University of Oklahoma
School of Computer Science/Graduate Program
C S 5293-001: Natural Language Processing (Text Analytics)
Spring 2025 (3 credits)

Instructor: Jie Cao <jie.cao@ou.edu>

Meetings:
Class: In-person, Tuesday/Thursday 1:30PM-2:45PM, Sarkeys Energy Ctr A0236
Office Hours: Hybrid, TBA
Final Exam: May 5, Monday, 1:30 PM -3:30 PM, Sarkeys Energy Ctr A0236

Course Prerequisite: C S 2413 or C S 2414 or C S 5005; C S 2813 or MATH 2513; and a statistics course from the departmentally approved list.

Course Description:
The course covers the design and implementation of algorithms to extract useful knowledge from text to support decision making, including methods to obtain data sets, understanding data formats, detecting duplicates, cleaning data sets, tagging, indexing, search, evaluating algorithms, classification, clustering, topic modeling, and entity resolution.

+ statistical models over linguistic structures (e.g., sequences, trees, and graphs), **vector semantics, neural networks, large language models,** etc.

Textbook

- Speech and Language Processing, Dan Jurafsky and James H. Martin, 3rd Edition, 2024. Available Online at https://web.stanford.edu/~jurafsky/slp3/ (Required for reading material)

Learning Outcomes:
You will be able to …

- To understand the problems, algorithms, methods of natural language processing, and their relation to both linguistics and statistics.
  - Understand the basic linguistic phenomena and fundamental NLP tasks
  - Use linguistic feature engineering and manipulate probability formula for statistic NLP models
  - Understand the neural representation, deep learning for NLP
- Be able to design, implement, and evaluate practical NLP systems
  - Analyze text via N-gram language models, and linguistic features.
  - Train and evaluate text classification models and improve them with neural networks.
  - Design and implement basic structured prediction methods for sequence, tree, and graph.
  - Improve model performance with neural modes and pretrained language models.
  - Control and evaluate large language models.
- Be able to understand research papers in NLP, contribute initial findings via collaboration.
  - Author a report to survey a research area for your groups' research goal.
  - Design initial experiments to compare existing methods and explore new ideas.

Course Expectations & Grading: Students are required to take two in-class exams: one midterm and one final. There will be no makeup exams except in cases of emergencies. There will be a set of assignments that each student should individually complete, as well as a group project per two-students. The course letter grade will be assigned based on the overall percentage: >= 90 (A), >=80 and < 90 (B), >=70 and < 80 (C), >=60 and < 70 (D), and <60 (F). The allocation of percentages is given below:

	Percentages
Midterm Exam	15%
Final Exam	25%
Assignments	40%
Course Project	20%
Class Attendance	+5%
Participation	+3%

We've found that participation and attendance have a substantial effect on your learning outcomes. Both the scores of attendances and participation are bonus instead of any penalty. We will randomly choose three classes to check attendance, if you have checked in all the three selected classes, you will get an extra 5-point class attendance score, otherwise, you will get 0 bonus score. The score of participation (3%) is to encourage collaboration and sharing your ideas in class, online forum, and office hour, which can be assigned in a range from [0,3]. So, the highest score is 108, not 100.

Course Policies

Attendance: You are expected to attend all the class lectures. In cases of sickness, we want to work with you to identify alternative arrangements, but you must alert us before the class period you miss. As explained above, you will get 5% bonus score if you pass the 3 attendances check for randomly selected 3 classes. Your participation score will be considered based on your engagement in class, online forums, and office hour.

Canvas: We'll be using Canvas for our learning platform. Readings, assignments, and general course information will be posted there. Assignment submissions and exams will be done through Canvas as well. We'll use Piazza (an integrated module in Canvas) for online class discussion and communication. The quickest way to get answers to simple questions is to post them on the Piazza forum.

Private Emails: Only matters of personal interest should be directed to email instead of any online forum, e.g., informing the instructors of an extended personal illness.

Readings: The course schedule lists a set of readings for each lecture day, mainly from the 3rd Edition of Speech and Language Processing by Dan Jurafsky and James H. Martin. You are responsible for this material before class begins. You don't need to buy the book because the PDF chapters are legally free, and you can download the latest version from the textbook website.

