Course Syllabus

Jump to Today

MSDS 453 - Natural Language Processing

Course Description

This course explores cutting-edge developments in computational linguistics and machine learning, with a focus on deep learning techniques. Students work with unstructured and semi-structured text, transforming text into numerical vectors and converting higher-dimensional vectors into lower-dimensional ones for analysis and modeling. The course covers parts-of-speech parsing, information extraction, semantic processing, text classification, sentiment analysis, text embeddings, topic modeling, text summarization and generation, and question answering. Students explore large-scale language models, particularly generative pretrained transformers (GPTs). This is a project-based course with extensive programming assignments.

Course Objectives

By the end of this course, you will be able to:

- Understand the fundamentals of natural language processing (NLP), the challenges and applications of NLP, and how Artificial Intelligence has changed the landscape of NLP.
- Understand issues related to data acquisition and preprocessing of text. Be able to acquire data and perform preprocessing of text including normalization.
- Assess data quality with respect to the task and methods used to solve the task.
- Write code that can extract entities and their relationships from text; be able to identify, characterize, and apply methods for entity co-resolution; identify and select strategies for complex concept extraction; and apply appropriate text vectorization methods (specifically Google's Word2Vec and Doc2Vec). Develop methods for semantic processing of text.
- Identify, select, and apply both clustering and classification algorithms, together with other forms of machine learning including both supervised and unsupervised, including generative machine learning methods such as latent semantic analysis (LSA) and latent Dirichlet Allocation (LDA). Select, apply, and evaluate methods for sentiment analysis.
- Apply and use deep learning-based language models to perform common NLP-based tasks for improved performance.
- Develop (at a basic level) methods for text generation and summarization using neural models.
- Understand the fundamentals of question answering and commonsense reasoning in NLP.
- Develop methods that make use of models such as ChatGPT to enable natural language understanding and Chatbot technologies.

Prerequisites

Prerequisites: (1) MSDS 420-DL Database Systems or CIS 417 Database Systems Design and Implementation and (2) MSDS 422-DL Practical Machine Learning or CIS 435 Practical Data Science Using Machine Learning.

Suggested Prior Coursework and Competencies

We will use certain deep learning algorithms and processes. These often require mathematical notation; specifically, vector and matrix operations. **All students need to be conversant with basic vector and matrix algebra.**

Other Useful Knowledge: you will need a **working knowledge of Python**. You may be able to carry out your work using R. However, example codes and related support will all be in Python, with the addition of scikit-learn and TensorFlow, as well as Keras.

Diversity Statement

As educators and learners, we must share a commitment to diversity and equity, removing barriers to education so that everyone may participate fully in the community. In this course, we respect and embrace the unique experiences that brought each person here, including backgrounds, identities, learning styles, ways of expression, and academic interests. The broad spectrum of perspectives represented by our students enriches everyone's experiences, and we strive to meet each perspective with openness and respect.

Required and Optional Resources

Required Readings (all accessible through the course; nothing to purchase through the bookstore)

- Dan Jurafsky, D. and James H. Martin, J.. 2009. ***Speech and Language Processing*** (http://www.cs.colorado.edu/~martin/SLP/Updates/1.pdf) (2nd ed) (Just chapter 1) (Free)
- Dan Jurafsky, D. and James H. Martin, J. 2019. ***Speech and Language Processing*** *(https://web.stanford.edu/~jurafsky/slp3/)* (3rd ed. draft) (Free)
- Brownlee, J. 2019. ***Deep Learning for Natural Language Processing: Develop Deep Learning Models for Natural Language in Python.*** (https://machinelearningmastery.com/deep-learning-for-nlp/%C2%A0) An electronic book with accompanying source code available directly from the author.
- Lane, H., C. Howard, and H. M. Hapke. 2019. ***Natural Language Processing in Action: Understanding, Analyzing, and Generating Text with Python*** (https://www.manning.com/books/natural-language-processing-in-action) . Shelter Island, N.Y.: Manning. [ISBN-13: 978-1617294631]. **Code.** (https://github.com/totalgood/nlpia)
- Tunstall L., Werra L., and Wolf T. 2022. ***Natural Language Processing with Transformers*** (https://learning.oreilly.com/library/view/natural-language-processing/9781098136789/%C2%A0) . [ISBN-13: 978-1098136796]
- Raaijmakers S. 2022. ***Deep Learning for Natural Language Processing*** *(https://learning.oreilly.com/library/view/deep-learning-for/9781617295447/%C2%A0)* . [ISBN-13: 978-1617295447]

