Michigan Technological University

Course Syllabus

CS 4821/CS 5831- Data Mining / Adv. Data Mining
Spring 2024
Updated: Jan. 2, 2024

Course Identification

Course Number:	CS4821 & CS 5831
Course Name:	Data Mining & Advanced Data Mining
Course Location:	Noblet G002
Class Times:	Monday, Wednesday, Friday: 11:00am-11:50am
Prerequisites:	Statistics - MA2710, MA2720, MA3710
	Linear Algebra - MA2330
	Coding Experience - CS2321, (MA3740 + CS1121 or CS1131)
	Optional: Database Experience - MIS3100, CS3425

Instructor Information

Instructor:	Laura Brown
Office:	Rekhi 307
E-mail:	lebrown@mtu.edu (include CS4821 or CS5831 in subject)
Discussion:	Ed `https://edstem.org/us/courses/50862/discussion/`
Office Hours:	TBA

Course Description

Data mining focuses on extracting knowledge from large data sources (related fields include statistics and machine learning). The course introduces data mining concepts, methodology (measurement, visualization, evaluation), algorithms (classification/regression, clustering, association rules), and applications (recommender systems, web mining, text mining ...).

Course Policies

The instructor will conduct this class in an atmosphere of mutual respect. We are all members of an academic community where it is our shared responsibility to cultivate a climate where all students/individuals are valued and where both they and their ideas are treated with respect. You should expect that if your conduct during class seriously disrupts the atmosphere of mutual respect expected in this class, you would not be permitted to participate further.

CS4821 - Course Learning Objectives

Upon successful completion of this course, students will be able to

- Describe data mining concepts and techniques

- Apply the techniques of clustering, classification, association finding, and visualization on sample and real-world data sets
- Select and apply data mining tools to real-world problems

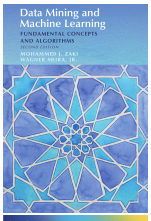
CS5831 - Course Learning Objectives

Upon successful completion of this course, students will be able to

- Describe advanced data mining concepts and techniques
- Apply the techniques of clustering, classification, association finding, and visualization on sample and real-world data sets
- Select and apply data mining tools to real-world problems
- Formulate a real-world data mining problem: find, get, and clean, select and implement appropriate methods, evaluate the performance of the methods, etc.
- Discuss and communicate the results and methods of applying data mining techniques to a problem

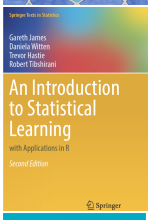
Course Resources

Required Textbook

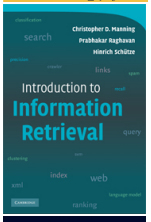


Data Mining & Machine Learning: Fundamental Concepts and Algorithms, 2nd Ed.
by Zaki and Meira, Jr.
Available online at: `https://dataminingbook.info/`

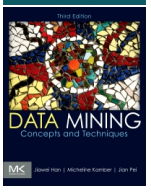
Supplemental Textbooks



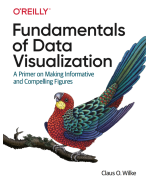
Introduction to Statistical Learning, 2nd Ed.
by James, Witten, Hastie, Tibshirani
Available online at `https://statlearning.com/`



Introduction to Information Retrieval
by Manning, Raghavan, Schutze
Available online at
`https://nlp.stanford.edu/IR-book/information-retrieval-book.html`

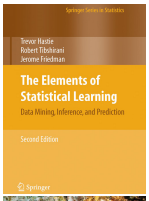


Data Mining: Concepts and Techniques, 3rd Ed.
by Han, Kamber, and Pei
Can be found online

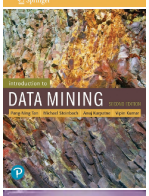


Fundamentals of Data Visualization, 1st Ed.
by Claus Wilke
Available online at `https://clauswilke.com/dataviz/`

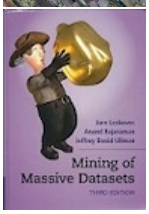
Recommended Reference Books



Elements of Statistical Learning, 2nd Ed.
by Hastie, Tibshirani and Friedman
Available online at `https://hastie.su.domains/ElemStatLearn/`



Introduction to Data Mining, 2nd Ed.
by Tan, Steinbach, Karpatne, and Kumar
Available online at `https://www-users.cs.umn.edu/~kumar001/dmbook/index.php`



Mining of Massive Datasets, 3rd Ed.
by Leskovec, Rajaraman, and Ullman
Available online at `http://i.stanford.edu/~ullman/mmdsn.html`

Grading Scheme

Grading Policy

Scores will be kept in Canvas. Grades will be calculated as a percentage of points received weighted for each part of the course:

CS4821

Course Component	Perc.
Assignments	65%
Exam	20%
Presentations	15%

CS5831

Course Component	Perc.
Assignments	60%
Exam	20%
Final Project	20%

Grade Changes

Any request for regrades should be made in writing (e-mail or Gradescope regrade request) to the instructor within one week of returning the material. No grade changes will be made after that point except in the case of an arithmetic error in summing points or the grade was recorded incorrectly in the grade book.

