IOM 528 – DATA WAREHOUSING, BUSINESS INTELLIGENCE 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

Emergency Contact number: 213-740-0172

TA: Gayatri Ratnaparkhi
Office: HOH 400 D (Hoffman Hall)
email: 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 is about how companies apply two new technologies, data warehousing (DW) and data mining (DM, including business intelligence, BI) to empower their employees, and build and manage a customer-centric business model. Besides learning the strategic role DW and DM plays in an enterprise, you will also get a close-up look at DW and DM by working on cases and gaining hands-on experience using software tools. Students taking this class will get an overview of the technologies of DW and BI/DM from a managerial perspective.

Fortune 500 companies such as American Express and Wal-mart have accumulated a great deal of data from their day to day business. Data warehouse is the technology that integrates the data collected from various sources that include transaction processing systems and e-commerce data collecting systems. Collecting and integrating data is just the first step. What is really critical is information, knowledge and insight. So the question is, what is the utility of the data? How can one use data in managing customer relationship and empowering employees? How can one uncover patterns and relationships hidden in organizational databases? These issues are addressed by a fast growing body of research and applications, broadly known as business intelligence.
(BI)/DM. These technologies draw their strengths from the fields of information technology, statistics, machine learning and artificial intelligence.

In summary, managers need to understand the strategic values of their company's information assets. DW, BI and DM are cornerstones

**Course Objectives:**

To develop an understanding of the strategic values of various concepts behind warehousing and mining data. To develop an understanding of concepts in DW, BI, and DM, and to gain hands-on experience of some DW, BI and DM software tools.

After taking this class, students should be able to:

- Understand the basic terms that are used in DW, On Line Analytical Process (OLAP), BI and DM
- Communicate to Information Technology workers their business perspective in terms of the language of DW and DM
- Choose appropriate tools for specific purposes of storing, integrating and analyzing data (business consideration, and technical consideration
- Use tools provided in class to perform simulated tasks in warehousing data
- Use tools to perform BI/DM activities on moderately large data sets in case studies. Students should be proficient with at least one of these software tools
- Articulate and present the results of their analyses and the business implications of these results
- Gain inference from your analysis, from a statistical point of view.

**Structure of lectures:**

IOM 528 will be organized in a way that include some combination of the following: lectures, case-based class discussion, group project, computer lab work, and maybe guest speakers.

**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 528 Course Pack, Data Warehousing Business Intelligence 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!)
- *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 4, 2004, Tuesday 7.00-9.00 pm.

**Grading.**

- Midterm 20% (take-home)
- Project + case presentation 20%
- Final 40%
- Homework + case studies 20%

**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 Session.** 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 will either Data Mining or Data Warehousing.

- **Lecture 1: Overview**

DATA WAREHOUSING (DW):

- **Lecture - DW1: A Strategic View**
  
  Data to knowledge to results, Davenport et al., Cal. Mgt Review 2001
  Strategic View of DW and CRM, Swift, 2002.

  **Case Study**: Canadian Tire

- **Lecture - DW2: A Tactical View**
  
  Westerman, Chapter 1, 10, 11.
  DW Components, Berson & Smith, 1997

- **Lecture - DW3: Technology of DW**
  
  Westerman, Chapter 6, 7.
  Relational DB, Computer World 2001
  Normalization, Whitehorn & Marklyn, 1998
  Mass Movement, Russom Intelligent Enterprise, 2001

  **Case Study**: Walmart

- **Lecture – DW4: Dimensionally Designed DW (I)**
  

- **Lecture – DW5: Dimensionally Designed DW (II)**
  
  Hotel Occupany Star Schema, Adamson & Venerable, 1998

  **Case Study**: Star scheme (notes)

- **Lecture - DW6: OLAP and Business Intelligence**
  
  Data Driven Decision Support, Dhar & Stein, 1997
  Business Intelligence Pays Dividends, Baron, Information Week, 2000
• Lecture - DW7: Web-Based OLAP and Business Reporting

OLAP Goes Online, Baron, Information Week, 1999.

DATA MINING:

• Lecture – DM1: Data Mining: an Overview of Application and Privacy Issues

Mining data: Wasserman, Region Review 2000
None of Your Business, Stepanek, Business Week Online, 2000

Case Study: Capital One

• Lecture – DM2: Using Data Mining Techniques for Personalization

Collaborative Filtering, Heylighen, 1999.
Beyond Personalization. Brobst & Rarey, Teradata review, 2000

Case Study: Firefly Network (now part of Microsoft)

• Lecture – DM3: Decision Tree and Rule-Based Systems in Business Applications

Decision Trees, Berry & Linoff, 1997
Case Study: Vermont Country Store

• Lecture – DM4: Decision Tree and Neural Network in Business Applications

Making Brain Waves, Baatz, CIO Magazine 1995

• Lecture – DM5: Understanding Neural Network & Data Mining Cases

Artificial Neural Network, Berry & Linoff, 1997.

Case Study: Real estate pricing model for houses in Rochester, MN.

• Lecture – DM6: Putting Things Together: CRM & Relationship Technology
A framework for CRM, Winer, Cal. Mgt Review

**Case Study** : Mail Boxes Etc.

- **Lecture – DM7: Special Topics**