DATA WAREHOUSING AND DATA MINING

Subject Code: **A70520**
Regulations: R15 – JNTUH
Class: IV Year B.Tech CSE I Semester

Department of Computer Science and Engineering

BHARAT INSTITUTE OF ENGINEERING AND TECHNOLOGY

Ibrahimpatnam - 501 510, Hyderabad

DATA WAREHOUSING AND DATA MINING (A70520)
COURSE PLANNER

I. Course Objective

This course will introduce the concepts of data ware house and data mining, which gives a complete description about the principles, used, architectures, applications, design and implementation of data mining and data ware housing concepts.

II. PRE- REQUISITES:

The knowledge of following subject is essential to understand the subject:

1. **Understand** the concepts of Data Ware housing and Data Mining Concepts.
2. **Explain** the methodologies used for analysis of data
3. **Describe** various techniques which enhance the data modeling.
4. **Discuss** and Compare various approaches with other techniques in data mining and data ware housing

III. COURSE OBJECTIVE:

1	Be familiar with mathematical foundations of data mining tools..
2	Understand and implement classical models and algorithms in data warehouses and data mining
3	Characterize the kinds of patterns that can be discovered by association rule mining, classification and clustering.
4	Master data mining techniques in various applications like social, scientific and environmental context.
5	Develop skill in selecting the appropriate data mining algorithm for solving practical problems.

IV. COURSE OUTCOME:

S.No	Description	Bloom's Taxonomy Level
1	**Understand** the functionality of the various data mining and data warehousing component	**Knowledge, Understand**
2	**Appreciate** the strengths and limitations of various data mining and data warehousing models	**Apply, Create**
3	**Explain** the analyzing techniques of various data	**Analyze**
4	**Describe** different methodologies used in data mining and data ware housing.	**Analyze**
5	**Compare** different approaches of data ware housing and data mining with various technologies.	**Evaluating**

Course Outcome :

V. HOW PROGRAM OUTCOMES ARE ASSESSED:

Program Outcomes (PO)		Level	Proficiency assessed by
PO1	**Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems related to Computer Science and Engineering.	**3**	Assignments

PO2	**Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems related to Computer Science and Engineering and reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.	**3**	Assignments
PO3	**Design/development of solutions**: Design solutions for complex engineering problems related to Computer Science and Engineering and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.	**2**	Assignments
PO4	**Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.	**3**	Assignments
PO5	**Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.	**--**	--
PO6	**The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the Computer Science and Engineering professional engineering practice.	**1**	Assignments
PO7	**Environment and sustainability**: Understand the impact of the Computer Science and Engineering professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.	**2**	--
PO8	**Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.	**-**	--
PO9	**Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.	**-**	--
PO10	**Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.	**-**	--
PO11	**Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.	**-**	--
PO12	**Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.	**2**	Research

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High) - : None

VI. HOW PROGRAM SPECIFIC OUTCOMES ARE ASSESSED

Program Specific Outcomes (PSO)	Level	Proficiency assessed by

PSO1	**Foundation of mathematical concepts:** To use mathematical methodologies to crack problem using suitable mathematical analysis, data structure and suitable algorithm.	**3**	Lectures, Assignments
PSO2	**Foundation of Computer System:** The ability to interpret the fundamental concepts and methodology of computer systems. Students can understand the functionality of hardware and software aspects of computer systems.	**2**	Lectures, Assignments
PSO3	**Foundations of Software development:** The ability to grasp the software development lifecycle and methodologies of software systems. Possess competent skills and knowledge of software design process. Familiarity and practical proficiency with a broad area of programming concepts and provide new ideas and innovations towards research.	--	--

1: Slight (Low) **2: Moderate (Medium)** **3: Substantial (High)** **- : None**

COURSE CONTENT:

JNTUH SYLLABUS

UNIT – I

Data Warehouse: Introduction to Data Ware House, Differences between operational data base systems and data Ware House, Data Ware House characteristics, Data Ware House Architecture and its components, Extraction-Transformation-Loading, Logical (Mulit-Dimensional), Data Modeling, Schema Design, star and snow-Flake Schema, Fact Constellation, Fact Table, Fully Addictive, Semi-Addictive, Non-Addictive Measures; Fact-Less-Facts, Dimension Table characteristics; Fact-Less-Facts, Dimension Table characteristics; OLAP cube, OLAP Operations, OLAP Server Architecture-ROLAP, MOLAP and HOLAP.

