



DATA MINING (CS702PC) COURSE PLANNER

I. COURSE OVERVIEW:

At the end of the course the student should be in a position to

1. Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.
2. This course will introduce the concepts of data mining, which gives a complete description about the principles used, architectures, applications, design and implementation of data mining.
3. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge- driven decisions.
4. Analyze data sets in order to predict future trends useful for data science.

II. PRE-REQUISITES:

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

1. Understand the concepts of 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.

III. COURSE OBJECTIVES:

- | |
|---|
| 1. Learn data mining concepts understand association rules mining |
| 2. Discuss classification algorithms learn how data is grouped using clustering techniques |
| 3. To develop the abilities of critical analysis to data mining systems and applications. |
| 4. To implement practical and theoretical understanding of the technologies for data mining |
| 5. To understand the strengths and limitations of various data mining models; |

IV. COURSE OUTCOMES:

CLO Code	CLO's	At the end of the course, the student will have the ability to:	Bloom's Taxonomy Levels	POs / PSOs mapped
CS702PC.01	CLO1	Ability to understand the types of the data to be mined and present a general classification of tasks and primitives to integrate a data mining system.	L2: UNDERSTAND	PO1, PO2, PO3, PSO2
CS702PC.02	CLO2	Apply preprocessing methods for any given raw data.	L3: DEMONSTRATE	PO1, PO2, PO3, PSO2
CS702PC.03	CLO3	Extract interesting patterns from large amounts of data.	L5: EVALUATE	PO2, PO3, PO4, PO5, PSO1, PSO2
CS702PC.04	CLO4	Discover the role played by data mining in various fields	L6: DESIGN	PO3, PO4, PO5, PSO1, PSO2
CS702PC.05	CLO5	Choose and employ suitable data mining algorithms to build analytical applications	L2: UNDERSTAND	PO2, PO3, PO4, PO5, PSO1, PSO2
CS702PC.06	CLO6	Evaluate the accuracy of supervised and unsupervised models and algorithms.	L5: EVALUATE	PO1, PO2, PO3, PSO2



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.	2.6	Mini Projects
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.	1.4	Lectures, Assignments, Exams
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.	2.4	Mini Projects
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.	2	--
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.	-	--
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.	-	Lectures, Assignments, Exams
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	-	Mini Projects



Program Outcomes (PO)		Level	Proficiency assessed by
	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.	2.2	Lectures, Assignments, Exams
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.	2.2	Lectures, Assignments, Exams

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.	2.4	Mini Project
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.	2.6	Lectures, Assignments, Exams
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.	2	Mini Project

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

VII. SYLLABUS:

UNIT – I

Data Mining: Data–Types of Data–, Data Mining Functionalities– Interestingness Patterns– Classification of Data Mining systems– Data mining Task primitives –Integration of Data mining system with a Data warehouse–Major issues in Data Mining–Data Preprocessing.

UNIT – II

Association Rule Mining: Mining Frequent Patterns–Associations and correlations – Mining Methods– Mining Various kinds of Association Rules– Correlation Analysis– Constraint based Association mining. Graph Pattern Mining, SPM.

UNIT – III



Classification: Classification and Prediction – Basic concepts–Decision tree induction–Bayesian classification, Rule–based classification, Lazy learner.

UNIT – IV

Clustering and Applications: Cluster analysis–Types of Data in Cluster Analysis–Categorization of Major Clustering Methods– Partitioning Methods, Hierarchical Methods– Density–Based Methods, Grid–Based Methods, Outlier Analysis.

UNIT – V

Advanced Concepts: Basic concepts in Mining data streams–Mining Time–series data–Mining sequence patterns in Transactional databases– Mining Object– Spatial– Multimedia–Text and Web data – Spatial Data mining– Multimedia Data mining–Text Mining– Mining the World Wide Web.

