

Course Syllabus for SIADS 632: Data Mining II

Course Overview and Prerequisites

This course extends Data Mining I and introduces additional data representations and tasks involved in mining real world data, with a particular focus on sequence modeling, time series analysis, and mining data streams. It introduces how to extract patterns, compute similarities/distances of data, and make predictions under these data representations.

Instructor and Course Assistants

Instructor: Qiaozhu Mei - qmei@umich.edu

Course Assistant: Yumou Wei - yumouwei@umich.edu

Quan Nguyen - quangu@umich.edu

Course Communication Expectations

- Contacting instructor and course assistant: Course channel in Slack
- Email response time: 24 - 48 hours
- Slack response time: 24 - 48 hours
- Office hours:
 - Qiaozhu Mei (**Thursdays 10am-11am**, Ann Arbor, Michigan time-Eastern Daylight Time - EDT, UTC -4)
 - Yumou Wei (**Mondays 9pm-10pm**, Ann Arbor, Michigan time-Eastern Daylight Time - EDT, UTC -4)
 - Quan Nguyen (**Fridays 4pm-5pm**, Ann Arbor, Michigan time-Eastern Daylight Time - EDT, UTC -4)

How to Get Help

If you have questions concerning the degree program, encounter a technical issue with Coursera, or issues using Slack, please submit a report to the ticketing system at umsimadshelp@umich.edu.

If you have an issue specific to the Coursera environment, you can also begin a [live chat session](#) with Coursera Technical Support (24/7) or view [Coursera troubleshooting guides](#). (you may be asked to log in to your Coursera account).

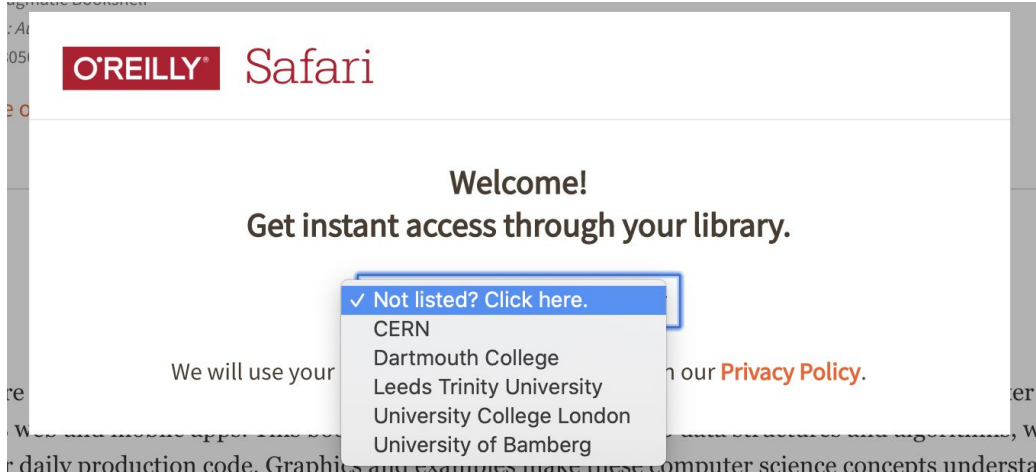
For questions regarding course content, refer to the **Communications Expectations** section below.