UNIT -II

Introduction to Data Mining: Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing- Data Cleaning, Missing Data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binaryzation , Data Transformation; Measures of similarity and dissimilarity-Basics.

UNIT – III

Association Rules: Problem Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation, APRIORI Algorithm, The Partition Algorithms, FP-Growth Algorithms, Compact Representation of Frequent Item Set-Maximal Frequent Item Set, Closed Frequent Item Set.

UNIT -IV

Classification: Problem definition, General Approaches to solving a classification problem, Evaluation of Classifiers, Classification techniques, Decision trees-Decision Tree Construction, Methods for expressing attribute test conditions, Measures for Selecting the Best split, Algorithm for Decision tree Induction, Naïve-Bayes Classifier, Bayesian Belief Networks; K-nearest neighbor classification-Algorithm and characteristics.

UNIT – V

Clustering: Problem Definition, Clustering overview, Evaluation of clustering algorithms, Partitioning clustering K-Means Algorithm, K-Means Additional Issues, PAM Algorithm, Hierarchical Clustering-Algorithm- Agglomerative Methods and Divisive Methods, Basic Agglomerative Hierarchical Clustering Algorithm, Specific techniques, Key Issues in Hierarchical Clustering, Strengths and weakness, Outlier Detection

GATE SYLLABUS: NOT APPLICABLE

IES SYLLABUS: NOT APPLICABLE

LESSON PLAN:

Lecture	Week	Topic	Course Learning Outcome	Reference
	UNIT - 1			
1.	1	**Introduction to Data warehouse**		Text Book No. 1
2.		Difference between operational database systems and data warehouses	**Analyze** about the operational database management and data warehouse	
3.		Data warehouse Characteristics, Data warehouse Architecture and its components	**Create** the architecture of Data warehouse	
4.		Extraction-Transformation-Loading, Logical(Multi-Dimensional)	**Understanding** Knowledge of data warehouse architecture	
		Tutorial		
5.	2	Data Modeling, Schema Design	**Creating** a Data Model	
6.		Star and snow-Flake schema, Fact Consultation	**Creating** the different types of schemas	
7.		Fact Table, Fully Adictive, Semi-Addictive, on-Addictive Measures	**Remembering** different types of Fact Tables	
8.		Fact-Less-Facts, Dimension Table Characteristics,	**Remembering** different types of Dimension Tables	
		Tutorial		
9.	3	OLAP Cube	**Creating** a OLAP cube	
10.		OLAP Operations,OLAP-Server Architecture-ROLAP	**Applying** different types of OLAP operations	
11.		MOLAP and HOLAP	**Applying** different types of OLAP operations	
12.		Review of Unit-I		
		Mock Test – I		
		UNIT – 2		
13.	4	Introduction to Data Mining, Definition of Data Mining	**Understanding** and **Remembering** the definition of data Mining	Text Book No. 1
14.		KDD, Challenges	**Understanding** KDD	
15.		Data Mining Task	**Analyzing** the different data mining Task	
16.		Data Preprocessing, Data Cleaning	**Understanding** the process of DDta preprocessing and Data Cleaning	
		Tutorial		
17.	5	Data Cleaning	**Applying** the data cleaning process	

