CSCI 5832: Natural Language Processing, Fall 2022

Course Information

- **Staff**
  - Instructor: Prof. James H. Martin
  - TAs: Rehan Ahmed and Adam Wiemerslage
  - CM: Satviki Pathak
- **Time and Location**
  - Tuesday-Thursday 12:30-1:45 in Atlas 100 (and on Zoom)
- **Office hours**
  - TBD

Natural Language Processing

The field of Natural Language Processing (NLP) is aimed at getting computers to perform useful and interesting tasks involving spoken and written human language. NLP is sometimes referred to as *Computational Linguistics* to emphasize the fact that it involves the combination of Computer Science with insights from Linguistics (the scientific study of human language). Practical applications of NLP include question answering, machine translation, information extraction, and interactive dialog systems (both written and spoken). Modern NLP systems rely heavily on machine learning methods based on probability, linear algebra and calculus.

Course Topics (Subject to Change)

- Words, word counting, and lexicons
- Probabilistic language modeling
- HMMs and part of speech tagging
- Text classification with language models
  - Naive Bayes
- Text classification with single layer neural networks
  - Logistic Regression
- HMMs and CRFs for part of speech tagging and named entity recognition
- Vector semantics
- Neural Networks
- Recurrent Neural Networks
- Causal and Bidirectional language models (GPT and BERT)
- Machine Translation
- Dialog Systems
- Dependency parsing

- Discourse and co-reference

Course Details

Readings: We'll be using draft chapters from the 3rd Edition of *Speech and Language Processing* by Dan Jurafsky and James H. Martin. You **should not buy** the current edition; draft pdfs of the chapters from the new edition are available from the textbook website.

Canvas: We'll be using Canvas for our learning platform. Readings, assignments, and general course information will be posted there.   Homework submissions and exams will be done through Canvas as well.

Piazza: We'll be using Piazza for class discussion and communication. The quickest way to get answers to simple questions is to post it on the Piazza forum. Instead of mailing me or the staff your questions, please post your questions on Piazza. If your question is of a private nature Piazza allows you to send private messages to the instructors. The signup link for our class forum is on the Canvas page.

Assignments: We will have both written and programming assignments. These will be available on and submitted through Canvas. Programming assignments will be done in Python 3.

Shared Task: The final assignment will be done in groups and will involve the use of data and evaluation methods from a standard NLP task to be determined later in the course.

Quizzes: There will be 3 quizzes during the semester. These will be administered on Canvas outside of class. They will consist of true/false, multiple choice, and short computational questions. Dates are on the class calendar on Canvas.

Grading: Grades will be based on a cumulative score as follows:
- Assignments (40%)
- Quizzes (40%; equally weighted)
- Shared task (15%)
- Engagement (5%)

The quizzes will assess your understanding of the material covered in class and in the readings. Not everything in the readings will be covered in class, so don't rely on lectures alone to prepare for the quizzes.

Engagement points will be based on attendance (informally assessed), useful participation on Piazza, participation in class, and anything else that indicates an active engagement with the class.

Collaboration: The collaboration policy is straightforward:

- You may discuss homework assignments with anyone. You are encouraged to discuss solutions with your classmates. However, unless specified otherwise the writing assignments and code that you turn in should be written entirely on your own. For written assignments that involve research and analysis, cite all your sources.
- It's fine to use StackExchange and the like for programming techniques and the like. But don't crawl around Github, StackExchange, Reddit, Chegg, etc. trying to find specific solutions to the programming assignments.
- When in doubt, ask: If you have doubts about this policy or would like to discuss specific cases, please ask the instructor.

Distance Sections: There are online sections for the course for remote students. Lectures will be streamed and recorded and a Zoom link will be provided for synchronous participation in the class. Recordings will be linked on Canvas as they become available. The synchronous zoom sessions and the recordings are available to everyone to use. **Local students in the 001 section are not required to come to campus**. Participating via Zoom (synchronous or recorded) is acceptable for classroom attendance. **However, due to the classroom capacity, students in the online sections should refrain from coming to class early in the semester.** We will revisit this after the first few weeks of class as enrollment and in-class attendance stabilizes (diminishes). If it looks like there's plenty of room I'll let you know.