Reference Textbooks and Online Resources

- Bird, S., E. Klein, and E. Loper 2009. *Natural Language Processing with Python*. Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-0-596-51649-9] **Available online.** (http://www.nltk.org/book/%C2%A0)
- Büttcher, S., C. L. A. Clarke, and G. V. Cormack 2010. *Information Retrieval: Implementing and Evaluating Search Engines*. Cambridge, Mass..: MIT Press. [ISBN-13: 978-0-262-02651-2]
- Cai, S., S. Bileschi, E.D. Nielsen, and F. Chollet 2020. *Deep Learning with JavaScript: Neural Networks in TensorFlow.js*. Shelter Island, N.Y.: Manning [ISBN-13 9781617296178] **Code.** (https://github.com/tensorflow/tfjs-examples)
- Chollet, F. 2018. *Deep Learning with Python. Shelter Island*, N.Y.: Manning. [ISBN-13: 978-1617294433] **Code.** (https://github.com/fchollet/deep-learning-with-python-notebooks.git%C2%A0)
- Chollet, F. and J.J. Allaire 2018. *Deep Learning with R.* Shelter Island, N.Y.: Manning. [ISBN-13: 9781617295546] **Documentation and code examples.** (https://tensorflow.rstudio.com/)
- Foster, D. 2019. *Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play.* Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-1492041948] **Code.** (https://github.com/davidADSP/gdl_code%20)
- Géron, A. 2019. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (second ed.), Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-1492032649] **Code.** (https://github.com/ageron/handson-ml2)
- Izenman, A. J. 2008. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. New York: Springer*. [ISBN-13: 978-0-387781884] Available from the **Springer collection.** (http://link.springer.com.turing.library.northwestern.edu/%C2%A0)
- Lane, H., C. Howard, and H. M. Hapke 2019. *Natural Language Processing in Action: Understanding, Analyzing, and Generating Text in Python*. Shelter Island, N.Y.: Manning. [ISBN-13: 978-1617294631] **Code.** (https://github.com/totalgood/nlpia%C2%A0)
- Manning, C.D., P. Raghavan, and H. Schutze 2008. *Introduction to Information Retrieval*. New York: Cambridge University Press. [ISBN-13: 978-0-521-86571-5] **Available online.** (https://nlp.stanford.edu/IR-book/%C2%A0)
- Mihalcea, R. and D. Radev 2011. *Graph-based Natural Language Processing and Information Retrieval.* Cambridge, UK: Cambridge University Press. [ISBN-13: 978-0521896134]
- Mitchell, R. 2018. *Web Scraping with Python: Collecting More Data from the Modern Web* (second ed.). Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-1491985571] **Source code.** (https://github.com/REMitchell/python-scraping)
- Müller, A. C. and S. Guido 2017. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. Sebastopol, Calif.: O'Reilly. [ISBN-13: 9781449369415] **Source code.** (https://github.com/amueller/introduction_to_ml_with_python%C2%A0)
- Patel, A. A. 2019. *Hands-on Unsupervised Learning using Python: How to Build Applied Machine Learning Solutions from Unlabeled Data.* Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-1-492-03564-0] **Code.** (https://github.com/aapatel09/handson-unsupervised-learning%C2%A0)
- Russell, M. A. and M. Klassen 2019. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More* (third ed.). Sebastopol, Calif.: O'Reilly. [ISBN-13: 978-

1491985045] **Code.** (https://github.com/mikhailklassen/Mining-the-Social-Web-3rd-Edition%C2%A0)

- Weiss, S. M., N. Indurkhya, and T. Zhang 2015. *Fundamentals of Predictive Text Mining* (second ed.). New York, Springer. [ISBN-13: 978-1447167495] Available from the **Springer collection** (http://link.springer.com.turing.library.northwestern.edu/%C2%A0%C2%A0) .

Course Reserves and Optional Readings and Resources

Selected readings are available through the Course Reserves. For assistance with Course Reserves, use electronic mail: **e-reserve@northwestern.edu** (mailto:e-reserve@northwestern.edu) . Visit Northwestern's **Ask A Librarian** (http://www.library.northwestern.edu/about/library-administration/departments-offices/reference-department/ask-librarian) for questions.

The **Springer Collection** (https://link-springer-com.turing.library.northwestern.edu/%C2%A0%C2%A0) is especially useful in this course.