GATE SYLLABUS: NOT APPLICABLE

IES SYLLABUS: NOT APPLICABLE

VIII. LESSON PLAN:

Lecture No.	Unit No.	Topics to be covered	Link for PPT	Link for PDF	Link for Small Projects/ Numericals(if any)	Course learning outcomes	Teaching Methodology	Reference
1	I	* Intoduction to Data Mining and OBE	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	Data Mining vs Machine Learning	CLO2	<ul style="list-style-type: none"> • Chalk & Talk • LCD / PPT • Assignments • Videos (NPTEL) 	T1
2		Data-Types of Data				CLO2		T1
3		Data Mining Functionalities- Interestingness Patterns				CLO1		T1
4		Classification of Data Mining systems				CLO2		T1
5		Data mining Task primitives, Integration of Data mining system with a Data warehouse				CLO2		T1
6		Major issues in Data Mining-Data Preprocessing.				CLO2		T1
7	Mock Test#1							
8	II	Tutorial Class(Bridge Class/ Revision Class)/ Student Presentation	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	Fake news detection	CLO2	<ul style="list-style-type: none"> • Chalk & Talk • LCD / PPT • Assignm 	T1
9		Mining Frequent	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing		CLO3		T1



		Patterns	?usp=sharing	kdDb?usp=sh						
10		Associations and correlations		aring			CLO3	ents • Videos (NPTEL)	T1	
11		*Apriori Algorithm					CLO3		T1	
12		Mining Methods					CLO2		T1	
13		Mining Various kinds of Association Rules					CLO2		T1	
14		Correlation Analysis- Constraint based Association mining					CLO1		T1	
15		Graph Pattern Mining, SPM					CLO1		T1	
16		Tutorial Class(Bridge Class/ Revision Class)/ Student Presentation					CLO1		T1	
17	III	Classification and Prediction- Basic concepts	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing			CLO5		T1	
18		Decision tree induction						CLO5	T1	
19	III	Bayesian classification					CLO5	• Chalk & Talk • LCD / PPT • Assignments • Videos (NPTEL)	T1	
20		Rule-based classification	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing			CLO5		T1	
21		Lazy learner							CLO5	T1
22		Tutorial Class(Bridge Class/ Revision Class)/ Student Presentation				Diabetes prediction			CLO5	T1
23		Cluster analysis- Types of Data in Cluster Analysis							CLO5	T1
24	IV	Categorization of Major Clustering Methods	https://drive.google.com/drive/folders/1	https://drive.google.com/drive/folders			CLO2		T1	



25	Partitioning Methods	DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing		CLO2		T1	
26	Hierarchical Methods				CLO2		T1	
27	Density-Based Methods				CLO2		T1	
28	Grid-Based Methods				CLO3		T1	
29	Outlier Analysis.				CLO3		T1	
30	Tutorial Class(Bridge Class/ Revision Class)/ Student Presentation				CLO3		T1	
31	Basic concepts in Mining data streams				CLO3		T1	
32	Mock Test#2							
33	Mining Time-series data				CLO1		T1	
34	Mining sequence patterns in Transactional databases				CLO4		T1	
35	Mining Object-Spatial-Multimedia-Text and Web data	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing	https://drive.google.com/drive/folders/1DCP8SdamAp917REUKXsXaAgNFqRFkdDb?usp=sharing		CLO4	<ul style="list-style-type: none"> • Chalk & Talk • LCD / PPT • Assignments • Videos (NPTEL) 	T1	
36	Spatial Data mining- Multimedia Data mining				CLO4			T1
37	Text Mining- Mining the World Wide Web				CLO4			T1
38	Tutorial Class(Bridge Class/ Revision Class)/ Student Presentation				CLO2			T1



TEXT BOOKS:

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 Steinbach, Pearson Education.
3. Data mining Techniques and Applications, Hongbo Du Cengage India Publishing

REFERENCES:

1. Data Mining Techniques, Arun K Pujari, 3rd Edition, Universities Press.
2. Data Mining Principles & Applications – T.V Sveresh Kumar, B.Esware Reddy, Jagadish S Kalimani, Elsevier.
3. Data Mining, Vikaram Pudi, P Radha Krishna, Oxford University Press

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

Course	Program Outcomes												Program Specific Outcomes		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	3	3	2	2	2	-	1	1	1	1	2	1	1	1	1
2	3	3	3	3	3	1	1	1	2	1	2	2	2	2	3
3	3	3	3	3	3	3	2	2	2	2	3	3	2	3	3
4	3	3	3	3	3	2	2	2	2	2	2	2	2	3	3
AVG	3	3	2.75	2.75	2.75	1.5	1.5	1.5	1.75	1.5	2.25	2.0	1.75	2.25	2.5

