ISE 540 Text Analysis
Section: 31720 D
Fall 2024 Mon, Wed 10:00am-11:50am; THH 119

Instructor: Dr. Maryam Pishgar
Email: pishgar@usc.edu
Phone number: 949-656-0063
Mon & Wed 10:00-11:50 AM or by appointment - Professor Pishgar's office - GER 216

Teaching Assistant: Hexin Li
Email: hexinli@usc.edu
Office hours: Tue 10:00 AM - 11:00 AM
(By appointment)

Teaching Assistant: Rachel Monger
Email: rmonger@usc.edu
Office hours: Mon 1:00 PM - 2:00 PM
https://usc.zoom.us/j/92850391313

Catalog Course Description

Foundations, techniques, applications, and algorithms for conducting predictive analytics on problems that involve significant text data, including webpages, social media, and 'natural language' documents. Topics include applied natural language processing, large language models (LLMs), information retrieval, and the semantic web.

Expanded Course Description

Students will learn the practical aspects of the techniques needed to build predictive analytical systems over text data. Today, many of these systems are applications of machine learning, including supervised and unsupervised learning, as well as advanced approaches utilizing large language models (LLMs). Topics include information retrieval (including search and indexing), natural language processing (including information extraction and entity linking), knowledge discovery, and the application of LLMs for text generation and understanding. The class will be run as a fast-paced lecture course with lots of student participation and significant hands-on experience. The class will occasionally feature guest lecturers with advanced knowledge in some of the covered topical areas, including LLMs.

Learning Objectives and Outcomes

The learning objectives for this course are:

• Understand the fundamentals and limitations of building predictive analytics systems for real-world problems involving text data;

• Understand the different aspects of text data (including structured and unstructured data, proprietary and public data, and social media data) from the lens of Big Data (4 Vs of volume, veracity, velocity and variety);

• Understand the different components in a predictive analytics ecosystem, including differences in input data (e.g., website vs. social media), evaluation metrics, cloud and infrastructure, and algorithmic tradeoffs;

• Gain an appreciation of both theory and practice in doing predictive analytics on text data, and apply course techniques to an actual project designed in a team setting;

• Understand how to structure a text analytics problem, and reason about the validity, utility and tradeoffs of competing solutions in real-world settings

Prerequisite(s): An undergraduate-level course on statistics is a minimum prerequisite, since we will be regularly relying on statistical methods like significance testing, normal distributions etc.
Recommended Preparation: Knowledge of a programming language such as Python is desirable, some background in predictive analytics and AI . An Engineering Data Analytics course like ISE 529 is highly recommended but not required. Unless an exception is sought with good reason, we will use Python as the programming language for assignments.

Course Notes

The course will be run as a lecture class with student participation strongly encouraged. The first 2-3 weeks of the course are structured as a quickstart to provide a primer on fundamentals, followed by deeper presentations and more technical material for the remainder of the course. Note that this is not an engineering data analytics course: we will not be going into depth into the theory and math of machine learning or statistics. Students will be expected to review relevant aspects of such material (I will post regular and accessible pointers) before coming to class. All of the course materials, including lecture slides and homeworks will be posted online on brightspace. The class projects are a significant aspect of this course.

Technological Proficiency and Hardware/Software Required

All assignments and lectures will assume electronic access to brightspace. Programming assignments will be in Python, which is freely available.

Required Readings and Supplementary Materials

There is no required textbook. I will be posting all relevant material online on brightspace.

Attendance and Participation Policy

5% of the overall grade would constitute for class attendance and participation.

Description and Assessment of Assignments

Approximately 7 assignments will be given to students.

Midterm and Final Exam

Final Project

Grading Breakdown

Assignment	% of Grade
Assignment	30
Midterm Exam	20
Final Exam	20
Final Project	2
Participation	5

Grading Scale

Course final grades will be determined using the following scale

A 95-100
A- 90-94
B+ 87-89
B 83-86
B- 80-82
C+ 77-79
C 73-76
C- 70-72
D+ 67-69
D 63-66
D- 60-62
F 59 and below

Assignment Submission Policy

Homework assignments are due at 11:59pm on the due date and should be submitted in Blackboard. You can submit homework up to one week late, but you will lose 25% of the possible points for the assignment. After one week, the assignment cannot be submitted.

