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);
- Text mining (topic modeling, word embedding, computing journalism);
- 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/99012/pages/instructions-on-course-project).

Another important aspect of this course is the literature review, which includes reading the most recent top-tier conference research papers, as well as surveying different techniques of high impact in the context of a particular data mining topic, e.g., clustering, deep neural network. This will be reflected in the discussions set up during the course.

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 text documents, e.g., language model construction, topic modeling, and categorization;
- 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, 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, ...}

Course Topics

Students should finish the first part (association rule mining, clustering, and classification) prior to moving to the text/graph mining part. The first part of this course covers the most fundamental principles for data mining regardless of the underlying data type. While the other two parts can be considered as two important fields of data mining applications in practice.

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
- **Mining of Massive Dataset, 3rd ed.** Jure Leskovec, Anand Rajaraman, Jeff Ullman. (University of Stanford, CS246)

Note that all reading materials will be uploaded for free and easy access.

Hardware and Software Requirements

The minimum requirements can be found at

[https://www.memphis.edu/uofmglobal/services/technology/requirements.php](https://www.memphis.edu/uofmglobal/services/technology/requirements.php)
Assessment and Grading

Assessment aspects and weights

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

- Quizzes: 15%
- Assignment: 30%
- Discussion & Paper review: 15%
- Course Project: (40%)
  - Proposal: 5%
  - Presentation: 15%
  - Report: 15%
  - Deliverables: 5%

Note: A set of coding assignments will be created and released as the course proceeds. Coding assignments are completely optional. It is designed for students who are more comfortable in coding and interested in implementing some classic algorithms themselves. Note that one coding assignment is equivalent to two written assignments. Thus, a student who completes one coding assignment can arbitrarily skip two written assignments.

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:

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

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

10 Written Assignments, 300 points in total (30%)

5 Quizzes, 150 points in total (15%)

Discussion, 150 points in total (15%)
Project Proposal 50 points (5%), Report 150 points (15%), Presentation 150 points (15%), Deliverables 50 points (5%)

Total: 1000 Points

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. Meanwhile, some assignments will have different questions for 7000- and 8000-level enrollments.

Each coding assignment is 60 points. So you can do five coding assignments to meet the total weight of the assignment assessment, which is 300 points.

Class Participations

Students are expected to communicate with the instructor as a learning resource, students must check the course bulletin board frequently for announcements, and students must actively participate in threaded discussion events.

Punctuality

All assignments are due before midnight (11:59 p.m. CDT) of the due date specified upon release.

The discussion participation deadline will be announced upon the creation of the discussion topic.

The course project consists of four parts: a proposal, a report, a presentation (a recorded video), and deliverables. The proposal will be due by the end of the 3rd week. The presentation will be due by the end of week 14. Deliverables and the project report will be due by Dec. 10.

Late submission is allowed but penalties will be applied. Let T be the number of hours after the deadline:

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

Course Ground Rules

- Students are expected to communicate with other students in team projects, learn how to navigate in Canvas, and keep abreast of course announcements;
- Students must use the assigned university e-mail address rather than a personal e-mail address;
- Students must observe course netiquette at all times.
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 emoticons 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

Discussion Groups

- Review the discussion threads thoroughly before entering the discussion. Be a lurker then a discussant.
- Try to maintain threads by using the "Reply" button rather than starting a new topic.
- Do not make insulting or inflammatory statements to other members of the discussion group. Be respectful of others' ideas.
- Be patient and read the comments of other group members thoroughly before entering your remarks.
- Be cooperative with group leaders in completing assigned tasks.
- Be positive and constructive in group discussions.
- Respond in a thoughtful and timely manner.

Student Health

Students who have a positive COVID-19 test should contact the Dean of Students at deanofstudents@memphis.edu.

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.
Please read in particular, the section about "Academic Dishonesty"

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/)(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 oie@memphis.edu. Complaints can be submitted online at [File a Complaint](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 oie@memphis.edu. The full text of the policy can be found at [GE2030 - Non-Discrimination and Anti-harassment](https://memphis.policietech.com/dotNet/documents/?docid=430&public=true).
Technology Requirements

The following is a list of the minimum requirements to use our learning management system. Some courses will have more advanced requirements.

- Access to a reliable, high-speed Internet connection (DSL or Cable).
- Test your device to ensure it is compatible with our LMS (Learning Management System) using the [System Check Wizard](https://memphis.instructure.com/courses/40191/file_contents/course%20files/d2l/systemCheck).
- Open PDF files using the free downloadable PDF software.
- Access Flash-based content with [Adobe Flash Player](https://get.adobe.com/flashplayer/) (free).
- Use Microsoft Office for document creation (available for students via [umapps.memphis.edu](http://umapps.memphis.edu/)).


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 both by individual email communication and by posting both notification and nature of change(s) on the course bulletin board.

Technical Support

Call the Helpdesk: 901-678-8888

Online Helpdesk: [https://www.memphis.edu/umtech/service_desk/contact.php](https://www.memphis.edu/umtech/service_desk/contact.php)