

INF 496/596: Intermediate Special Topics in Informatics Social Media Analysis (3 Credits)

Semester: Fall 2019

Day/Time: T 5:45-8:35

Location: Husted Hall 004

Prerequisites: INF 108 or permission of instructor

Instructor: Dr. Sam Jackson

Contact: sdjackson@albany.edu

Office Location and Hours: Draper Hall 304a T/Th 4:05-5:00

Course Description, Purpose and Goals

Course Description:

Between Facebook, YouTube, Twitter, and countless others, social media have become part of many people's daily lives. As people use these digital platforms to connect with friends and family, learn about breaking news, or watch the latest cat video, they leave records of what they do. Researchers and analysts have discovered that these records can be a rich source of data. For example, perhaps we can map the spread of the flu based on web searches. Perhaps we can determine motivations of those who commit violent crimes based on their social media posts. Perhaps we can use social media analytics to find people who need help during disaster scenarios.

This course introduces students to a range of methods of collecting and analyzing social media (and other internet-based) data to investigate social phenomena. Students will learn about the kinds of analyses that can be performed on social media data. Additionally, the course explores legal and ethical implications of analyzing this internet-based data.

This course focuses on providing hands-on experience using Python tools to collect and analyze data. Assessment will be based on activities in which students perform data collection and analysis, culminating in a final project in which students will collect and analyze data to answer a question of their choosing.

Student Learning Outcomes:

Upon completion of this course, students will be able to:

- Articulate the opportunities and limitations of social media data for research and investigation,
- Identify and work through legal and ethical implications of social media analysis,
- Collect data using APIs and web scraping, and
- Conduct a variety of types of analysis using social media data.

Course sessions: We will use class meetings for a mixture of lecture, discussion, group activities, and time to work on assignments individually. As this course is coding-heavy, we will use computers in most class sessions. Students can use classroom computers or their own laptops.

Required Readings: This course does not have a required textbook. Required readings will be available on Blackboard or provided in class.

Additional readings and exercises may be added during the semester.

Course Delivery Structure, Schedule, & Grading

Grading: This course is A-E graded and the grades are determined based on graded assignments:

| | | | | |
|------------------|------------------|------------------|------------------|-----------|
| | 86.5 - 89.4 = B+ | 76.5 - 79.4 = C+ | 66.5 - 69.4 = D+ | |
| 92.5 - 100 = A | 82.5 - 86.4 = B | 72.5 - 76.4 = C | 62.5 - 66.4 = D | <59.5 = E |
| 89.5 - 92.4 = A- | 79.5 - 82.4 = B- | 69.5 - 72.4 = C- | 59.5 - 62.4 = D- | |

Assessments: Unless specified otherwise, assignments are submitted via Blackboard by 11:59pm on their due date.

1. Python exercise (10%): Sept 10

This assignment functions as a check to make sure you have enough experience with Python to successfully complete this class. If you struggle with this exercise, you should be prepared to spend some time honing your coding skills; if you *really* struggle with it, you might want to reconsider taking this course this semester.

2. Final project description (10%): Oct 1

For this assignment, you will write a two-page (double-spaced, 12pt font, 1" margins) description of your general plan for your final project. This assignment should answer the following questions: What question do you want to answer? What site(s) might you collect data from? What kind of data (or metadata) might you use? How will you ensure that this project is carried out in an ethically and legally responsible manner?

This assignment prompts you to start thinking about your final project early in the semester; the details of your final project are likely to change over the course of the semester, though, and that is perfectly fine.

3. Technique exercises (45%): due throughout the semester

In each of these exercises, you will practice the techniques covered in class. Group work is not permitted for these exercises. For each exercise, you will start by writing a two paragraph plan of what you intend to do. Then, you will attempt to carry out that task. Finally, you will write a reflection on that process (two pages double-spaced max). You will submit your work plan, any code you write along with any output generated, and your reflection. The lowest grade on these exercises will be dropped.

- a. Scraping exercise: October 8
- b. API exercise: October 15
- c. Metadata analysis exercise: November 5
- d. Network analysis exercise: November 12
- e. Supervised machine learning exercise: November 19
- f. Unsupervised machine learning exercise: November 26
- g. Data visualization exercise: December 3

4. Final project (35%): Dec 15

Undergraduate students

For the final project for this class, you will pull together all of the elements from throughout the semester into one larger analysis project. The assignment consists of four components:

- 1) A 3-6 page paper (double-spaced, 12 pt font, 1" margins) in which you introduce the project, describe the data and analysis performed, and describe the conclusions drawn from this analysis;
- 2) The Python code you write to collect and process your data;
- 3) The output of that data collection and processing code;

- 4) The Python code you write to do your analysis;
- 5) Any output of that analysis code.

