

Spring 2003

# ISE 599: Datamining- Concepts and Applications

Class Times: 5-6:15 PM MW

## CONTACT

Satish Bukkapatnam  
Daniel J. Epstein Industrial and Systems Engineering Department  
GER 203

Ext: 09549  
e-mail: [satish@usc.edu](mailto:satish@usc.edu)  
<http://www-nmcl.usc.edu>

## OBJECTIVES AND EXPECTATIONS

The recent advancements in information technologies have spurred industries to computerize many aspects of their operations. As a consequence, large datasets from various subsystems of an enterprise have begun to be collected and accumulated at a dramatic pace. These large data sets have off late assumed unmanageable proportions, and new technologies, collectively known as *datamining*, are imperative to judiciously tapping the information buried in these large datasets so that modern day industries can reap the full benefits of computerization. Datamining refers to a multipurpose abstraction of information from large data sources. Today, datamining is fast emerging as a core component of many industries, especially those related to Finance, Banking, Manufacturing, Biosystems, Biotechnology, and Information Systems and Services.

The main objective of this course is to provide students with the basic datamining concepts and applications that can enable them to set up and manage an industrial datamining system, as well as to conduct advanced studies on the subject. Students will specifically learn:

- Broad classes of datamining technologies
- Diverse datamining applications in today's industrial environments
- Challenges associated with processing large-scale industrial data
- Principles of data preparation including those to address missing and noisy information
- Basics of clustering, classification and association
- Dynamic datamining techniques

The class will combine lectures (on Wednesdays) and Lab (on Mondays). Lectures will focus on learning the concepts and principles, and lab sessions will focus on computer-based assignments to implement and exercise the concepts. All computer assignments will be based in Matlab environment as it is fast emerging as the tool of choice to perform many datamining tasks, including rapid data analysis and visualization. Weekly assignments will extend the in-class exercises and will in addition provide new conceptual perspectives. The course will also feature a project that shall consolidate the learning achieved during the course.

## SYLLABUS

<i>Week</i>	<i>Topic</i>
1-2	Introduction and Preliminary Background
3	Datamining Taxonomy
4	Datamining Issues and Review of Current Tools
5-6	Data preparation methods
7-10	Static Datamining- Clustering, Classification and Association
8	<b>MIDTERM</b>
11-13	Dynamic Datamining
14-15	Advanced Topics and Presentations

## BOOKS AND REFERENCE MATERIALS

Most of the instructional content will be extracted from the following sources:

- Jiawei Han, Micheline Kamber, [Data Mining : Concepts and Techniques](#), Morgan Kaufmann, August 2000,
- David J. Hand, Heikki Mannila and Padhraic Smyth, [Principles of Data Mining](#), MIT Press, Fall 2000.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman, [The Elements of Statistical Learning: Data Mining, Inference, and Prediction](#), Springer Verlag, 2001.

In addition supplementary lecture notes will be provided.

About the instructor: The instructor's expertise lies in dynamic data mining. His experiences include devising neuro-genetic methods for clustering, classification, and estimation.