Lecture	Week	Topic	Course Learning Outcome	Reference
18.		Dimensionality Reduction	**Analyzing** Dimensionality Reduction	
19.		Feature Subset Selection	**Understanding** Subset selection	
20.		Discretization and Binaryzation	**Understanding** Discretization and Binaryzation	
		Tutorial		
21.	6	Data Transformation	**Analyzing** Data Transformation	
22.		Measures of Similarity-Basics	**Analyzing** Similarity – Basics and Dissimilarity Basics	
23.		Measures of Dissimilarity-Basics		
24.		Revision of Unit-II		
		Tutorial		
		UNIT – 3		
25.	7	Association Rules :Problem Definition	**Understanding** Association Rules	Text Book No. 1
26.		Frequent Item Set Generation	**Analyzing** Frequent Item set Generation	
27.		The APRIORI Principle	**Understanding** APRIORI principal, support and confidence Measures	
28.		Support and Confidence Measures		
		Tutorial		
29.	8	Association Rule Generation	**Evaluating Association Rule**	
30.		APRIORI Algorithm	**Analyzing** APRIORI Algorithm	
31.		Revision		
32.		Revision		
		Tutorial		
		I Mid Examinations (Week 9)		
		UNIT– 3 Contd.		
33.	10	The Partition Algorithms	**Understanding** the Partition Algorithms	Text Book No. 1, 3
34.		FP-Growth Algorithms	**Understanding** FP-Growth Algorithms	
35.		Compact Representation of Frequent Item-Set-Maximal Frequent Item Set	**Creating** a Representation of Frequent Item-set and Maximal, Closed Frequent Item set	
36.		Closed Frequent Item Set , Revision of Unit-III		
		Tutorial		
		UNIT – 4		
37.	11	Classification: Problem Definition	**Understanding the Classification**	Text Book No.

Lecture	Week	Topic	Course Learning Outcome	Reference
38.		General Approaches to Solving a Classification Problem	**Evaluating** the Classification problem	
39.		Evaluation of Classifiers ,Classification Techniques	**Evaluating** the classifiers and classification Techniques	
40.		Decision-Trees-Decision tree Construction	**Creating** Decision Tree	
		Tutorial		
41.	12	Methods for Expressing attribute test conditions	**Understanding** attribute test conditions	
42.		Measures for selecting the best split	**Evaluating** for selecting the best split	
43.		Algorithm for Decision tree induction		
44.		Naïve-Bayes Classifier	**Understanding** Naïve – Bayes Classifier	
		Tutorial		
45.	13	Bayesian Belief Networks	**Understanding** Bayesian Belief Networks	
46.		K-Nearest Neighbor classification-Algorithm and characteristics	**Understanding** K-Nearest Neighbor classification-Algorithm	
47.		K-Nearest Neighbor classification-Algorithm and characteristics Continuation		
48.		Revision		
		Mock Test – II		
	UNIT – 5			
49.	14	Clustering: Problem Definition	**Understanding c**lustering	Book No. 1, 2, 4
50.		Clustering Overview, Evaluation of Clustering Algorithms	**Evaluating** Clustering Algorithms, Partitioning, K-Means Algorithm	
51.		Partitioning Clustering		
52.		K-Means Algorithm		
		Tutorial		
53.	15	K-Means Additional Issues	**Understanding** K-Means	
54.		K-Means Additional Issues Cont…		
55.		PAM Algorithm	**Evaluating** PAM Algorithm	
56.		PAM Algorithm Continuation		
57.	16	Hierarchical clustering Methods	**Understanding** Hierarchical clustering Methods	
58.		Hierarchical clustering		

Lecture	Week	Topic	Course Learning Outcome	Reference
		Methods Continuation		
59.		Agglomerative Hierarchical clustering Algorithm	**Understanding and Evaluating** Agglomerative Hierarchical clustering Algorithm	
60.		Specific Techniques		
61.	17	Issues in Hierarchical clustering	**Analyzing** Issues in Hierarchical clustering	
62.		Issues in Hierarchical clustering Continuation		
63.		Strengths and Weakness	**Analyze** Strengths and Weakness of clustering	
64.		Outlier Detection	**Understanding** Outlier Detection	
		Tutorial		
		II Mid Examinations (Week 18)		

SUGGESTED BOOKS:

TEXT BOOK:

1) Data Mining-Concepts and Techniques- Jiawei Han, Micheline Kamber, Morgan Kaufmann Publishers, Elsevier, 2 Edition, 2006.
2) Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbanch, Pearson Education.