Graduate students

For the final project for this class, you will pull together all of the elements from throughout the semester into one larger analysis project. The assignment consists of four components:

- 1) A 5-10 page paper (double-spaced, 12 pt font, 1" margins) in which you introduce the project, provide some review of existing research related to the topic, describe the data and analysis performed, and describe the conclusions drawn from this analysis;
- 2) The Python code you write to collect and process your data;
- 3) The output of that data collection and processing code;
- 4) The Python code you write to do your analysis;
- 5) Any output of that analysis code.

In addition, graduate students will present their in-progress projects in the last week of class.

➤ **Extra credit: Showcase presentation (5%): Dec 4**

You may choose to present your final project during the iCEHC Showcase on Dec 4. You will create a poster based on your project, and you will present your poster at Showcase.

You must sign up to do this by Oct 25.

This is the only opportunity for extra credit in this course.

Course outline:

| DATE | TOPIC | READINGS | ASSIGNMENTS |
|---------------------|--|--|---------------------------|
| 27-Aug | Course Introduction, Intro to JupyterHub | | |
| 3-Sep | Python Review | Rogel-Salazar, " Python: For Something Completely Different "* | |
| 10-Sep | Platforms, Users, and Online Activity | McCay-Peet and Quan-Haase, " What is Social Media and What Questions Can Social Media Research Help Us Answer? " | Python exercise |
| 17-Sep | Data (and Metadata) | Berger, " Evil in a Haystack " | |
| 24-Sep | Ethics and the Law | Golbeck, " Legal Issues "; Omand et al. " #Intelligence " | |
| 1-Oct | Data Collection: Scraping | Ignatow and Mihalcea, " Web Crawling and Scraping " | Final project description |
| 8-Oct | Data Collection: APIs | Freelon, " Computational Research in the Post-API Age " | Scraping exercise |
| NO CLASS OCTOBER 15 | | | API exercise |

| | | | |
|--------|---------------------------------|---|--|
| 22-Oct | Data Cleaning, Processing, etc. | Readings TBD | |
| 29-Oct | Metadata Analysis | Readings TBD | |
| 5-Nov | Network Analysis | Hogan, “ Online Social Networks ” | Metadata analysis exercise |
| 12-Nov | Supervised Machine Learning | Readings TBD | Network analysis exercise |
| 19-Nov | Unsupervised Machine Learning | Readings TBD | Supervised machine learning exercise |
| 26-Nov | Data Visualization | Healy, “ Look at Data ” | Unsupervised machine learning exercise |
| 3-Dec | Workshop/Application | | Data visualization exercise |

* Optional readings

Grade Appeals: Students who feel that their exams or assignments have been graded incorrectly should follow a three-step procedure. First, the student carefully reads the exam or assignment and identifies the precise problem(s) with the grading. Second, the student sends a written appeal explaining why their answer was appropriate to the instructor; **you must wait 24 hours after receiving your grade to submit an appeal.** Third, the instructor regrades the exam or assignment, taking the appeal into account. Note that the instructor may assign a higher or lower grade following such an appeal.

The Course Environment

Instructor Availability: I will be available for student consultation during office hours and by email. I endeavor to respond to email in a timely manner, but I generally do not respond to email after 5pm or on the weekend. If you do not hear back from me after 2 business days, please send me a follow up email. Email should be used in a professional manner. Every message should contain a meaningful subject, a polite greeting, and a closing. If an email contains an attachment, there should be a description of what that attachment is and why you are sending it. If I receive an attachment with no explanation, I will not open it!

Students and professors should be professional at all times. Faculty should be addressed as Prof. XXX or Dr. XXX. Disrespect in any form in any CEHC class will not be tolerated.

Attendance Policy: Regular attendance is recommended and generally related to the grade attained, but attendance will not be taken during class. However, many class meetings will include dedicated time to work on assignments during which I will be available to answer questions.

Extensions and Late Assignments: Each student can have an extension for any one assignment other than the final project, no questions asked. But you **must** arrange this with me at least 24 hours before the assignment is due.

Other than this one free extension, unexcused late assignments will be penalized 10% per day (so an assignment submitted two days late that would have been a 100 will be an 80). Students missing an assignment without prior approval of the instructor (or documentation of an emergency medical situation) will receive a "0" for that assignment unless they have a valid and documented excuse. UAlbany's medical excuse policy can be reviewed at: http://www.albany.edu/health_center/medicalexcuse.shtml.

