.	Remember	BTL-1
4.	Analyze the various Pattern Evaluation method.	Analyze	BTL-4
5.	Compare Monotonic and Anti-monotonic constraints.	Understand	BTL-2
6.	How would you illustrate key distribution center?	Apply	BTL-3
7.	Analyze the constraint based frequent pattern mining.	Analyze	BTL-4
8.	List the categories of pattern mining constraints.	Remember	BTL-1
9.	Define correlation and market basket analysis.	Remember	BTL-1
10.	Evaluate the classification using Frequent patterns.	Evaluate	BTL-5
11.	Are all patterns generated are interesting and useful? Give reasons to justify.	Understand	BTL-2
12.	How will you Apply pattern mining in Multilevel space?	Apply	BTL-3
13.	Discuss association rule mining. List the two interesting measures of an association rule.	Understand	BTL-2
14.	Explain how you will generate association rules from frequent item sets.	Analyze	BTL-4
15.	Compare the advantages of FP growth algorithm over apriori algorithm	Analyze	BTL-2
16.	Formulate the principle frequent item set and closed item set.	Create	BTL-6
17.	How will you apply FP growth algorithm in Data mining?	Apply	BTL-3
18.	How would you explain the principle of Apriori algorithm? How can the efficiency of an Apriori algorithm be improved?	Evaluate	BTL-5
19.	Generalize on Mining Closed and Max Patterns.	Create	BTL-6
20.	List the uses of multilevel mining with reduced support.	Remember	BTL-1
PART B			
1.	Discuss about constraint based association rule mining with examples and state how association mining to correlation analysis is dealt with. (13)	Apply	BTL-3
2.	(i) Explain in detail Constraint Based Frequent Pattern mining. (7) (ii) Summarize on Classification using frequent Patterns. (6)	Evaluate	BTL-5
3.	(i)What is interestingness of a pattern? What approach would you design to mine interestingness of patterns? (6) (ii) Design and discuss in detail about integration of data mining system with a data warehouse? (7)	Create	BTL-6

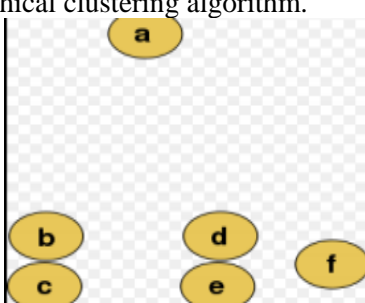
4.	Discuss the various Pattern evaluation methods and compare their measures. (13)	Understand	BTL-2																
5.	(i) Compare Classification and Prediction. (3) (ii) Explain the issues regarding classification and prediction. (3) (iii) Write and explain the algorithm for mining frequent item sets without candidate generation. (7)	Analyze	BTL-4																
6.	Find all frequent item sets for the given training set using Apriori and FP growth respectively. Compare the efficiency of the two mining processes (13) TID ITEMS BROUGHT 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 }	Apply	BTL-3																
7.	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																
8.	Describe about constraint and correlation based association mining. (13)	Remember	BTL-1																
9.	(i) Describe 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)	Remember	BTL-1																
10.	Define Market Basket Analysis. Describe about Frequent Itemsets, Closed Itemset and Association Rules. (13)	Remember	BTL-1																
11.	(i) Examine in detail the Pattern Mining in Multilevel Associations. (7) (ii) Describe in detail Pattern mining in Multidimensional Associations. (6)	Remember	BTL-1																
12.	(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																
13.	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 Analyze on the multilevel association rules and find if you could relate any relation on from the above application. (13)	Analyze	BTL-4																
14.	Explain and Apply the Apriori algorithm for discovering frequent item sets of the table. (13) <table><tr><td>Trans ID</td><td>Items Purchased</td></tr><tr><td>101</td><td>Milk,bread,eggs</td></tr><tr><td>102</td><td>Milk,juice</td></tr><tr><td>103</td><td>Juice,butter</td></tr><tr><td>104</td><td>Milk,bread,eggs</td></tr><tr><td>105</td><td>Coffee,eggs</td></tr><tr><td>106</td><td>Coffee</td></tr><tr><td>107</td><td>Coffee,Juice</td></tr></table>	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	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.			
PART C				
1.	Compare the Pattern Evaluation Measures and summarize the various pattern evaluation methods. (15)		Analyze	BTL-4
2.	i)What is interestingness of a pattern ? (5) ii)Summarize the various classification methods using frequent patterns. (10)		Evaluate	BTL-5
3.	Analyze the various Frequent Itemset mining method with examples. (15)		Analyze	BTL-4
4.	Generalize how pattern mining is done in multilevel and multidimensional space with necessary examples. (15)		Create	BTL-6
UNIT IV CLASSIFICATION AND CLUSTERING				
Decision Tree Induction - Bayesian Classification – Rule Based Classification – Classification by Back Propagation – Support Vector Machines — Lazy Learners – Model Evaluation and Selection-Techniques to improve Classification Accuracy. Clustering Techniques – Cluster analysis-Partitioning Methods - Hierarchical Methods – Density Based Methods - Grid Based Methods – Evaluation of clustering – Clustering high dimensional data- Clustering with constraints, Outlier analysis-outlier detection methods.				
PART A				
1.	Identify what changes you make to solve the problem in cluster analysis.		Remember	BTL-1
2.	Formulate the role of application and challenges in clustering.		Create	BTL-6
3.	Evaluate what information is used by outlier detection method.		Evaluate	BTL-5
4.	Define Data pruning. State the need for pruning phase in decision tree construction.		Remember	BTL-1
5.	Classify the typical phases of outlier detection methods.		Analyze	BTL-4
6.	Pointout why a cluster has to be evaluated?		Analyze	BTL-4
7.	What is Naïve Bayesian classification? How is it differing from Bayesian classification?		Analyze	BTL-4
8.	Distinguish between Classification and clustering.		Understand	BTL-2
9.	Classify the hierarchical clustering methods.		Understand	BTL-2
10.	Name the features of Decision tree induction.		Understand	BTL-2
11.	How would you evaluate accuracy of a classifier?		Evaluate	BTL-5
12.	List the challenges of outlier detection.		Remember	BTL-1

