NGM COLLEGE

DEPARTMENT OF COMPUTER SCIENCE (SF)

18PCS1E1– DATA MINING

K1 Level Questions

1. …….. is a comparison of the general features of the target class data objects against the general features of objects from one or multiple contrasting classes.

a) Data Characterization b) Data selection c) Data Classification**d) Data discrimination**

2. In neural how can connectons between different layers be achieved?

a) interlayer b) intralayer**c) both interlayer and intralayer** d) either interlayer or intralayer

3. The operation of outstar can be viewed as?

a) content addressing the memory

b) memory addressing the content

c) either content addressing or memory addressing

d) both content & memory addressing

4_____oriented and is used for data analysis by knowledge workers, including managers, executives, and analysts.

A) OLAP B) OLTP C) **Both of the above** D) None of the above

5.Querying of unstructured textual data is referred to as

A. Information access
B. Information updation
C. Information manipulation
D. Information retrieval

6. A manual component to data mining, consists of preprocessing data to a form acceptable to

A. Variables
B. Algorithms
C. Rules
D. Processes

7. Patterns that can be discovered from a given database, can be of

A. One type only

B. No specific type
C. More than one type
D. Multiple type always

8. Analysis tools precompute summaries of very large amounts of data, in order to give

A. Queries response
B. Data access
C. Authorization
D. Consistency

9. **Query tools are**

A) reference to the speed of an algorithm, which is quadratically dependent on the size of the data

b) Attributes of a database table that can take only numerical values.

c) Tools designed to query a database.

d)None of these

10.**Noise is**

A.A component of a network

B.In the context of KDD and data mining, this refers to random errors in a database table.

C.One of the defining aspects of a data warehouse

D.None of these

Unit: 2

11.Informationis

a. Data
b. ProcessedData
c. Manipulated input
d. Computeroutput

12.Data by itself is not usefulunless

e. It ismassive
f. It is processed to obtaininformation
g. It is collected from diversesources

h. It is properly stated

13.For taking decisions data mustbe

a Very accurate

b Massive

c Processed correctly

d Collected from diversesources

14.Strategic information is neededfor

a Day to day operations

b Meet government requirements c Long rangeplanning

d Short rangeplanning

15.Strategic information is required by

a Middle managers

b **Linemanagers**

c Top managers

d Allworkers

16.Tactical information is neededfor

a Day to day operations

b Meet government requirements c **Long rangeplanning**

d Short rangeplanning

17.Tactical information is requiredby

a **Middle anagers**

b Linemanagers

c Top managers

d Allworkers

18.Operational information is needed for

a Day to day operations

b Meet government requirements c **Long rangeplanning**

d Short rangeplanning

19. Operational information is requiredby

a Middle managers

b Linemanagers

c Top managers

d Allworkers

20. Statutory information is needed for

a Day to day operations

b Meet government requirements

c Long rangeplanning

d Short rangeplanning

UNIT:3

21.Volume of strategic informationis

a Condensed

b Detailed

c s**ummarized**

d Irrelevant

22.Volume of tactical informationis

a **Condensed**

b Detailed

c Summarized

d relevant

23.Volume of operational informationis

a Condensed

b Detailed

c Summarized

d **Irrelevant**

24.Strategic informationis

a Haphazard

b **Wellorganized**

Unstructured

d Partly structured

25.Tactical informationis

a Haphazard

b Well organized

c Unstructured

d **Partly structured**

26.Operational informationis

a Haphazard

b Well organized

c Unstructured

d Partly structured

27.Match and find best pairing for a Human Resource ManagementSystem

(i)Policies ongiving bonus (iv)Strategic information

(ii)Absenteereduction (v)Tactical information

(iii)Skillsinventory (vi)Operational Information a

(i) and(v)

b (i) and(iv)

c (ii) and(iv)

d (iii) and(v)

28.Match and find best pairing for a Production ManagementSystem

(i) Performance appraisalofmachines (iv)Strategic information todecide on replacement

(ii)Introducingnewproduction (v)Tactical information technology

(iii)Preventivemaintenance schedules (vi)Operational informationfor machines

a (i) and(vi)

b (ii) and(v)

c (i) and(v)

d (iii) and(iv)

29.Match and find best pairing for a Production ManagementSystem

(i) Performance appraisalofmachines (iv)Strategic information todecide on replacement

