NATIONAL INSTITUTE OF TECHNOLOGY
JAMSHEDPUR, JH – 831014
Department of Computer Science and Engineering

SPRING SEMESTER 2012-2013

Course Handout

Course Code: CS 505

Course Name: Data Mining and Data Warehousing

Faculty: Bhaskar Mondal

Course description:

Introduction to data mining and data warehousing

Know about the Data: data objects and attributes, Basic statistical tendency of data, visualization, measuring similarity and dissimilarity.

Data Preprocessing: Data Cleaning, Data Integration, Data Transformation and Data Reduction, Data Summarization Based Characterization.

Modeling a data warehouse: Concepts, Data warehouse life cycle, building a data warehouse, Data Warehousing Components, Data Warehousing Architecture. On Line Analytical Processing, Categorization of OLAP Tools.

Introduction to Data mining and knowledge discovery: Data Mining Functionalities, Steps In Data Mining Process, Architecture of a Typical Data Mining Systems, Classification of Data Mining Systems, Overview of Data Mining Techniques

Mining Frequent Pattern and Association Rules: concepts, frequent item mining methods, pattern evaluation method.

Classification and Prediction, Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Other Classification Methods, Prediction

Clusters Analysis: Types of Data in Cluster Analysis, Categorization of Major Clustering Methods

Applications of Data Mining: Social Impacts of Data Mining, Mining WWW, Mining Text Database, Mining Spatial Databases

Scope:

- Mining different kinds of knowledge in databases
- Knowledge at multiple levels of abstraction
- Incorporation of background knowledge
- Data mining query languages and ad hoc data mining
- Presentation and visualization of data mining results
- Handling noisy or incomplete data
- Pattern evaluation—the interestingness problem
- Efficiency and scalability of data mining algorithms
- Parallel, distributed, and incremental mining algorithms

Objective:

To learn the methods of mining Knowledge from the historical data bases and the management of Data for the purpose of retrieving knowledge and the business needs. Steps for data mining like Identify target datasets and relevant fields, Data cleaning, Remove noise and outliers, Data transformation.

Model evaluation using frequent pattern mining, Association rule, Classification and clustering decision support systems.

- Understand the need for analysis of large, complex, information-rich data sets.
- Identify the goals and primary tasks of the data mining process.
- Describe the root of data mining technology.
- Recognize the iterative character of a data process and specify its basic steps.
- Explain the influence of data quality on a data mining process.
- Establish the relation between data warehousing and data mining.

Course Plan

Lecture No.	Learning Objective	Topics to be covered	Refer to Chapter, see (Book)
1-2	Introduction to data mining and data warehousing	Introduction	[T1]
3-5	• data objects and attribute types: numeric, nominal, binary, categorical, special etc. • Basic statistical tendency of data, visualization, measuring similarity and dissimilarity.	Know about the Data	[T1]
6-8	• Data Cleaning • Data Integration • Data Transformation and Data Reduction, Data Summarization Based Characterization. The ETL Process.	Data Preprocessing	[T1]
9-13	• Concepts • Modeling a data warehouse, Data Marts • Data cubes • Data warehouse life cycle • building a data warehouse • Data Warehousing Components • Data Warehousing Architecture.	Data warehouse	[T1]
14-15	• OLTP and OLAP systems • On Line Analytical Processing • Multidimensional OLAP: Star, snowflake and fact constellation schema. • Categorization of OLAP Tools: MOLAP, ROLAP, HOLAP.	OLTP	[T1]
16-17	• Data Mining Functionalities • Steps In Data Mining Process • Architecture of a Typical Data Mining Systems • Classification of Data Mining Systems • Overview of Data Mining Techniques	Introduction to Data mining and knowledge discovery	[T1]
18-23	• Concepts: transactional data set analysis • Frequent item mining methods: Apriori, Partition method, Pincer search, FP Tree Growth. • Generating the association rule from frequent pattern. • Pattern evaluation method.	Mining Frequent Pattern and Association Rules	[T2]

24-30	• Issues Regarding Classification and Prediction • Classification by Decision Tree Induction: induction, attributes selection methods. • Bayesian Classification • Rule based Classification • Prediction	Classification and Prediction	[T2]
31-36	• Types of Data in Cluster Analysis • Categorization of Major Clustering Methods: k-medoids(PAM), k-means, BIRCH, DBSCAN etc. • Evaluation of clustering.	Clusters Analysis	[T1] [T2]
37-42	• Social Impacts of Data Mining • Mining the web (social networks) • Mining Text Database • Mining Spatial Databases	Applications of Data Mining	[T1] Some hand out will be provided.

Text Book:

[**T1**] **Data Mining Concepts and technique** by Jiawei Han, Micheline Kamber, Jian Pei; *Morgan Kaufmann ISBN: 9789380931913*
[T2] **Data Mining Techniques** by Arun K Pujari; *University Press*

Evaluation Scheme:

EC No.	Evaluation Component	Duration	Weightage	Data & Time	Nature of Component
1.	Test I	60 min	20%		Closed Book
2.	Test II	60 min	20%		Closed Book
3.	End Sem Exam	3 Hrs	40%		Closed Book
4.	Assignments		10%		Take home
5.	Surprise Quizzes	5 min	10%		Closed Book

You May Meet Me:

Every day 5:00pm.

You may mail me at bm.6779@gmail.com; always mention your Roll Number followed by Subject at the subject field.