ADAN/ADEC 7431 Natural Language Processing (NLP))
Woods College of Advancing Studies, Boston College
Spring 2025, 13 Jan-13 May

Instructor Name: Dr. Larry Fulton
E-mail: lawrence.fulton@bc.edu
Phone: 850-844-3178
Best contact method: email
Office Hours: Sat & Sun morning, text 850-844-3178 first
Meeting Room: https://bccte.zoom.us/my/doclarry2 and library
Class Location: Online weekly and O'Neill Library 257 on 1 February (2-6 PM), 22 March (2-6 PM), 13 May (2-6 PM)
Online Class Meeting Time(s): Wednesdays, 7-9 PM

Course Description

This course introduces students to natural language processing (NLP) concepts and methods. Students will learn how to conduct both supervised and unsupervised NLP. The course covers 1) the basics of NLP, 2) text (document) classification, 3) text summarization, 4) text similarity & clustering, 5) semantic analysis, 6) sentiment analysis, and 7) deep learning approaches (Recurrent Neural Networks and transformer-based architecture.

Course Format: online

Textbooks (with ISBN) & Readings (Required but all Freely Available through BC Library)

Dong, Z. S., Meng, L., Christenson, L., & Fulton, L. (2021). Social media information sharing for natural disaster response. Natural hazards, 107, 2077-2104.

Ozdemir, S. (2023). Quick Start Guide to Large Language Models: Strategies and Best Practices for Using ChatGPT and Other LLMs. Addison-Wesley Professional.(O'Reilly Media)

Sabharwal, N., Agrawal, A., (2021). Hands-on Question Answering Systems with BERT: Applications in Neural Networks and Natural Language Processing, 15-39.

Sarkar, D. (2019). Text analytics with Python: a practitioner's guide to natural language processing (pp. 1-674). Bangalore: Apress.

Summers, L., Shallenberger, A. N., Cruz, J., & Fulton, L. V. (2023). A Multi-Input Machine Learning Approach to Classifying Sex Trafficking from Online Escort Advertisements. Machine Learning and Knowledge Extraction, 5(2), 460-472.

Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). Practical natural language processing: A comprehensive guide to building real-world NLP systems. O'Reilly Media.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations (pp. 38-45).

Other equipment /material requirements
Python 3.X or Google Co-Lab

Textbooks & Readings (Recommended)
Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.".

Canvas

Canvas is the Learning Management System (LMS) at Boston College, designed to help faculty and students share ideas, collaborate on assignments, discuss course readings and materials, submit assignments, and much more - all online. Your course will make significant use of Canvas this semester; you should familiarize yourself with this useful tool. For more information and training resources for using Canvas, see https://bcservices.bc.edu/service/canvas. In the case of any technical difficulties or concerns, please contact canvas@bc.edu or 617-552-HELP (4357) for immediate

assistance. Canvas requires particular computer specifications and Wi-Fi access. It is essential that you plan accordingly.

Course Outcomes
By the end of this course, students will be able to do the following.
1. Build and interpret NLP models for document classification.
2. Build and interpret NLP models for text summarization.
3. Build and interpret NLP models for text similarity and clustering.
4. Build and interpret NLP models for sentiment and semantic analysis.
5. Build and interpret NLP models using deep learning models.

Assessments and Grading Policy/Weighting
In this course, there are four graded components.

- 16 discussions (Objectives 1 through 5)
- 2 homework projects (Objectives 1-3 and Objectives 1-5, respectively)
- 1 midterm (Objectives 1 through 5)
- 1 final (Objectives 1 through 5)

Discussions (16 x 1% = 16%). Post your first response (written or video) to Discussion #1 early in the learning week by WEDNESDAY at 11:59 pm EST, then respond to at least two other posts (text only) by SUNDAY at 11:59 pm EST. No late posts are accepted. The rubric follows:

Criteria	Not completed	Below Graduate Level	Graduate Level
The initial post answers the requirement	0	.25	.5
Follow-up discussions contribute to learning	0	.25	.5
Total	0	.5	1

Homework Assignments (2 x 15% = 30%). Homework assignments are due in the week assigned by SUNDAY at 11:59 pm EST. All assignments require the use of Python and Jupyter for professional formatting.

Midterm Assignment (24%). You will complete a take-home, applied NLP midterm project. The rubric and deadline for this midterm will be posted to Canvas.

Final Assignment (30%). During the week of finals, you will compete in an applied NLP final project. The rubric and deadline for this assignment will be posted to Canvas.

Deadlines and Late Work
You may only submit late work up to 5 days after the assignment's due date. The penalty without a priori coordination is 20% / day. I do not accept late midterm assignments, final assignments, or discussions.

Course Assignments
Most students should spend nine hours each week working on mastering the content in this course. The weekly schedule and assignments follow.

Assessment Grading Breakdown Course Component	*Percentage*
Discussion Participation (16 total x 1% each)	16%
Assignments (2 total x 15% each)	30%
Midterm Assignment (24%)	24%
Final Assignment (30%)	30%

I do not provide extra credit for graduate students.

The graduate grading system for Woods Colleg follows.

