

**University of Southern California**  
**MARSHALL SCHOOL OF BUSINESS**  
**Spring, 2004**  
**Course Guidelines & Syllabus**

**IOM 428 – DATA WAREHOUSING AND DATA MINING**

Instructor: Dr. Arif Ansari  
Office: HOH 400 D  
Office Hours: Tuesday and Thursday 12.00-1.00  
Office phone: (213) 821-5521  
Email: [aansari@marshall.usc.edu](mailto:aansari@marshall.usc.edu)  
**Emergency Contact number: 213-740-0172**

TA: Gayatri Ratnaparkhi  
Office: HOH 400 D (Hoffman Hall)  
email: [ratnapar@usc.edu](mailto:ratnapar@usc.edu)  
Office hours: TBA

**COURSE OBJECTIVES**

- To develop an understanding of the various concepts and tools behind data warehousing and mining data for business intelligence.
- To develop quantitative skills pertinent to the analysis of data from huge corporate data warehouses

**Overview:**

This course provides an overview of two of the newest and hottest technologies in the area of information science: data warehousing (DW) and data mining (DM). We also plan to spend sometime on a third topic: On-Line Analytical Processing (OLAP), yet another important area of tremendous growth. Many large companies such as American Express and Wal-mart have accumulated a great deal of data from their day to day business. DW is the technology that integrates the data collected from various sources of transaction processing systems that record day to day business. Collecting data is just the first step. Companies really want information – knowledge and insight. So, the next question is, how can one uncover patterns and relationships hidden in organizational databases? Specifically, what can they learn from the data about how to please their customers, how to target their most profitable customers, how to optimally allocate their resources and how to minimize their loss such as those incurred from fraud. The size and complexity of data in a data warehouse, however, could be overwhelming. When there are millions of trees, how can one draw meaningful conclusion about the forest? (a quote from a Two Crows Corporations Report)

## Why study Data Mining (DM) and Data Warehousing (DW)?

DM is a cutting edge information technology that decision-makers and analysts use to extract valuable information (e.g. patterns, trends) from large databases. According to the latest statistics, both industries are growing at double digit rates for the past few years. DW, in particular, has turned into a billion dollar business and has stimulated strong growth in data mart – specialized DW for specific purposes. DW, data mart, and DM have attracted a lot of attention in the business world lately because of their great potential payoff. As an example, American Express reported a 15-20% increase in credit-card purchase after using DM to improve upon targeting its market. Because Fortune 500 companies are investing heavily in this technology and smaller companies such as restaurant chains are beginning to catch up, we expect employment opportunities for students who have backgrounds in DW and DM to be strong in the next five years.

### Course description:

The following is a sample of issues that I will try to address in this course. At the end of the course, you should have a framework for understanding them.

- what is a data warehouse, its design and function
- what is a data mart , its increasing use in DW
- what is the relationship between DW and DSS
- what is a multidimensional database
- how does the latest database technology OLAP work
- how DM enhances decision support systems
- which DM tool is appropriate for a particular business application
- how market can be segmented via unsupervised learning
- how loan officers make their decision using supervised learning programs
- how decision trees can help managers identify their prospective customers
- how you can use the latest visualization software to present your complex analysis results without showing tons of numbers
- how can neural networks "learn" from examples and help manager make intelligent business decisions

This class is designed in such a way that only limited mathematical and statistical background is required. Learning and understanding underlying DW concepts, studying cases, applying DM ideas and methods to business data, and communicating ideas and solutions will be our main theme. Technical details of selected DM methods will be discussed. Students are expected to try out new software for various business applications. A **Project** is required for this class.

**Course Materials.** The following items will be necessary for completion of reading assignments and homework.

- The first book is a standard Data Mining, Introductory and Advanced topics (Margaret H. Dunham) book, focused on business applications that we will use for our readings.
- IOM 428 Course Pack, Data Warehousing and Data Mining (This reader is non-returnable. It cannot be exchanged for cash or credit. Please be sure you are permanently enrolled in the class before purchasing!)
- *JMP Start Statistics: A Guide to Statistical and Data Analysis Using JMP and JMP-IN Software* (manual and software) by J. Sall and A. Lehman, Duxbury Press, Belmont, CA, 1996. (\* Only for certain applications)
- *Class notes.*

Class notes for this class will be available on blackboard. You should familiarize yourself with these notes before they are covered in class. You will be using different softwares to describe and analyze data.

**Important dates:**

*Class Registration:*

January 30: Last day to register and add class

January 30: Last day to drop a class without a mark of “W”

April 9: Last day to drop a class with a mark of “W”

*Midterm exams:*

TBA

*Final Exams:*

May 6, 2004, Thursday 11.00-1.00 pm.

**Grading.**

There will be 1 midterm and 1 final exam. They are close-book.

Midterm - 20%.

Final - 25%.

Project - 15%.

A group presentation of result on a project on issues in DW/BI/DM or from an application of DW/BI/DM. Size of group depends on class size.

Quizzes - 20% .

There will be 2 quizzes. They are close-book.

Homework - 12% . There will be homework assignments.

Class participation – 8%. The grades will be assigned as per the participation and in the class quizzes.

## **Homework**

Homework assignments will be distributed via blackboard. Homework is extremely important to your learning the material in the class. Homework assignments may be discussed with members of your team ( 2 or 3 students) . You have the following objectives on your homework assignments:

- Answer the question you were asked.
- Argue clearly and concisely that your answer is correct.

We will judge your homework assignments by how clearly you communicate and understand the material. Remember that nothing conveys clear thinking like clear writing. The definition of clear writing includes the appropriate use of and reference to computer output. If you examined certain graphs and/or printouts when arriving at your solution then include that output in your report so that the reader can follow your logic to your conclusion.