NLP and Python

Check out entries for natural language processing and Python. Additional resources of note include the Springer book series Text, Speech, and Language Technology, which includes 46 volumes published between 1997 and 2015. Use your NUID log-in from Northwestern University to access **this series.** (https://link-springer-com.turing.library.northwestern.edu/bookseries/6636%C2%A0%C2%A0)

Additional online references include proceedings from the Association for Computational Linguistics. Of special note are those from the annual Conference on Empirical Methods in Natural Language Processing:

- **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing** (https://aclanthology.org/volumes/2021.emnlp-main/)
- **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations** (https://aclanthology.org/volumes/2022.emnlp-demos/)
- **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing** (https://aclanthology.org/volumes/2022.emnlp-main/)

Assignment Overview and Grading Breakdown

Grading and feedback turnaround will be one week from the due date. You will be notified if turnaround will be longer than one week. The discussion forums and written assignments will be graded based on specific grading guidelines.

Assignments: There are no quizzes or exams. Rubrics are provided for assignments.

Deliverable	***Description***	***Week Assigned***	***Week Due***	***Total Points***

Class Participation	Graded discussion threads (10 points per week)	Beginning of Week	End of Week	100 Points
Project Topic	Choose Project Topic	Week 2	Week 3	0 Points
Assignment 1	First Research/Programming Assignment	Week 3	Week 5	100 Points
Project Data	Acquire Data for Project	Week 4	Week 6	0 Points
Assignment 2	Second Research/Programming Assignment	Week 5	Week 7	100 Points
Project Proposal	Project Check-in		Week 7	0 Points
Assignment 3	Third Research/Programming Assignment	Week 7	Week 9	100 Points
Project	Final Project Report		Week 10	200 Points
	Total			600 Points

Grading Scale

Grade	Percentage	Total Points
A	94%–100%	558– 600 points
A-	90%–93%	546 – 557 points
B+	87%–89%	522 – 545 points
B	84%–86%	498 – 521 points

B-	80%–83%	480 – 497 points
C+	77%–79%	462– 479 points
C	74%–76%	438 – 461 points
C-	70%–73%	420 – 437 points
F	Below 70%	0 – 419 points

The School of Professional Studies does not award D grades in graduate coursework.

Late Work Policy

Students should provide written notification of late assignment work 24 hours prior to the deadline. A grace day is allowed for those who provide late work notification. Late papers will be subject to point reductions.

Online Communication and Interaction Expectations

Discussion Forums

The purpose of the discussion boards is to allow students to freely exchange ideas. It is imperative to remain respectful of all viewpoints and positions and, when necessary, agree to respectfully disagree. While active and frequent participation is encouraged, cluttering a discussion board with inappropriate, irrelevant, or insignificant material will not earn additional points and may result in receiving less than full credit. Frequency matters, but contributing content that adds value is paramount. Please remember to cite all sources and to avoid plagiarism. Although we will use Chicago style (*CMOS* 17) for written work in this course, you are not required to use any particular style when citing or referencing in your discussion posts. Please post your viewpoints first and then discuss others' viewpoints.

The quality of your posts and how others view and respond to them are most valued. A single statement mostly implying "I agree" or "I do not agree" is not counted as a post. Explain, clarify, politely ask for details, provide details, persuade, and enrich communications for a great discussion experience. Please note, there is a requirement to respond to at least two fellow class members' posts.

Online Communication Etiquette

Beyond interacting with your instructor and peers in discussions, you will be expected to communicate by Canvas message, email, and sync session. Your instructor may also make themselves available by phone or text. In all contexts, keep your communication professional and

respect the instructor's posted availability. To learn more about professional communication, please review the **Connecting with Faculty (https://www.northwestern.edu/academic-support-learning/academic-resource-directory/find-strategies-for-success/connecting-with-faculty.html)** guide.

Just as you expect a response when you send a message to your instructor, please respond promptly when your instructor contacts you. Your instructor will expect a response within two business days. This will require that you log into the course site regularly and set up your notifications to inform you when the instructor posts an announcement, provides feedback on work, or sends you a Canvas message. For guidance on setting your notifications, please review **How do I set my Canvas notification settings as a student? (https://community.canvaslms.com/docs/DOC-10624)** Check your Northwestern email account regularly, or forward your u.northwestern email to an account you check frequently.

Participation and Attendance

This course follows the asynchronous distance learning approach of Northwestern University School of Profession Studies. The course does not meet at a particular time each week. Learning objectives and assessments are supported through classroom elements that can be accessed at any time. To measure class participation (or attendance), your participation in threaded discussion boards is required, graded, and paramount to your success in this course.