Zoom info

- https://cuboulder.zoom.us/j/98349387895
- Meeting ID: 983 4938 7895
- You need to be logged in to Zoom through your CU identikey to access the class.

REQUIRED CAMPUS SYLLABUS STATEMENTS

Classroom Behavior

Both students and faculty are responsible for maintaining an appropriate learning environment in all instructional settings, whether in person, remote or online. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. For more information, see the classroom behavior policy, the Student Code of Conduct, and the Office of Institutional Equity and Compliance.

Requirements for COVID-19

As a matter of public health and safety, all members of the CU Boulder community and all visitors to campus must follow university, department and building requirements and all public health orders in place to reduce the risk of spreading infectious disease. CU Boulder currently requires COVID-19 vaccination and boosters for all faculty, staff and students. Students, faculty and staff must upload proof of vaccination and boosters or file for an exemption based on medical, ethical or moral grounds through the MyCUHealth portal.

The CU Boulder campus is currently mask-optional. However, if public health conditions change and masks are again required in classrooms, students who fail to adhere to masking requirements will be asked to leave class, and students who do not leave class when asked or who refuse to comply with these requirements will be referred to Student Conduct and Conflict Resolution. For more information, see the policy on classroom behavior and the Student Code of Conduct. If you require accommodation because a disability prevents you from fulfilling these safety measures, please follow the steps in the "Accommodation for Disabilities" statement on this syllabus.

If you feel ill and think you might have COVID-19, if you have tested positive for COVID-19, or if you are unvaccinated or partially vaccinated and have been in close contact with someone who has COVID-19, you should stay home and follow the further guidance of the Public Health Office (contacttracing@colorado.edu). If you are fully vaccinated and have been in close contact with someone who has COVID-19, you do not need to stay home; rather, you should self-monitor for symptoms and follow the further guidance of the Public Health Office (contacttracing@colorado.edu). {Faculty: insert your procedure here for students to alert you about absence due to illness or quarantine. Because of FERPA student privacy laws, do not

require students to state the nature of their illness when alerting you. Do not require "doctor's notes" for classes missed due to illness; campus health services no longer provide "doctor's notes" or appointment verifications.}

Accommodation for Disabilities

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the [Disability Services website](). Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition, see [Temporary Medical Conditions]() on the Disability Services website.

Preferred Student Names and Pronouns

CU Boulder recognizes that students' legal information doesn't always align with how they identify. Students may update their preferred names and pronouns via the student portal; those preferred names and pronouns are listed on instructors' class rosters. In the absence of such updates, the name that appears on the class roster is the student's legal name.

Honor Code

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the [Honor Code](). Violations of the Honor Code may include, but are not limited to: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to Student Conduct & Conflict Resolution (honor@colorado.edu); 303-492-5550). Students found responsible for violating the [Honor Code]() will be assigned resolution outcomes from the Student Conduct & Conflict Resolution as well as be subject to academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found on the [Honor Code website]().

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation

CU Boulder is committed to fostering an inclusive and welcoming learning, working, and living environment. University policy prohibits sexual misconduct (harassment, exploitation, and assault), intimate partner violence (dating or domestic violence), stalking, protected-class discrimination and harassment, and related retaliation by or against members of our community

on- and off-campus. These behaviors harm individuals and our community. The Office of Institutional Equity and Compliance (OIEC) addresses these policies, and individuals who believe they have been subjected to misconduct can contact OIEC at 303-492-2127 or email cureport@colorado.edu. Information about university policies, reporting options, and support resources can be found on the OIEC website.

Please know that faculty and graduate instructors have a responsibility to inform OIEC when they are made aware of any issues related to these policies regardless of when or where they occurred to ensure that individuals impacted receive information about their rights, support resources, and resolution options. To learn more about reporting and support options for a variety of concerns, visit Don't Ignore It.

Religious Holidays

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, please inform us within the first two weeks of the semester of any religious observances that will conflict with required work in the class. We will work with you to make appropriate accommodations for each situation.

See the campus policy regarding religious observances for full details.