X. QUESTION BANK

UNIT-1

Short Answer Questions

QUESTIONS	Blooms taxonomy level	Course Outcome
1. Define data mining?	Understand	CO1
2. Explain the functionalities of data mining?	Understand	CO1
3. Interpret the major issues in data mining?	Knowledge	CO1
4. Name the steps in knowledge discovery?	Knowledge	CO1
5. Distinguish between data ware house and data mining?	Analyze	CO1

Long Answer Questions

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?	Understanding	CO1
--	----------------------	------------



d. Describe the steps involved in data mining when viewed as knowledge discovery process?		
2. Discuss briefly about data smoothing techniques?	Creating	CO1
3. List and describe the five primitives for specifying the data mining tasks?	Analyzing	CO1
4. Define data cleaning? Express the different techniques in handling the missing values?	Understanding	CO1
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	CO1

UNIT-2

Short Answer Questions

QUESTIONS	Blooms taxonomy level	Course Outcomes
1. Explain the frequent item set?	Understanding	CO2
2. Explain about maximal frequent items set and closed item set?	Knowledge	CO2
3. Name the steps in association rule mining?	Understand	CO2
4. Explain the efficiency of APRIORI algorithm	Analyze	CO2
5. Define item set? Interpret the support and confidence rules for item set A and item set B?	Understand	CO2

Long Answer Questions

1. Discuss which algorithm is an influential algorithm for mining frequent item sets for Boolean association rules? Explain with an example?	Analysis	CO2
2. Describe the FP-growth algorithm with an example?	Analysis	CO2
3. Explain how to mine frequent item sets using vertical data format?	Understand	CO2
4. Explain how to mine the multi dimensional association rules from relational data bases and data ware houses?	Understand	CO2
5. Explain the APRIORI algorithm with an example?	Analysis	CO2

UNIT-3

Short Answer Questions

QUESTIONS	Blooms taxonomy level	Course Outcomes
1. State classification and define regression analysis?	Understand	CO2
2. Name the steps in data classification and define training tuple?	Knowledge	CO2
3. Explain the IF-THEN rule in classification?	Analysis	CO3
4. What is tree pruning and define the Naïve Bayes classification?	Knowledge	CO3
5. Explain the decision tree?	Understand	CO3



Long Answer Questions

1.Explain about the classification and discuss with an example?	Analysis	CO2
2.Summarize how does tree pruning work? What are some enhancements to basic decision tree induction?	Understanding	CO2
3.Describe the working procedures of simple Bayesian classifier?	Analysis	CO3
4.Discuss about Decision tree induction algorithm?	Evaluate	CO3
5.Explain about IF-THEN rules used for classification with an example and also discuss about sequential covering algorithm?	Knowledge	CO3

UNIT-4

Short Answer Questions

QUESTIONS	Blooms taxonomy level	Course Outcomes
1.Define clustering?	Knowledge	CO3
2.Illustrate the meaning of cluster analysis?	Knowledge	CO3
3.Explain the different types of data used in clustering?	Knowledge	CO4
4.Explain the fields in which clustering techniques are used?	Understand	CO4
5.State the hierarchical methods?	Understand	CO4

Long Answer Questions

1.Discuss various types of data in cluster analysis?	Analysis	CO3
2.Explain the categories of major clustering methods?	Understand	CO3
3.Explain in brief about k-means algorithm and partitioning in k-means?	Analysis	CO4
4.Describe the different types of hierarchical methods?	Knowledge	CO4
5.Discuss about the outliers? Explain the weakness and strengths in hierarchical clustering methods?	Knowledge	CO4

UNIT-5

Short Answer Questions

QUESTIONS	Blooms taxonomy level	Course Outcomes
1.Define Web mining and text mining?	Knowledge	CO4
2.Write a short note on web content mining.	Understand	CO4
3.What are the features of Unstructured text mining.	Knowledge	CO4
4. Write a short note on web structure mining.	Understand	CO4
5.Write a short note on web usage mining.	Understand	CO4

Long Answer Questions

1.Explain about authoritative and Hub pages?	Knowledge	CO4
2.Give taxonomy of web mining activities.For what purpose web usage mining is used?	Understand	CO4
3. what activities are involved in web usage mining?	Knowledge	CO4
4.Explain Episode rule discovery for texts.	Knowledge	CO4
5.Write a short note on Text clustering.	Understand	CO4



Objective Questions:

UNIT-1

1. The Synonym for data mining is
(a)Data warehouse (b)**Knowledge discovery in database** (c)ETL (d)Business intelligence
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-2

1. Association rules are always defined on_____.
A. Binary attribute. B. Single attribute. C. Relational database. D. Multidimensional attribute.
2. _____ is data about data.
A. Metadata. B. Microdata. C. Minidata D. Multidata.
3. Which of the following is the data mining tool?
A. C. **B. Weka.** C. C++. D. VB.
4. Capability of data mining is to build _____ models.
A. Retrospective. B. Interrogative. **C. Predictive.** D. Imperative.
5. The _____ is a process of determining the preference of customer's majority.
A. Association. **B. Preferencing.** C. segmentation. D. classification.