Real-time, synchronous meetings (Sync Sessions) are scheduled for Wednesday or Thursday evenings on selected weeks. Sync Sessions are conducted with Zoom. While your attendance is highly encouraged, it is not required. You will not be graded on your attendance or participation. Sync Sessions will be recorded by the instructor for educational purposes. These recordings will be shared only with students enrolled in the course and will be deleted at the end of the quarter. Your instructor will communicate how you can access the recordings.

Unauthorized student recording of classroom or other academic activities (including advising sessions or office hours) is prohibited. Unauthorized recording is unethical and may also be a violation of university policy and state law. Students requesting the use of assistive technology as accommodation should contact AccessibleNU. Unauthorized use of classroom recordings – including distributing or posting them – is also prohibited.

Under the University's Copyright Policy, faculty own the copyright to instructional materials – including those resources created specifically for the purposes of instruction, such as syllabi, lectures and lecture notes, and presentations. Students cannot copy, reproduce, display or distribute these materials. Students who engage in unauthorized recording, unauthorized use of a recording or unauthorized distribution of instructional materials will be referred to the appropriate University office for follow-up. Refer to the weekly schedule at the end of this syllabus for details about weekly learning objectives, required and optional readings, assignments, and Sync Sessions.

Study Groups (optional)

A structure for optional student study groups will be set up to foster a collaborative learning environment. Zoom is available as a conferencing tool. See the People/Study Groups menu and tab

items of Canvas.

Each student is encouraged to join a study group. It may make sense to join a group based on time zone (Eastern/Central versus Mountain/Pacific) and preferred computing environment (Windows 10, Mac OSX, or Google Colaboratory). It is recommended that each study group consists of no more than four students.

Faculty Observer

Please note that this course may have a faculty observer for this term. The observer is present in the Canvas site and some Zoom sessions exclusively for training purposes. They will not be responsible for or engage in any in-class interactions, student assessment or grading, or any other aspect of course delivery. If you have questions or concerns about the faculty observer, please contact your instructor.

Student Support Services

AccessibleNU

This course is designed to be welcoming to, accessible to, and usable by everyone, including students who are English-language learners, have a variety of learning styles, have disabilities, or are new to online learning. Be sure to let me know immediately if you encounter a required element or resource in the course that is not accessible to you. Also, let me know of changes I can make to the course so that it is more welcoming to, accessible to, or usable by students who take this course in the future.

Northwestern University and **AccessibleNU (https://www.northwestern.edu/accessiblenu/)** are committed to providing a supportive and challenging environment for all undergraduate, graduate, professional school, and professional studies students with disabilities who attend the university. Additionally, the university and AccessibleNU work to provide students with disabilities and other conditions requiring accommodation a learning and community environment that affords them full participation, equal access, and reasonable accommodation. The majority of accommodations, services, and auxiliary aids provided to eligible students are coordinated by AccessibleNU, which is part of the **Dean of Students Office (https://www.northwestern.edu/studentaffairs/dos/)**.

SPS Student Services

The Student Services advising team assists students in a variety of matters during their time at SPS. Each program has an assigned adviser to help students with academic planning (course selection, degree planning, adding/dropping courses), policies, and administrative procedures, and to serve as a guide to resources at SPS and the greater Northwestern community. Student advisers use both proactive advising to keep students on track and intrusive advising to help resolve issues or concerns.

Student advisers can serve as a resource in student issues including, but not limited to, providing guidance to a student in danger of failing class or who may be looking to drop a class, making contact

with a student who is not responsive to messages from faculty, and assisting a student experiencing family or medical issues that are impacting their academics. The advising team can be reached at **spsacademicadvising@northwestern.edu (mailto:spsacademicadvising@northwestern.edu)** , and more information on the academic advisers and available student services can be found on the SPS Student Services page.

Academic Support Services

Northwestern University Library

As one of the leading private research libraries in the United States, Northwestern University Library serves the educational and information needs of its students and faculty as well as scholars around the world. Visit the **Library About (https://www.library.northwestern.edu/about)** page for more information or contact Distance Learning Librarian Tracy Coyne at 312-503-6617 or **tracy-coyne@northwestern.edu (mailto:tracy-coyne@northwestern.edu)** .