(ii)Introducingnewproduction (v)Tactical informationtechnology

(iii)Preventivemaintenance schedules (vi)Operational informationfor machines

a (iii) and(vi)

b (i) and(iv)

c (ii) and(v)

d None of the above

30.Match and find best pairing for a Materials ManagementSystem

(i) Developingvendorperformance	(iv) Strategicinformation measures
(ii) Developing vendorsforcritical	(v) Tactical information items
(iii)List of items rejected fromavendor	(vi)Operationalinformation

a (i) and(v)

b (ii) and(v)

c (iii) and(iv)

d (ii) and(vi)

UNIT: 4

31.Match quality of information and how it is ensured using the followinglist

QUALITY	HOWENSURED
(i) Brief	(iv) Unpleasant information nothidden
(ii) Relevant	(v) Summarize relevantinformation
(iii) Trustworthy	(vi) Understands user needs

a (ii) and(vi)

b (i) and(iv)

c (iii) and(v)

d (ii) and(iv)

32.The quality of information which does not hide any unpleasant information is knownas

a Complete

b Trustworthy

c Relevant

d **None of the above**

33.The quality of information which is based on understanding userneeds

a Complete

b Trustworthy

c Relevant

d None of the above

34.Every record stored in a Master file has a key fieldbecause

a it is the most importantfield

b **it acts as a unique identification**

c it is the key to thedatabase

d it is a very concisefield

35.The primary storage medium for storing archival datais

a floppy disk

b magnetic disk

c **magnetic tape**d

CD-ROM

36.Master files are normally storedin

a a hard disk

b a tape

c CD –ROM

d computer's mainmemory

37.Master file is a filecontaining

a all masterrecords

b all records relevant to the application

c a collection of dataitems

d historical data of relevance to theorganization

38.Edit program is required to

a authenticate data entered by an operator

b **format correctly inputdata**

c detect errors in input data

d expedite retrieving inputdata

39.Data rejected by edit programare

a corrected and re-entered

b removed from processing

c **collected for lateruse**

d ignored duringprocessing

40.Online transaction processing is usedbecause

a it isefficient

b disk is used for storingfiles

c it can handle random queries.

d **Transactions occur**

inbatches

41.On-line transaction processing is usedwhen

i) it is required to answer randomqueries

ii) it is required to ensure correctprocessing

iii) all files are availableon-line

iv) all files are stored using hard disk

a i ,ii

b i, iii

c ii ,iii, iv

d i , ii,iii

UNIT:5

42.Off-line data entry is preferablewhen

i) data should be entered withouterror

ii) the volume of data to be entered islarge

iii) the volume of data to be entered issmall

iv) data is to be processed periodically

a i, ii

b ii, iii

c ii, iv

d iii,iv

43.Batch processing is usedwhen

i) response time should be short

ii) data processing is to be carried out at periodicintervals

iii) transactions are in batches

iv) transactions do not occurperiodically

a i ,ii

b i ,iii,iv

c ii ,iii

d i , ii,iii

44.Batch processing is preferred over on-line transaction processingwhen

i) processing efficiency is important

ii) the volume of data to be processed islarge

iii) only periodic processing is needed

iv) a large number of queries are to be processed

a i ,ii

b i, iii

c ii ,iii

d i , ii,iii

45.A management information system is onewhich

a is required by all managers of anorganization

b processes data to yield information of value in tacticalmanagement c provides operationalinformation

d allows better management oforganizations

46.Data mining is used to aidin

a operationalmanagement

b **analyzing past decision made by managers**

c detecting patterns in operationaldata

d retrieving archivaldata

47.Data mining requires

a large quantities of operational data stored over a period of time

b lots of tacticaldata

c **several tape drives to store archival data**

d large mainframecomputers

48.Data mining can not be doneif

a **operational data has not beenarchive**d

b earlier management decisions are notavailable

c the organization islarge

d all processing had been only batch processing

49.Decision support systems are usedfor

a Management decisionmaking

b Providing tactical information to management

c **Providing strategic information to management**

d Better operation of an organization

50.Decision support systems are usedby

a Linemanagers.

b Top-level managers.

c Middle level managers.

d **Systemusers**

51.Decision support systems are essentialfor

a **Day–to-day operation of an organization**.

b Providing statutoryinformation.

c Top level strategic decisionmaking.

d Ensuring that organizations are profitable.

