

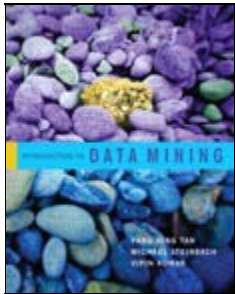
**CSCE 5380 – Data Mining  
Fall 2010**

**Instructor:** Dr. Yan Huang  
**Office:** NTDP F251  
**Office Hours:** TTH 09:00am-10:00am  
**Email:** huangyan 'at' unt.edu

**Time and Place:** TTh 11:00-12:20pm, NTDP B192  
**Class Website:** <http://www.cs.unt.edu/~huangyan/5380>

**TA:** Indrakanti, Saratchandra (SaratchandraIndrakanti@my.unt.edu)  
**TA Office Hours:** Fridays: 12:00 pm - 4:00 pm  
**TA Office:** F205

**Text Book:** **Introduction to Data Mining**



**Pang-Ning Tan**, *Michigan State University*  
**Michael Steinbach**, *University of Minnesota*  
**Vipin Kumar**, *University of Minnesota*

ISBN: 0-321-32136-7  
Publisher: Addison-Wesley

**Recommended readings:**

- J. Han and M. Kamber (2000) *Data Mining: Concepts and Techniques*, Morgan Kaufmann
- U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, The MIT Press, 1996
- U. Fayyad, G. Grinstein, and A. Wierse, *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann, 2001
- D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*, MIT Press, 2001.
- T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, 2001
- T. M. Mitchell, *Machine Learning*, McGraw Hill, 1997.
- S. M. Weiss and N. Indurkha, *Predictive Data Mining*, Morgan Kaufmann, 1998
- H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 2001

**Grading Scheme:**

Assignments:	15%
Project:	30%
Exam I:	15%
Exam II:	15%
Final :	15%
Class participation:	10%

Each assignment will specify the material to be turned in. Assignments are due before class on the due date. Assignments may be turned in up to 3 days late, with a penalty of 10% for each day late. No credit will be given after 3 days.

**Course Objectives:** This course will provide a broad and rapid introduction to the field of data mining. We will provide a general introduction to main data mining tasks, e.g. classification, clustering, association rules, graph mining, sequential pattern mining, and outlier detection as well as some of the latest

developments, e.g. mining spatial data and web data.

**Exams:** Exams are closed book and closed notes. One page cheating sheet (8x11, any font size, front and back) is allowed. The exams are NOT accumulative or comprehensive. No exams will be given at other times without documented excuses.

- Policies:**
1. Each program and homework assignment must be worked on individually. A submission carries with it an implicit statement that the submission is your own work. You may discuss the requirements and syntactical issues, but not solutions or designs. Cheating behavior will be handled seriously. Specifically, first time violation will result in a zero mark on that assignment, project, or exam. Second time violation will result in a failing grade. All students involved in a collusion will be treated the same.
  2. Student behavior that interferes with an instructor's ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Center for Student Rights and Responsibilities to consider whether the student's conduct violated the Code of Student Conduct. The university's expectations for student conduct apply to all instructional forums, including university and electronic classroom, labs, discussion groups, field trips, etc. The Code of Student Conduct can be found at [www.unt.edu/csrr](http://www.unt.edu/csrr)
  3. Please be on time especially when there are assignments due at the beginning of the class. Assignments will not be accepted 10 minutes after the class starts and will be considered as one day late.

**Disabilities Accommodations:** Disabilities Accommodation: The University of North Texas complies with Section 504 of the 1973 Rehabilitation Act and with the Americans with Disabilities Act of 1990. The University of North Texas provides academic adjustments and auxiliary aids to individuals with disabilities, as defined under the law. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you have a disability requiring accommodation, please see the instructor and/or contact the Office of Disability Accommodation at 940-565-4323 during the first week of class.

**SETE Evaluation:** The Student Evaluation of Teaching Effectiveness (SETE) is a requirement for all organized classes at UNT. This short survey will be made available to you at the end of the semester. Please make sure you take this chance to comment on how this class is taught.

**Tentative Schedule:**

- Week 1 Introduction
- Week 2 Exploratory Analysis and data mining tools
- Week 3 Classification
- Week 4 Alternative classification techniques
- Week 5 Exam I
- Week 6,7 Clustering
- Week 8 Associate rule
- Week 9 Association rule, sequential pattern
- Week 10 Exam II
- Week 11 Graph mining
- Week 13 Anomaly detection
- Week 14 Advanced topics
- Week 15 Project presentations