REFERENCES:

1) Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
2) Data Ware Housing Fundamentals, Pualraj Ponnaiah, Wiley Student Edition.
3) The Data Ware House Life Cycle Toolkit- Ralph Kimball, Wiley Student Edition.
4) Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University.

IX. MAPPING COURSE OUTCOMES LEADING TO THE ACHIEVEMENT OF PROGRAM OUTCOMES AND PROGRAM SPECIFIC OUTCOMES:

	Program Outcomes												Program Specific Outcomes		
	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10	PO 11	PO 12	PSO 1	PSO 2	PSO 3
CO 1	3	3	-	-	-	-	-	-	-	-	-	-	2	-	-
CO 2	-	2	-	-	-	-	-	-	-	-	-	-	-	2	-
CO 3	-	-	2	3	-	-	-	-	-	-	-		2	-	-
CO 4	-	-	-	3	-	-	-	-	-	-	-	2	3	-	-
CO 5	2	2	-	-	-	-	-	-	-	-	-			2	
AVG	2.5	2.5	2	3	0	-	-	-	-	-	-	2	3.5	2	0

DESCRIPTIVE QUESTIONS:

UNIT-I

Short Answer Questions

S. No	Question	Bloom Level
1	Explain the definition of Data Ware House	Understand
2	List the key features of Data Ware house?	Knowledge
3	Define Data Mart?	Knowledge
4	List the various multi dimensional models?	Knowledge
5	Name the OLAP operations and explain about various schemas?	Analyze

Long Answer Questions

S. No	Question	Bloom Level
1	Distinguish between the data ware house and databases? How they are similar?	Understand
2	Discuss briefly about the architecture of Data Ware house and its components?	Knowledge
3	Demonstrate the efficient processing of OLAP queries?	Analyze
4	Explain about Multi dimensional data models?	Knowledge
5	Discuss briefly with an example about multi dimensional schemas?	Analyze

UNIT-2

Short Answer Questions

S. No	Question	Bloom Level
1	Define data mining?	Understand
2	Explain the functionalities of data mining?	Understand
3	Interpret the major issues in data mining?	Knowledge
4	Name the steps in knowledge discovery?	Knowledge
5	Distinguish between data ware house and data mining?	Analyze

Long Answer Questions

S. No	Question	Bloom Level
1	Describe Data Mining? In your answer explain the following: a. Is it another hype? b. Is it simple transformation of technology developed from databases, statistics and machine learning? c. Explain how the evolutions of database technology lead to data mining? d. Describe the steps involved in data mining when viewed as knowledge discovery process?	Understanding
2	Discuss briefly about data smoothing techniques?	Creating
3	List and describe the five primitives for specifying the data mining tasks?	Analyzing
4	Define data cleaning? Express the different techniques	Understanding

	in handling the missing values?	
5	Explain mining of huge amount of data (eg: billions of tuples) in comparison with mining a small amount of data (Eg: data set of few hundred of tuples).	**Analyzing**

UNIT-3

Short Answer Questions

S. No	Question	Bloom Level
1	Explain the frequent item set?	**Understanding**
2	2. Explain about maximal frequent items set and closed item set?	**Knowledge**
3	Name the steps in association rule mining?	**Understand**
4	Explain the efficiency of APRIORI algorithm	**Analyze**
5	Define item set? Interpret the support and confidence rules for item set A and item set B?	**Understand**

Long Answer Questions

S. No	Question	Bloom Level
1	Discuss which algorithm is an influential algorithm for mining frequent item sets for Boolean association rules? Explain with an example?	**Analysis**
2	Describe the FP-growth algorithm with an example?	**Analysis**
3	Explain how to mine frequent item sets using vertical data format?	**Understand**
4	Explain how to mine the multi dimensional association rules from relational data bases and data ware houses?	**Understand**
5	Explain the APRIORI algorithm with an example?	**Analysis**

