Course Description

Data is a key source of intelligence and competitive advantage for businesses. With the explosion of electronic data and the demand for better and faster decisions, the role of data driven intelligence is becoming central in organizations. Data mining is the process of converting the raw data into useful knowledge required to support decision-making. It automates the process of knowledge discovery, making us orders of magnitude more productive in our search for useful information than we would be otherwise. It also increases the confidence with which we can make business decisions. Virtually every business organization these days is in the process of exploring and implementing business intelligence solutions to core business problems. This course is essential for anyone interested in understanding how to get the maximum value from data, especially when abundant data are available. The course covers various techniques, applications and software used for data mining.

The course is recommended for students interested in understanding the techniques and applications of data mining and acquiring hands-on skills for making intelligent business decisions in data-rich organizations. No prior knowledge is required for taking this course. There are no prerequisites for this course.

Course Web site:


We will make extensive use of the Web site in this course. Important information for the class (e.g., announcements, lecture notes and other handouts) will appear on the Web site. You should make a habit of checking it regularly.

Note: This syllabus may change during the quarter. The course Web site will always have the most up-to-date syllabus.
Textbook:

Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Second Edition

Software:

We will use the database software Microsoft Access (optional) and a free data mining software WEKA (download from http://www.cs.waikato.ac.nz/ml/weka). No programming skill is needed.

Office Hours: Knock on my door, make an appointment, or just give me a call.

Grading:

<table>
<thead>
<tr>
<th>Components</th>
<th>Grades</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Participation</td>
<td>15%</td>
</tr>
<tr>
<td>Homework</td>
<td>60%</td>
</tr>
<tr>
<td>Term Project</td>
<td>25%</td>
</tr>
</tbody>
</table>

Participation: Make sure you read the cases carefully before each class, and actively participate in discussion and offer insightful comments in class. The following scheme is used to grade student participation in each class:

3: Excellent
2: Good
1: OK
0: Absent with good reason and advance notification
-3: Absent with no reason

Homework:

There are 6 homework assignments in total. Each homework assignment may include problem solving, data analysis and case discussion. You should work on them individually.
Term Project:

Each group is responsible for choosing a company to study its data utilization situation and make recommendations about how its data should be used for generating business intelligence. There are two phases. In phase 1, which is due in the middle of the quarter, each group needs to turn in a 2-3 page write-up about the company they picked and generally describe its data situation. In the last class, each group presents its project to the entire class. In the final project report, each group needs to follow detailed guidelines to analyze the company’s data strategies, and make detailed recommendations based on the techniques and strategies we learned from the class. Group size can vary between 1 ~ 3 students. See detailed term project description on Page 6.

Class Schedule: (Subject to change)

<table>
<thead>
<tr>
<th>No.</th>
<th>Davis</th>
<th>Sac</th>
<th>Bay</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Jan 4</td>
<td>Jan 6</td>
<td>Jan 15</td>
<td>Course Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Jan 11</td>
<td>Jan 13</td>
<td>Jan 15</td>
<td>Modeling Data in Organizations</td>
</tr>
<tr>
<td>3</td>
<td>Jan 18</td>
<td>Jan 20</td>
<td>Jan 29</td>
<td>Building Business Intelligence using Queries, Data Warehouse</td>
</tr>
<tr>
<td>4</td>
<td>Jan 25</td>
<td>Jan 27</td>
<td>Jan 29</td>
<td>Market Basket Analysis &amp; Association Rules, CRM</td>
</tr>
<tr>
<td>5</td>
<td>Feb 1</td>
<td>Feb 3</td>
<td>Feb 12</td>
<td>Market Segmentation &amp; Clustering, Prepare data</td>
</tr>
<tr>
<td>6</td>
<td>Feb 8</td>
<td>Feb 10</td>
<td>Feb 12</td>
<td>Prediction &amp; Classification – Decision Tree</td>
</tr>
<tr>
<td>7</td>
<td>Feb 15</td>
<td>Feb 17</td>
<td>Feb 26</td>
<td>Personalization &amp; Nearest Neighbor</td>
</tr>
<tr>
<td>8</td>
<td>Feb 22</td>
<td>Feb 24</td>
<td>Feb 26</td>
<td>Financial Forecasting &amp; Neural Networks</td>
</tr>
<tr>
<td>9</td>
<td>Mar 1</td>
<td>Mar 3</td>
<td>Mar 12</td>
<td>Link Analysis &amp; Web mining</td>
</tr>
<tr>
<td>10</td>
<td>Mar 8</td>
<td>Mar 10</td>
<td>Mar 12</td>
<td>Term project presentations</td>
</tr>
</tbody>
</table>

