Course Overview and Prerequisites

This course introduces students to working with language data. Students will learn how to use common techniques and software libraries to turn unstructured text into structured information for use with other data science techniques, and will become familiar with key concepts relating to the semantics and syntax of language.

The prerequisites for SIADS 655 include:

- SIADS 542: Supervised Learning (C- or better)
- SIADS 643: Machine Learning Pipelines (C- or better)

Instructor and Course Assistants

Instructors:

Noha Ghannam (nmghanna@umich.edu Opens in a new tab) Adjunct Lecturer in Information, School of Information

Alexis Castellanos (acastel@umich.edu Opens in a new tab) Adjunct Lecturer in Information, School of Information

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 Opens in a new tab .

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

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

Course Communication Expectations

Slack is the preferred communication tool for this course. We will post announcements and regular supplementary material on Slack; please try to monitor the Slack channels for the course regularly.

If you have questions about course content (e.g. lecture videos or assignments), please make sure to use Slack. Instructor and course assistant response time to **Slack** messages will aim to be within 24 hours. The course staff is based in Eastern Time, and we are more likely to be around during daytime. Please plan accordingly as you work on your assignments. If you wish to direct-message us with questions, please message **all** of us so we can respond more quickly.

Personal communication that may involve sensitive information may be emailed directly to the instructor or course assistant. If you email the instructor or course assistant, please include SIADS 655 in the email subject. Instructor and course assistants' response time to **email** messages will be within 24 hours.

Textbook Information

This course will use the following textbook:

Jurafsky, Dan, and Martin, James H. Speech and Language Processing : An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Third Edition draft. Opens in a new tab

Learning Outcomes

- Train a text classifier using simple text features for any task
- Use the basic forms of text data and how to work with each in a machine learning setting
- Evaluate text-based models
- Work with word vectors in classification settings
- Describe and use the structure of language (syntax) and how to make use of it in applications
- Train word vector models
- Recognize that words have different meanings that can be distinguished
- Use basic operations in NLP software libraries like Gensim and NLTK
- Be aware of many NLP applications and how they relate to the content covered in this course
- Be aware of parsing algorithms
- Be aware of graphical models like HMMs and CRFs
- Be aware of social, pragmatic, and common-sense knowledge in text

Course Schedule

- This course begins on **Tuesday, April 2, 2024,** and ends on **Monday, April 29, 2024, at 11:59 p.m. EST.**
- Assignments 1 and 2 will be due on Monday, April 15, 2024 at 11:59 pm EST.
- Assignments 3 and 4 will be due on Monday, April 29, 2024 at 11:59 pm EST.

Week 1: Build and train multiple standard text classifiers, learn how to turn text into meaningful features, understand how hyper-parameters affect performance, and plot results for different amounts of training data to see how textual variability affects performance.

Week 2: Train a word2vec model on data, compare off-the-shelf vectors with the trained vectors in a classification task, use those vectors to measure word similarity in an evaluation.

Week 3: Write programs that use part of speech tagging, dependency parsing, and coreference models to extract simple information from a text corpus and analyze that information.

Week 4: Train a word sense disambiguation (WSD) system using classifiers, compare performances for unsupervised and supervised WSD systems, and, train a classifier to predict social/pragmatic information from text (e.g., sarcasm)

Weekly Office Hours via Zoom (Ann Arbor, MI 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. Passcode for Zoom sessions is 655.

- **Alexis Castellanos:** Fridays at 4:00 pm EST
- **Noha Ghannam**: Wednesday at 4:00 pm EST

Grading

Course Item	Percentage of Final Grade	Due Date

Week 1 Notebook Assignment	25%	Monday, April 15, 2024 at 11:59 pm EST
Week 2 Notebook Assignment	25%	Monday, April 15, 2024 at 11:59 pm EST
Week 3 Notebook Assignment	25%	Monday, April 29, 2024 at 11:59 pm EST.
Week 4 Notebook Assignment	25%	Monday, April 29, 2024 at 11:59 pm EST.
Total	100%	

Letter Grades, Course Grades, and Late Submission Policy

This course requires you to complete all four assignments for full credit. You'll see from the above grading schema that half of the assignments will be due in the second week of class, and the other half will be due at the end of the session. **Do not procrastinate - it is your responsibility to manage your time and submit assignments in a timely manner.**

If due to mitigating circumstances you need an extension, please let us know at least 96 hours ahead of time by contacting all of the course staff on Slack.

Unless you have already arranged for an extension, any assignments submitted after the respective due dates will be awarded zero points.

The grading scale for this course is as follows:

A+	**97%**
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 Opens in a new tab section of the UMSI Student Handbook. (access to Student Orientation course required).

Accommodations

Refer to the Accommodations for Students with Disabilities Opens in a new tab section of the UMSI Student Handbook.

Use the Student Application Form in Accommodate Opens in a new tab 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 Opens in a new tab document to learn accessibility tips for Jupyter Notebooks.

Library Access

Refer to the U-M Library's information sheet Opens in a new tab on accessing library resources from off-campus. For more information regarding library support services, please refer to the U-M Library Resources Opens in a new tab 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 Opens in a new tab for a listing of resources for students.

Student Services

Refer to the Introduction to UMSI Student Life Opens in a new tab 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. For more details, consult the Jupyter Notebook FAQ Opens in a new tab . Be aware that due to software version differences, you may get slightly different output on your own computer vs on Coursera. We will grade according to whatever output is produced on Coursera.