NOORUL ISLAM COLLEGE OF ENGINEERING, KUMARACOIL.

DEPARTMENT OF INFORMATION TECHNOLOGY

CS 1004 – DATA WAREHOUSING AND DATA MINING

2 MARKS QUESTIONS AND ANSWERS

1.What are the uses of statistics in data mining?

Statistics is used to
*to estimate the complexity of a data mining problem;
*suggest which data mining techniques are most likely to be successful; and
*identify data fields that contain the most "surface information".

2. What is the main goal of statistics?

The basic goal of statistics is to extend knowledge about a subset of a collection to the entire collection.

3. What are the factors to be considered while selecting the sample in statistics?

The sample should be
*Large enough to be representative of the population.
*small enough to be manageable.
*accessible to the sampler.
*free of bias.

4. Name some advanced database systems.

Object-oriented databases,Object-relational databases.

5. Name some specific application oriented databases.

Spatial databases,
Time-series databases,
Text databases and multimedia databases.

6. Define Relational datbases.

A relational databases is a collection of tables,each of which is assigned a unique name.Each table consists of a set of attributes(columns or fields) and usually stores a large set of tuples(records or rows).Each tuple in a relational table represents an object identified by a unique key and described by a set of attribute values.

7.Define Transactional Databases.

A transactional database consists of a file where each record represents a transaction.A transaction typically includes a unique transaction identity number(trans_ID), and a list of the items making up the transaction.

8.Define Spatial Databases.

Spatial databases contain spatial-related information.Such databases include geographic(map) databases,VLSI chip design databases, and medical and satellite image databases.Spatial data may be represented in raster format, consisting of n-dimensional bit maps or pixel maps.

9.What is Temporal Database?

Temporal database store time related data .It usually stores relational data that include time related attributes.These attributes may involve several time stamps,each having different semantics.

10.What is Time-Series databases?

A Time-Series database stores sequences of values that change with time,such as data collected regarding the stock exchange.

11.What is Legacy database?

A Legacy database is a group of heterogeneous databases that combines different kinds of data systems,such as relational or object-oriented databases,hierarchical databases,network databases,spread sheets,multimedia databases or file systems.

12. What is learning?

Learning denotes changes in the system that enables the system to do the same task more efficiently the next time.

Learning is making useful changes or modifying what is being experienced.

13. Why machine learning is done?

To understand and improve the efficiency of human learning.

To discover new things or structure that is unknown to human beings.

To fill in skeletal or computer specifications about a domain.

14. Give the components of a learning system.

1 Critic
2 Sensors
3 Learning Element
4 Performance Element
5 Effectors
6 Problem generators.

15. Give some of the factors for evaluating performance of a learning algorithm.

1 Predictive accuracy of a classifier.
2 Speed of a learner
3 speed of a classifier
4 Space requirements

16. What are the steps in the data mining process?

a. Data cleaning

b. Data integration
c. Data selection
d. Data transformation
e. Data mining
f. Pattern evaluation
g. Knowledge representation

17. Define data cleaning

Data cleaning means removing the inconsistent data or noise and collecting necessary information

18. Define data mining

Data mining is a process of extracting or mining knowledge from huge amount of data.

20. Define pattern evaluation

Pattern evaluation is used to identify the truly interesting patterns representing knowledge based on some interesting measures.

21. Define knowledge representation

Knowledge representation techniques are used to present the mined knowledge to the user.

22. What is Visualization?

Visualisation is for depiction of data and to gain intuition about data being observed. It assists the analysts in selecting display formats, viewer perspectives and data representation schema

23. Name some conventional visualization techniques

- Histogram
- Relationship tree
- Bar charts
- Pie charts
- Tables etc.

24. Give the features included in modern visualisation techniques

a. Morphing
b. Animation
c. Multiple simultaneous data views
d. Drill-Down
e. Hyperlinks to related data source

25. Define conventional visualisation

Conventional visualisation depicts information about a population and not the population data itself

26. Define Spatial Visualisation

Spatial visualisation depicts actual members of the population in their feature space

27.What is Descripive and predictive data mining?

Descriptive datamining describes the data set in a concise and summarative manner and presents interesting general properties of the data.

Predictive datamining analyzes the data in order to construct one or set of models and attempts to predict the behavior of new data sets.

