

Syllabus

Course Number and Title: CIS6930 - 4188X / CIS4930 - 4465X Data Mining

Description: This course is a graduate level survey of concepts, principles and techniques related to data mining. Students will become acquainted with both the strengths and limitations of various data mining techniques like Classification, Association analysis and Cluster analysis.

Instructor: Dr. Sanjay Ranka, Professor, Computer and Information Science and Engineering department, University of Florida

- *Office hours:* To Be Decided
- *Office:* E432 CSE Bldg, University of Florida, Gainesville, FL 32611-6120
- *E-mail:* ranka AT cise.ufl.edu
- *Phone:* (352) 392-6838

Teaching Assistant: Manas Somaiya

- *Office hours:* To Be Decided
- *Location:* E309 CSE Bldg
- *E-mail:* mhs AT cise.ufl.edu

Pre-requisites: Basic competency in Algorithms, Programming, Databases and Statistics.

Announcements: It is your responsibility to check [announcements](#) at least once every week.

Textbooks:

- *Introduction to Data Mining* by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, 2005, ISBN: 0321321367
- *Data Mining: Concepts and Techniques* by Jiawei Han and Micheline Kamber, 2000, ISBN: 1558604898

Topics to be covered:

1. Introduction: What is data mining, What is not data mining, Challenges, Other issues
2. Data: Types of data, Data quality, Data preprocessing
3. Classification: Problem definition, General approach, Decision tree induction, Rule based classifiers, Nearest neighbour classifiers, Bayesian classifiers, Artificial neural networks, Support vector machine, Ensemble methods, Model evaluation
4. Association analysis: Problem definition, Frequent itemset generation, Rule generation, Challenges, Interestingness measures, Generalization of association patterns
5. Cluster analysis: Introduction, Similarity and distance, Density, Characteristics of clustering algorithms, Center based clustering techniques, Hierarchical clustering, Density based clustering, Other clustering techniques, Scalable clustering algorithms, Cluster evaluation
6. Visualization: Introduction, General concepts, Visualization techniques

Evaluation:

- Homework Assignments: 10% of the letter grade
- 2 In-class Exams: 30% of the letter grade *each*
- A term project will constitute the remaining 30% of the letter grade.

- No Final Exam
- Letter grade will be based on a curve

Course Policies:

- There will be no makeup exams (Exceptions may be made for medical emergencies).
- Attending lectures is not mandatory. However, you are responsible for all announcements and course material discussed in the class.
- While you may discuss general ideas with others, when you actually write them, you should do so alone. The actual written solutions to the homework assignments and project (code or otherwise) should be entirely your own and individual work. Also, you are expected to indicate on your homework the names of those people with whom you discussed the homework.
- You will be asked to sign the following statement on all exams in this course: *On my honor, I have neither given nor received unauthorized aid on this examination.*

Academic Dishonesty: Read [Academic Honesty Guidelines](#) as posted by Dean of Students Office. All academic dishonesty cases will be handled through the University of Florida Honor Court procedures as documented by the Office of Student Services, P202 Peabody Hall. You may contact them at 392-1261 for "Student Judicial Process: Guide for Students" pamphlet.

Students with Disabilities: Students requesting classroom accommodation must first register with the [Dean of Students Office](#). The Dean of Students Office will provide documentation to the student who must then provide this documentation to the Instructor when requesting accommodation.

Miscellaneous: Please make it to the class on time. If you are late, enter the class and settle down quietly, so that you do not disturb the class. Please turn your cell phone to OFF or VIBRATE mode.