Weekly Readings or Textbook Information

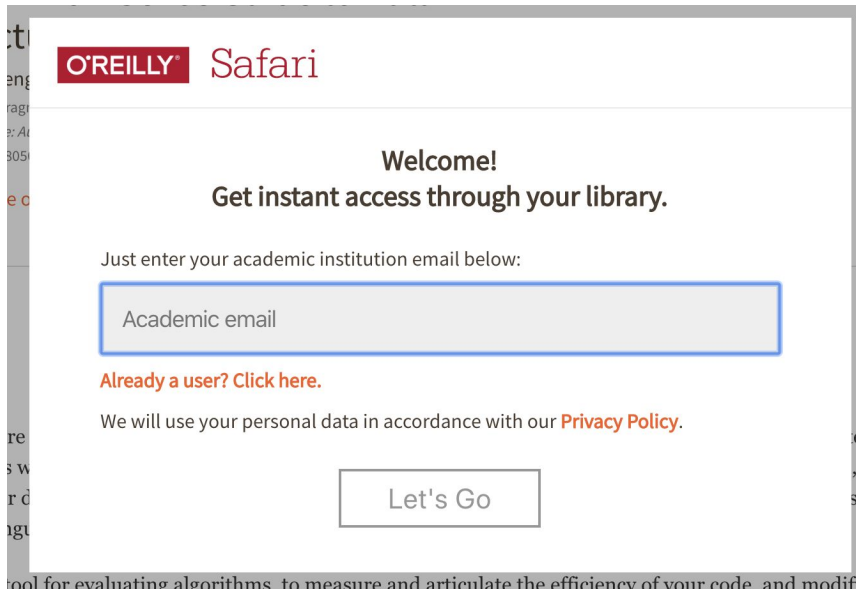
- [Speech and Language Processing](#) (3rd ed. draft) Dan Jurafsky and James H. Martin. ([Chapter 3](#) and [Appendix Chapter A](#)).
- [Data Mining: the Textbook](#), Charu C. Aggarwal. (Chapter 12 and Chapter 14).

To access the required textbooks, simply click on the links above, which will direct you to the U-M Library website. Scroll down the page and click on *Available Online* (some texts will have multiple online options, but it is recommended you choose **Safari books online**).

After you are directed to the textbook, you will see an O'Reilly pop-up window asking for you to select your institution. U-M is not an available option, so you will need to select the option "Not Listed? Click Here".



You will be prompted to input your U-M email address (no password required).



Learning Outcomes

- Be able to formulate real world data as sequences, time series, or data streams.
- Be able to formulate a real world problem as sequence prediction and solve it using N-Gram language models.
- Be aware of how Hidden Markov Models work.
- Extract patterns from time series data, including trends, seasons, cycles, and outliers.
- Measure similarity between time series.
- Conduct time series forecasting using autoregressions.
- Articulate the restriction of data streams and strategies for mining data streams.
- Implement Reservoir sampling, Bloom filter, and lossy counting.
- Name real world applications of these data representations and methods.

Course Schedule

- **This course begins on February 3, 2021 and ends on March 2, 2021.**
- Weekly assignments will be **due on Wednesday at 11:59 pm** (Ann Arbor, Michigan time-Eastern Daylight Time - EDT, UTC -4).

Weekly Office Hours via Zoom (Ann Arbor, Michigan time):

Your instructor will hold weekly, synchronous office hours using the video-conferencing tool, Zoom. The schedule of office hours can be found by clicking on the **Live Events** link in the left-hand navigation menu. Additionally, all office hours will be recorded and archived so that you can retrieve them at a later date. Archived office hours can be found in the respective module of the course.

Grading

Course Item	Percentage of Final Grade	Due
Week 1 Programming Assignment Group	25% (40%, 30%, 30% per assessment)	Wednesday 2/10 11:59pm
Week 2 Programming Assignments	25% (50% per assessment)	Wednesday 2/17 11:59pm
Week 3 Programming Assignment	25% (50% per assessment)	Wednesday 2/24 11:59pm
Week 4 Programming Assignments	25% (50% per assessment)	Wednesday 3/3 11:59pm
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.

The grading scale for this course is as follows:

A	93%
A-	90%
B+	87%

B	83%
B-	80%
C+	77%
C	73%
C-	70%
D+	67%
D	63%
D-	60%
F	0%

Academic Integrity/Code of Conduct

Refer to the [Academic and Professional Integrity](#) section of the UMSI Student Handbook. (access to Student Orientation course required).

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.

Accessibility

Refer to the [Screen reader configuration for Jupyter Notebook Content](#) document to learn accessibility tips for Jupyter Notebooks.

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).

Technology Tips

- Recommended Technology
 - This program requires Jupyter Notebook for completion of problem sets and Adobe or other PDF viewer for reading articles.
- Working Offline

- While the Coursera platform has an integrated Jupyter Notebook system, you can work offline on your own computer by installing Python 3.5+ and the Jupyter software packages, including pyspark. For more details, consult the Jupyter Notebook FAQ.