28.What is Data Generalization

It is process that abtracts a large set of task-relevent data in a database from a relatively low conceptual to higher conceptual levels

2 approachs for Generalization

1)Datacube approach

2)Attribute-oriented induction approach

29.Define Attribute Oriented Induction

These method collets the task-relevant data using a relational database query and then perform generalization based on the examination in the relevant set of data.

30. What is Jack Knife?

It's a bias reduction tool for eliminating low order bias from an estimator. The essence of the procedure is to replace the original 'n' observations by 'n' more correlated estimates of the quantity of interest. These are obtained by systematically leaving out one or more observations and re-computing the estimator.

31. What is boot strap?

An interpretation of the jack knife is that the construction of pseudo value is based on repeatedly and systematically sampling with out replacement from the data at hand. This lead to generalized concept to repeated sampling with replacement called boot strap.

32. View of statistical approach?

Statistical method is interested in interpreting the model. It may sacrifice some performance to be able to extract meaning from the model structure. If accuracy is acceptable then the reason that a model can be decomposed in to revealing parts is often more useful than a 'black box' system, especially during early stages of investigation and design cycle.

33. What are the assumptions of statistical analysis?

The assumptions are

- Residuals
- Diagnostics
- Parameter Covariance

34. What is the use of Probabilistic graphical model?

Probabilistic graphical model are a frame work for structuring, representation and decomposing a problem using the notation of conditional independence.

35. What is the importance of Probabilistic graphical model?

- They are a lucid representation for a variety of problems, allowing key dependencies with in a problem to be expressed and irrelevancies to be ignored
- It performs problem formulation and decomposition
- Helps in designing a learning algorithm
- It identifies valuable knowledge
- It generates explanation

36. Define Deterministic models?

Deterministic models, which takes no account of random variables, but gives precise, fixed reproducible output.

37. Define Systems and Models?

System is a collection of interrelated objects and Model is a description of a system. Models are abstract, and conceptually simple.

38. How do you choose the best model?

All things being equal, the smallest model that explains the observations and fits the objectives that should be accepted. In reality, the smallest means the model should optimizes a certain scoring function (e.g. Least nodes, most robust, least assumptions)

39. Principles of Qualitative Formulation

- Model Simplification
- Minimize state variables
- Convert a variable into a constant aggregate state variable
- Make stronger assumptions
- Remove temporal complexity
- Remove spatial complexity

40. General properties of Boolean Networks

- Fixed topology
- Dynamic synchronous Node States
- Gate function
- Synergetic behavior

41. What is clustering?

Clustering is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters.

42. What are the requirements of clustering?

* Scalability
* Ability to deal with different types of attributes
* Ability to deal with noisy data
* Minimal requirements for domain knowledge to determine input parameters
* Constraint based clustering
* Interpretability and usability

43. State the categories of clustering methods?

*Partitioning methods
*Hierarchical methods
*Density based methods
*Grid based methods
*Model based methods

44. What is linear regression?

In linear regression data are modeled using a straight line. Linear regression is the simplest form of regression. Bivariate linear regression models a random variable Y called response variable as a linear function of another random variable X, called a predictor variable.

$$Y = a + b X$$

45. State the types of linear model and state its use?

Generalized linear model represent the theoretical foundation on which linear regression can be applied to the modeling of categorical response variables. The types of generalized linear model are

(i) Logistic regression
(ii) Poisson regression

46. What are the goals of Time series analysis?

1.Finding Patterns in the data
2.Predicting future values

47. What is smoothing?

Smoothing is an approach that is used to remove nonsystematic behaviors found in a time series. It can be used to detect trends in time series.

48. What is lag?

The time difference between related items is referred to as lag.

49. Write the preprocessing steps that may be applied to the data for classification and prediction.

a. Data Cleaning
b. Relevance Analysis
c. Data Transformation

50. Define Data Classification.

It is a two-step process. In the first step, a model is built describing a pre-determined set of data classes or concepts. The model is constructed by analyzing database tuples described by attributes. In the second step the model is used for classification.

51. What are Bayesian Classifiers?

Bayesian Classifiers are statistical classifiers. They can predict class member-ship probabilities, such as the probability that a given sample belongs to a particular class.

52. Describe the two common approaches to tree pruning.

In the prepruning approach, a tree is "pruned" by halting its construction early. The second approach, postpruning, removes branches from a "fully grown" tree. A tree node is pruned by removing its branches.