NGM COLLEGE

DEPARTMENT OF COMPUTER SCIENCE (SF)

18PCS1E1– DATA MINING

Unit:1

1. What are the uses of multi feature cubes?
   Multi feature cubes, which compute complex queries involving multiple dependent aggregates at multiple granularity. These cubes are very useful in practice. Many complex data mining queries can be answered by multi feature cubes without any significant increase in computational cost, in comparison to cube computation for simple queries with standard data cubes.
2. Compare OLTP and OLAP Systems.
   If an on-line operational database systems is used for efficient retrieval, efficient storage and management of large amounts of data, then the system is said to be on-line transaction processing. Data warehouse systems serves users (or) knowledge workers in the role of data analysis and decision-making. Such systems can organize and present data in various formats. These systems are known as on-line analytical processing systems.
3. What is data warehouse metadata?
   Metadata are data about data. When used in a data warehouse, metadata are the data that define warehouse objects. Metadata are created for the data names and definitions of the given warehouse. Additional metadata are created and captured for time stamping any extracted data, the source of the extracted data, and missing fields that have been added by data cleaning or integration processes.
4. 4.Explain the differences between star and snowflake schema.
   The dimension table of the snowflake schema model may be kept in normalized Form to reduce redundancies. Such a table is easy to maintain and saves storage space.
5. In the context of data warehousing what is data transformation? `
   In data transformation, the data are transformed or consolidated into forms appropriate for mining. Data transformation can involve the following: Smoothing, Aggregation, Generalization, Normalization, Attribute construction
6. Define Slice and Dice operation.
   The slice operation performs a selection on one dimension of the cube resulting in A sub cube. The dice operation defines a sub cube by performing a selection on two (or) more dimensions.
7. What is data warehouse?
   A data warehouse is a repository of multiple heterogeneous data sources organized under a unified schema at a single site to facilitate management decision making. (Or) A data warehouse is a subject-oriented, time-variant and nonvolatile collection of data in support of -making process.
8. Differentiate fact table and dimension table.
   Fact table contains the name of facts (or) measures as well as keys to each of the related dimensional tables. A dimension table is used for describing the dimension. (e.g.) A dimension table for item may contain the attributes item_ name, brand and type.

9. What is descriptive and predictive data mining?
   Descriptive data mining, which describes data in a concise and summarative manner and presents interesting general properties of the data. Predictive data mining, which analyzes data in order to construct one or a set of models and attempts to predict the behavior of new data sets.Predictive data mining, such as classification, regression analysis, and trend analysis.
10. What is the need for preprocessing the data?
    Incomplete, noisy, and inconsistent data are commonplace properties of large real world databases and data warehouses. Incomplete data can occur for a number of reasons. Attributes of interest may not always be available, such as customer information for sales transaction data. Other data may not be included simply because it was not considered important at the time of entry. Relevant data may not be recorded due to a misunderstanding, or because of equipment malfunctions. Data that were inconsistent with other recorded data may have been deleted. Furthermore, the recording of the history or modifications to the data may have been overlooked. Missing data, particularly for tuples with missing values for some attributes, may need to be inferred.

Unit:2

11. What is parallel mining of concept description? (OR) What is concept description?
    Data can be associated with classes or concepts. It can be useful to describe individual classes and concepts in summarized, concise, and yet precise terms. Such descriptions of a class or a concept are called class/concept descriptions. These descriptions can be derived via (1) data characterization, by summarizing the data of the class under study (often called the target class) in general terms, or (2) data discrimination, by comparison of the target class with one or a set of comparative classes (often called the contrasting classes), or (3) both data characterization and discrimination
12. Mention the various tasks to be accomplished as part of data pre-processing.
    Data cleaning 2.Data Integration 3.Data Transformation 4. Data reduction
13. What is data cleaning?
    Data cleaning means removing the inconsistent data or noise and collecting necessary information of a collection of interrelated data
14. Define Data mining.
    THE amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mini
15. What are the types of concept hierarchies?
    A concept hierarchy defines a sequence of mappings from a set of low-level concepts to higher-level, more general concepts. Concept hierarchies allow specialization, or drilling down, where by concept values are replaced by lower-level concepts.
16. Define frequent set and border set.
    A set of items is referred to as an itemset. An itemset that contains k items is a kitemset. The set Of computer, antivirus software is a 2-itemset. The occurrence frequency of an itemset is the number of transactions that contain the itemset. This is also known, simply, as the frequency, support count, or count of the itemset.Where each variation involves in slightly different way.The variations, where nodes indicate an item or itemset that has been

examined, and nodes with thick borders indicate that an examined item or itemset is frequent