Due Dates:

<table>
<thead>
<tr>
<th></th>
<th>Davis</th>
<th>Sacramento</th>
<th>Bay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework 1</td>
<td>Jan 11</td>
<td>Jan 13</td>
<td>Jan 29</td>
</tr>
<tr>
<td>Homework 2</td>
<td>Jan 18</td>
<td>Jan 20</td>
<td>Jan 29</td>
</tr>
<tr>
<td>Group Decision</td>
<td>Jan 18</td>
<td>Jan 20</td>
<td>Feb 12</td>
</tr>
<tr>
<td>Homework 3</td>
<td>Jan 25</td>
<td>Jan 27</td>
<td>Feb 12</td>
</tr>
<tr>
<td>Homework 4</td>
<td>Feb 1</td>
<td>Feb 3</td>
<td>Feb 12</td>
</tr>
<tr>
<td>Project Phase 1</td>
<td>Feb 1</td>
<td>Feb 3</td>
<td>Feb 26</td>
</tr>
<tr>
<td>Homework 5</td>
<td>Feb 8</td>
<td>Feb 10</td>
<td>Feb 26</td>
</tr>
<tr>
<td>Homework 6</td>
<td>Mar 1</td>
<td>Mar 3</td>
<td>Mar 12</td>
</tr>
<tr>
<td>Final Project Presentation</td>
<td>Mar 8</td>
<td>Mar 10</td>
<td>Mar 12</td>
</tr>
<tr>
<td>Final Project Report</td>
<td>Mar 15</td>
<td>Mar 17</td>
<td>Mar 19</td>
</tr>
</tbody>
</table>

Note: Homework is due before class starts. Please hand in hard copies. You will get 20% less for your grade for each day your homework is late (i.e. you get 0 after 5 days).
Reading list:

1. Course Introduction

Cases:
- A Case Study in Business Data Mining – Bank of America (textbook pp. 22-25)
- A Wireless Communication Company Makes the Right Connections (textbook pp. 34-39)
- Neural Networks and Decision Trees Drive SUV Sales (textbook pp. 39-42)

Textbook: Chapter 1, 2, 3

2. Modeling Data in Organizations

Textpak:
- Modeling Data in the Organization (available from Study.net)

3. Building Business Intelligence using Queries, Data Warehouse

Optional Reading:
- Introduction to Structured Query Language (on course website)
- Microsoft Access tutorial (link on course website)

4. Market Basket Analysis, Association Rules

Textbook: Chapter 4, 9

Additional reading posted on course website

5. Market Segmentation, Clustering

Cases:
- Clustering Town – The Boston Globe (textbook pp. 374-380)

Textbook: Chapter 11, 17

6. Predication, Classification, Decision Tree

Textbook: Chapter 6

Additional reading posted on course website

7. Personalization, k-Nearest Neighbors

Cases:
- Polyphonic HMI: Mixing Music and Math (HBS case, available from Study.net)
Textbook: Chapter 8, 12

8. Financial Prediction, Neural Networks

Textbook: Chapter 7

Additional reading posted on course website

9. Link Analysis, Web Mining, Search Engine Marketing

Cases:
- MedNet.com Confronts “Click-Through” Competitionet (HBS case, available from Study.net)
- Google Advertising (HBS case, available from Study.net)

Textbook: Chapter 10
Term Project

Grading:

The term project in total counts 25% of the total grade.
Phase 1 report: 5%
Class presentation: 5%
Final report: 15%

This project is intended to provide you with valuable opportunity to be involved with a company’s data-driven decisions. The main goal is to thoroughly understand how a company leverages its data for business intelligence, and to connect what you learn from this class with the real problems in order to make valuable recommendations to the company. The project contains the following steps.

Step 1: Form a group

Please email me your decision by the due date.

Step 2: Identify a company to study

This company can be your previous employer, current employer or any other company that is willing to discuss the related information with you.

