DEPARTMENT: Behavioral, Social & Health Education Sciences

COURSE NUMBER: 760R/GRAD 700R (+lab)

CREDIT HOURS: 4 credits

SEMESTER: Fall 2022

COURSE TITLE: Reducing Drug-Related Harms using Internet-Based "Big Data": Machine Learning and AI Methods

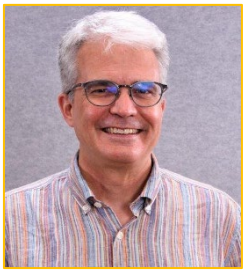
DATES: Wednesdays @ 10:30 AM – 1:20 PM (2-hr lecture and 1-hr lab)

LOCATION: Rollins School of Public Health | Grace Crum Rollins Building, Rm. L45

INSTRUCTORS:

Abeed Sarker, Ph.D.
Assistant Professor
Department of Biomedical Informatics
School of Medicine
Email: abeed.sarker@emory.edu

Hannah Cooper, ScD
Professor
Department of Behavioral, Social and Health Education Sciences
Rollins School of Public Health
Email: hcoope3@emory.edu



Lance A. Waller, Ph.D.
Professor
Department of Biostatistics and Bioinformatics
Rollins School of Public Health
Email: lwaller@emory.edu

Please feel free to contact us anytime via email with questions related to the course or your shared interest in substance use-related harm. We will endeavor to respond within 48 hours.

COURSE DESCRIPTION:

This course will prepare students to conduct ethical, rigorous, and theoretically-informed analyses of "big data" (machine learning/social media) in the context of research and interventions into intersecting crises of substance use disorders (SUDs) and drug-related harms. It will apply the strengths of social and behavioral sciences – including a focus on theory and validity – to the emerging field of advanced data analytics.

This course is one of two courses on analyzing "big data" to study and intervene in drug-related harms (the other course is entitled "Reducing Drug-Related Harms Using Big Data: Administrative, geospatial and network sources"). We recommend, but do not require, taking this course second, or after taking the other course.

COURSE COMPETENCIES:

Across each component/module trainees will learn to:

- Design and conduct theoretically-informed and data-driven analyses of distributions and ecologies of SUD-related harms by applying advanced data science methods to internet-based data.
- Design and conduct theoretically-informed analyses assessing policies and programmatic interventions that may affect SUD-related harms and services by applying advanced data science methods to internet-based data.
- Communicate findings to select stakeholder communities to strengthen efforts to end SUD-related harms.
- Critically apply principles of the ethical and responsible conduct of research.

PRE-REQUIREMENTS:

This course requires students to develop skills across multiple programming environments. It is recommended that students be familiar with database formatting using R or SQL via SAS. Prior completion of the following courses/equivalents is also needed to successfully complete this module:

1. Regression (examples: BIOS 501, BSHES 700), and
2. At least one statistical programming course, such as SAS (examples: BIOS 501) or R (examples: BIOS 544)

COURSE LEARNING OBJECTIVES

Upon completion of this course, the student will be able to:

1. Design theoretically guided analyses describing distributions and ecologies of SUD-related harms and services using appropriate data science methods for internet-based data.
2. Conduct theoretically guided analyses describing distributions and ecologies of SUD-related harms and services using appropriate data science methods for internet-based data.
3. Design theoretically guided analyses of policies and programmatic interventions that may affect SUD-related harms and services using appropriate data science methods for internet-based data.
4. Conduct theoretically guided analyses of policies and programmatic interventions that may affect SUD-related harms and services using appropriate data science methods for internet-based data.
5. Compare the rigor (e.g., validity) of various data science methods as tools to study SUD-related harms and services.
6. Communicate the rationale, methods, findings, and conclusions of theoretically guided analyses of internet-based data describing distributions and ecologies of SUD-related harms and services to diverse audiences.
7. Communicate the rationale, methods, findings, and conclusions of theoretically guided analyses of policies and programmatic interventions involving internet-based data that may affect SUD-related harms and services to diverse audiences.
8. Assess ethical issues posed by each data science method and consider the responsible conduct of related analyses, particularly as applied to SUD-related research.

