Syllabus
COMP 595DM: Data Mining
Spring 2011

Course Information

Lecture textbook:	Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, 2nd Edition, Morgan Kaufmann, 2006
Lab textbook:	Joseph Adler, R in a Nutshell: A Desktop Quick Reference O'Reilly, 2009
Instructor:	Taehyung (George) Wang, *twang@csun.edu*
Class website	http://www.csun.edu/~twang/595DM
Class number	18489
Lecture/Lab:	JD3520/JD2211, F 11:00 – 13:45
Office hour:	JD4447, M 15:15- 14:15, W 11:30 – 12:30 or by appointment on Fridays
Midterm exam:	To be announced
Final exam:	To be announced
Prerequisite:	COMP380/L

Course Description

A study of the concepts, principles, techniques, and applications of data mining. Topics include data preprocessing, data warehousing and OLAP technology, Apriori algorithm for mining frequent patterns, classification methods (such as decision tree induction, Bayesian classification, neural networks, support vector machines, genetic algorithms), clustering methods (such as *k*-means algorithm, hierarchical clustering methods, self-organizing feature map), and applications and trends in data mining such as Web, finance, telecommunication, biology and medicine, science and engineering. Social impacts of data mining, such as privacy and data security issues, are also discussed.

Course Objectives

Upon successful completion of the course the student will:

- Be able to understand the concepts, strategies, and methodologies related to the design and construction of data mining
- Be able to comprehend several data preprocessing methods
- Be able to utilize data warehouses and OLAP for data mining and knowledge discovery activities
- Be able to determine an appropriate mining strategy for given large dataset
- Be able to apply appropriate mining techniques to extract unexpected patterns and new rules that are "hidden" in large databases
- Be able to obtain knowledge of current data mining applications

Course Requirements

Homework Assignments

Several homework assignments will be given to assess your understanding of class materials. Homework solutions should be submitted before the lecture starts on the due date. Presentation along with Power Point Slide will be asked and counted for the grade.

Group Project

As one of the course requirements, the success of the group project is necessary. The students must successfully finish their group project on time. A set of final-version deliverables is required to be

submitted at the end of semester. These deliverables shall be consistent, complete, and correct. The topic of the group project should be approved no later than Week 3.

Group Term Paper

The result of a group project will be shown in a term paper. The format of the term paper will be available at the class web site.

Midterm Exams

There will be two midterms. The format of the exam will be announced in advance. No makeup exam is allowed unless you have a legal document.

Final Exam

The scope of final exam is comprehensive. The format will be announced before final exam week. No makeup exam is allowed unless you have a legal document.

Course Policies

Academic Dishonesty

"Cheating or plagiarism in connection with an academic program at a campus is listed in Section 41301, Title V, California Code of Regulations, as an offense for which a student may be expelled, suspended, or given a less severe disciplinary sanction." - (http://www.csun.edu/catalog/)

Grading

The portion of each grading component and the grading formula are as follows:

Grading components	Percentage
Homework and lab assignments	25%
Class performance	5%
Group project and term paper	30%
Two midterm exams	20%
Final exam	20%
Total	100%

Grade	Grade cutoff
A	>= 93.00 %
A-	>= 90.00 %
B*	80 – 89.99%
C**	70 – 79.99%
D***	60 – 69.99%
F	< 60%

* (B+ >= 87% B >= 83% B- >= 80%)
** (C+ >= 77% C >= 73% C- >= 70%)
*** (D+ >= 67% D >= 63% D- >= 60%)

If class performance is not good, class rank shall be considered for final grade. Attendance and peer evaluation will be counted.

Course Schedule and Outline

	Lecture	Lab
Week 1 (1/28)	Ch. 1	Organize teams. Introduction to R
Week 2 (2/4)	Ch. 2	Ch. 1, 2 and 3
Week 3 (2/11)	Ch. 5 (Association Rule)	Ch. 4, 5 and 6. Present project topics
Week 4 (2/18)	Midterm I	Ch. 7 and 8.
Week 5 (2/25)	Ch. 6 (Decision Tree)	Ch. 9 and 10.
Week 6 (3/4)	Ch. 6 (Bayesian Classification)	Ch. 11 and12. Requirement/Specification
Week 7 (3/11)	Ch. 6 (Neural Network)	Ch. 13. Data collection
Week 8 (3/18)	Ch. 6 (SVM)	Ch. 14. Data preprocessing
Week 9 (3/25)	Ch. 3 (Multi-dimensional data)	Ch. 15. Data preprocessing
Week 10 (4/1)	Ch. 3 (Data warehouse)	Data mining
Week 11 (4/8)	Spring Break	
Week 12 (4/15)	Midterm II	Data mining
Week 13 (4/22)	Ch. 7 (Clustering)	Evaluation
Week 14 (4/29)	Ch. 7 (Clustering)	Presentation
Week 15 (5/6)	Ch. 8 (Time-series data mining)	Presentation
Week 16 (5/13)	Review for Final Exam	Project presentation
Week 17	Final Exam	

Note that the course schedule and outline can be changed without advance notice.