Program-Specific Library Guides

- **Information Systems (http://libguides.northwestern.edu/informationsystemsguide)**
- **Data Science (https://libguides.northwestern.edu/predictiveanalyticsguide)**

Additional Library Resources

- **Connectivity: Campus wireless and off-campus access to electronic resources (https://www.library.northwestern.edu/visit/technology/internet-access.html)**
- **Getting available items: Delivery to long-distance patrons (https://www.library.northwestern.edu/find-borrow-request/requests-interlibrary-loan/getting-available-items.html)**
- **Quick access to major newspapers (https://libguides.northwestern.edu/newssources)**
- **Reserve a library study room (https://northwestern.libcal.com/)**
- **Resources for data analysis (https://libguides.northwestern.edu/c.php?g=114906)**
- **Sign up for an in-person or online research consultation appointment (https://www.library.northwestern.edu/research/research-support/research-consultation-appointment.html)**
- **Social science data resources (https://libguides.northwestern.edu/c.php?g=114894)**

Recommendations about Writing

We use the seventeenth edition of *The Chicago Manual of Style* (2019) as our standard reference for citations and English grammar and reference citations.

In his essay "Politics and the English Language," George Orwell (1946) provided six rules of clear writing:

1. Never use a metaphor, simile, or other figure of speech which you are used to seeing in print.
2. Never use a long word where a short one will do.
3. If it is possible to cut a word out, always cut it out.
4. Never use the passive where you can use the active.

5. Never use a foreign phrase, a scientific word, or a jargon word if you can think of an everyday English equivalent.
6. Break any of these rules sooner than say anything outright barbarous.

Zinzer (2012) provides additional writing advice.

Assignment papers are to be structured as formal research reports for academic journals. In addition to an abstract that provides an executive overview of the research, major sections of the research report answer questions as follows:

- Introduction***. Why*** did you engage in this research?
- Literature Review***. Who*** else has conducted research like this?
- Methods. ***How*** did you conduct the research?
- Results. ***What*** did you learn from the research?
- Conclusions***. So, what*** will the research mean to management?

Audio overview (https://northwestern.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=15434363-7b7e-43c7-9ab2-ac0100f4fd4d%C2%A0%C2%A0) regarding structure of papers.

References on Writing

Merriam-Webster's Collegiate Dictionary (eleventh ed.), 2008. Springfield, Mass.: Merriam-Webster.

Orwell, G., 1946, April. ***Politics and the English Languag*** **(http://www.orwell.ru/library/essays/politics/english/e_polit%C2%A0)** e. Horizon.

Rodale, J. I, L. Urdang, and N. LaRoche, 1978. *The Synonym Finder.* Emmaus, Pa.: Rodale Press.

The Chicago Manual of Style **(http://www.chicagomanualofstyle.org/home.html%20%C2%A0)** (seventeenth ed.), 2019. Chicago: University of Chicago Press.

Zinsser, W., 2012. *On Writing Well: An Informal Guide to Writing Nonfiction* (Thirtieth Anniversary Edition). New York: Harper Perennial.

The Writing Place

The Writing Place is Northwestern's center for peer writing consultations. Consultations are free and available to anyone in the Northwestern community: undergraduates, graduate students, faculty, or staff. To book an appointment, go to **The Writing Place (http://www.writing.northwestern.edu/)** website.

The Math Place

The Math Place is a free tutorial service provided to students currently enrolled in Northwestern University's School of Professional Studies courses or in other Northwestern University courses. Students of all levels can benefit from the individual tutoring provided from this service, whether they are taking undergraduate- or graduate-level courses. To book an appointment, go to **The Math Place (https://www.library.northwestern.edu/libraries-collections/schaffner-library/math-place.html)** website.

SPS Learning Studios

Learning studios are available to students who would like additional support in commonly used tools and topics such as statistics, Excel, and coding in Python or R. An instructor is available to answer your questions as you work through self-paced content and exercises. Students can self-enroll for free by visiting the SPS **Academic Services (https://sps.northwestern.edu/student-services/resources.php)** page.

Read&Write Gold

Read&Write Gold is an optional text reading and writing program with numerous beneficial features. Originally developed to assist users with print disabilities, such as visual impairments, dyslexia, ADHD, etc., this program provides a wide array of tools to assist with reading, writing, and note-taking. One of the most useful tools is the text-to-speech function, which students may use to convert digital text into an audio format.

Read&Write Gold is available for free to all Northwestern students, faculty, and staff. Visit the **Northwestern IT site on Read&Write Gold (https://www.it.northwestern.edu/software/readwritegold/index.html)** for more information about the software as well as instructions on how to download it.