53. What is a "decision tree"?

It is a flow-chart like tree structure, 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.

Decision tree is a predictive model. Each branch of the tree is a classification question and leaves of the tree are partition of the dataset with their classification.

54. What do you meant by concept hierarchies?

A concept hierarchy defines a sequence of mappings from a set of low-level concepts to higer-level, more general concepts. Concept hierarchies allow specialization,or drilling down ,where by concept values are replaced by lower-level concepts.

55. Where are decision trees mainly used?

Used for exploration of dataset and business problems
Data preprocessing for other predictive analysis
Statisticians use decision trees for exploratory analysis

56. How will you solve a classification problem using decision trees?

a. **Decision tree induction:**

Construct a decision tree using training data

b. For each $\mathbf{t_i \in D}$ apply the decision tree to determine its class
t_i - tuple
D - Database

57. What is decision tree pruning?

Once tree is constructed , some modification to the tree might be needed to improve the performance of the tree during classification phase.

The pruning phase might remove redundant comparisons or remove subtrees to achieve better performance.

58. Explain ID3

ID3 is algorithm used to build decision tree. The following steps are followed to built a decision tree.

a. Chooses splitting attribute with highest information gain.
b. Split should reduce the amount of information needed by large amount.

59. What is Association rule?

Association rule finds interesting association or correlation relationships among a large set of data items which is used for decision-making processes. Association rules analyzes buying patterns that are frequently associated or purchased together.

60. Define support.

Support is the ratio of the number of transactions that include all items in the antecedent and consequent parts of the rule to the total number of transactions. Support is an association rule interestingness measure.

61. Define Confidence.

Confidence is the ratio of the number of transactions that include all items in the consequent as well as antecedent to the number of transactions that include all items in antecedent. Confidence is an association rule interestingness measure.

62. How are association rules mined from large databases?

Association rule mining is a two-step process.

- Find all frequent itemsets.
- Generate strong association rules from the frequent itemsets.

63. What is the classification of association rules based on various criteria?

1. Based on the types of values handled in the rule.
    a. Boolean Association rule.
    b. Quantitative Association rule.
2. Based on the dimensions of data involved in the rule.
    a. Single Dimensional Association rule.
    b. Multi Dimensional Association rule.
3. Based on the levels of abstractions involved in the rule.
    a. Single level Association rule.
    b. Multi level Association rule.
4. Based on various extensions to association mining.
    a. Maxpatterns.
    b. Frequent closed itemsets.

64. What is Apriori algorithm?

Apriori algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules using prior knowledge. Apriori algorithm uses prior knowledge of frequent itemset properties and it employs an iterative approach known as level-wise search where k-itemsets are used to explore (k+1)-itemsets.

65. What are the advantages of Dimensional modelling?

- Ease of use.

- High performance
- Predictable,standard framework
- Understandable
- Extensible to accomodate unexpected new data elements and new design decisions

66. Define Dimensional Modelling?

Dimensional modelling is a logical design technique that seeks to present the data in a standard framework that intuitive and allows for high-performance access.It is inherently dimensional and adheres to a discipline that uses the relational model with some important restrictions.

67. What comprises of a dimensional model?

Dimensional model is composed of one table with a multipart key called fact table and a set of smaller tables called dimension table.Each dimension table has a single part primary key that corresponds exactly to one of the components of multipart key in the fact table.

68. Define a datamart?

Data mart is a pragmatic collection of related facts,but does not have to be exhaustive or exclusive.A datamart is both a kind of subject area and an application.Data mart is a collection of numeric facts.

69. What are the advantages of a data modelling tool?

- Integrates the datawarehouse model with other corporate data models.
- Helps assure consistency in naming.
- Creates good documentation in a variety of useful formats.
- Provides a reasonably intuitive user interface for entering comments about objects.

70. What is datawarehouse performance issue?

The performance of a data warehouse is largely a function of the quantity and type of data stored within a database and the query/data loading work load placed upon the system.

71. What are the types of performance issue?

1.Capacity planning for the data warehouse
2.data placement techniques within a data warehouse
3.Application Performance Techniques.
4. Monitoring the Data Warehouse.

72. What is Data Inconsistency Cleaning?

This can be summarized as the process of cleaning up the small inconsistencies that introduce themselves into the data. Examples include duplicate keys and unreferenced foreign keys.