Grading System

Your final score will determine your letter grade.

Letter Grade	Score	Grade points	Rating
A	>93%	4.0	Excellent
AB	[88% – 93%)	3.5	Very Good
B	[83% – 88%)	3.0	Good
BC	[78% – 83%)	2.5	Above Average
C	[73% – 78%)	2.0	Average
CD	[68% – 73%)	1.5	Below average
D	[60% – 68%)	1.0	Inferior
F	<60%	0.0	Failure
I	Incomplete; used when a student is unable to complete a segment of the course because of circumstances beyond the student's control.		
X	Conditional, with no grade points per credit; given only when the student is at fault in failing to complete a minor segment of a course, but in the judgment of the instructor does not need to repeat the course. It must be made up by the close of the next semester or the grade becomes a failure (F). A (X) grade is included in the grade point average calculation as a (F) grade.		

Course Components

Assignments

There will be ∼5-7 assignments to be completed throughout the semester. Most assignments will have two parts a written and a programming part.

- The first two assignments are to be completed individually, each student is expected to create and turn-in their own solution by the due date listed.
- Subsequent assignments will allow students to work together in groups.
- Assignments should be typed and in the file format requested in the assignment description.
- Use of any allowed sources other than the textbook, slides, and resources posted on Canvas should be acknowledged (below is a discussion on academic integrity for allowable resources).
- Your work should be neat, formatted well, and follow examples given in class.
- It is your responsibility to ensure the correct assignment/version is uploaded in the correct file format. Late penalties will apply until this is done.

Late Policy
All late submissions will receive a grade of zero with the following exceptions:

- Each student has **8 Assignment late days** to be used in the semester.
  - For any given assignment, you can use a maximum of 2 late days.
  - Each late day used allows an assignment to be turned in a day late without penalty. For example, an assignment due Monday may be turned in on Wednesday, using

2 late days, or an assignment due on Friday may be turned in on Saturday, using 1 late day.
  - After using you late days, no submissions will be accepted.
- Technology issues unless they are campus or city-wide are not to be used as an excuse for late work.
- Excused absences planned and known in advance, should have assignments submitted ahead of time.

The policy is set to account for routine illness, travel, or assignments due in other courses. No other late assignments will be accepted unless a prior arrangement is made with me or I receive an excuse from the Dean of Students office.

Presentations - CS 4821 only

Presentations will be led by groups of the students. As the presenter, you should be prepared to talk about the main problems and methods discussed in the paper. More details will be given in the Canvas assignment.

Final Project - CS 5831 only

Students will work together to frame a data mining problem, select methods, apply those methods, and evaluate their performance. Results of the project will be discussed and communicated (i.e., report, presentation, and/or poster). More details will be given in the Canvas assignments.

Exams

There will be one exam for the class worth 20% of the course grade and will cover topics from the textbook, lectures, and assignments. The exam is expected given in Week 11-12 of the course.

Makeup exams

A makeup exam will not be given without prior arrangements. Excused absences as described in the Student Handbook must be e-mailed to the instructor 5 business days prior to the test date. Unplanned events (illness, emergencies) should be brought to attention as soon as possible to make alternative arrangements. Unexcused missed exams will result in a zero for that exam.

Collaboration and Cheating

Adapted from CMU's Center for Teaching Excellence & Educational Innovation and A. Mueller
Assignments

Assignments are opportunities to learn and demonstrates knowledge of concepts, skills and techniques. All submissions should be your own work (or group's work when allowed).

Students are encouraged to discuss course materials and assignments with other students to understand the material of this course, but not produce assignment solutions. Discussion should be limited to general approaches.

Here are some examples of acceptable collaboration:

- Clarifying ambiguities or vague points in class handouts, textbooks, or lectures.
- Discussing or explaining the general class material.
- Providing assistance with general Python/R/Matlab questions, in using the system facilities, or with understanding Python/R/Matlab tools.
- Discussing the code that we give out on the assignment.
- Discussing the assignments to better understand them.
- Getting help on programming issues which are clearly more general than the specific assignment (e.g., what does a particular error message mean?).

You are not allowed to share code or assignment responses between submissions or submission groups. For homeworks submitted individually, each individual is required to write their own solution. For homeworks submitted in groups, a single write-up should be submitted. Collaboration is not permitted for any of the exams.

Any attempts to game the system, exploit loopholes, or submission of work that is not yours is academic misconduct (even if that specific example is not listed below).
Examples of unacceptable collaboration and academic misconduct are:

- Copying (program or assignment) files from another person or source, including retyping their files, changing variable names, etc.
- Using a solution you find on the web, and copying it into your solutions is cheating.
- Allowing someone else to copy your code or written assignment, either in draft or final form.
- Copying from another student during an exam, quiz, or midterm. This includes receiving exam-related information from a student who has already taken the exam.
- Writing, using, or submitting a program that attempts to alter or erase grading information or otherwise compromise security.
- Looking at someone else's files containing draft solutions, even if the file permissions are incorrectly set to allow it.
- Receiving help from students who have taken the course in previous years.
- Lying to course staff.
- Use of any course materials (or software) from previous years.
- Use of any AI tools or software to generate your submissions.