Academic Integrity at Northwestern

Students are required to comply with university regulations regarding academic integrity. If you are in doubt about what constitutes academic dishonesty, speak with your instructor or graduate coordinator before the assignment is due and/or examine the university website. Academic dishonesty includes, but is not limited to, cheating on an exam, obtaining an unfair advantage, and plagiarism (e.g., using material from readings without citing or copying another student's paper). Failure to maintain academic integrity will result in a grade sanction, possibly as severe as failing and being required to retake the course, and could lead to a suspension or expulsion from the program. Further penalties may apply. For more information, visit **The Office of the Provost's Academic Integrity page (https://www.northwestern.edu/provost/policies/academic-integrity/)**.

Some assignments in SPS courses may be required to be submitted through Turnitin, a plagiarism detection and education tool. You can find **an explanation of the tool here (https://community.canvaslms.com/docs/DOC-1799)**.

You may also enjoy this video from Snoop Dogg that he created for a professor. **Snoop Dogg says read the syllabus! (https://www.youtube.com/watch?v=aOlvB2YtAhY)**

Course Technology

This course will involve a number of different types of interactions. These interactions will take place primarily through the Canvas system. Please take the time to navigate through the course and become familiar with the course syllabus, structure, and content and review the list of resources below.

Systems Requirements for Distance Learning

Students and faculty enrolled in SPS online classes should have access to a computer with the **Minimum System Requirements (https://sps.northwestern.edu/download/downloads_files.php)** .

Canvas

The **Canvas Student Center (https://canvas.northwestern.edu/courses/5666)** includes information on communicating in Canvas, navigating a Canvas course, grades, additional help, and more.
The **Canvas at Northwestern (https://www.it.northwestern.edu/education/login.html)** website provides information related to getting to know Canvas at Northwestern and getting Canvas support.
The **Canvas Student Guide (https://guides.instructure.com/m/4212)** provides tutorials on all the features of Canvas. For additional Canvas help and support, you can always click the Help icon on the global navigation menu within Canvas to begin a live text chat with Canvas support, contact Canvas Support by phone, or ask the Canvas Chatbot a question.

The **Canvas Accessibility Statement (https://community.canvaslms.com/docs/DOC-2061)** and **Canvas Privacy Policy (https://www.canvaslms.com/policies/privacy)** are also available.

Python Programming and Anaconda

This course uses the Python programming language and Python packages especially useful in natural language processing, including NLTK, Scikit Learn, Gensim, and spaCy. For neural networks and deep learning, we use TensorFlow and Keras. Python can be downloaded at **Anaconda (https://www.anaconda.com/download/)** . Download and install a Python 3.7.x version of Anaconda. Additional documentation on downloading and installing Anaconda can be found at **Anaconda Documentation (https://docs.anaconda.com/anaconda/)** . After installation, you should be able to access the Anaconda Navigator. To work with a Python-aware editor, consider Anaconda Spyder, Visual Studio Code, or Sublime Text. Jupyter notebooks may be utilized for various assignments. Juypter, like Spyder, can be launched via the Anaconda Navigator. Check out the Python learning studio at the SPS **Academic Services (https://www.sps.northwestern.edu/masters/resources.php#studio)** page.

Zoom

We will use Zoom for optional synchronous meetings. The **Zoom Support page (https://support.zoom.us/hc/en-us)** provides additional guidance for using Zoom, and the **Zoom for Students in Canvas (https://canvas.northwestern.edu/courses/5666/pages/zoom-for-students-in-canvas)** page has guidance specifically for students.

The **Zoom Privacy Policy (https://zoom.us/privacy)** and information on **Accessibility Features on Zoom (https://zoom.us/accessibility)** are also available.

Please note that any scheduled synchronous meetings are optional. While your attendance is highly encouraged, it is not required, and you will not be graded on your attendance or participation. These synchronous sessions will be recorded, so you will be able to review the sessions after their initial date.

Panopto

Videos in this course may be hosted in Panopto. If you have not used Panopto in the past, you may be prompted to log in to Panopto for the first time and authorize Panopto to access your Canvas account. You can learn more about using Panopto and log in to Panopto directly by visiting the Panopto guide on the **Northwestern IT Resource Hub (https://digitallearning.northwestern.edu/use-panopto-now)** . Depending on the assignment requirements of this course, you may be asked to create videos using Panopto in addition to viewing content that your instructor has provided through Panopto.