73. What is Column Level Cleaning?

This involved checking the contents of each column field and ensuring it conforms to a set of valid values. For example convert EBCDIC to ASCII.

74. What is Bottleneck Detection?

Bottleneck Detection is the process where by the database administrator will detect for what reason the database performance level has reached a plateau. To increase the performance from this plateau may require the addition of more hardware resources or reconfiguration of the system software(O/S, RDBMS or Application).

75. What is back room Meta data?

Back room Meta data is process related, and it guides the extraction, cleaning and loading process.

76. What is Front Room Meta data?

It is more descriptive and it helps query tools and report writers function smoothly.

77. What are the specifications of source system Meta data?

- Repositories
- Source Scheme
- Copy Books
- Spread Sheet Sources
- Lotus notes database

78. What is active Meta data?

Active Meta data is Meta data that drives a process rather than documents.

79. What is Meat data catalogue?

Meat data catalogue is a single common storage point for information that drives the entire warehouse process.

80. Why do you need data warehouse life cycle process?

Data warehouse life cycle approach is essential because it ensures that the project pieces are brought together in the right order and at the right time.

81. What are the steps in the life cycle approach?

*Project Planning
*Business Requirements definition
*Data track: Dimensional modeling, Physical Design, Data Staging Design & Development
*Technology track: Technical Architecture design, Product Selection & Installation
*Application track: End user Application Specification, End user Application Development
*Deployment
*Maintenance & Growth
*Project Management

82.Merits of Data Warehouse.

*Ability to make effective decisions from database
*Better analysis of data and decision support
*Discover trends and correlations that benefits business

*Handle huge amount of data.

83.What are the characteristics of data warehouse?

*Separate
*Available
*Integrated
*Subject Oriented
*Not Dynamic
*Consistency
*Iterative Development
*Aggregation Performance

84.List some of the DataWarehouse tools?

*OLAP(OnLine Analytic Processing)
*ROLAP(Relational OLAP)
*End User Data Access tool
*Ad Hoc Query tool
*Data Transformation services
*Replication

85.Explain OLAP?

The general activity of querying and presenting text and number data from DataWarehouses, as well as a specifically dimensional style of querying and presenting that is exemplified by a number of "OLAP Vendours" .The OLAP vendours technology is nonrelational and is almost always biased on an explicit multidimensional cube of data.OLAP databases are also known as multidimensional cube of databases.

86.Explain ROLAP?

ROLAP is a set of user interfaces and applications that give a relational database a dimensional flavour.ROLAP stands for Relational Online Analytic Processing.

87.Explain End User Data Access tool?

End User Data Access tool is a client of the data warehouse.In a relational data warehouse,such a client maintains a session with the presentation server,sending a stream of separate SQL requests to the server.Evevtually the end user data access tool is done with the SQL session and turns around to present a screen of data or a report,a graph,or some other higher form of analysis to the user.An end user data access tool can be as simple as an Ad Hoc query tool or can be complex as a sophisticated data mining or modeling application.

88.Explain Ad Hoc query tool?

A specific kind of end user data access tool that invites the user to form their own queries by directly manipulating relational tables and their joins.Ad Hoc query tools ,as powerful as they are ,can only be effectively used and understood by about 10% of all the potential end users of a data warehouse.

89.Name some of the data mining applications?

- Data mining for Biomedical and DNA data analysis
- Data mining for Financial data analysis
- Data mining for the Retail industry
- Data mining for the Telecommunication industry

90.What are the contribution of data mining to DNA analysis?

- Semantic integration of heterogeneous,distributed genome databases
- Similarity search and comparison among DNA sequences
- Association analysis: identification of co-occuring gene sequences
- Path analysis: linking genes to different stages of disease development
- Visualization tools and genetic data analysis

91.Name some examples of data mining in retail industry?

- Design and construction of data warehouses based on the benefits of data mining
- Multidimensional analysis of sales,customers,products,time and region
- Analysis of the effectiveness of sales campaigns
- Customer retention-analysis of customer loyalty
- Purchase recommendation and cross-reference of item

92. Name some of the data mining applications

- Data mining for Biomedical and DNA data analysis
- Data mining for Financial data analysis
- Data mining for the Retail industry
- Data mining for the Telecommunication industry

93. What is the difference between "supervised" and unsupervised" learning scheme.

In data mining during classification the class label of each training sample is provided, this type of training is called supervised learning (i.e) the learning of the model is supervised in that it is told to which class each training sample belongs. Eg.:Classification

In unsupervised learning the class label of each training sample is not known and the member or set of classes to be learned may not be known in advance. Eg.:Clustering

94. Discuss the importance of similarity metric clustering? Why is it difficult to handle categorical data for clustering?