A (4.00), A- (3.67)
B+ (3.33), B (3.00)
B- (2.67)
C (2.00)
F (.00)

All students can access final grades through Agora after each semester's grading deadline. Students who complete course evaluations can access grades earlier. Most students will spend nine hours each week working on mastering the content in this course. The schedule and assignments follow.

Course Assignments

Most students will spend nine hours each week working on mastering the content in this course. Some weeks will require more time. Some weeks will require less time. The schedule and assignments follow. S1=Sarkar Text, S2=Sabharwal Text, V=Vajjala Text, O=Ozdemir Text.

Week	Begins	Ends	Outcomes	Learning Resources	Assignments	Discussion	Class
1	13-Jan	19-Jan	Discuss applications of NLP	S1 Chapter 1, S2 Chapter 1		1 and 2	Online Synch
2	20-Jan	26-Jan	Use Python to build ANN for NLP	S1 Chapter 2, S2 Chapter 2		3	Online Synch
3	27-Jan	2-Feb	Prepare text for NLP	S1 Chapter 3, S2 Chapters 3-4		4	Sat Feb 1, 8 AM-1 PM
4	3-Feb	9-Feb	Build text classification models	S2 Chapter 5, Dong et al. article		5	Online Synch
5	10-Feb	16-Feb	Built text summarization and topic models	S2 Chapter 6	#1	6	Online Synch
6	17-Feb	23-Feb	Build text similiarity models	S2 Chapter 7		7	Online Synch
7	24-Feb	2-Mar	Build text clustering models	S2 Chapter 8		8	Online Synch
Break	3/3/2025	3/9/2025	Break	Break	[redacted]		
8	10-Mar	16-Mar	Midterm	[redacted]			
9	17-Mar	23-Mar	Conduct semantic analysis	S2 Chapter 9	Midterm	9	Sat Mar 22, 8 AM-1 PM
10	24-Mar	30-Mar	Conduct sentiment analysis	S2 Chapter 10		10	Online Synch
11	31-Mar	6-Apr	Build deep learning models for NLP	S2 Chapter 11, Summers et al. article	#2	12	Online Synch
12	7-Apr	13-Apr	Build BERT models	S1 Chapter 4		13	Online Synch
Easter	14-Apr	20-Apr	[redacted]				
13	21-Apr	27-Apr	Build Q&A system	S1 Chapter 5		14	Online Synch
14	28-Apr	4-May	Use BERT for other NLP requirements	S1 Chapter 6		15	Online Synch
15	5-May	11-May	Understanding LLMs	B Chapter 1-2		16	Online Synch
Finals	12-May	18-May	Final Project/Presentation	[redacted]	Final		Tue 13 May, 8 AM-1 PM

Classroom Policies

Participation

Participating in class is an important component of learning. Students are expected to participate and complete all discussions, assignments, and assessments. All discussions are due in the week assigned. The first response is due by Wednesday at midnight ET, and follow-up contributions are due by Sunday While this is an asynchronous course, participating in online class discussion boards is an important component of learning. These discussion boards will take the place of our in-class interactions and should be used as both a space to provide opinions and thoughts, while also posing questions or asking for clarification on theories and concepts.

Consistent with BC's commitment to creating a learning environment that is respectful of persons of differing backgrounds, we believe that every reasonable effort should be made to allow members of the university community to observe their religious holidays without jeopardizing their academic status. Students are responsible for reviewing course syllabi as soon as possible, and for communicating with the instructor promptly regarding any possible conflicts with observed religious holidays. Students are responsible for completing all class requirements for days missed due to conflicts with religious holidays.

Submitting Papers

All writing assignments are to be submitted through Canvas no later than 11:59 PM on the paper due date. Please submit all documents in either .doc or .docx format.

Late Work Policies

Students must complete an initial discussion board post on Friday by 11:59 PM for the week they are due. If a student fails to submit an initial discussion board post they will receive a zero for that week's discussion board, regardless of if they post responses to classmates.

Due to the compressed nature of this course, you may only submit late work up to 5 days after the assignment's due date. The penalty without a priori coordination is 20% / day. I do not accept late examinations or discussion posts.

Communication

You may contact me by email and I will respond within 24 hours.

Boston College Policies

Academic Integrity

I have no tolerance for cheating in any form. Such will earn the student an automatic zero. Academic misconduct includes (but is not limited to): cheating on exams; using material from the internet without citing it; plagiarizing any part of work done by someone else; and submitting substantially similar work for two courses without consent. It is your responsibility to familiarize yourself with Boston College policy on academic integrity. If you have any questions, always consult with me.

Accommodations

Boston College is committed to providing accommodations to students, faculty, staff, and visitors with disabilities. Specific documentation from the appropriate office is required for students seeking accommodation in Woods College courses. Advanced notice and formal registration with the appropriate office are required to facilitate this process. There are two separate offices at BC that coordinate services for students with disabilities:

- The Connors Family Learning Center (CFLC) coordinates services for students with LD and ADHD.
- The Disabilities Services Office (DSO) coordinates services for all other disabilities.

Find out more about BC's commitment to accessibility at www.bc.edu/sites/accessibility.