Collaboration must stop short of copying answers.

Assignment submissions will be checked using software to detect cheating. Any violations will be reported to the Dean of Students and Office of Student Affairs.

Final Projects:
Final projects are to be completed individually or assigned groups. All work and products (presentations, papers, posters) are to be completed by you (or your group). Examples of academic misconduct:

- Purchase papers or have someone write a paper for you.
- Make use of a system / tool to write a paper for you.

- Copying ideas, data, or wording without citing your source.

Academic Misconduct:
Academic misconduct in any form will not be tolerated. Evidence of misconduct will result in zero credit for the assignment, drop in one final course grade level (A $\rightarrow$ B, BC $\rightarrow$ CD), and notification with the Office of Academic and Community Conduct. Further evidence will result in immediate failure of the course and again reporting of the misconduct to the University.

Resources on Academic Integrity

- General Resources:
  - `https://www.mtu.edu/conduct/integrity-center/students/`
  - `http://integrity.mit.edu/`
- Avoiding Plagarism:
  - `https://owl.purdue.edu/owl/avoiding_plagiarism/index.html`
  - `https://writing.wisc.edu/handbook/assignments/quotingsources/`
  - `https://guides.lib.unc.edu/plagiarism/purpose`
    The tutorial walks through many issues with plagiarism and gives short quiz questions to gauge or knowledge (note, when talking about academic consequences for plagiarism they refer to UNC's Honor Code and Student Judicial Process that may be different than Michigan Tech's).

Getting Assistance:
If you have questions on the course, I ask you to consider the following options:

- Stop by the instructor's office hours
- Post a general query on the Ed Discussionboard for everyone. Example questions are:
  - ✔ "I don't understand concept X"
  - ✔ "Within problem Y we need to use the definition of Z, can you help explain this again"

  Questions should not be:
  - ✘ "How do I solve problem X"
  - ✘ "What's the answer to problem Y"
- Post a private query on the Ed Discussionboard about your specific submission.
- Email the instructor a clear and detailed question
- Schedule a meeting with the instructor

For more details on academic integrity, please visit the Academic Integrity Resource Center `https://www.mtu.edu/conduct/integrity-center/students/`.

University Policies

Academic Integrity

Academic regulations and procedures are governed by University policy. Academic misconduct cases will be handled in accordance the University's policies.
http://www.mtu.edu/conduct/integrity-center/students/
https://www.mtu.edu/senate/policies-procedures/list-policies/policy109.1.pdf

Assessment Statement

Student work products (exams, essays, projects, etc.) may be used for purposes of university, program, or course assessment. All work used for assessment purposes will not include any individual student identification.
https://www.mtu.edu/senate/policies-procedures/list-policies/policy312.1.pdf

Disability Services

If you have a disability that could affect your performance in any class or that requires an accommodation under the Americans with Disabilities Act, please contact your instructor or Disability Services at 487-1494 as soon as possible so that appropriate arrangements can be made.

Any student requiring accommodations due to a documented disability must provide the instructor of the course notification of needed accommodations no later than five business days prior to the use of the accommodations. In situations where fewer than five days' notice is given, the instructor is encouraged, but not obligated, to provide accommodations. The instructor will determine, in consultation with the Testing Center in the Jackson Center for Teaching and Learning, whether these accommodations can be met.
https://www.mtu.edu/success/disability/

Institutional Equity and Inclusion:

The Office of Institutional Equity has asked that you be made aware of the following:

Michigan Technological University complies with all federal and state laws and regulations regarding discrimination, including the Americans with Disabilities Act of 1990.

Michigan Tech has a policy of affording equal opportunity to all of its employees, students, applicants for employment, and applicants for admission without regard to race, religion, color, national origin, age, sex, sexual orientation, gender identity, height, weight, genetic information, or marital status, disabled veteran status, veteran status, or disability. http://www.mtu.edu/equity/

Mental Health

Michigan Tech is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help or to find additional resources, contact Counseling Services at 906-487-2538 or visit the Counseling Services website, http://www.mtu.edu/counseling.

Veterans / Military

Veterans and active duty military personnel with special circumstances (e.g., upcoming deployments, drill requirements, disabilities) are welcomed and encouraged to communicate

these, in advance if possible, to their instructor(s).
https://www.mtu.edu/registrar/students/veterans/

Equal Opportunity, Discrimination, or Harassment Statement

https://www.mtu.edu/bot/governance/policies/chapter5/sections/5.01-5.05.html
For other concerns about discrimination, you may contact your advisor, Chair/Dean of your academic unit, Dean of Students Office at 487-2212 or The Office of Institutional Equity at 487-3310.