# 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.

#### **Instructional Team**

Instructor: Paramveer Dhillon - dhillonp@umich.edu

Course Support: Anmol Panda - anmolp@umich.edu; Yachuan Liu - yachuan@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 hour sessions will be recorded for the benefit of students who are unable to join at these times. Password to join any Office hours is **632**

# 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 <u>umsimadshelp@umich.edu</u>.

If you have an issue specific to the Coursera environment, you can also begin a <u>live chat session</u> with Coursera Technical Support (24/7) or view <u>Coursera troubleshooting guide</u>s. (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

- <u>Speech and Language Processing</u> (3rd ed. draft) Dan Jurafsky and James H. Martin. (<u>Chapter 3</u> and <u>Appendix Chapter A</u>).
- 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**).

| <u>Available online</u> | <u>Get This</u> | N/A |
|-------------------------|-----------------|-----|
|-------------------------|-----------------|-----|

Access to the Safari books online online version restricted; authentication may be required: 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).

| ļ   | o'reilly' Safari   |  |  |  |
|-----|--|--|--|--|
|     | Welcome!<br>Get instant access through your library.                               |  |  |  |
|     | Just enter your academic institution email below:                                  |  |  |  |
|     | Academic email   |  |  |  |
|     | Already a user? Click here.  |  |  |  |
|     | We will use your personal data in accordance with our <b>Privacy Policy</b> .      |  |  |  |
|     | Let's Go   |  |  |  |
| for | r maluating algorithms, to measure and articulate the officiency of your code, and |  |  |  |

## 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 1, 2023 and ends on February 28, 2023.

 Weekly assignments will be due on Wednesdays 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. All office hours will have a passcode of **632.** Weekly Office Hours Schedule is below.

- Paramveer Dhillon: Wednesdays, 10-11 AM
- Anmol Panda: Mondays, 10-11 AM

# Grading

| Course Item                            | Percentage of Final Grade             | Due                     |
|--|---------------------------------------|-------------------------|
| Week 1 Programming<br>Assignment Group | 25% (40%, 30%, 30% per<br>assessment) | Wednesday 2/8, 11:59pm  |
| Week 2 Programming<br>Assignments      | 25% (50% per assessment)              | Wednesday 2/15, 11:59pm |
| Week 3 Programming<br>Assignment       | 25% (50% per assessment)              | Wednesday 2/23, 11:59pm |
| Week 4 Programming<br>Assignments      | 25% (50% per assessment)              | Wednesday 3/1, 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 <u>MADS Assignment Submission and Grading Policies</u> 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:

| А  | 93% |
|----|-----|
| A- | 90% |
| B+ | 87% |
| В  | 83% |

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

# Academic Integrity/Code of Conduct

Refer to the <u>Academic and Professional Integrity</u> section of the UMSI Student Handbook. (access to Student Orientation course required).

#### Accommodations

Refer to the <u>Accommodations for Students with Disabilities</u> section of the UMSI Student Handbook (access to the Student Orientation course required). Use the <u>Student Intake Form</u> to begin the process of working with the University's Office of Services for Students with Disabilities.

## Accessibility

Refer to the <u>Screen reader configuration for Jupyter Notebook Content</u> document to learn accessibility tips for Jupyter Notebooks.

## Library Access

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

## Student Mental Health

Refer to the University's <u>Resources for Stress and Mental Health website</u> 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.