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. Similarity metric is important because it is used for outlier detection. The clustering algorithm which is main memory based can operate only on the following two data structures namely,

A) Data matrix

B) Dissimilarity matrix

So it is difficult to handle categorical data.

95. Mention atleast 3 advantages of Bayesian Networks for data analysis. Explain each one.

a) Bayesian Network is a graphical representation of unknown knowledge that is easy to construct and interpret.

b) the representation has formal probabilistic semantics, making it suitable for statistical manipulation.

c) The representation is used for encoding uncertain expert knowledge in expert systems.

96. Why do we need to prune a decision tree? Why should we use a separate pruning data set instead of pruning the tree with the training database?

When a decision tree is built, many of the branches will reflect animation in the training data due to noise or outliers. Tree pruning methods are needed to address this problem of overfitting the data.

97. Explain the various OLAP operations.

a) Roll-up: The roll-up operation performs aggregation on a data cube, either by climbing up a concept hierarchy for a dimension.

b) Drill-down: It is the reverse of roll-up. It navigates from less detailed data to more detailed data.

c) Slice: Performs a selection on one dimension of the given cube, resulting in a subcube.

98. Discuss the concepts of frequent itemset, support & confidence.

A set of items is referred to as itemset. An itemset that contains k items is called k-itemset. An itemset that satisfies minimum support is referred to as frequent itemset.

Support is the ratio of the number of transactions that include all items in the antecedent and consequent parts of the rule to the total number of transactions.

Confidence is the ratio of the number of transactions that include all items in the consequent as well as antecedent to the number of transactions that include all items in antecedent.

99. Why is data quality so important in a data warehouse environment?

Data quality is important in a data warehouse environment to facilitate decision-making. In order to support decision-making, the stored data should provide information from a historical perspective and in a summarized manner.

100. How can data visualization help in decision-making?

Data visualization helps the analyst gain intuition about the data being observed. Visualization applications frequently assist the analyst in selecting display formats, viewer perspective and data representation schemas that faster deep intuitive understanding thus facilitating decision-making.

101. What do you mean by high performance data mining?

Data mining refers to extracting or mining knowledge. It involves an integration of techniques from multiple disciplines like database technology, statistics, machine learning, neural networks, etc. When it involves techniques from high performance computing it is referred as high performance data mining.

16 Marks Questions

1. Explain the various datamining issues?
   Ans:
   Explain about
   - Knowledge Mining
   - User interaction
   - Performance
   - Diversity in datatypes

2. Explain the datamining functionalities?
   Ans:
   The datamining functionalities are:
   - Concept class description
   - Association analysis
   - Classification and prediction
   - Cluster Analysis
   - Outlier Analysis

3. Explain the different types of data repositories on which mining can be performed?
   Ans:
   The different types of data repositories on which mining can be performed Are:
   - Relational Databases
   - DataWarehouses
   - Transactional Databases
   - Advanced Databases
   - Flat files
   - World Wide Web

4. Explain the architecture of data warehouse.

   --steps for the design and construction of DW
   Top-down view- data source view-data warehouse view-business query view
   --3tier DW architecture

5. Explain indexing techniques of OLAP data, with example.
   Bitmap indexing and join indexing

6. What is Data Mining? Explain the steps in Knowledge Discovery?
   Ans:

Datamining refers to extracting or mining knowledge from large amount of data. The steps in knowledge discovery are:

- Data cleaning
- Data integration
- Data selection
- Data transformation
- Data mining
- Pattern Evolution
- Knowledge Discovery.

