Course Syllabus for SIADS 532: Data Mining I

Course Overview and Prerequisites

This course will introduce basic concepts and tasks of data mining. It focuses on how to formally represent real world information as basic data types (itemsets, matrices, and sequences) that facilitate downstream analytics tasks. Students will learn how to characterize each type of data through pattern extraction and similarity measures.

Course prerequisites: Mathematics Methods for Applied Data Science, Data Manipulation

Instructor and Course Assistants

- Instructor: Qiaozhu Mei
- Co-instructors: Rachel Wyatt, Emily Schemanske , Amanda Hardy

Communication Expectations

Contacting instructor and course assistants: Course channel in Slack (siads532_wi24_004)
Email response time: 24 - 48 hours
Slack response time: 24 - 48 hours
For Direct Messages, please tag both instructors.
Office hours: see ***Course Schedule*** below

Required Textbook

1. Data mining: concepts and techniques. Han, J., Pei, J. and Kamber, M. (3rd Edition) [DMCT]
2. Mining of massive datasets. Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman. (2nd Edition) [MMDS]

Online access to these textbooks are provided through the University of Michigan Library. You may be asked to sign in with your UMich uniquename and password to access these materials.

Technology Requirements (unique to this course)

None

- *If you encounter a technical issue with Coursera or admin issues with slack, please submit a report to the ticketing system at umsimadshelp@umich.edu*

Accessibility

Screen reader configuration for Jupyter Notebook Content

Learning Outcomes

1. Understand the basic concepts of knowledge discovery from data.
2. Describe the basic computational tasks of data mining, including associations, retrieval, classification, clustering, ranking, prediction, and outlier detection.
3. Formulate real world data as item sets, matrices, or sequences and articulate the pros and cons.
4. Extract patterns and associations from itemsets, matrices, and sequences.
5. Compute similarities/distances of item sets, vectors, and sequences.
6. Know how these techniques are applied to different domains.

Course Schedule

This course **begins on Tuesday, April 2, 2024** and **ends on Monday, April 29, 2023=4**.

Weekly **Quizzes** and **Programming Assignments will be due on Tuesday at 11:59 pm except for last week when assessments are due on Monday, April 29** (time zone = Ann Arbor, Michigan = Eastern Time).

Schedule of Weekly Office Hours via Zoom (time zone = Ann Arbor, Michigan = Eastern Time):

- Rachel Wyatt: 4 -5 pm Monday
- Qiaozhu Mei: 10 - 11 am Wednesday **(First Week is 9 - 9: 30 am, Tuesday)**
- Emily Schemanske : 8 - 9 am Friday
- Amanda Hardy:
- Passcode to all office hours are **532**
- Access via Live Events from the course menu

Grading

Course Assignment	**Percentage of Final Grade**
Wk 1 Quiz Part 1 - Basic Concepts	2.5%
Wk 1 Quiz Part 2 - Data Representations	3.75%
Wk 1 Quiz Part 3 - Solving a Real World Problem	13.75%
Wk 1 Assignment 1 - Patterns and Similarities	5%

Wk 2 Assignment 2 - Part 1 - Itemset Data	1.25%
Wk 2 Assignment 2 - Part 2 - Frequent Itemsets	5%
Wk 2 Assignment 2 - Part 3 - Apriori under the Hood	10%
Wk 2 Assignment 2 - Part 4 - Evaluating Frequent Itemsets	3.75%
Wk 2 Assignment 2 - Part 5 - Itemset Similarity	5%
Wk 3 Assignment 3 - Part 1 - Vector and Matrix Data	1.25%
Wk 3 Assignment 3 - Part 2 - Vector Similarity and Distance	11.25%
Wk 3 Assignment 3 - Part 3 - Finding Similar Restaurants	6.25%
Wk 3 Assignment 3 - Part 4 - Singular Value Decomposition	6.25%
Wk 4 Assignment 4 - Part 1 - Sequence Data and Ngrams	10%
Wk 4 Assignment 4 - Part 2 - Edit Distance	7.5%
Wk 4 Assignment 4 - Part 3 - Shingling and Near Duplicates	5%
Wk 4 Assignment 4 - Part 4 - Application: Sequence Alignment	2.5%
Total	**100%**

Note: All assignments are required to earn credit for this course.

Letter Grades, Course Grades, and Late Submission Policy

Refer to the MADS Assignment Submission and Grading Policies section of the UMSI Student Handbook (access to Student Orientation course required)

For this course, the late submission policy is 15% reduction if assignment is turned in one day late, 30% reduction if two days late, 50% if reduction three days, and a zero (0) if four or more days late.

Note: we understand that the fluid Covid-19 situation may have various influences on our working schedule and conditions. If you have specific needs, please let us know and we'd be happy to discuss and accommodate them on top of the default late submission policy.

Accommodations

Refer to the Accommodations for Students with Disabilities section of the UMSI Student Handbook (access to the Student Orientation course required).

Use the Student Intake Form to begin the process of working with the University's Office of Services for Students with Disabilities.

Help Desk(s): How to Get help

- Degree program questions or general help - umsimadshelp@umich.edu
- Coursera's Technical Support (24/7) - https://learner.coursera.help/

Library Access

Refer to the U-M Library's information sheet on accessing library resources from off-campus. For more information regarding library support services, please refer to the U-M Library Resources section of the UMSI Student Handbook (access to the Student Orientation course required).

Student Mental Health

Refer to the University's Resources for Stress and Mental Health website for a listing of resources for students.

Student Services

Refer to the Introduction to UMSI Student Life section of the UMSI Student Handbook (access to the Student Orientation course required).