The **Panopto Privacy Policy (https://www.panopto.com/privacy/)** and information on the **Accessibility Features on Panopto (https://support.panopto.com/s/article/Learn-About-Accessibility-Features)** are also available.

Required Technical Skills

Students in an online program should be able to do the following:

- Communicate via email and Canvas discussion forums.
- Use web browsers and navigate the World Wide Web.
- Use the learning management system Canvas.
- Use integrated Canvas tools (e.g., Zoom, Panopto, Course Reserves).
- Use applications to create documents and presentations (e.g., Microsoft Word, PowerPoint).
- Use applications to share files (e.g., Box, Google Drive).

Required Digital Literacy Skills

In order to be successful in an online course, students should be able to locate, evaluate, apply, create, and communicate information using technology.

Students in this online course should be able to do the following:

- Create, name, compose, upload, and attach documents.
- Download, modify, upload, and attach document templates.
- Create, name, design, and upload presentations.
- Access and download Course Reserve readings; read and review PDF documents.
- Access and use a digital textbook.
- Record and upload video taken with a webcam or smartphone.
- Use the library website for scholarly research tasks.
- Search the internet strategically and assess the credibility of internet sources.
- Participate in threaded discussions by contributing text responses, uploading images, and sharing links.
- Coordinate remote work with peers, which may include contacting each other by email, phone, video conference, or shared document.
- Follow directions to engage with a remote proctor by text, webcam, and audio.
- Use a video player to review content, including pausing and restarting video.

Technical Help and Support

The **SPS Help Desk (https://help.sps.northwestern.edu/)** is available for faculty, students, and staff to support their daily IT needs. For additional technical support, contact the **Northwestern IT Support Center (https://www.it.northwestern.edu/supportcenter/)** .

Permissions

Instructional Materials

This course was developed in partnership with distance learning staff in the School of Professional Studies at Northwestern University. Every effort has been made to responsibly acquire instructional materials for this class by adhering to copyright law, obtaining permission from copyright holders, selecting open educational resources (OERs) and Creative Commons (CC) materials, and using citations to credit the work of others.

The same is expected of students in this course. Please review the Academic Integrity statement for more information.

Sharing Course Content

Content within this course—including assignment descriptions, exam questions, and other course components—may not be distributed outside of the course, either to other students or on the internet more broadly.

Student Ownership of Content

Students retain ownership of all content developed while completing this course as dictated by the university **Copyright Policy (https://www.invo.northwestern.edu/invention-disclosure/policies-forms/copyright-policy/)** (“copyright ownership resides with the Creator(s) of copyrightable works”).

Per the Family Educational Rights and Privacy Act (**FERPA (https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html)**), if your instructor wishes to share your work with future students, your permission must be obtained in writing.

Your instructor may limit access to the course after a cutoff date. When you complete the course, please ensure that you have saved all work. You may not be able to return to the course to download your submissions.

Course Summary:

Date	Details	Due
Fri Mar 29, 2024	**2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205379&include_contexts=course_194633)**	7:30pm to 9pm
Mon Apr 1, 2024	**D.0: Introductions**	due by 10:29am

Date	Details	Due
	(https://canvas.northwestern.edu/courses/194633/assignments/1378870)	
	D.1: NLP and AI (10 Points) (https://canvas.northwestern.edu/courses/194633/assignments/1378094)	due by 10:29am
	D.1: Select and Post Your Documents (10 pts) (https://canvas.northwestern.edu/courses/194633/assignments/1378097)	due by 10:29am
Fri Apr 5, 2024	2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205380&include_contexts=course_194633)	7:30pm to 9pm
Mon Apr 8, 2024	D.2: Compare Extraction Engine Results (https://canvas.northwestern.edu/courses/194633/assignments/1378092)	due by 10:29am
Fri Apr 12, 2024	2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205381&include_contexts=course_194633)	7:30pm to 9pm
Mon Apr 15, 2024	A.1: First Vectorized Representation (https://canvas.northwestern.edu/courses/194633/assignments/1378098)	due by 10:29am
	D.3: First Vectorized Representation (https://canvas.northwestern.edu/courses/194633/assignments/1378090)	due by 10:29am
Fri Apr 19, 2024	2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205382&include_contexts=course_194633)	7:30pm to 9pm
Mon Apr 22, 2024	D.4: Implicit Clusters, Semantic Analysis, and Topic Models (https://canvas.northwestern.edu/courses/194633/assignments/1378087)	due by 10:29am
	P.1: Choose Project Topic and Submit a Description (https://canvas.northwestern.edu/courses/194633/assignments/1378101)	due by 10:29am