Assignments: We will have both written and programming assignments. These will be made available through and submitted through Canvas. Programming assignments will be done in Python, along with associated packages and libraries.
Shared Task or Project Proposal: The final project will be done in groups. It could be selected from a list of given shared task, or you could propose your own research project. **No double dipping projects across multiple classes**. In any case, the data and evaluation methods should be **accessible and ready to use.**

Lateness Policy: Assignments are typically due at 12:00pm(noon), an hour before the start of class. It's important to get assignment done on time so that you can follow the subsequent lectures. We will use the timestamp on Canvas as the submission time. **Assignments will be accepted up to 24 hours after the deadline** but will be assessed a 10% penalty. That is, if your assignment is late and scores 90, then your actual grade will be 81 = 90 - 9. **Assignments will *not* be accepted 24 hours after the deadline.** We will be strict about this policy: A submit time of 12:01 PM, you will face a 10% penalty! An assignment with no submission within 24 hours after the deadline will be scored 0. This may sound harsh, but we must draw a line somewhere. **Exceptions**: All submissions are subject to the late policy stated here. We understand, however, that certain factors may occasionally interfere with your ability to hand in work on time. If that factor is an extenuating circumstance such as a medical condition, we ask you to provide documentation directly issued by the University, and we will try to work out an agreeable solution with you.

Exams: There will be two exams during the semester. These will be administered in class. They contain true/false, multiple choice, and short computational questions. Dates are on Canvas's class calendar. The quizzes will assess your understanding of the material covered in class and the readings. Not everything in the readings will be covered in the classroom lecture, so don't rely on lectures alone to prepare for the quizzes. You could organize a double-sided one-page notes (US Letter Size) for the exams, no other devices or materials are allowed.

Incompletes: The grade of "I" is intended for the rare circumstance when a student who has been successful in a class has an unexpected event shortly before the class's end. We will not consider giving a student a grade of "I" unless the following three conditions have been met:

1. It is within two weeks of the end of the semester.
2. The student has a grade of C or better in the class.
3. The reason that the student cannot complete the class is properly documented and compelling.

Classroom Conduct: Because cell phones and laptops can distract substantially from the classroom experience, students are asked not to use either during class except when required as part of a classroom exercise. Disruptions of the class will also not be permitted. In the case of disruptive behavior, we may ask that you leave the classroom and charge you with violating the Student Code of Responsibilities and Conduct. Examples of disruptive behavior include:

- Allowing a cell phone or pager to repeatedly beep audibly.
- Playing music or computer games during class in such a way that they are visible or audible to other class members.
- Exhibiting erratic or irrational behavior.
- Behavior that distracts the class from the subject matter or discussion.

- Making physical or verbal threats to a faculty member, teaching assistant, or class member.
- Refusal to comply with faculty or teaching assistant direction.

Proper Academic Conduct: The overall goal of this course is your learning. In order to demonstrate that you have reached this goal, the work you turn in needs to be your own. This includes putting written work into your own words and citing your sources, as appropriate to avoid plagiarism. If you work in a group, seek assistance from a tutor, use a resource on campus, and/or use online resources (including AI software), the work you turn in must be your own, demonstrating your own understanding of the material that you have gained through the learning process. For example,

1. You may not look at or share any solution with others.
2. If you discuss a solution with anyone, you must document their names in your assignment.
3. You **must document in your solution** if you use external resources (e.g. Canvas, StackOverflow.com, LLMs, or any other AI tools).

If you have questions about academic integrity or plagiarism, please ask: my aim is to foster an environment where you can learn and grow, while also maintaining academic honesty and a clear representation of your learning and ideas. Penalties for serious offenses include a zero on the assignment and egregious offenses can even result in expulsion from the university, so it is important to understand expectations.
Visit the OU Integrity Office for more information on what constitutes plagiarism.

Make sure that your computer account is properly protected. Use an appropriate password, and do not give your friends access to your account or computer system. Do not leave printouts, computers, or thumb drives around a laboratory where others might access them.

Programming projects will be checked by software designed to detect collaboration. This software is extremely effective and has withstood repeated reviews by the campus judicial processes.

Upon the first documented occurrence of inappropriate collaborative work or of taking a solution from a network resource, the instructors will report the academic misconduct to the Campus Judicial Coordinator. The procedure is documented in the University of Oklahoma Academic Misconduct Code (http://integrity.ou.edu). The provider and receiver of a solution will be treated equally in the misconduct process.

