

CSE 801B: Introduction to Data Mining

Fall 2024

Course Information

This course will cover core topics in data mining and their applications. The topics covered include classification, association, clustering, and anomaly detection. The course aims at helping the students understand different techniques that can be applied for different problems and their limitations, evaluate the results and select the appropriate methods when faced with a new problem. It will give students hands-on experience applying those techniques by implementing a complete solution using one or more data mining software packages.

Class Information

Lecture hours: Tue/Thu 12:40pm - 2:00pm

Lecture room: Wells Hall A128

Instructor

Zijun Cui

Email: cuizijun@msu.edu

Office location: Engineering building 2212

Office hour: By appointment via email

Teaching Assistant

Francisco Santos

Email: santosf3@msu.edu

Office location: Engineering building 3312

Office hour: Tuesday 3-4pm; Wednesday 3-4pm

Recommended Textbooks

- Introduction to Data Mining
By: Pang-Ning Tan, Michael Steinbach, Vipin Kumar
- Data Mining: Practical Machine Learning Tools and Techniques **
Morgan Kaufmann series in data management systems
By: Ian Witten, Eibe Frank, Mark Hall
- Data Mining for Business Analytics **
By: Galit Shmueli, Peter C. Bruce, Nitin R. Patel

- Python Data Science Handbook
By: Jake Vanderplas

***Electronic resource available through MSU libraries website: <http://www.lib.msu.edu/>*

Class Structure (subject to changes as class progresses)

Week	Date	Tuesday	Thursday
1	8/26	1-Introduction (HW0 out)	2-Data Preprocessing Part 1
2	9/2	3-Data Preprocessing Part 2	4-Classification (Decision Trees) (HW1 out)
3	9/9	5-Classification (SVM)/Project proposal	6-Classification (Bayes Networks, Bayes)
4	9/16	7-Classification (KNN)	8-Classification (Neural Network) (HW2 out; HW1 due)
5	9/23	9-Classification (Ensemble; classifier comparison) (Project proposal Due)	10-Classification (class imbalance, Multi-class)
6	9/30	11-Association Mining	12-Association Mining (HW3 out; HW2 due)
7	10/7	Exam 1	13-Sequence and Graph Mining
8	10/14 *last day to drop	14-Regression	15-Text Mining (HW4 out; HW3 due)
9	10/21	No class (Fall break)	16-Clustering
10	10/28	17-Clustering – Hierarchical (Project progress report Due)	18-Clustering – Density
11	11/4	19-Clustering – Miscellaneous	20-Anomaly Detection (HW5 out; HW4 due)
12	11/11	21-Time Series Mining	22-General Tips
13	11/18	Exam Review	Exam 2
14	11/25	No Class (Projects)	No Class (Thanksgiving)
15	12/2	Project Presentations	Project Presentations

Grading Policy

The student's final grade will be based on points earned on the exams, homework assignments and final project. The grade will be based on the following scheme:

Homework (5)	35%
Exams (2)	30%
Project	35%

Course Policies

- **[Complete Assignments]** Assignments for this course will be submitted electronically through D2L unless otherwise instructed. Assignments must be submitted by the given deadline or special permission must be requested from the instructor before the due date. Extensions will not be given beyond the next assignment except under extreme circumstances.
- **[Late submission]** Homework assignments are due by the due date. Late submissions with no extension permission will be deducted 50% if submitted within two days after the deadline. Assignments with no extension permission submitted more than two days after the deadline will not be accepted.
- **[Plagiarism]** You are encouraged to form study groups to learn the materials in class. However, all submitted assignments (including computer programs) must be your own work. If plagiarism is detected, students will automatically receive a 0 for the grade and will be reported to the university.
- **[Attendance]** Attendance at all regularly scheduled class meetings is a requirement of this course.
- **[AI tools]** You are welcome to use generative AI tools (e.g. ChatGPT, etc.) in this class as doing so aligns with the course learning goal. You are responsible for the information you submit based on an AI query. Remember, AI is not likely to generate a response that would be seen as quality work. No AI tools are allowed during exams and the final presentation.

Inform Your Instructor of Any Accommodations Needed

Michigan State University is committed to providing equal opportunity for participation in all programs, services and activities. Requests for accommodations by persons with disabilities may be made by contacting the Resource Center for Persons with Disabilities at 517-884-RCPD or on the web at rcpd.msu.edu. Once your eligibility for an accommodation has been determined, you will be issued a verified individual services accommodation (“VISA”) form. Please present this form to me at the start of the term and/or two weeks prior to the accommodation date (test, project, etc). Requests received after this date will be honored whenever possible.

Academic Honesty

Article 2.3.3 of the Academic Freedom Report states that "The student shares with the faculty the responsibility for maintaining the integrity of scholarship, grades, and professional standards." In addition, the Computer Science department adheres to the policies on academic honesty as specified in General Student Regulations 1.0, Protection of Scholarship and Grades; the allUniversity Policy on Integrity of Scholarship and Grades; and Ordinance 17.00, Examinations. (See Spartan Life: Student Handbook and Resource Guide and/or the MSU Web site: www.msu.edu). Therefore, unless authorized by your instructor, you are expected to complete all course assignments, including lab work, quizzes, tests and exams, without assistance from any source.

The Spartan Code of Honor Academic Pledge

As a Spartan, I will strive to uphold values of the highest ethical standard. I will practice honesty in my work, foster honesty in my peers, and take pride in knowing that honor in ownership is worth more than grades. I will carry these values beyond my time as a student at Michigan State University, continuing the endeavor to build personal integrity in all that I do.