COURSE MATERIALS:

Mandatory Synchronous Class Sessions
Each class section will meet with Professor Cooper, Professor Waller, or Professor Sarker.

Textbook

- **There is no required textbook.**
- It is your responsibility to review the required readings for each week. These same readings will be used for a weekly asynchronous assignment.

Technology

- This course uses synchronous meetings and delivery of asynchronous content delivered via Canvas. Click here for a PDF of the Emory College Online technology requirements.
- Lecture slides will be made available to all students via Canvas.
- Zoom meeting links will be made available prior to class if/when it is determined that a session/sessions will not meet in person.

The Canvas Learning Management System

- This course will use a Canvas site for communication and posting of course materials (e.g. documents, exams, assignments, lecture slides, supplemental readings, Kahoot surveys, etc.).
- It is your responsibility to check this site regularly to stay up-to-date on announcements and assignments.
- Computer specifications for Canvas
- Canvas Resources for Students

Office Hours Tools

- Office hours for each professor will be held in-person each week and will be relayed at the start of the semester and posted on Canvas.

Library Resources & Online Videos

- This course will refer to scientific publications that can be accessed on the internet. Emory University Libraries provides access to all databases online.
- A guide to library research tools is online.
- This course will also use videos that are shown in-class or as recommended viewing. Some videos will be available on the internet via YouTube.

OVERALL COURSE POLICIES

Attendance

Attendance at assigned class is mandatory – you will be marked absent if you miss class. Students are expected to be active learners and participants. Evidence of this includes:

- Attending and being on-time. Please see the Absence Policy for more details on how this affects grading.
- Should the circumstances of the pandemic require that we switch the course format to zoom meetings, students are required to be visually present during Zoom meetings. In such instances:
  - The camera on your device should be turned on. Your mic will be muted. A raised hand should be used to indicate you have a question/comment relevant to the material being presented.
  - Your zoom name should match your real name. This will allow TAs to confirm your attendance for the duration of each lecture.
- Being engaged by asking questions and contributing to class discussions. Students will be able to submit anonymous questions during each class session.

Assignments

- Assignments include answering specific questions after reviewing and interpreting the assigned readings and completing laboratory assignments each week.
- Assignments must be written in your own words. Submitted assignments must have a Canvas plagiarism/"similarity score" < 25%. See here for info.
- Assignments must be submitted via Canvas, per instructions, by the due date indicated at the time of distribution.
  - Canvas will not allow late assignments to be submitted.
  - Assignments should be submitted well in advance of the 11:59 PM deadline.
  - Assignments cannot be made up at a later time. *Exceptions to this policy will require explicit permission of the instructor in writing, prior to the due date.*
- Questions about assignments should be emailed to the professors.

Grading Policy (+ Extra Credit Opportunity)

- Earned points at the end of the semester are used to determine each student's grade.
- Appeals to final grade decisions should be submitted (in writing) to the Director of Graduate Studies in RSPH.

Lectures & Zoom Videos of Synchronous sessions (in case of switch to online learning)

- Lectures and other classroom presentations via video conferencing and other materials posted on Canvas are for the sole purpose of educating the students enrolled in the course.
  - *These lectures and videos are the property of the instructors. The release of such information (including but not limited to directly sharing, screen capturing, or recording content) is strictly prohibited, unless the instructors state otherwise. Doing so without the permission of the instructor will be considered an Honor Code violation, and may also be a violation of state or federal law, such as the Copyright Act.*
- All University policies remain in effect for students participating in remote education.
- Videos will only show the PowerPoint slides and include the voice of the instructor.
- Videos will become available several days after class in order to allow time for processing and editing.
- Videos are not a stand-in for synchronous class attendance, which is required (see attendance policy).