13.	Define decision tree induction.	Remember	BTL-1
14.	Show the intrinsic methods in cluster analysis.	Apply	BTL-3
15.	How would you show your understanding about pessimistic pruning?	Apply	BTL-3
16.	Give why pruning is needed in decision tree?	Understand	BTL-2
17.	What inference can you formulate with Bayes theorem?	Create	BTL-6
18.	Demonstrate the Bayes classification methods.	Apply	BTL-3
19.	Define Lazy learners and eager learners with an example.	Remember	BTL-1
20.	How do you explain the extrinsic and intrinsic methods?	Remember	BTL-1
PART B			
1.	(i) Analyze the Requirements of clustering in Data Mining. (8) ii) Pointout the desirable properties of Clustering algorithm. (5)	Analyze	BTL-4
2.	(i)Write the difference between CLARA and CLARANS. (8) (ii) Discuss the different hierarchical methods in cluster analysis. (5)	Understand	BTL-2
3.	What is clustering? Describe in detail about the features of K-means partitioning method. (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.	(i)What is Classification? What are the features of Bayesian classification? Explain in detail with an example. (8) (ii)Explain how the Bayesian Belief Networks are trained to perform classification. (5)	Evaluate	BTL-5
6.	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
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) Discuss the different types of data in cluster analysis. (5) (ii) Interpret following clustering algorithm using examples. (8) 1. K.means 2. K-medoid.	Understand	BTL-2
9.	(i) Demonstrate in detail about model based clustering methods. (7) (ii)Illustrate the concepts (6) 1. CLIQUE 2. DBSCAN	Apply	BTL-3
10.	i) Generalize the Bayes theorem of posterior probability and explain the working of a Bayesian classifier with an example. (9) ii) Formulate rule based classification techniques. (4)	Create	BTL-6
11.	i) Explain the hierarchical based method for cluster analysis. (7) ii) Explain in detail about density based methods. (6)	Analyze	BTL-4

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) Explain algorithm for constructing a decision tree from training samples. (9) (ii) Write Bayes theorem. (4)	Analyze	BTL-4
14.	i) Describe in detail about the following Classification methods. (6) (a) Bayesian classification (b) Fuzzy set approach (c) Genetic algorithms. ii) Describe in detail Classification by Back propagation. (7)	Remember	BTL-1

PART C

1.	<p>i)Develop a clustering high dimensional data. (8)</p> <p>(ii)Consider five points { X_1, X_2,X_3, X_4, X_5} with the following coordinates as a two dimensional sample for clustering: $X_1 = (0,2.5)$; $X_2 = (0,0)$; $X_3= (1.5,0)$; $X_4 = (5,0)$; $X_5 = (5,2)$ Compose the K-means partitioning algorithm using the above data set. (7)</p>	Create	BTL-6										
2.	<p>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. Analyze the K-means algorithm to show the three cluster centers after the first round of execution and the final tree clusters. (15)</p>	Analyze	BTL-4										
3.	<p>Discuss the single dimensional Boolean association rule mining for transaction database. Evaluate the below transaction database. (15)</p> <table><tr><th>Transaction ID</th><th>Items Bought</th></tr><tr><td>2000</td><td>A,B,C</td></tr><tr><td>1000</td><td>A,C</td></tr><tr><td>4000</td><td>A,D</td></tr><tr><td>5000</td><td>B,E,F</td></tr></table> <p>Let minimum support 50% and minimum confidence 50% We have $A \Rightarrow C$ (50% , 66.6%) $C \Rightarrow A$ (50%, 100 %)</p>	Transaction ID	Items Bought	2000	A,B,C	1000	A,C	4000	A,D	5000	B,E,F	Evaluate	BTL-5
Transaction ID	Items Bought												
2000	A,B,C												
1000	A,C												
4000	A,D												
5000	B,E,F												
4.	<p>Explain hierarchical clustering in detail. Analyse the below diagram and draw the dendrogram using hierarchical clustering algorithm. (15)</p> 	Create	BTL-6										