17. How is association rule mined from large databases?
    Suppose, however, that rather than using a transactional database, sales and related information are stored in a relational database or data warehouse. Such data stores are multidimensional, by definition. For instance, in addition to keeping track of the items purchased in sales transactions, a relational database may record other attributes associated with the items, such as the quantity purchased or the price, or the branch location of the sale. Additional relational information regarding the customers who purchased the items, such as customer age, occupation, credit rating, income, and address, may also be stored.
18. What is over fitting and what can you do to prevent it?
    Tree pruning methods address this problem of over fitting the data. Such methods typically use statistical measures to remove the least reliable branches. An unpruned tree and a pruned version of it. Pruned trees tend to be smaller and less complex and, thus, easier to comprehend. They are usually faster and better at correctly classifying independent test data (i.e., of previously unseen tuples) than unpruned trees.
19. List the techniques to improve the efficiency of Apriori algorithm.
    Hash based technique Transaction Reduction Portioning Sampling Dynamic item counting
20. What is FP growth?
    FP-growth, which adopts a divide-and-conquer strategy as follows. First, it compresses the database representing frequent items into a frequent-pattern tree, or FP-tree,which retains the itemset association information. It then divides the compressed database into a set of conditional databases (a special kind of projected database),

Unit:3

21. What is tree pruning?
    Tree pruning attempts to identify and remove such branches, with the goal of improving classification accuracy on unseen data.
22. List the requirements of clustering in data mining.
    Mining data streams involves the efficient discovery of general patterns and dynamic changes within stream data. For example, we may like to detect intrusions of a computer network based on the anomaly of message flow, which may be discovered by clustering data streams, dynamic construction of stream models, or comparing the current frequent patterns with that at a certain previous time.
23. What is classification?
    Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data (i.e., data objects whose class label is known).
24. What is the objective function of the K-means algorithm?
    The k-means algorithm takes the input parameter, k, and partitions a set of n objects into k clusters so that the resulting intra cluster similarity is high but the inter cluster similarity is low. Cluster similarity is measured in regard to the mean va
25. The naïve Bayes classifier makes what assumption that motivates its name?

Studies comparing classification algorithms have found a simple Bayesian classifier known as the naïve Bayesian classifier to be comparable in performance with decision tree and selected neural network classifiers.

26. Define outliers.
    List various outlier detection approaches. (May/June 2010) A database may contain data objects that do not comply with the general behavior or model of the data. These data objects are outliers. Most data mining methods discard outliers as noise or exceptions. These can be categorized into four approaches: the statistical approach, the distance-based approach, the density-based local outlier approach, and the deviation-based approach.
27. What is meant by hierarchical clustering?
    A hierarchical method creates a hierarchical decomposition of the given set of data objects. A hierarchical method can be classified as being either agglomerative or divisive, based on how the hierarchical decomposition is formed
28. What is Association based classification?
    Association-based classification, which classifies documents based on a set of associated, frequently occurring text patterns. Notice that very frequent terms are likely poor discriminators. Thus only those terms that are not very frequent and that have good discriminative power will be used in document classification. Such an association-based classification method proceeds as follows: First, keywords and terms can be extracted by information retrieval and simple association analysis techniques. Second, concept hierarchies of keywords and terms can be obtained using available term classes, such as WordNet, or relying on expert knowledge, or some keyword classification systems.
29. What do you go for clustering analysis?
    a. Clustering can be used to generate a concept hierarchy for A by following either a top down splitting strategy or a bottom- up merging strategy, where each cluster forms a node of the concept hierarchy. In the former, each initial cluster or partition may be further decomposed into several sub clusters, forming a lower level of the hierarchy. In the latter, clusters are formed by repeatedly grouping neighboring clusters in order to form higher- level concepts.

30. What are the requirements of cluster analysis?
    a. Scalability Ability to deal with different types of attributes Discovery of clusters with arbitrary shape Minimal requirements for domain knowledge to determine input parameters Ability to deal with noisy data Incremental clustering and insensitivity to the order of input records High dimensionality Constraint-based clustering

    Unit:4

31. What is mean by cluster analysis?
    a. A cluster analysis is the process of analyzing the various clusters to organize the different objects into meaningful and descriptive object.

