

B. Course Syllabus

Course Description

Aimed at extracting useful and interesting patterns and knowledge from large data repositories such as databases and the Web, the field of data mining integrates techniques from databases, statistics, and artificial intelligence. This course will provide a broad overview of the field and focus on a series of advanced topics. The following topics will be covered:

- Knowledge discovery in databases (association rule, clustering, classification);
- Stream data mining;
- Graph mining (PageRank, frequent graph patterns, summarization, linkage prediction);

A major component of this course is a project, which should be targeting a real-world problem and have multiple data mining techniques applied. Besides a list of suggested topics, students are encouraged to propose their own project topic. More details on the project proposal and the expected outcome can be found in the [Course Project module \(https://memphis.instructure.com/courses/142609/modules/452293\)](https://memphis.instructure.com/courses/142609/modules/452293).

Course Objectives

At the end of the course, students will be able to:

- Understand the process of modeling/transforming a real-world problem to an automatic machine computable knowledge discovery framework;
- Learn the principle of data mining techniques, including association rule mining, clustering, and classification;
- Apply clustering/classification techniques to a real-world application, e.g., a movie recommendation system;
- Evaluate the mining outcome using different metrics;
- Apply data mining to graph data (e.g., social networks), e.g., computing PageRank and predicting linkage;

Prerequisites

The official pre-requisite of the course is COMP 3160. However, as Data Mining is a diverse field, it draws on different aspects of the knowledge in fields such as Databases, Artificial Intelligence, and

Statistics. The following is a checklist of material that will be used in the course. It is OK if you do not know all of them, but do try to read up on your own.

- Basic computer algorithms (COMP 4030)
- Undergraduate level statistics/probability (ISDS 2710/MATH 4611)
- Database systems (COMP 7115/ISDS 7605)

Programming skill requirement:

- Familiar with any of {C, C++, Java, Python, Matlab, R, JavaScript, Rust, ...}

Textbooks, Supplementary Materials, Hardware and Software Requirements

There is no required textbook for this course. Some recommended but not mandatory extra reading materials are from the following textbook (online digital version available):

- **Data Mining: Concepts and Techniques, 3rd ed.** Jiawei Han, Micheline Kamber, and Jian Pei. Morgan Kaufmann Series in Data Management Systems Morgan Kaufmann Publishers, July 2011. ISBN 978-0123814791
- **Introduction to Data Mining, 2nd ed.** Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar. Pearson Publisher, 2019. ISBN 978-0133128901
- **Mining of Massive Dataset, 3rd ed.** Jure Leskovec, Anand Rajaraman, Jeff Ullman. (University of Stanford, CS246)

Assessment and Grading

Assessment aspects and weights

As a graduate-level course, the assessment of a student's performance consists of the following parts:

- In-class quizzes: 18%
- Assignments: 32%
- Attendance: 10%
- Course Project: (40%)
 - Proposal: Complete/Incomplete
 - Presentation: 10%
 - Report: 20%
 - Deliverables: 10%

Grading SCALE

We will calculate final letter grades in two different ways; then each student will receive the higher of the two grades. One way is a fixed grading scale, with the following cutoffs:

- $A : \geq 92$; $B : \geq 80$; $C : \geq 68$; $D : \geq 56$; $F : < 56$

The other way is a curve, with the following percentages of students receiving each grade:

- $A : 18\%$; $B : 28\%$; $C : 28\%$; $D : 20\%$; $F : 6\%$

Note that any student with truly exceptional performance will be awarded an A+. Grade F will be given to a student who clearly did not put any effort into the course.

Assignments and Participation

Assignments and Projects

4 Written Assignments (32%)

7 In-class Quizzes, only top-6 scores are counted (18%)

Attendance (10%)

Project Proposal, Report (20%), Presentation (10%), Deliverables (10%)

[Note: For students enrolled in the 8000-level course, you are expected to work on more challenging \(research-intensive\) course projects as you are all Ph.D. students.](#)

Class Participations

Attendance is mandatory and will count for 10% of the total score. All quizzes will be in-class and no makeup credit will be granted.