UNIT-4

Short Answer Questions

S. No	Question	Bloom Level
1	State classification and define regression analysis?	**Understand**
2	Name the steps in data classification and define training tuple?	**Knowledge**
3	Explain the IF-THEN rule in classification?	**Analysis**
4	What is tree pruning and define the Naïve Bayes classification?	**Knowledge**
5	Explain the decision tree?	**Understand**

Long Answer Questions

S. No	Question	Bloom Level
1	Explain about the classification and discuss with an example?	**Analysis**
2	Summarize how does tree pruning work? What are some enhancements to basic decision tree induction?	**Understanding**
3	Describe the working procedures of simple Bayesian classifier?	**Analysis**
4	Discuss about Decision tree induction algorithm?	**Evaluate**
5	Explain about IF-THEN rules used for classification with an example and also discuss about sequential covering algorithm?	**Knowledge**

UNIT-5

Short Answer Questions

S. No	Question	Bloom Level
1	Define clustering?	**Knowledge**
2	Illustrate the meaning of cluster analysis?	**Analysis**
3	Explain the different types of data used in clustering?	**Knowledge**
4	Explain the fields in which clustering techniques are used?	**Understand**
5	State the hierarchical methods?	**Knowledge**

Long Answer Questions

S. No	Question	Bloom Level
1	Discuss various types of data in cluster analysis?	**Analysis**
2	Explain the categories of major clustering methods?	**Understand**
3	Explain in brief about k-means algorithm and portioning in k-means?	**Analysis**
4	Describe the different types of hierarchical methods?	**Knowledge**
5	Discuss about the outliers? Explain the weakness and strengths in hierarchical clustering methods?	**Knowledge**

OBJECTIVE QUESTIONS :(JNTUH)

UNIT-1

1. _____ is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decisions.

A. Data Mining.
B. Data Warehousing.
C. Web Mining.
D. Text Mining.

ANSWER: B

2. The data Warehouse is _____.

A. read only.
B. write only.
C. read write only.
D. none.

ANSWER: A

3. Expansion for DSS in DW is _____.

A. Decision Support system.
B. Decision Single System.
C. Data Storable System.
D. Data Support System.

ANSWER: A

4. The important aspect of the data warehouse environment is that data found within the data warehouse
is_____.

A. subject-oriented.
B. time-variant.
C. integrated.
D. All of the above.

ANSWER: D

5. The time horizon in Data warehouse is usually _____.

A. 1-2 years.
B. 3-4years.

C. 5-6 years.
D. 5-10 years.
ANSWER: D
6. The data is stored, retrieved & updated in _____.
UNIT-2
1. The Synonym for data mining is
(a) Data warehouse
(b) **Knowledge discovery in database**
(c) ETL
(d) Business intelligence
(e) OLAP.
2. Data transformation includes which of the following?
A. A process to change data from a detailed level to a summary level
B. A process to change data from a summary level to a detailed level
C. **Joining data from one source into various sources of data**
D. Separating data from one source into various sources of data
3. Which of the following process includes data cleaning, data integration, data transformation, data selection, data mining, pattern evaluation and knowledge presentation?
A. **KDD** process
B. ETL process
C. KTL process
D. None of the above
4. At which level we can create dimensional models?
(a) Business requirements level
(b) **Architecture models level**
(c) Detailed models level
(d) Implementation level
(e) Testing level.
5. What are the specific application oriented databases?
A. Spatial databases,
B. Time-series databases,
C. **Both a & b**
D. None of these
UNIT-3
1. Association rules are always defined on _____.
A. Binary attribute.
B. Single attribute.
C. Relational database.
D. Multidimensional attribute.
ANSWER: A
2. _____ is data about data.
A. Metadata.
B. Microdata.
C. Minidata
D. Multidata.
ANSWER: A
3. Which of the following is the data mining tool?
A. C.
B. Weka.
C. C++.