Step 3: Turn in Phase 1 report

Each group needs to turn in a 2-3 page write-up about the company they picked and generally describe its data situation. By the time you turn in your phase 1 report, you should have a rough idea about the amount and depth of the information the company allows you to access, and please describe that in the report too. Please turn in a hard copy of the report on the due date.

Step 4: Class Presentation

Your presentation should at least include the companies’ background, how its data is handled, how business intelligence is generated from the data and your recommendations. Every member of the group should be involved in the presentation.

Step 5: Turn in the final project report

When writing your final report, try to touch upon the 11 steps of data mining methodology in Chapter 3 of the textbook, or use these steps to structure your report if necessary. Feel free to structure your report differently if you feel it can help you bring out the insights better. You can decide to focus more on the strategy side of data mining or technical side of data mining. When analyzing the current situation of the company, please discuss the strength and weakness of its current practice, and how business
intelligence is derived from data. When making recommendations, please try to back up your recommendations as much as possible. The report can also include information such as the software used, data flow within the organization, and other valuable observations.

You don’t need to do data analysis in order to fulfill the requirement of the project. If you are able to obtain the company’s data and have interests in analyzing them, you can choose option 2 for 469.

If you are able to finish your report before the last class, please give me a hard copy before your class presentation. Otherwise, you can email me the report before the due date.
469

469 is intended to build your hands-on experience based on the materials you learned in 269. You can choose to do one of the two projects described below to fulfill the requirement of 469. The project can be done individually or in a group of 2-4 students.

For each option, there are two phases. Phase 1 is due before the 7th lecture. Phase 2 is due one week after the last lecture.

**Option 1 (Using queries in a database to build business intelligence):**

This option is intended to provide you with valuable hands on experience in designing and implementing a real world database system application, and as such, you are encouraged to develop such a system to address managerial issues you face at your workplace. In this project, you should identify an application in your workplace and develop a database system for it. If you are not able to find a proper application related to your work, you should come and talk to me before you choose anything else.

The project has 2 phases.

**Phase 1: Database design**

1. Pick an application, describe the application.
2. Why a database is needed for this application?
3. Draw Entity-Relationship diagrams for the application. Indicate the assumptions and constraints of the ER diagram. Convert the ER diagram to relational tables. Discuss the possible problems of the design.

Please turn in a 2-3 page report for this phase.

**Phase 2: Implementation in Microsoft Access.**

1. Properly refine your phase 1 according to my suggestions. You are allowed to make other changes to phase 1 if desirable.
2. Implement the tables in Microsoft Access.
3. Populate the tables.
4. Each group should write 5 queries. For each query, you should explain what business intelligence can be derived. Please write your 5 most insightful queries.

After you complete Phase 2, please submit your final project report. In the final report you turn in, you should clearly state the background of the application, the business problem you are addressing, the technical challenges you encountered, the design, the queries, and the potential uses of the database. The focus of the report should be on the business intelligence you can derive from your queries. For each query, please elaborate the story behind it.
When you submit the final report, please submit the Access file together with the Word file. Each group should only submit one copy of their report.

**Option 2 (Using data mining methods to build business intelligence):**

Each group identifies an application, and locates a relevant data set (you can acquire the data set from a company or from the Internet), then further comes up with interesting questions which can be answered based on the data set using data mining techniques learned in class. Finally, you should perform data analysis on the data set and report results.

This option is intended to provide you with valuable *hands on* experience in designing and implementing a *real world* data mining application, and as such, you are encouraged to be involved in the entire data mining process including 1) find a problem/application that interest you, 2) manage to get the data by yourself, 3) come up with questions that are important to your problem/application, 4) use the techniques you learned in class to answer those questions, 5) search for interesting patterns and models from the results, and 6) evaluate your results.

**Phase 1 report**

In the report you submit, please include the following information:

1) Description of the problem/application
2) Description of the data set, names and description of the attributes
3) Questions you want to address using data mining techniques
4) The business value of your planned analysis

**Phase 2: Turn in the final project report**

The deliverable will be a report documenting data preparation, techniques used, models produced, and evaluation of the models.

When writing your final report, please provide as much details as possible about your implementation and solution. When you submit your final report, please also include other files you want me to look at.

Your final report will be evaluated by the intensity of your analysis, creativity of your approach, interpretation of your results, and the business insights.

You might encounter a lot of problems when you start to process the data and build models using the software, so please start early and ask me for advice when you have difficulties.