

56:138 Knowledge Discovery and Management

Summer 2001

Course URL: <http://css.engineering.uiowa.edu/~ie138/>

Objectives: Introduction to data analysis methods, data mining tools and techniques, data engineering, data warehousing, and evolutionary computation. Case studies illustrating applications of knowledge discovery and management in engineering, medicine, and service applications are discussed.

Textbook: J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Academic Press, San Diego, CA, 2001 (Available at IMU Bookstore).

Reference: W.H. Inmon, R.H. Terdeman, and C. Imhoff, *Exploration Warehousing*, John Wiley, 2000.

Instructor: Andrew Kusiak, Professor
4312 Seamans Center
Email: andrew-kusiak@uiowa.edu
<http://www.icaen.uiowa.edu/~ankusiak>
Tel: 319 - 335 5934
Fax: 319-335 5424

Instructor Office Hours:
1:30 - 3:00 pm TTh
4312 Seamans Center

Class Time: 3:00 - 6:10 pm TTh

Classroom: 3321 Seamans Center

Dates to Remember:

Tutorial topic selected (visit <http://css.engineering.uiowa.edu/~ie138/TTopics138.doc>):

- June 27

Tutorial presentations:

- Week of July 9

Tutorial reports due:

- July 12

Project presentations:

- Begin July 19

Project reports due:

- July 21

Final exam due:

- TBA

Due to compressed class schedule no extensions will be granted.

Course Contents:

1. Introduction (Chapter 1 + the material posted on the web)
2. Classification and prediction (Chapter 7 + web material)
3. Cluster Analysis (Chapter 8 + web material)
4. Data preprocessing (Chapter 3)
5. Mining association rules (Chapter 6)
6. Data warehousing (Chapter 2)
7. Concept description (Chapter 5)
8. Applications and trends in data mining (Chapter 10)
9. Knowledge management (web material)
10. Case studies
11. DM Software

Course grading scheme

- | | |
|---|-----|
| <input type="checkbox"/> Tutorial Paper | 25% |
| <input type="checkbox"/> Project | 50% |
| <input type="checkbox"/> Final Exam | 25% |

Both the tutorial and the project result are to be presented in class.

Tutorial/Project grading scheme

- 30% tutorial/project presentation
- 60% tutorial/project content
- 10% attendance of tutorial/ project presentations

TUTORIAL PAPER (Teams of 1 - 2 students)

The tutorial paper is to describe a data mining algorithm or a procedure and be illustrated with an example. The list of algorithms that can be selected for tutorial development is available at <http://css.engineering.uiowa.edu/~ie138/TTopics138.doc>. The tutorial may be developed by individual student of a group of two students and is to be presented in class.

PROJECT

The project is to be conducted by an individual student or a team of two students. Each student or a team may select one of the following two types of projects.

A. Software Demonstration Project (Teams of 1 - 2 students)

Search the web and select a freeware or a trial version commercial data mining software. You need to get familiar with the software, select two examples, and run them with the software.

A typical sequence of activities in type A project is as follows:

- Search the www (You may begin by visiting the Knowledge Discovery Clinic <http://www.icaen.uiowa.edu/~ankusiak/KDC/main.htm>)
- Identify a data mining tool
- Get familiar with the tool
- Prepare presentation material
- Classroom demonstration (tool description and demonstration with examples)
- Prepare project report

B. Software Development Project (Teams of 1 - 2 students)

The student(s) will develop software for one of the data mining algorithms discussed in class. Improvements to the existing algorithms will be an asset. The code should be written in one of the following languages C, VB, C#, or JAVA and a user-friendly interface should be developed. Web implementations of the software are encouraged.

FINAL EXAM

TBA

TUTORIAL REPORT FORMAT

The tutorial report should be prepared as an MS Word file and should contain figures and tables that are necessary to make the report complete. Be concise in your writing and consult technical writing references as needed.

The body of the term project report should include:

1. Introduction
2. Algorithm description
3. Problem data (two examples)
4. Solution Procedure
5. Results

PROJECT REPORT FORMAT

The project report should be prepared on a word processor and should contain figures and tables that are necessary to make the report complete. Be concise in your writing and consult technical writing references as needed.

The body of the term project report should include:

A. Application Project

1. Introduction
2. Tool description
3. Problem data (two examples)
4. Results

B. Software Development Project

1. Introduction
2. Algorithm description
3. User's manual
4. Example problems (2)
5. Computer code description

The code should run on the Windows platform and be developed with ICEAN supported software.

Tutorial/Project Guidelines

Effort (Time) estimate

It is expected that each student should spend not less than 20 hours on the tutorial and 30 hours on the project.

Tutorial/Project presentation time estimate

Not more than 20 minutes per presentation.

Tutorial/Project report submission requirements

Each team (student) should submit the following items (one per tutorial/project):

- ❑ Hard copy of the tutorial/project report
- ❑ Power point slides used in your classroom presentation
- ❑ File with the tutorial/project report.
For software development projects, submit also the source and executables and specify the computer hardware and software needed to run the program. All software must be developed Windows operating system. Please email your message to andrew-kusiak@uiowa.edu and title it RE: 56:138 Tutorial/Project.

REFERENCES ON TECHNICAL WRITING

- [1] Olsen, L. A. and Huckin, T. N. (1983). *Principles of Communication for Science and Technology*, McGraw-Hill Book Company, New York.
- [2] Sherlock, J. (1985). *A Guide to Technical Communication*, Allyn and Bacon, Inc., Newton, MA.