AI Policy: Throughout this course, you will develop the skills needed to effectively use generative AI as an aid in your learning and in preparation for our changing field. Follow assignment instructions carefully, as they will guide you in what you are permitted to use Generative AI for in each assignment. Where Generative AI is used, you must appropriately acknowledge its use (including using quotation marks for direct quotes) and provide a statement describing how and why it was used, and how you verified the accuracy of its output. Usage of Generative AI outside of the scope of what is explicitly defined in our assignments, and without acknowledgment, will be considered a violation of the academic integrity policy for this course.

University Policies

Mental Health Support Services:
Support is available for any student experiencing mental health issues that are impacting their academic success. Students can either been seen at the University Counseling Center (UCC) located on the second floor of Goddard Health Center or receive 24/7/365 crisis support from a licensed mental health provider through TELUS Health. To schedule an appointment or receive more information about mental health resources at OU please call the UCC at 405-325-2911 or visit University Counseling Center. The UCC is located at 620 Elm Ave., Room 201, Norman, OK 73019.

Title IX Resources and Reporting Requirement
The University of Oklahoma faculty are committed to creating a safe learning environment for all members of our community, free from gender and sex-based discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking, in accordance with Title IX. There are resources available to those impacted, including: speaking with someone confidentially about your options, medical attention, counseling, reporting, academic support, and safety plans. If you have (or someone you know has) experienced any form of sex or gender-based discrimination or violence and wish to speak with someone confidentially, please contact OU Advocates (available 24/7 at 405-615-0013) or University Counseling Center (M-F 8 a.m. to 5 p.m. at 405-325-2911). Because the University of Oklahoma is committed to the safety of you and other students, and because of our Title IX obligations, I, as well as other faculty, Graduate Assistants, and Teaching Assistants, are mandatory reporters. This means that we are obligated to report gender-based violence that has been disclosed to us to the Institutional Equity Office. This means that we are obligated to report gender-based violence that has been disclosed to us to the Institutional Equity Office. This includes disclosures that occur in: class discussion, writing assignments, discussion boards, emails and during Student/Office Hours. You may also choose to report directly to the Institutional Equity Office. After a report is filed, the Title IX Coordinator will reach out to provide resources, support, and information and the reported information will remain private. For more information regarding the University's Title IX Grievance procedures, reporting, or support measures, please visit Institutional Equity Office at 405-325-3546.

Reasonable Accommodation Policy
The University of Oklahoma (OU) is committed to the goal of achieving equal educational opportunity and full educational participation for students with disabilities. If you have already established reasonable accommodations with the Accessibility and Disability Resource Center (ADRC), please submit your semester accommodation request through the ADRC as soon as possible and contact me privately, so that we have adequate time to arrange your approved academic accommodations.
If you have not yet established services through ADRC, but have a documented disability and require accommodations, please complete ADRC's pre-registration form to begin the registration process. ADRC facilitates the interactive process that establishes reasonable accommodations for students at OU. For more information on ADRC registration procedures, please review their Register with the ADRC web page. You may also contact them at (405)325-3852 or adrc@ou.edu, or visit www.ou.edu/adrc for more information.
Note: disabilities may include, but are not limited to, mental health, chronic health, physical, vision, hearing, learning and attention disabilities, pregnancy-related. ADRC can also support students experiencing temporary medical conditions.

Religious Observance

It is the policy of the University to excuse the absences of students that result from religious observances and to reschedule examinations and additional required classwork that may fall on religious holidays, without penalty.

Persons with Disability or Special Accommodation or Accommodation for any reason:

The Accessibility and Disability Resource Center is committed to supporting students with disabilities to ensure that they can enjoy equal access to all components of their education. This includes your academics, housing, and community events. If you are experiencing a disability, a mental/medical health condition that has a significant impact on one or more life functions, you can receive accommodations to provide equal access. Possible disabilities include, but are not limited to, learning disabilities, AD(H)D, mental health, and chronic health. Additionally, we support students with temporary medical conditions (broken wrist, shoulder surgery, etc.) and pregnancy. To discuss potential accommodations, please contact the ADRC at 730 College Avenue, (ph.) 405.325.3852, or adrc@ou.edu.