Written Communication with the Instructor

- We will respond to written communications within 48 business hours.
- Use email as the primary mode for communication, especially if your email will contain personal information, such as grades, attendance, illness, etc.
    - As a general rule, email communications with instructors should be conducted in a professional manner:
        - An email and all subsequent replies should include appropriate salutation and valediction statements and your signature.
        - The body must not always be formal, but it should be written in a respectful tone (i.e. If you would hesitate to say something to someone's face, do not write it in an e-mail.).
        - Strive for clarity; after reading the email once, the recipient should be able to understand the purpose and context of the email and your expected or desired response.
        - Strive for brevity; lengthy questions and/or explanations are more appropriate for a synchronous online meeting.
        - The email should contain correct grammar and punctuation, and it should not contain terms or abbreviations that are adapted specifically for text messaging or social media.

LGS & RSPH POLICIES

Laney Graduate Student Handbook

The Laney Graduate School Handbook (https://gs.emory.edu/handbook/) is the official reference for graduate students and others regarding the administrative and procedural policies, as well as the rules and regulations, of the Laney Graduate School. If you have questions about specific policies, please contact the appropriate LGS staff member for assistance.

Accessibility and Accommodations

As the instructors of this course we endeavor to provide an inclusive learning environment. However, if you experience barriers to learning in this course, do not hesitate to discuss them with us and the Office of Accessibility Services (OAS). Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact the OAS. It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed.

Students who registered with OAS and have a letter outlining their academic accommodations are strongly encouraged to coordinate a meeting time with me to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible.

Contact Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at http://equityandinclusion.emory.edu/access/students/index.html

Emory COVID-19 Policies

All Emory University students, faculty, and staff are required to be fully vaccinated for COVID-19. Campus members who have an approved vaccination exemption or anyone not yet fully vaccinated are required to conduct regular screening tests. At this time, testing is required weekly, however, frequency could increase depending on community prevalence. Testing procedures and scheduling information are available on Emory Forward's testing information page.

Please visit Emory University's website on Navigating COVID-19 to stay informed of the latest guidance and policies regarding COVID-19.

Honor Code

You are bound by Emory University's Student Honor and Conduct Code. Students enrolled in Laney Graduate School (LGS) will follow the LGS Policies and Honor Code. Students enrolled in RSPH will follow the policies and honors of RSPH. The LGS honor code is available at: https://gs.emory.edu/handbook//honor-conduct-grievance/honor/index.html.

RSPH requires that all material submitted by a student fulfilling his or her academic course of study must be the original work of the student. Violations of academic honor include any action by a student indicating dishonesty or a lack of integrity in academic ethics. *Academic dishonesty refers to cheating, plagiarizing, assisting other students without authorization, lying, tampering, or stealing in performing any academic work, and will not be tolerated under any circumstances.*

The RSPH Honor Code states: "Plagiarism is the act of presenting as one's own work the expression, words, or ideas of another person whether published or unpublished (including the work of another student). A writer's work should be regarded as his/her own property." (http://www.sph.emory.edu/cms/current_students/enrollment_services/honor_code.html)

COURSE EVALUATION

	Course Points Distribution
Attendance & Participation	**10%**
Assignments	**40%**
Laboratory tasks & Asynchronous learning	**20%**
Readings & Discussion	**30%**
Course Total:	**100%**

93 and above	A
90 to 92.99	A-
87 to 89.99	B+
83 to 86.99	B
80 to 82.99	B-
70 to 79.99	C
0 to 69.99	F

COURSE OVERVIEW

This course will focus on different techniques that are used to study and reduce drug-related harms. It will involve a series of lectures and labs that will help students tackle a large body of literature and a BIG data project at the end of the semester.

In this course, students will be introduced to the emerging field of internet-based data mining, including social media mining, for public health surveillance. Students will understand the value of social media big data and its utility, and will learn to utilize methods such as natural language processing for extracting knowledge from the data. Students will also learn to appreciate the ethical complexities of conducting research using public social media data.

