

**CS 530 – Data Mining
Spring 2017
Course Syllabus**

Instructor: Erik Linstead (linstead@chapman.edu)

Time and Place: Tues/Thurs 2:30-3:45pm, Demille 100

Office Hours: After class and by appointment

Overview: An introduction and exploration of a wide variety of algorithms and techniques for data mining and machine learning. Students will learn to analyze large datasets for structure, and build models using supervised and unsupervised approaches to pattern recognition. In addition to understanding the mathematical and theoretical basis for data mining algorithms, students will implement non-trivial machine learning tools in C++, Matlab, R, or Python. The course will also introduce students to the basics of database systems, including relational and document-based systems, as repositories for data to be mined.

Prerequisites: CS 510 or permission of instructor.

Units: CPSC 530 is a 3 unit course.

Required Text: *Data Mining: The Textbook* by Aggarwal. (Available free online)

Course materials: All course materials will be made available via the course site on Blackboard when possible. Blackboard will also be used for submitting assignments, viewing grades, etc.

Homework, Exams, and Grading (subject to change):

Homework will consist of problem sets and programming assignments. Problem sets should be typeset neatly using LaTeX.

For CS 530, absolutely no late work will be accepted.

There will also be the usual midterm and final, which must be taken on the date specified. In the case of a well-documented, unavoidable conflict, I will do my best to accommodate you.

Homework and programming assignments will count for 20% of the course grade, the midterm for 35%, and the final exam for 45% .

Note: To receive a passing grade in the course you must receive a passing grade (>60%) in every component (assignments, exams) of the course.

Course Learning Objectives:

1. Understand issues related to database management and information retrieval
2. Appreciate the specific requirements for scientific databases;
3. Become aware of data standards and modeling
4. Understand database architectures and management for scientific applications
5. Demonstrate ability to apply the proper data mining technique for a given scientific problem
6. Contrast data mining techniques in terms of performance and scalability
7. Design and deploy performance-minded database schemas and systems
8. Hypothesize data models from visualizations produced from common data mining algorithms
9. Solve data-driven computational problems using a combination of data management and data analysis techniques.

Program Learning Objectives:

1. Graduates will develop quantitative reasoning skills which will enable them to: a. solve problems by utilizing extrapolation, approximation, precision, accuracy, rational estimation and statistical validity, b. create quantitative models to describe natural phenomena.
2. Graduates will be able to apply the principles of computational science to scientific problems. Students will develop critical thinking, end to end problem-solving, and data analysis skills. With these skills, they will be able to: a. collect, process and analyze data, b. use mathematics and computing to solve scientific problems
3. Graduates will be able to apply principles of applied mathematics to scientific problems to:
 - a. evaluate the accuracy of approximations,
 - b. interpret the results of calculations
4. Graduates will be able to apply principles of computer technology and computer science to scientific problems.
 - a. use high performance computer architectures including clusters and supercomputers. Create programs to manipulate and analyze data on high performance computer systems.
 - b. construct solutions to scientific problems using parallel algorithms and data structures.
 - c. analyze the performance of algorithms.

Equity and Diversity:

Chapman University is committed to ensuring equality and valuing diversity. Students and professors are reminded to show respect at all times as outlined in Chapman's Harassment and Discrimination Policy: <http://tinyurl.com/CUHarassment-Discrimination>. Any violations of this policy should be discussed with the professor, the Dean of Students and/or otherwise reported in accordance with this policy.

Chapman University's Academic Integrity Policy:

"Chapman University is a community of scholars that emphasizes the mutual responsibility of all members to seek knowledge honestly and in good faith. Students are responsible for doing their own work and academic dishonesty of any kind will be subject to sanction by the instructor/administrator and referral to the university Academic Integrity Committee, which may impose additional sanctions including expulsion. Please see the full description of Chapman University's policy on Academic Integrity at www.chapman.edu/academics/academicintegrity/index.aspx."

Chapman University's Students with Disabilities Policy

"In compliance with ADA guidelines, students who have any condition, either permanent or temporary, that might affect their ability to perform in this class are encouraged to contact the Disability Services Office. If you will need to utilize your approved accommodations in this class, please follow the proper notification procedure for informing your professor(s). This notification process must occur more than a week before any accommodation can be utilized. Please contact Disability Services at (714) 516-4520 or visit www.chapman.edu/students/student-health-services/disability-services if you have questions regarding this procedure or for information or to make an appointment to discuss and/or request potential accommodations based on documentation of your disability. Once formal approval of your need for an accommodation has been granted, you are encouraged to talk with your professor(s) about your accommodation options. The granting of any accommodation will not be retroactive and cannot jeopardize the academic standards or integrity of the course."