Punctuality

All assignments are due before midnight (11:59 p.m. CDT) of the due date specified upon release. Late submission is allowed but penalties will be applied. Let T be the number of hours after the deadline:

- 10% points deducted if $24 \geq T > 0$;
- 25% points deducted if $48 \geq T > 24$;
- 50% points deducted if $72 \geq T > 48$;
- Submissions after 72 hours will not be accepted.

Guidelines for Communication

Email

- Always include a subject line.
- Remember without facial expressions some comments may be taken the wrong way. Be careful in wording your emails. The use of emotions might be helpful in some cases.
- Use standard fonts.
- Do not send large attachments without permission.
- Special formatting such as centering, audio messages, tables, HTML, etc. should be avoided unless necessary to complete an assignment or other communication.
- Respect the privacy of other class members

Student Health

In case of any medical or health-related condition that prevents attending the class, please inform the instructor at your earliest convenience and submit doctor's notes timely (within a week of absence).

Plagiarism and Integrity

Plagiarism, cheating, and other forms of academic dishonesty are prohibited. Students guilty of academic misconduct, either directly or indirectly, through participation or assistance, are immediately responsible to the instructor of the class in addition to other possible disciplinary sanctions which may be imposed through the regular institutional disciplinary procedures. Expectations for academic integrity and student conduct are described in detail on the website of the Office of Student Judicial and Ethical Affairs <http://saweb.memphis.edu/judicialaffairs>  (<http://saweb.memphis.edu/judicialaffairs>). Please read in particular, the section about "Academic Dishonesty".

If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor's discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the Office of Student Accountability for further disciplinary action. For further information on the UofM code of student conduct and academic discipline procedures, please refer to <http://www.memphis.edu/studentconduct/misconduct.htm>

Within this class, you are welcome to use foundation models (ChatGPT, GPT, DALL-E, Stable Diffusion, Midjourney, GitHub Copilot, and anything after) in a totally unrestricted fashion, for any purpose, at no penalty. However, you should note that all large language models still have a tendency to make up incorrect facts and fake citations, code generation models have a tendency to produce inaccurate outputs, and image generation models can occasionally come up with highly offensive products. **You will be responsible for any inaccurate, biased, offensive, or otherwise unethical content you submit regardless of whether it originally comes from you or a foundation model. If you use a foundation model, its contribution must be acknowledged in the hand-in;** you will

be penalized for using a foundation model without acknowledgment. Having said all these disclaimers, the use of foundation models is encouraged, as it may make it possible for you to submit assignments with higher quality, in less time.

Students With Disabilities

Qualified students with disabilities will be provided reasonable and necessary academic accommodations if determined eligible by disability services staff at the University of Memphis. Prior to granting disability accommodations in this course, the instructor must receive written verification of a student's eligibility for specific accommodations from the disability services staff. It is the student's responsibility to initiate contact with [Disability Resources for Students](https://www.memphis.edu/drs/)  (<https://www.memphis.edu/drs/>) (DRS) and to follow the established procedures for having the accommodation notice sent to the instructor.

Sexual Misconduct and Domestic Violence Policy

This policy specifically addresses sexual misconduct which includes dating violence, domestic violence, sexual assault, and stalking. The policy establishes procedures for responding to Title IX-related allegations of sexual misconduct. Complaints can be reported to the Office for Institutional Equity (OIE). You may contact OIE by phone at 901.678.2713 or by email at oiememphis.edu (<mailto:oiememphis.edu>). Complaints can be submitted online at [File a Complaint](https://www.memphis.edu/oie/complaint.php)  (<https://www.memphis.edu/oie/complaint.php>). OIE's office is located at 156 Administration Building.

Non-Discrimination and Anti-Harassment Policy

University policy prohibits discrimination and harassment based on protected characteristics and classes. Complaints of discrimination and harassment can be reported to the Office for Institutional Equity (OIE). You may contact OIE by phone at 901.678.2713 or by email at oiememphis.edu. The full text of the policy can be found at [GE2030 - Non-Discrimination and Antiharassment](https://memphis.policytech.com/dotNet/documents/?docid=430&public=true)  (<https://memphis.policytech.com/dotNet/documents/?docid=430&public=true>).

Syllabus Changes

The instructor reserves the right to make changes as necessary to this syllabus. If changes are necessitated during the term of the course, the instructor will immediately notify students of such changes by posting both the notification and the nature of the change(s) on the course bulletin board.