32. Define CLARANS. CLARANS
    a. (Cluster Large Applications based on Randomized Search) to improve the quality of CLARA we go for CLARANS. It Draws sample with some randomness in each step of search. It overcome the problem of scalability that K-Medoids suffers from.

33. What is meant by web usage mining?
    a. Web usage mining is the process of extracting useful information from server logs i.e. users history. Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data

34. Define BIRCH,ROCK and CURE. BIRCH
    a. (Balanced Iterative Reducing and Clustering Using Hierarchies): Partitions objects hierarchically using tree structures and then refines the clusters using other clustering methods. It defines a clustering feature and an associated tree structure that summarizes a cluster. The Produce spherical Cluster and may produce unintended cluster.
    b. ROCK(RObust Clustering using links): Merges clusters based on their interconnectivity. Great for categorical data. Ignores information about the looseness of two clusters while emphasizing interconnectivity.
    c. CURE(Clustering Using Representatives): Creates clusters by sampling the database and shrinks them toward the center of the cluster by a specified fraction. Obviously better in runtime but lacking in precision.

35.What is the use of the knowledge base?

Knowledge base is domain knowledge that is used to guide search or evaluate the interestingness of resulting pattern. Such knowledge can include concept hierarchies used to organize attribute /attribute values in to different levels of

36. What is meta learning.

Concept of combining the predictions made from multiple models of data mining and analyzing those predictions to formulate a new and previously unknown

37.What is the purpose of Data mining Technique?

It provides a way to use various data mining tasks.

38. Define Predictive model.

It is used to predict the values of data by making use of known results from a different set of sample data.

39. Data mining tasks that are belongs to predictive model

Classification Regression Time series analysis

40. Define descriptive model

It is used to determine the patterns and relationships in a sample data. Data mining tasks that belongs to descriptive model: Clustering Summarization Association rules Sequence

Unit:5

41.What is meant by pattern?

Pattern represents knowledge if it is easily understood by humans; valid on test data with some degree of certainty; and potentially useful, novel,or validates a hunch about which the used

was curious. Measures of pattern interestingness, either objective or subjective, can be used to guide the discovery process.

42. How is a data warehouse different from a database?

Data warehouse is a repository of multiple heterogeneous data sources, organized under a unified schema at a single site in order to facilitate management decision-making. Database consists of a collection of interrelated data.

43.When we can say the association rules are interesting?

Association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Users or domain experts can set such thresholds

44.How are association rules mined from large databases?

I step: Find all frequent item sets:

II step: Generate strong association rules from frequent item set

45.What is the purpose of Apriori Algorithm?

Apriori algorithm is an influential algorithm for mining frequent item sets for Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent item set properties.

46.. Give few techniques to improve the efficiency of Apriori algorithm

Hash based technique

Transaction

Reduction Portioning

Sampling

Dynamic item count

47.What is Decision tree?

A decision tree is a flow chart like tree structures, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions. The top most in a tree is the root node.

48. What is Attribute Selection Measure?

The information Gain measure is used to select the test attribute at each node in the decision tree. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split

49.Define Pre Pruning

A tree is pruned by halting its construction early. Upon halting, the node becomes a leaf. The leaf may hold the most frequent class among the subset samples.

50. Define Post Pruning.

Post pruning removes branches from a “Fully grown” tree. A tree node is pruned by removing its branches. Eg: Cost Complexity Algorithm

NGM COLLEGE

DEPARTMENT OF COMPUTER SCIENCE (SF)

18PCS1E1– DATA MINING

Unit:1

1. Explain the differences between Knowledge discovery and data mining.
2. What are the application areas of data Mining?
3. List out different sources of information.
4. What type of benefit you might hope to get from data mining?
5. How can Data Mining help business analyst?
6. Explain the differences between Knowledge discovery and data mining.
7. What are the application areas of data Mining?
8. List out different sources of information.
9. What type of benefit you might hope to get from data mining?
10. How can Data Mining help business analyst?

Unit:2

11. Classification is supervised learning. Justify.
12. Explain different classification Techniques.
13. Entropy is an important concept in information theory. Explain its significance in mining context.
14. What are over fitted models? Explain their effects on performance.
15. Explain Naive Baye's Classification.
16. List out the differences between OLTP and OLAP.
17. Discuss the various schematic representations in multidimensional model.
18. Explain the OLAP operations I multidimensional model.
19. Write notes on metadata repository.
20. Write short notes on VLDB.