Computer output should be clearly labeled and referred to in the text. Ideally, the output should be placed in a figure close to the textual reference. Including large sections of computer output without reference in the text is a signal to the TA that you are not sure what is important and what is not and will likely count against your grade.

If you believe that an error has been made in the grading of your homework you may ask to have it regarded. Please be specific about the problem. If you are still concerned after this process you may come and see me.

If you do not agree with the TAs grading, you may appeal your solution to me. Note, however, that I will review your *entire* assignment and will include in my assessment of your grade your oral arguments as well. *I am a tougher grader than the TA, so be prepared when you see me. I reserve the right to adjust your grade up or down as I see fit.*

**Review Sessions.** There will be a review session before the exams.

**Academic Integrity.** Academic dishonesty of any type will not be tolerated in this class. Students who find this statement ambiguous should consult the Student Conduct Code, page 83, of the USC *SCampus* handbook.

A comment about writing the assignments up individually and working in teams: You can work together in teams to discuss the problems and concepts. However, you are required to write up the assignments individually. This means that all the words in your assignments are your own, and you generate all of your own computer output and graphs.

Now, while correct solutions will have very similar or even the same computer output, no two answers should be phrased the same way. If I find two or more assignments that are highly similar, I will at a minimum give the homework a zero, and may refer the incident to the Dean. *Do not test me on this policy.*

## **STUDENTS WITH DISABILITIES**

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me as early in the semester as possible. DSP is located in STU 301 and is open 8:30 am - 5:00 pm, Monday through Friday. The phone number for DSP is 213 740-0776.

## **Tentative Schedule:**

The course will start with either Data Mining or Data Warehousing.

- **Lecture 1: Overview**

---

### **DATA WAREHOUSING (DW):**

- **Lecture - DW1 : Introduction**
  - Data Warehousing, Data Mining & OLAP. Berson / Smith Chapter 1: Introduction to Data Warehousing. pg. 3-21
  - "Along the Infoban--Data warehouses" L. Fisher, Strategy & Business
  - "A Data Warehouse Comes of Age". T. Marshall, Teradata Review, Fall 1998
- **Lecture - DW2 : Its Components**
  - Data Warehousing, Data Mining & OLAP. Berson / Smith Chapter 1: Introduction to Data Warehousing. pg. 3-21
  - "A Data Warehouse Comes of Age". T. Marshall, Teradata Review, Fall 1998
- **Lecture - DW3 : The Walmart Example**
  - Data Warehousing Using the Walmart Model. Westerman, Chapter 1

- **Lecture - DW4 : The Star Schema**  
- Data Warehousing Design Solutions. Adams / Venerable. Chapter 1: The Business Driven Data Warehouse. Pg 1-28
- **Lecture - DW5 : Applying the star schema**  
Notes
- **Lecture - DW6 : Examples of the Star Schema**  
Same as Lecture – DW5
- **Lecture - DW7 : Technical Construction**  
Westerman, chapter 8.
- **Lecture - DW 8 : The Multidimensional Model and OLAP (I): Introduction**  
- Seven Methods for Transforming Corporate Data into Business Intelligence. Dhar/Stein. Chapter4 Data-Driven Decision Support. pg. 30-50.  
- Fresh data, Rice, eWeek, 2001
- **Lecture - DW 9 : The Multidimensional Model and OLAP (II): Examples**  
Same as Lecture – DW7
- **Lecture - DW 10 The Multidimensional Model and OLAP (III): MOLAP and ROLAP**  
- Data Warehousing, Data Mining & OLAP. Berson / Smith Chapter 13 Online Analytical Processing. pg 247-266
- **Lecture - DW 11 : The Multidimensional Model and OLAP (IV) : Web-Based Reporting**  
OLAP Goes Online, Baron, Information Week, 1999

---

#### DATA MINING (DM):

- **Lecture - DM1 : Personalization and Customization (I) : Introduction**  
- Punting on personalization, Tweney, Business 2.0, 2000.  
- Web Personalization, Quellette, ComputerWorld, 1999.  
- Collaborative Filtering Heylighen, 1999.
- **Lecture - DM2 : Personalization and Customization (II) : Underlying Methods**  
- Case-Based Reasoning, Watson, 1997, pg. 23-28
- **Lecture – DM3 : Personalization and Customization (III): Integration**

- Beyond Personalization, Brobst & Rarey, Teradata Review, 2000
- **Lecture - DM4 : Personalization and Customization (IV) : CRM**
  - The Customer rules, Bergent & Kazimer-Schockley, Intelligent Enterprise, 2001.
  - Personalization tools dig deep, Colkin, Information Week, 2001.
- **Lecture – DM5 : Decision Tree (I): Overview**
  - Data Mining Techniques. Berry / Linoff. Chapter 12: Decision Trees. pg. 243-265
- **Lecture –DM6 : Decision Tree (II): Methods and Applications**
  - Case Study: Mail Order/Retail. Techguide, 1997
- **Lecture –DM7 : Decision Tree (III): Examples**
  - Same as Lectures DM5 & DM6
- **Lecture – DM8 : Neural Network (I): Overview**
  - Data Mining Techniques. Berry / Linoff. Chapter 13: Artificial Neural Networks. pg. 286-305
  - "A Gentle Introduction to Neural Networks". Hank Simon. DM Review. Decision Support Systems and Intelligent Systems (5th edition). Turban / Aronson. pg. 687 - 693
- **Lecture DM9 : Neural Network (II): Methods and Applications**
  - Same as lecture-DM8
- **Lecture DM10 : Neural Network (III): Examples**
  - Same as lecture-DM8
- **Lecture – DM11: Special Topics**