LECTURE SCHEDULE OVERVIEW

Date	Topic	Instructor
Aug-24	Course Introduction	Cooper
Aug-31	Theoretical background to natural language processing, machine learning and internet-based data	Sarker
Sep-7	Introduction to natural language processing	Sarker
Sep-14	Fundamental problems in natural language processing of internet-based data	Sarker
Sep-21	Rule-based methods	Sarker
Sep-28	Introduction to machine learning for natural language processing	Sarker
Oct-5	Machine learning for natural language processing II	Sarker
Oct-26	Evaluation of machine learning methods	Sarker
Nov-2	Drawing insights from social media data	Sarker
Nov-9	Future directions and ethical considerations	Sarker
Nov-16	Theory, Data, Mathematics, and the Alignment Problem	Waller
Nov-23	NO CLASS – THANKSGIVING HOLIDAY	
Nov-30	Looping back full circle: Advanced Data Analytics and Behavioral Science	Waller

DETAILED LECTURE, LAB, ASSIGNMENT AND READING SCHEDULE

Readings are to be completed prior to the date of the corresponding class.

Course Component	Date, Instructor & Details
Aug-24 (Hannah Cooper)	
Class session	• Overview of the course • The epidemiology of drug-related harms • Theories on drug-related harms
Pre-Course Assignment	• Explore *this website*, creating 1 table/chart that are of interest to them, and write a 1 paragraph description of what you see, and bring it to class for discussion.
Pre-Course Readings	• Overdose trends - Read *this report* carefully. • HIV - Scan *this report*. • Hepatitis C - Read *this paper* carefully.
Aug-31 (Abeed Sarker)	
Class session	Theoretical background to natural language processing, machine learning & internet-based data • First hour: Theoretical foundations of natural language processing and machine learning • Second hour: Application of natural language processing and machine learning to internet-based data (theoretical concepts)
Lab	Students will learn how to use the Twitter developer account and R to access real-time data • R and the Twitter API • Twitter developer accounts for research
Asynchronous learning	Tutorial for Twitter data access
Readings	**Importance of novel surveillance methods for SUD** Kolodny A, Frieden TR. Ten Steps the Federal Government Should Take Now to Reverse the Opioid Addiction Epidemic. JAMA. 2017 Oct 24;318(16):1537-1538. doi: 10.1001/jama.2017.14567. Erratum in: JAMA. 2019 Sep 24;322(12):1215. PMID: 29049522. **Use of social media data for substance use and related topics** Broniatowski DA, Paul MJ, Dredze M. Twitter: big data opportunities. Science. 2014 Jul 11;345(6193):148. doi: 10.1126/science.345.6193.148-a. PMID: 25013052. Kazemi DM, Borsari B, Levine MJ, Dooley B. Systematic review of surveillance by social media platforms for illicit drug use. J Public Health (Oxf). 2017 Dec 1;39(4):763-776. doi: 10.1093/pubmed/fdx020. PMID: 28334848; PMCID: PMC6092878.

Course Component	Date, Instructor & Details
Sep-07 (Abeed Sarker)	
Class session	Introduction to natural language processing • First & second hour: Natural language processing basics
Lab	Students will learn how to do basic natural language processing tasks using R
Asynchronous learning	Introduction to basic natural language processing using R
Readings	**Social media-based analysis of SUD data** Hammond AS, Paul MJ, Hobelmann J, Koratana AR, Dredze M, Chisolm MS. Perceived Attitudes About Substance Use in Anonymous Social Media Posts Near College Campuses: Observational Study. JMIR Ment Health. 2018 Aug 2;5(3):e52. doi: 10.2196/mental.9903. PMID: 30072359; PMCID: PMC6096169. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, Asch DA, Schwartz HA. Facebook language predicts depression in medical records. Proc Natl Acad Sci U S A. 2018 Oct 30;115(44):11203-11208. doi: 10.1073/pnas.1802331115. Epub 2018 Oct 15. PMID: 30322910; PMCID: PMC6217418. Spadaro A, Sarker A, Hogg-Bremer W, Love JS, O'Donnell N, Nelson LS, Perrone J. Reddit discussions about buprenorphine associated precipitated withdrawal in the era of fentanyl. Clin Toxicol (Phila). 2022 Jun;60(6):694-701. doi: 10.1080/15563650.2022.2032730. Epub 2022 Feb 4. PMID: 35119337.
Sep-14 (Abeed Sarker)	
Class session	Fundamental problems in natural language processing of internet-based data. Data/text mining pipelines for internet-based and social media data • First hour: Challenges of natural language processing for internet-based and social media data • Second hour: Practical systems for natural language processing of internet-based data
Lab	Students will continue to solve basic natural language processing problems using R
Asynchronous learning	Natural language processing, rule-based methods and machine learning
Readings	**Strategies for deriving insights about substance use from social media** Hanson CL, Burton SH, Giraud-Carrier C, West JH, Barnes MD, Hansen B. Tweaking and tweeting: exploring Twitter for nonmedical use of a psychostimulant drug (Adderall) among college students. J Med Internet Res. 2013 Apr 17;15(4):e62. doi: 10.2196/jmir.2503. PMID: 23594933; PMCID: PMC3636321. Chan B, Lopez A, Sarkar U. The Canary in the Coal Mine Tweets: Social Media Reveals Public Perceptions of Non-Medical Use of Opioids. PLoS One. 2015 Aug 7;10(8):e0135072. doi: 10.1371/journal.pone.0135072. PMID: 26252774; PMCID: PMC4529203. Graves RL, Tufts C, Meisel ZF, Polsky D, Ungar L, Merchant RM. Opioid Discussion in the Twittersphere. Subst Use Misuse. 2018 Nov 10;53(13):2132-2139. doi: 10.1080/10826084.2018.1458319. Epub 2018 Apr 16. PMID: 29659320; PMCID: PMC6314840.