Additional Policies

It is my expectation that students make every effort to attend every class, and quizzes will be designed to enforce this policy. There will also be a strict no-cellphone policy. Since the class is virtual this fall, additional course guidelines are noted on the next page. Readings for each class are posted below as links. Students must do these readings before coming to class. These readings are particularly important as you navigate your career in today's competitive economy, and are generally from industrial sources that will

help you be informed on subject matter. Occasionally, quizzes will be given at the beginning of class and may involve the readings for that class day as test material.

Course Schedule: Breakdown

- **Python Text Basics**
- **NLP Python Basics**
- **Parts of Speech Tagging**
- **Text Classification**
- **Semantics and Sentiment Analysis**
- **Topic Modeling**
- **Maybe ChatBots and Advanced Material**

SCHEDULE

Week of	**Topic**
August 26	Syllabus day
August 28	Intro to the Text Analysis, python basics (Rachel and Hexin)
September 2	**Labor Day Holiday**
September 4	Text Analysis Basics

September 9	Introduction to NLP
September 11	NLP Basics
September 16	Text Classification
September 18	Deep Learning for NLP basics
September 23	RNN, LSTM modeling basics
September 25	Sentiment and Semantic Analysis
September 30	Review for Midterm/ class activity
October 2	**Midterm Exam**
October 7	Topic Modeling, Parts of Speech Tagging
October 9	LLM, Chat GPT, Encoder decoder (Maryam, Hexin, Rachel)
October 14	LLM, BERT (TBD)(Maryam, Hexin, Rachel)
October 16	Introduction on using supercomputer

October 21	Web scraping (Maryam, Hexin, Rachel)
October 23	Intro to Final Project(Maryam, Hexin, Rachel)
October 28	Word Embedding (Maryam, Hexin, Rachel)
October 30	Interview Questions
November 4	**Prelim Presentation**
November 6	**Prelim Presentation**
November 11	**Veterans Day Holiday**
November 13	Final Exam Review/ class activity
November 18	**Final Exam**
November 20	Building Chatbot (Maryam, Hexin, Rachel)
November 25	Consult about the final presentation
November 27	**Thanksgiving Holiday**

December 2	**Final Project Presentation**
December 4	**Final Project Presentation**

Statement on Academic Conduct and Support

Systems Academic Conduct:

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion ofplagiarism in SCampus in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific- misconduct.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the Office of Equity and Diversity http://equity.usc.edu or to the Department of Public Safety http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. The Center for Women and Men http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage http://sarc.usc.edu describes reporting options and other resources.

Support Systems:

Student Health Counseling Services - (213) 740-7711 – 24/7 on call engemannshc.usc.edu/counselingFree and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call suicidepreventionlifeline.org Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, seven days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 – 24/7 on call
engemannshc.usc.edu/rsvp Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office ofEquity and Diversity (OED) | Title IX - (213) 740-5086
equity.usc.edu, titleix.usc.edu
Information about how to get help or help a survivor of harassment or discrimination, rights ofprotected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following protected characteristics: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations.

Bias Assessment Response and Support - (213) 740-2421
studentaffairs.usc.edu/bias-assessment-response-support Avenue to report incidents of bias, hate crimes, and microaggressions for appropriate investigation and response.

The Office ofDisability Services and Programs - (213) 740-0776
Dsp.usc.edu Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test-taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Support andAdvocacy - (213) 821-4710
studentaffairs.usc.edu/ssa Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101
diversity.usc.edu Information on events, programs, and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various student resources.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
dps.usc.edu, emergency.usc.edu
Emergency assistance and avenues to report a crime. Latest updates regarding safety, including how instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department ofPublic Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call
Dps.usc.edu
Non-emergency assistance or information.