Date	Details	Due
	P.2: Acquire Data for Project (https://canvas.northwestern.edu/courses/194633/assignments/1378102)	due by 10:29am
Fri Apr 26, 2024	2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205383&include_contexts=course_194633)	7:30pm to 9pm
Mon Apr 29, 2024	D.5: Classification (https://canvas.northwestern.edu/courses/194633/assignments/1378088)	due by 10:29am
	A.2: Clustering and Classification (https://canvas.northwestern.edu/courses/194633/assignments/1378099)	due by 10:29am
Fri May 3, 2024	2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205384&include_contexts=course_194633)	7:30pm to 9pm
Mon May 6, 2024	D.6: Information Extraction (https://canvas.northwestern.edu/courses/194633/assignments/1378096)	due by 10:29am
	P.3: Project Proposal (https://canvas.northwestern.edu/courses/194633/assignments/1378103)	due by 10:29am
Fri May 10, 2024	2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205385&include_contexts=course_194633)	7:30pm to 9pm
Mon May 13, 2024	D.7: Natural Language Understanding (https://canvas.northwestern.edu/courses/194633/assignments/1378095)	due by 10:29am
	A.3: Using LLMs for Entity Extraction (https://canvas.northwestern.edu/courses/194633/assignments/1378100)	due by 10:29am
Fri May 17, 2024	2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205386&include_contexts=course_194633)	7:30pm to 9pm

Date	Details	Due
Mon May 20, 2024	D.8: Chatbots, conversational AI, and multi-modal NLP (https://canvas.northwestern.edu/courses/194633/assignments/1378093)	due by 10:29am
Fri May 24, 2024	2024SP_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1205387&include_contexts=course_194633)	7:30pm to 9pm
Mon May 27, 2024	D.9: Application of Deep Learning in NLP (https://canvas.northwestern.edu/courses/194633/assignments/1378091)	due by 10:29am
Mon Jun 3, 2024	D.10: Final Presentation of Your Work (https://canvas.northwestern.edu/courses/194633/assignments/1378089)	due by 10:29am
	P.4: Final Report (https://canvas.northwestern.edu/courses/194633/assignments/1378104)	due by 10:29am
	2023_MS_DSP453_NLP_Module 1 TA Session by Nitin (https://canvas.northwestern.edu/calendar?event_id=1203076&include_contexts=course_194633)	
	2023_MS_DSP453_NLP_Module 3 TA Session by Nitin (https://canvas.northwestern.edu/calendar?event_id=1203077&include_contexts=course_194633)	
	2023_MS_DSP453_NLP_Module 5 TA Session by Nitin (https://canvas.northwestern.edu/calendar?event_id=1203078&include_contexts=course_194633)	
	2023_MS_DSP453_NLP_Module 7 TA Session by Nitin (https://canvas.northwestern.edu/calendar?event_id=1203079&include_contexts=course_194633)	
	2023_MS_DSP453_NLP_Module 9 TA Session by Nitin	

Date	Details	Due
	(https://canvas.northwestern.edu/calendar?event_id=1203080&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1203075&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1203081&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1203082&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1203083&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1203084&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1203085&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (https://canvas.northwestern.edu/calendar?event_id=1203086&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing	

Date	Details	Due
	(https://canvas.northwestern.edu/calendar?event_id=1203087&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (TA Session- Rakesh) (https://canvas.northwestern.edu/calendar?event_id=1203088&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (TA Session- Rakesh) (https://canvas.northwestern.edu/calendar?event_id=1203089&include_contexts=course_194633)	
	2023FA_MS_DSP_453-DL_SEC61 Natural Language Processing (TA Session- Rakesh) (https://canvas.northwestern.edu/calendar?event_id=1203090&include_contexts=course_194633)	
	Discussion on A3 requirements (https://canvas.northwestern.edu/calendar?event_id=1203092&include_contexts=course_194633)	
	Encoder-Decoder system - Language translation model (https://canvas.northwestern.edu/calendar?event_id=1203091&include_contexts=course_194633)	
	HandsOn:- Encoder-Decoder system - Language translation model (https://canvas.northwestern.edu/calendar?event_id=1203093&include_contexts=course_194633)	
	Module 8 Sync Session: Chatbots, conversational AI, and multi-modal NLP (https://canvas.northwestern.edu/calendar?event_id=1203094&include_contexts=course_194633)	