Course Component	Date, Instructor & Details
Sep-21 (Abeed Sarker)	
Class session	Rule-based methods in natural language processing Data-centric methods for analyzing SUD-related texts from internet-based sources and machine learning • First hour: Rule-based methods text matching • Second hour: Regular expressions
Lab	• Connecting to Twitter API and analyzing search data using R
Asynchronous learning	Matching text patterns and information extraction
Assignment 1	Students will submit a report of their completed lab work. Specifically, building on the lab sessions, they will analyze some text-based data using natural language processing, and submit a 2-page report (plus figures, tables and references). Students will be provided a markdown file as template for running the analysis and the data needed for the analysis. The analysis will focus on generating and comparing frequency distributions of terms/phrases. Due date: End of the week of September 21. Completion points = 20 (i.e., 20% of module)
Readings	**Social media-based survey** Daniulaityte R, Lamy FR, Barratt M, Nahhas RW, Martins SS, Boyer EW, Sheth A, Carlson RG. Characterizing marijuana concentrate users: A web-based survey. Drug Alcohol Depend. 2017 Sep 1;178:399-407. doi: 10.1016/j.drugalcdep.2017.05.034. Epub 2017 Jun 29. PMID: 28704769; PMCID: PMC5567791. **Analysis of Reddit data** Zhan Y, Zhang Z, Okamoto JM, Zeng DD, Leischow SJ. Underage JUUL Use Patterns: Content Analysis of Reddit Messages. J Med Internet Res. 2019 Sep 9;21(9):e13038. doi: 10.2196/13038. PMID: 31502542; PMCID: PMC6786857. Sarker A, Al-Garadi MA, Ge Y, Nataraj N, Jones CM, Sumner SA. Signals of increasing co-use of stimulants and opioids from online drug forum data. Harm Reduct J. 2022 May 25;19(1):51. doi: 10.1186/s12954-022-00628-2. PMID: 35614501; PMCID: PMC9131693. Chancellor S, Nitzburg G, Hu A, Zampieri F, and Choudhury MD. Discovering Alternative Treatments for Opioid Use Recovery Using Social Media. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 124, 1–15. 2019. DOI: https://doi.org/10.1145/3290605.3300354