Absence due to religious observance: New York State Education Law (Section 224-a) - Campuses are required to excuse, without penalty, individual students absent because of religious beliefs, and to provide equivalent opportunities for make-up examinations, study, or work requirements missed because of such absences. Faculty should work directly with students to accommodate religious observances. Students should notify the instructor of record in a timely manner.

Withdrawal from the Course: The drop date for the Fall 2019 semester is November 4 for undergraduate students. That is the last date you can drop a course and receive a 'W'. It is your responsibility to take action by this date if you wish to drop the course. In particular, grades of "incomplete" will not be awarded to students because they missed the drop deadline.

Cell phones & laptops: To help everyone avoid distraction, please put away cell phones and headphones at the beginning of class. Please do not use computers for any purposes not related to class during class sessions -- such behavior can be a distraction for others in the class in addition to yourself.

Unexpected Life Events: Over the course of the semester, it is possible that things may happen in your life outside of this course that make it difficult for you to attend class or complete assignments on time. I urge you to reach out to me if something like this happens to you. I will work with you to make arrangements as appropriate. But you **must** tell me about these things as they happen. If you wait until the end of the semester to tell me about something that made it difficult for you to complete assignments due weeks prior, you will be stuck with the grades you received on those assignments.

Additional Information

Disability Policy: Reasonable accommodations will be provided for students with documented physical, sensory, systemic, cognitive, learning and psychiatric disabilities. If you believe you have a disability that warrants accommodation in this class, please notify the Director of the Disability Resource Center (Campus Center 137, 442-5490). That office will provide the course instructor with verification of your disability and will recommend appropriate accommodations. If language or some other barrier inhibits your performance, you should meet with the instructor during the first two weeks of class to devise a solution.

Academic Integrity: It is every student's responsibility to become familiar with the standards of academic integrity at the University. Claims of ignorance, of unintentional error, or of academic or personal pressures are not sufficient reasons for violations of academic integrity. See http://www.albany.edu/undergraduate_bulletin/regulations.html. Additional resources are available at <https://library.albany.edu/infolit/integrity>.

The following is a partial list of the types of behaviors that are defined as examples of academic dishonesty and are therefore unacceptable. Attempts to commit such acts also fall under the term academic dishonesty and are subject to penalty.

- Plagiarism
- Allowing other students to see or copy your assignments or exams
- Examining or copying another student's assignments or exams
- Lying to the professor about issues of academic integrity
- Submitting the same work for multiple assignments/classes without prior consent from the instructor(s)
- Forgery
- Sabotage
- Unauthorized Collaboration (just check first!)
- Falsification
- Bribery
- Theft, Damage, or Misuse of Library or Computer Resources

Any violation of academic integrity will result in a zero on that assignment and referral to the appropriate Departmental and University Committees, and additional grade sanctions may be imposed if appropriate.

All course material and documents developed by the instructor are copyrighted and may not be reproduced or distributed without express written permission.

Responsible Use of Information Technology:

<https://wiki.albany.edu/display/public/askit/Responsible+Use+of+Information+Technology+Policy>

Time Management: For every credit hour that a course meets, students should expect to work 3 additional hours outside of class every week (3 x 3 = 9). For a three-credit course you should expect to work 9 hours outside of class per week on average. Manage your time effectively to complete readings, assignments, and projects. Note that the amount of required reading and the assignments are not evenly distributed over the semester, so some weeks may require more than 9 hours of outside-of-class work and some weeks may require fewer.

Respect for Diversity: It is my intent that students from all diverse backgrounds and perspectives be well served by this course, that students' learning needs be addressed both in and out of class, and that the

diversity that students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups. In addition, if any of our class meetings conflict with your religious events, please let me know so that we can make arrangements for you.¹

Supplemental Materials

Resources on Social Media Research:

Center for Analysis of Social Media - Demos: <https://demos.co.uk/research-area/casm/>

Association of Internet Researchers: <http://aoir.org/>

VOX-Pol: <https://www.voxpol.eu/>

Student resources:

Student Health Services: https://www.albany.edu/health_center/

Counseling Center: https://www.albany.edu/counseling_center/

Student Emergency Fund: <https://www.albany.edu/studentaffairs/emergencyfund.shtml>

UAlbany library: <https://library.albany.edu/>

Library research resources: <http://libguides.library.albany.edu/c.php?g=537164&p=3677741>

¹ Respect for Diversity statement from <https://www.brown.edu/sheridan/teaching-learning-resources/inclusive-teaching/statements>