UNIT-3

1. Another name for an output attribute.
 - a. predictive variable
 - b. independent variable
 - c. estimated variable
 - d. 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.

Unit IV

Multiple Choice Questions

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.
2. K-nearest neighbor is one of the _____.
A. learning technique. B. OLAP tool. **C. purest search technique.** D. data warehousing tool.
3. Enrichment means _____.
A. adding external data. B. deleting data. C. cleaning data. D. selecting the data.
4. Clustering methods are _____.
A. Hierarchical. B. Agglomerative. C. PAM algorithm. D. K-nearest neighbor. **E. All the above**

UNIT-V

1. HITS abbreviation in Web Structure?
a. Hyperlink-Index Topic Search **b. Hyperlink-Induces Topic Search**
c. Hyperlink-Identification Text Search d. Hyperlink-Index Text Search
2. Preprocessing Web log activity is?
a. Count patterns that occur in sessions **b. Remove extraneous Information**
c. Count Page references d. Pattern Setting
3. Periodic Crawler defines?
a. Visits Portions of the Web b. Selectively searches the Web
c. Visits pages related to a particular subject d. Collect Information from visited pages
4. Which assigns relevance score to each page based on crawl topic?
a. Distiller b. Hub pages
c. Hypertext Classifier d. scores
5. What is main Objective of web mining?
a. Web Component, Score and Usage Mining b. Web Control, Text and Utility Mining
c. Web Content, Score and Utility Mining **d. Web Content, Structure and Usage**



Fill in the blanks:

Unit 1

1. **Data Mining** _____ predicts future trends & behaviors, allowing business managers to make proactive, knowledge-driven decisions
2. Data Cleaning is a process that removes ... **outliers**.....
3. The output of KDD is **useful information**
4. **Data Discrimination** 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
5. Strategic value of data mining is **time-sensitive**

Unit 2

1. _____ **Referencing** _____ is a process of determining the preference of customer's majority.
2. _____ **Data Mart** _____ is a metadata repository
3. The two steps in Apriori includes ... **join**..... and **prune**.....
4. FP Growth stands for **Frequent pattern growth**.....
5. Use normalization by decimal scaling to transform the value 35 for age **0.35**.....

Unit 3

1. **Classification** is the process of finding a model (or function) that describes and distinguishes data classes or concepts.
2. **Data mining** methods discard outliers as noise or exceptions.
3. **Prediction** also used for to know the unknown or missing values.
4. In a decision tree, leaf nodes represent **class labels or class distribution**.
5. **Decision Tree** is constructed in a top-down recursive divide-and-conquer manner.

Unit 4:

1. A **cluster analysis** is the process of analysing the various clusters to organize the different objects into meaningful and descriptive object.
2. ... **Agglomerative**..... clustering follows bottom up strategy
3. PAM means... "**partition around medoids**".....
4. Bayesian classifiers exhibited **high accuracy** and **speed** when applied to large databases.
5. Most data mining methods discard outliers as **noise or exceptions**.

Unit 5:

1. **Hub Pages** Contain links to many relevant pages
2. **PageRank, CLEVER** Techniques used in Web Structure Mining
3. **Weighting** is used to provide more importance to backlinks coming form important pages
4. PageRank equation **$PR(p)=c(PR(1)/N1 +...+PR(n)/Nn$**
5. What is the use of CLEVER? **Finding both Authoritative and Hub pages.**

XI. WEBSITES:



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

XII.EXPERT DETAILS:

1. Jiaweihan, Abel Bliss Professor, Department of Computer Science, Univ. of Illinois at Urbana-Champaign Rm 2132, Siebel Center for Computer Science
2. Michelinekamber, 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

XIII.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

XIV.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
10. Text mining

XV.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.