Adjustments for Pregnancy/Childbirth Related Issues

Should you need modifications or adjustments to your course requirements because of documented pregnancy-related or childbirth-related issues, please contact the Accessibility and Disability Resource Center at 405/325-3852 and/or the Institutional Equity Office at 405/325-3546 as soon as possible. Also, see the Institutional Equity Office FAQ on Pregnant and Parenting Students' Rights for answers to commonly asked questions.

Final Exam Preparation Period

Pre-finals week will be defined as the seven calendar days before the first day of finals. Faculty may cover new course material throughout this week. For specific provisions of the policy please refer to OU's Final Exam Preparation Period policy.

Emergency Protocol

During an emergency, there are official university procedures that will maximize your safety.
Severe Weather: If you receive an OU Alert to seek refuge or hear a tornado siren that signals severe weather.

1. Look for severe weather refuge location maps located inside most OU buildings near the entrances.
2. Seek refuge inside a building. Do not leave one building to seek shelter in another building that you deem safer. If outside, get into the nearest building.
3. Go to the building's severe weather refuge location. If you do not know where that is, go to the lowest level possible and seek refuge in an innermost room. Avoid outside doors and windows.
4. Get in, Get Down, Cover Up
5. Wait for official notice to resume normal activities.

Additional Weather Safety Information is available through the Department of Campus Safety.

The University of Oklahoma Active Threat Guidance

The University of Oklahoma embraces a Run, Hide, Fight strategy for active threats on campus. This strategy is well known, widely accepted, and proven to save lives. To receive emergency campus

alerts, be sure to update your contact information and preferences in the account settings section at one.ou.edu.
RUN: Running away from the threat is usually the best option. If it is safe to run, run as far away from the threat as possible. Call 911 when you are in a safe location and let them know from which OU campus you're calling from and location of active threat.
HIDE: If running is not practical, the next best option is to hide. Lock and barricade all doors; turn of all lights; turn down your phone's volume; search for improvised weapons; hide behind solid objects and walls; and hide yourself completely and stay quiet. Remain in place until law enforcement arrives. Be patient and remain hidden.
FIGHT: If you are unable to run or hide, the last best option is to fight. Have one or more improvised weapons with you and be prepared to attack. Attack them when they are least expecting it and hit them where it hurts most: the face (specifically eyes, nose, and ears), the throat, the diaphragm (solar plexus), and the groin.
Please save OUPD's contact information in your phone.
NORMAN campus: *For non-emergencies call (405) 325-1717. For emergencies call (405) 325-1911 or dial 911.*
TULSA campus: *For non-emergencies call (918) 660-3900. For emergencies call (918) 660-3333 or dial 911.*

Fire Alarm/General Emergency: (required)
If you receive an OU Alert that there is danger inside or near the building, or the fire alarm inside the building activates:

1. *LEAVE* the building. Do not use the elevators.
2. *KNOW* at least two building exits
3. *ASSIST* those that may need help
4. *PROCEED* to the emergency assembly area
5. *ONCE safely outside, NOTIFY first responders of anyone that may still be inside building due to mobility issues.*
6. *WAIT* for official notice before attempting to re-enter the building.

OU Fire Safety on Campus

Tentative Topics (Spring 2025)
https://www.ou.edu/registrar/academic-records/academic-calendars/spring-2025

- Course Introduction/Tokens, Words, Vocabularies, Lexicons
- Language Modeling with N-Grams
- Text Classification, Feature Engineering, Logistic Regression
- Sequence Labelling for Part of Speech and Named Entities
- Unsupervised Learning (Tagging with HMM), Topic Models, Clustering
- Symbolic Meaning Representation, Syntax, Semantics Structures (e.g., Trees and Graphs)
- Neural Representation Learning, Vector Semantics, Embeddings
- Recurrent Neural Networks and LSTM
- Attention, Transformers, State Space Models etc.
- Mask Language Model and Transfer Learning, Multi-task Learning
- Large Language Models and In-context Learning/Prompting
- RAG/Prompt Optimization/Instruction Finetuning
- Information Retrieval (Dataset Cleaning, Indexing, Search),

- Question Answering, Dialogue System, and Other NLP Applications