Unit:3

21. Explain the issues regarding classification and prediction?
22. Explain classification by Decision tree induction?
23. Write short notes on patterns?
24. Is the data warehouse a prerequisite for data mining? Does the Data warehouse helps data mining. If so in what ways?
25. List out few common provisions to be found in a good security policy.
26. Give reasons why the data warehouse must be back up. How is this different from an OLTP system.
27. Describe various phases of testing Data Warehouse.
28. List out five reasons why you think data quality is critical in a Data Warehouse.
29. Explain how Data Quality is much more than just Data Accuracy. Give an example.
30. What are the different functions of data mining?

Unit:4

31. What is Data Aggregation and Generalization?
32. What are the various sources for data warehouse?
33. riefly discuss the schemas for multidimensional databases.
34. Define data warehouse? Differentiate between operational database systems and data warehouses?
35. Explain the architecture of data warehouse.
36. Discuss Extraction-Transformation-loading with neat diagram?
37. Discuss schemas for multi-dimensional tables?
38. Discuss OLAP operations?
39. List the three important issues that have to be addressed during data integration.
40. How do you clean the data?

Unit:5

41. Describe the Architecture of a typical data mining system/Major Components?
42. How is a data warehouse different from a database? How are they similar?
43. List out Data mining tasks?
44. What do you mean by Attribute sub selection / Feature selection?
45. Describe Apriori algorithm?
46. Illustrate FP-Growth algorithm?
47. Define outliers. List various outlier detection approaches.
48. Compare clustering and classification
49. What is meant by hierarchical clustering
50. What is classification?

NGM COLLEGE

DEPARTMENT OF COMPUTER SCIENCE (SF)

18PCS1E1 – DATA MINING

UNIT:1

1. Discuss the need of human intervention in data mining process.
2. What steps you would follow to identify a fraud for a credit card company.
3. Explain the differences between " Explorative Data Mining" and "Predictive Data Mining" and give one example of each.
4. State three different application for which data mining techniques seem appropriate. Informally explain each application.
5. Explain briefly the differences between "classification" and ''clustering" and give an informal example of an application that would benefit from each techniques.
6. How can we handle missing values?
7. How is data warehouse different from a database? How are they similar?

UNIT:2

8. Can you briefly describe the four stages of knowledge discovery(KDD)? Can you describe the multi-tiered data warehouse architecture?
9. 9. Describe a data set for which sampling would actually increase the amount of work. In other words it would be faster to work on full data set.
10. 10. Is support as defined in correlation rule paper Downward closed? Why?
11. 11.How large is a contingency table for itemset of N items .
12. 12. (i) With a neat sketch explain the architecture of a data warehouse
    (ii) Discuss the typical OLAP operations with an example.
13. (i) Discuss how computations can be performed efficiently on data cubes.
    (ii) Write short notes on data warehouse meta data.
14. (i) Explain various methods of data cleaning in detail.
    (ii) Give an account on data mining Query language.

UNIT:3

15. How is Attribute-Oriented Induction implemented? Explain in detail.
16. Write and explain the algorithm for mining frequent item sets without candidate generation. Give relevant example.
17. Describe the essential features of decision trees in context of classification.
18. What are the advantages and disadvantages of decision tress over other classification methods?
19. Explain ID3 Algorithm.
20. Explain the methods for computing best split.
21. What is Clustering? What are different types of clustering?

UNIT:4

22. Explain different data types used in clustering.
23. Explain mining single –dimensional Boolean associated rules from transactional databases?
24. Explain apriori algorithm?
25. Explain how the efficiency of apriori is improved?
26. Explain frequent item set without candidate without candidate generation?
27. Explain mining Multi-dimensional Boolean association rules from transaction
28. Explain constraint-based association mining?\

UNIT:5

29. State two clustering method thst are used in "grid and density based method?
30. Give examples of four types of data quality problems.
31. How does the data warehouse differ from an operational system in uses and value.

32. State Dr. Codd’s guidelines for OLAP system, giving a brief description for each.
33. Name any three advantages of the STAR schema . Can you think of any disadvantages of STAR Schema.
34. What are hierarchies and categories as applicable to a dimension table.
35. Why is Dimension table wide and the Fact table is deep?
36. Describe the composition of primary keys for the dimension and fact table.
37. What is the STAR Schema? What are the Fact tables.
38. How is Dimensional Modelling different.
39. Discuss the major design issues that need to be addressed before proceeding with the data design.