

CSCI 4502/5502: Data Mining

Fall 2018

1 Course Title, Sections, Time, and Location

This course is titled “Data Mining” and is offered in three different sections:

- CSCI 4502-001: undergraduate
- CSCI 5502-001: graduate on-campus
- CSCI 5502-001B: graduate distance learning

Time & Location: TuTh 9:30am – 10:45am, ECCS 1B12

2 Instructor Information

- **Lecturer:** Qin (Christine) Lv,
Associate Professor, Associate Chair for Graduate Education, Department of Computer Science
 - **Email:** *Qin.Lv@Colorado.EDU*
 - **Office Location:** ECCR 1B24
 - **Office Phone:** (303) 492 - 8821
 - **Office Fax:** (302) 492 - 2844
 - **Homepage:** <http://www.cs.colorado.edu/~lv>
 - **Office Hours:** TBD
- **Graders:** TBD

3 Course Overview

This course Introduces basic data mining concepts and techniques for discovering interesting patterns hidden in large-scale data sets, focusing on issues relating to effectiveness and efficiency. Topics covered include data preprocessing, data warehouse, association, classification, clustering, outlier detection, and mining specific data types such as time-series, social networks, multimedia, and Web data.

3.1 Prerequisites

- CSCI 4502-001: CSCI 2270 with a minimum grade of C-
- CSCI 5502-001 and CSCI 5502-001B: graduate status

3.2 Textbook

Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber, and Jian Pei. 3rd edition, Morgan Kaufmann, 2011. ISBN-13: 978-0123814791

3.3 Grading

- Problem sets (35%): Work alone; extra questions for students taking CSCI 5502
- Midterm exam (25%): Closed book; in class exam; extra questions for students taking CSCI 5502
- Course project (40%): Work in groups; a self-defined project that applies data mining to real-world problems; recommended group size is 3-4; group can be a mixture of 4502/5502, on-campus/distance learning students; higher threshold for students taking CSCI 5502; project scope needs to be appropriate for group size and group composition (CSCI 4502 vs. CSCI 5502)
- Late policy: At most 2-day delay in submission with 20-point penalty each day

3.4 Class Schedule (Tentative)

Week 1 (8/28, 8/30)	Introduction
Week 2 (9/4, 9/6)	Data Preprocessing
Week 3 (9/11, 9/13)	Data Warehouse
Week 4 (9/18, 9/20)	Frequent Patterns
Week 5 (9/25, 9/27)	Classification
Week 6 (10/2, 10/4)	Project Proposal
Week 7 (10/9, 10/11)	Classification
Week 8 (10/16, 10/18)	Clustering
Week 9 (10/23, 10/25)	Clustering
Week 10 (10/30, 11/1)	Midterm Exam
Week 11 (11/6, 11/8)	Outlier Detection
Week 12 (11/13, 11/15)	Project Checkpoint
Week 13(11/20, 11/22)	Fall Break, Thanksgiving
Week 14 (11/27, 11/29)	Data Streams, Time-Series
Week 15 (12/4, 12/6)	Graphs, Social Networks, Web Data
Week 16 (12/11, 12/13)	Project Final Report

Course Policies

Academic Integrity All students of the University of Colorado at Boulder are responsible for knowing and adhering to the academic integrity policy of this institution. Violations of this policy may include: cheating, plagiarism, aid of academic dishonesty, fabrication, lying, bribery, and threatening behavior. All incidents of academic misconduct shall be reported to the Honor Code Council (honor@colorado.edu; 303-735-2273). Students who are found to be in violation of the academic integrity policy will be subject to both academic sanctions from the faculty member and non-academic sanctions (including but not limited to university probation, suspension, or expulsion). Other information on the Honor Code can be found at <http://www.colorado.edu/policies/honor.html> and at <http://honorcode.colorado.edu>

Classroom Behavior Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, color, culture, religion, creed, politics, veterans status, sexual orientation, gender, gender identity and gender expression, age, disability, and nationalities. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. See policies at <http://www.colorado.edu/policies/classbehavior.html> and at http://www.colorado.edu/studentaffairs/judicialaffairs/code.html#student_code

Disability If you qualify for accommodations because of a disability, please submit to your professor a letter from Disability Services in a timely manner (for exam accommodations provide your letter at least one week prior to the exam) so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities. Contact Disability Services at 303-492-8671 or by e-mail at dsinfo@colorado.ed. If you have a temporary medical condition or injury, see Temporary Medical Conditions: Injuries, Surgeries, and Illnesses guidelines under Quick Links at Disability Services website and discuss your needs with your professor.

Discrimination and Harassment The University of Colorado Boulder (CU-Boulder) is committed to maintaining a positive learning, working, and living environment. The University of Colorado does not discriminate on the basis of race, color, national origin, sex, age, disability, creed, religion, sexual orientation, or veteran status in admission and access to, and treatment and employment in, its educational programs and activities. (Regent Law, Article 10, amended 11/8/2001). CU-Boulder will not tolerate acts of discrimination or harassment based upon Protected Classes or related retaliation against or by any employee or student. For purposes of this CU-Boulder policy, "Protected Classes" refers to race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, or veteran status. Individuals who believe they have been discriminated against should contact the Office of Discrimination and Harassment (ODH) at 303-492-2127 or the Office of Student Conduct (OSC) at 303-492-5550. Information about the ODH, the above referenced policies, and the campus resources available to assist individuals regarding discrimination or harassment can be obtained at <http://hr.colorado.edu/dh/>

Religious Observance Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, students should notify the instructor at least one work before religious obligations that will interfere with required coursework. See full details at http://www.colorado.edu/policies/fac_relig.html