7. Explain the data pre-processing techniques in detail?

Ans:

The data preprocessing techniques are:

- Data Cleaning
- Data integration
- Data transformation
- Data reduction

8. Explain the smoothing Techniques?

Ans:

- Binning
- Clustering
- Regression

9. Explain Data transformation in detail?

Ans:

- Smoothing
- Aggregation
- Generalisation
- Normalisation
- Attribute Construction

10. Explain Normalisation in detail?

Ans:

- Min Max Normalisation
- Z-Score Normalisation
- Normalisation by decimal scaling

11. Explain data reduction?

Ans:

- Data cube Aggretion
- Attribute subset Selection
- Dimensional reduction
- Numerosity reduction

12. Explain parametric methods and non-parametric methods of reduction?

Ans:

Parametric Methods:

● Regression Model ● Log linear Model

Non-Parametric Methods

● Sampling ● Histogram ● Clustering

13. Explain Data Discretization and Concept Hierarchy Generation?

Ans:

Discretization and concept hierarchy generation for numerical data:

- Segmentation by natural partitioning
- Binning
- Histogram Analysis
- Cluster Analysis

14. Explain Datamining Primitives?

Ans:

There are 5 Data mining Primitives.They are:

- Task relevant data
- Kinds of knowledge to be mined
- Concept Hierarcies
- Interesting Measures
- Knowledge Presentation and Visualization Technique to be used for Discovery patterns

15. Explain Attribute Oriented Induction?

Ans:

Explain:

- Attribute oriented indution for data characterzation
- Algorithm
- Presentation of derived generalization
- Example

16. Explain Statistical measures in data bases?

- Measuring the central tendency
- Measuring the dispersion of data
- Graph displays

17. Explain the apriori algorithm for finding frequent itemsets?

Ans:

- Algorithm
- Example
- Explanation

18. Explain the apriori algorithm for finding frequent itemsets without candidate generation?

Ans:

- Algorithm

●Example
●Explanation

19. Explain Multilevel association rule?
Ans:

●Example
●Explanation
●Variations

20. Explain any one of the Data Mining tools?
Ans:
Explain about OLE-DB Data Mining Tool

21. Explain Multidimentional Database briefly?
Ans:
●Star schema
●Snowflake schema
●Fact constellation
Explain with examples for defining star, snowflake, fact constellation schema And Diagrams.

22. Explain Indexing with suitable examples?
Ans:
Bitmap Indexing
Join Indexing
Bitmapped join indexing

23. Explain the Back Propagation technique?
Ans:
● Definition
● Back Propagation Algorithm & diagram
●Example

24. Explain Bayessian classification?
Ans:
●Bayes Theorem with example and steps.
●Bayesian belief networks

25. Explain various classification methods?
- K-nearest neighbor
- Case-Based reasoning
- Genetics
- Fuzzy logic

26. Explain Classifier accuracy with examples?

- Hold out method
- K-fold cross-validation method
- Bagging
- Boosting
- Sensitivity
- Specificity
- Precision

27. Explain Partition Methods?

Ans:

Explain

●K-Means Partition

● K-Medoids Partition

●CLARANS method

With examples.

28. Explain Hierarchical method of classifications?

Ans:

Explain

● Agglomerative hierarchical clustering

●Divisive hierarchical clustering

● BIRCH

●Chameleon

●CURE

29. Explain classification by Decision tree induction?

Ans:

●Explain the steps in decision tree induction

●Generation of decision tree algorithm

● Example and diagram

● Tree pruning

30. Explain the Time Series Analysis?

Ans:

A time series database consists of sequences of values or events obtained over repeated measurements of time.

●Explain about trend analysis

-Consists of 4 components

1.long-term movement

2.cyclic movement

3.seasonal movement

4.random movement

●Explain with example problem

31. Explain the types of data in cluster analysis.

Data matrix—dissimilarity matrix

Interval scaled variables—Binary variables—Nominal, Ordinal and Ratio scaled variables

32. Explain Outlier analysis?

- Statistical based outlier detection
- Distance based outlier detection
- Deviation based outlier detection

33. Explain Mining complex types of data?

- Multidimensional analysis and descriptive mining
- Mining spatial databases
- Mining Multimedia databases
- Mining Text databases
- Mining Time-series and sequence data
- Mining WWW

34. Briefly explain about Data Mining Application?

Ans:

1.Financial Data Analysis
2.Retail Industry
3.Telecommunication Industry
4.Biological Data Analysis
5.Scientific Application

35. Explain social impacts of data mining?

- Innovators
- Early adopters
- Chasm
- Early majority
- Late majority
- Laggards

36. Explain Additional themes in data mining?

- Audio and visual mining
- Scientific and statistical data mining

CS 1004

DATA WAREHOUSING AND DATA MINING

QUESTION BANK

Compiled by
R. Mathu Soothana S. Kumar
Department of Information Technology
N.I. College of Engineering
Kumaracoil