Course Component	Date, Instructor & Details
Sep-28 (Abeed Sarker)	
Class session	Introduction to machine learning for natural language processing • First and second hour: Machine learning basics including text classification • Second hour: Meta-data—geolocation and time
Lab	• Students will run a text classification experiment following instructions/scripts provided
Asynchronous learning	Text classification using R
Readings	**NLP methods for social media data processing** Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, Upadhaya T, Gonzalez G. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform. 2015 Apr;54:202-12. doi: 10.1016/j.jbi.2015.02.004. Epub 2015 Feb 23. PMID: 25720841; PMCID: PMC4408239. Ovalle A, Goldstein O, Kachuee M, Wu ESC, Hong C, Holloway IW, Sarrafzadeh M. Leveraging Social Media Activity and Machine Learning for HIV and Substance Abuse Risk Assessment: Development and Validation Study. J Med Internet Res. 2021 Apr 26;23(4):e22042. doi: 10.2196/22042. PMID: 33900200; PMCID: PMC8111510. Patton DU, Lee FT, Frey WR, McKeown K, McGregor KA, Moss E. Contextual Analysis of Social Media: The Promise and Challenge of Eliciting Context in Social Media Posts with Natural Language Processing. Proc AAAI ACM Conf AI Ethics Soc. 2020 Feb;2020:337-342. doi: 10.1145/3375627.3375841. Epub 2020 Feb 7. PMID: 35265948; PMCID: PMC8902697. Conway M, Hu M, Chapman WW. Recent Advances in Using Natural Language Processing to Address Public Health Research Questions Using Social Media and ConsumerGenerated Data. Yearb Med Inform. 2019 Aug;28(1):208-217. doi: 10.1055/s-0039-1677918. Epub 2019 Aug 16. PMID: 31419834; PMCID: PMC6697505.

Course Component	Date, Instructor & Details
Oct-5 (Abeed Sarker)	
Class Session	Machine learning for natural language processing II • First and second hour: Review of text classification and more advanced strategies for text representation for machine learning
Lab	• Students will conduct the same text classification experiments as the previous week, but this time using more advanced representations
Asynchronous learning	Text representation and classification using R
Readings	**NLP methods for social media-based SUD surveillance** Yang YC, Al-Garadi MA, Love JS, Perrone J, Sarker A. Automatic gender detection in Twitter profiles for health-related cohort studies. JAMIA Open. 2021 Jun 23;4(2):ooab042. doi: 10.1093/jamiaopen/ooab042. PMID: 34169232; PMCID: PMC8220305. Kalyanam J, Katsuki T, R G Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. Addict Behav. 2017 Feb;65:289-295. doi: 10.1016/j.addbeh.2016.08.019. Epub 2016 Aug 17. PMID: 27568339. Preiss A, Baumgartner P, Edlund MJ, Bobashev GV. Using Named Entity Recognition to Identify Substances Used in the Self-medication of Opioid Withdrawal: Natural Language Processing Study of Reddit Data. JMIR Form Res. 2022 Mar 30;6(3):e33919. doi: 10.2196/33919. PMID: 35353047; PMCID: PMC9008522.
Oct-26 (Abeed Sarker)	
Class Session	Evaluation of machine learning methods
Lab	Students will evaluate machine learning methods and compare two datasets from social media data
Asynchronous learning	Intrinsic and extrinsic evaluations
	Advanced social media-based data mining pipelines Sarker A, DeRoos A, Perrone J. Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework. J Am Med Inform Assoc. 2020 Feb 1;27(2):315-329. doi: 10.1093/jamia/ocz162. PMID: 31584645; PMCID: PMC7025330. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of Codeine Misuse on Instagram: Content Analysis. JMIR Public Health Surveill. 2018 Mar 20;4(1):e22. doi: 10.2196/publichealth.8144. PMID: 29559422; PMCID: PMC5883072. Jha D, Singh R. SMARTS: the social media-based addiction recovery and intervention targeting server. Bioinformatics. 2019 Oct 24:btz800. doi: 10.1093/bioinformatics/btz800. Epub ahead of print. PMID: 31647520.