D. VB.

ANSWER: B

4. Capability of data mining is to build _____ models.

A. Retrospective.

B. Interrogative.

C. Predictive.

D. Imperative.

ANSWER: C

5. The _____is a process of determining the preference of customer's majority.

A. Association.

B. Preferencing.

C. segmentation.

D. classification.

ANSWER: B

UNIT -4

1. Another name for an output attribute.
    a. predictive variable
    a. independent variable
    b. estimated variable
    c. dependent variable
2. Classification problems are distinguished from estimation problems in that
    a. classification problems require the output attribute to be numeric.
    b. classification problems require the output attribute to be categorical.
    c. classification problems do not allow an output attribute.
    d. classification problems are designed to predict future outcome.
3. Which statement is true about prediction problems?
    a. The output attribute must be categorical.
    b. The output attribute must be numeric.
    c. The resultant model is designed to determine future outcomes.
    d. The resultant model is designed to classify current behavior.
4. Which statement about outliers is true?
    a. Outliers should be identified and removed from a dataset.
    b. Outliers should be part of the training dataset but should not be present in the test data.
    c. Outliers should be part of the test dataset but should not be present in the training data.
    d. The nature of the problem determines how outliers are used.
    e. More than one of a, b, c or d is true.
5. Which statement is true about neural network and linear regression models?
    a. Both models require input attributes to be numeric.
    b. Both models require numeric attributes to range between 0 and 1.
    c. The output of both models is a categorical attribute value.
    d. Both techniques build models whose output is determined by a linear sum of weighted input attribute values.
    e. More than one of a,b,c or d is true.

Answers to Unit IV

Multiple Choice Questions

1. d
2. b
3. c

4. d
5. d

UNIT -5

1. A trivial result that is obtained by an extremely simple method is called _____.
A. naive prediction.
B. accurate prediction.
C. correct prediction.
D. wrong prediction.

ANSWER: A

2. K-nearest neighbor is one of the _____.
A. learning technique.
B. OLAP tool.
C. purest search technique.
D. data warehousing tool.

ANSWER: C

3. Enrichment means ____.
A. adding external data.
B. deleting data.
C. cleaning data.
D. selecting the data.

ANSWER: A

4. _____ is an example for case based-learning.
A. Decision trees.
B. Neural networks.
C. Genetic algorithm.
D. K-nearest neighbor.

ANSWER: D, ANSWER: A, ANSWER: D

WEBSITES:

1. www.autonlab.org/tutorials : Statistical Data mining Tutorials
2. www- db.standford.edu /`ullman/mining/mining.html : Data mining lecture notes
3.ocw.mit.edu/ocwweb/slon-School-of-management/15-062Data- MiningSpring2003/course home/index.htm: MIT Data mining open courseware

EXPERT DETAILS:

1. Jiawei han, Abel Bliss Professor, Department of Computer Science, Univ. of Illinois at Urbana-Champaign Rm 2132, Siebel Center for Computer Science
2. Micheline kamber, Researcher,Master's degree in computer science (specializing in artificial intelligence) from Concordia University, Canada
3. Arun k pujari, Vice Chancellor, Central University Of Rajasthan - Central University Of Rajasthan

JOURNALS:

1. Data warehousing, data mining, OLAP and OLTP technologies are essential elements to support decision-making process in Industries
2. Effective navigation of query results based on concept hierarchy
3. Advanced clustering data mining text algorithm

LIST OF TOPICS FOR STUDENT SEMINARS:

1. Fundamentals of Data Mining
2. Data Mining functionalities
3. Classification of data mining system
4. Pre-processing Techniques
5. APRIORI Algorithm

6. FP-Growth Algorithm
7. Spatial data mining
8. Web mining
9. Trends and applications of data mining

CASE STUDIES / SMALL PROJECTS:

Case study-1:

Search queries on biomedical databases, such as PubMed, often return a large number of results, only a small subset of which is relevant to the user. Ranking and categorization, which can also be combined, have been proposed to alleviate this information overload problem. Results categorization for biomedical databases is the focus of this work. A natural way to organize biomedical citations is according to their MeSH annotations. First, the query results are organized into a navigation tree.