UNIT V WEKA TOOL

Datasets – Introduction, Iris plants database, Breast cancer database, Auto imports database - Introduction to WEKA, The Explorer – Getting started, Exploring the explorer, Learning algorithms, Clustering algorithms, Association–rule learners.

PART A			
1.	Define Weka tool.	Remember	BTL-1
2.	List the applications of Weka.	Remember	BTL-1
3.	Name the features of Weka tool.	Remember	BTL-1
4.	Analyze how attribute selection is achieved in Weka tool?	Analyze	BTL-4
5.	How will you apply supervised filter to test data?	Apply	BTL-3
6.	Compare the classifier algorithms supported by Weka.	Evaluate	BTL-5
7.	Define IRIS database.	Remember	BTL-1
8.	Demonstrate the method to convert files to ARFF format.	Apply	BTL-3
9.	How will you apply Weka tool to learning method?	Apply	BTL-3
10.	Generalize the tabs available in explorer window.	Create	BTL-6
11.	Formulate the procedure to load the data into the Explorer.	Create	BTL-6
12.	Analyze Association rule learners.	Analyze	BTL-4
13.	Discuss the use of Cluster and Associate panel.	Understand	BTL-2
14.	Define Multilayer perceptron.	Remember	BTL-1
15.	Classify the various clustering algorithms.	Understand	BTL-2
16.	Differentiate tree visualizer and data visualizer.	Understand	BTL-2
17.	Evaluate the use of supervised instance filters.	Evaluate	BTL-5
18.	Compare Multi-Instance Classifiers and Miscellaneous Classifiers.	Understand	BTL-2
19.	Define Bayesian logistic regression.	Remember	BTL-1
20.	Classify the file formats supported by Weka converters.	Analyze	BTL-4
PART B			
1.	Explain the need of Auto imports database with necessary implementations. (13)	Analyze	BTL-4
2.	Describe Datasets. Give the importance of datasets in Data Mining. (13)	Remember	BTL-1
3.	(i) Describe the steps involved in getting started with The Explorer. (7) (ii) Examine the output of building a decision tree. (6)	Remember	BTL-1
4.	Analyze the purpose of Neural Networks in learning algorithms. (13)	Analyze	BTL-4

5.	(i) Define Clustering algorithms. And explain the functions of several clustering algorithms. (7) (ii) Describe the function of Association-rule learners. (6)	Remember	BTL-1
6.	Examine how you will apply the various tabs of the explorer window? (13)	Apply	BTL-3
7.	(i) Explain the functionality of the Bayes classifier algorithms. (7) (ii) Describe the functions of several tree based classifier algorithm in Weka. (6)	Apply	BTL-3
8.	(i) Describe the methods for generating rules in classifiers. (7) (ii) Interpret the various functions used in classifier algorithms in Weka. (6)	Understand	BTL-2
9.	Summarize the use of Breast cancer dataset and its specifications. (13)	Evaluate	BTL-5
10.	Develop the implementation of Iris Plants Database in Weka and its uses. (13)	Create	BTL-6
11.	Distinguish lazy classifiers and miscellaneous classifiers. (13)	Understand	BTL-2
12.	(i) Summarize the clustering algorithms used in Weka. (7) (ii) Discuss in detail the various Association-rule learners. (6)	Understand	BTL-2
13.	Give an introduction on Weka Tool and analyze its importance. (13)	Analyze	BTL-4
14.	Define Learning Algorithms and explain the various classifier algorithms in Weka. (13)	Remember	BTL-1
PART C			
1.	Analyze the procedure for getting started with explorer, preparing and loading the data, building a decision tree and examine the output. (15)	Analyze	BTL-4
2.	Evaluate the use of Datasets and explain the procedure to implement any database in Weka. (15)	Create	BTL-6
3.	Conclude the use of Learning and Clustering Algorithms and explain its functionality. (15)	Evaluate	BTL-5
4.	(i) Formulate the importance of Weka & explain implementation procedure. (8) (ii) Generalize the three association-rule learners listed by Weka. (7)	Create	BTL-6