Course Component	Date, Instructor & Details
Nov-2 (Abeed Sarker)	
Class Session	Drawing insights from social media data
Lab	Students will perform thorough comparisons of two datasets, which will be part of assignment 2
Asynchronous learning	Advanced natural language processing and machine learning methods
Readings	**Geolocation-specific analysis of social media data and comparison with other sources of information** Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from Tweets: Estimating Misuse of Prescription Opioids in the USA from Social Media. J Med Toxicol. 2017 Dec;13(4):278-286. doi: 10.1007/s13181-017-0625-5. Epub 2017 Aug 22. PMID: 28831738; PMCID: PMC5711756. Sarker A, Gonzalez-Hernandez G, Ruan Y, Perrone J. Machine Learning and Natural Language Processing for Geolocation-Centric Monitoring and Characterization of Opioid-Related Social Media Chatter. JAMA Netw Open. 2019 Nov 1;2(11):e1914672. doi: 10.1001/jamanetworkopen.2019.14672. PMID: 31693125; PMCID: PMC6865282.

Course Component	Date, Instructor & Details
Nov-9 (Abeed Sarker)	
Class Session	First hour: Future directions and ethical considerations. Second hour: Lab and practical support. Assignment help.
Lab	This week will focus on providing students with technical and theoretical support for their lab work and assignment. Particular focus will be on refining machine learning methods and making sure the methods/findings are scientifically sound. • Lab work will focus on validating the methods developed.
Asynchronous learning	Combining machine learning, natural language processing and evaluation for social media-based datasets
Assignment 2	Students will submit a report outlining their findings of the data analysis. The data analysis will particularly compare how internet chatter differs between opioids and stimulants, also over different geolocations and time periods. The 3-page (maximum) report will discuss how the methods learned can form the foundations of learning from big internet-based data and also have to include a viewpoint-style paragraph regarding the ethical implications of their work. Due date: end of the week of November 30. Completion points = 20 (i.e., 20% of module)
Readings	**Ethics** Conway M, O'Connor D. Social Media, Big Data, and Mental Health: Current Advances and Ethical Implications. Curr Opin Psychol. 2016 Jun;9:77-82. doi: 10.1016/j.copsyc.2016.01.004. PMID: 27042689; PMCID: PMC4815031. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. J Med Internet Res. 2017 Oct 31;19(10):e353. doi: 10.2196/jmir.6426. PMID: 29089287; PMCID: PMC5686417. Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AY, Mayer MA, Merolli M, Hansen M. Ethical Issues of Social Media Usage in Healthcare. Yearb Med Inform. 2015 Aug 13;10(1):137-47. doi: 10.15265/IY-2015-001. PMID: 26293861; PMCID: PMC4587037.

Course Component	Date, Instructor & Details
Nov-16 (Lance Waller)	
Class session	• First hour: Historical review of interactions between theory, data, and ethics. • Second hour: Discussion.
Lab	• There will be no lab for this class session
Asynchronous learning	• The readings provide many topics for discussion, please read beforehand and come with questions and comments.
Readings	• Watch Brian Christian's lecture on the Alignment Problem • Read the first 21 pages of Lillian Lieber's *The Education of T.C. MITS* (originally published in 1944)…note this was written toward the end of World War II and represents an oddly wonderful semi-free verse poetic snap shot in time of mathematicians struggling with how mathematics could be both beautifully abstract and yet still contribute to the horrors of war.
Thanksgiving Week – NO CLASS	
Nov-30 (Lance Waller)	
Class Session	• First hour: Overview of types of data science and AI, and discussion of how these fit into a theory-based setting of behavioral science to reduce drug-related harms. • Second hour: Discussion.
Lab	• There will be no formal lab, instead the lecture will lead to a discussion on if, how, when data science approaches can provide novel insight in behavioral science and substance misuse.
Asynchronous learning	• Review the readings from the preceding session and define connections to behavioral science and reducing drug-related harms.
Readings	• No additional readings for this session.

LABORATORY RESOURCES

1. **Twitter** API pages:
   a. Access: https://developer.twitter.com/en/apply-for-access
   b. API uses: https://developer.twitter.com/en
2. CRAN Project page on rtweet (package for Twitter API on R): https://cran.r-project.org/web/packages/rtweet/rtweet.pdf
3. RBlogger tutorial on rtweet: https://www.r-bloggers.com/2019/10/twitter-data-analysis-in-r/
4. An introduction to analyzing Twitter data with R: https://data.library.virginia.edu/an-introduction-to-analyzing-twitter-data-with-r/.