
CSE 5095: Social Media Mining & Analysis

Course outline, Schedule, Logistics



General Information

CSE 5095	: Social Media Mining & Analysis
Instructor	: Swapna S. Gokhale
Phone	: 6-2772.
Email	: ssg@engr.uconn.edu
Office	: ITEB 237
Lecture time	: Synchronous, Distance Learning Tu/Th 6:30 pm - 7:45 pm https://uconn-cmr.webex.com/meet/ssg02003
Office hours	: Tu/Th 11:00 am - 12:00 pm : Or by appointment, always available by email.
Web page	: HuskyCT (All materials, and general announcements will be posted on the blackboard)

Course Goals

- ❑ Appreciation and motivation for social network mining and analysis as an emerging field that combines machine learning, statistical analysis, natural language processing, and visualization.
 - ❑ Different types of user-generated content on different social media platforms.
 - ◆ Novel ways in which this information can be harnessed.
 - ❑ Statistical analysis and data mining techniques
 - ❑ Various applications of social networking
 - ❑ Risks and negative impacts of social networking
-

◆ Course Goals



Mining - searching for hidden gems, in a deluge of information



Expected Learning Outcomes

- ❑ Learn about the various types of OSNs and their different objectives/purposes.
 - ❑ Learn the issues and challenges in handling data:
 - ◆ Outliers, noise, sampling, duplicates.
 - ◆ Aggregation, transformation, preprocessing.
 - ❑ Understand the several different kinds of information that can be mined from user-generated content on social media.
 - ◆ Sentiment analysis, opinion mining, user characterization and grouping, popularity and rating prediction, political bias, adverse drug reactions, etc.
 - ❑ Learn about the APIs and other tools available to collect this content.
 - ❑ Gain hands-on experience in applying data mining algorithms to real-life social media data.
 - ❑ Cultivate & hone your written and oral communication skills
-



Course Logistics

- Course has three components:
 - ◆ Lectures (by the instructor)
 - ◆ Paper presentations by the students
 - ◆ Project presentations by the students
 - Approximately 1/3-1/3-1/3 split between the three components.
-



Component #1 - Lectures & Assignments

- ❑ Lecture #0 - Schedule and logistics
 - ❑ Lecture #1 - Overview of papers and projects
 - ❑ Lectures #2-#3 -- Overview of social media platforms.
 - ❑ Lectures #4-#7 - Data mining essentials.
 - ◆ Data issues, supervised and unsupervised machine learning, performance evaluation
-



Component #1 - Lectures & Assignments

- ❑ Two types of assignments based on the lectures.
 - ❑ Short, written assignments, open-ended, designed to:
 - ◆ Elucidate and share your experiences about how you use social media, and your level of engagement
 - ◆ Seek to understand and illustrate how social media has permeated all aspects of our lives.
 - ◆ Bring some humor..
 - ◆ Share your reflections and experience about the class.
 - ❑ Small examples to practice and reinforce the basic data mining techniques covered in the class.
 - ◆ Ensure that the students are keeping with how to apply these techniques.
-



Component #2 -- Paper Presentations

- ❑ No standard text book (this is a very new and rapidly evolving field), read a series of current papers on research topics related to Social Media Mining and Analysis and discuss these in the class.
 - ❑ Each student will present one group of papers:
 - ◆ Each group has two related papers.
 - ◆ Papers and abstracts have already been uploaded to HuskyCT, feel free to browse these.
 - ❑ Paper presentations will begin week of Sept. 27, 2021.
 - ❑ Paper presentations are done by individual students.
-



Component #2 - Paper Presentations

- ❑ Non-presenting students are expected to attend/review these presentations.
 - ❑ Each student must prepare a 2-page synopsis of any 5 topics based on their review.
 - ❑ Synopsis should not include the content of the papers, it must critique the papers:
 - ◆ Synopsis must be submitted within a week (on HuskyCT) after the topic is presented.
 - ◆ Presentations will be uploaded right after the in-class session.
 - ◆ Synopsis are to be submitted as "assignments" on HuskyCT, submission for each assignment will close after a week.
 - ◆ Synopsis are to be written individually.
 - ◆ Rubric for the synopsis will also be shared in the next class.
-

Component #3 - Project

- ❑ Students will complete a semester-long project.
 - ◆ Projects will try to gauge online reactions to offline events related to Covid-19 using social media data ("Conversations during Covid")
 - ❑ Project descriptions and preliminary data have been uploaded to HuskyCT.
 - ◆ https://www.dropbox.com/sh/pz3kqm5h2xix6e8/AAATLJt2__WWlMUGGMrUZX_a?dl=0
 - ❑ Projects should be done individually or in groups of 2:
 - ◆ I recommend working in groups of 2, but some students with prior Python or R experience may be able to complete the project on their own.
 - ◆ Unless there are extraordinary circumstances, groups larger than 2 students will not be allowed.
 - ◆ Choose your partners, and your project.
-

Component #3 -- Project

- ❑ “Cookbook” or suggested steps for data analysis are provided in a sample paper; they will also be briefly summarized in the next class.
 - ◆ Each team is expected to complete at least these suggested steps by the end of the semester.
 - ◆ Any additional analysis is always welcome!
 - ◆ About half of the steps should be completed by the midterm.
 - ❑ Project presentations:
 - ◆ Each team will prepare two presentations for their project.
 - ◆ Midterm presentation long (time -- November).
 - ◆ Final presentations (during the last week or the exam week).
 - ◆ Midterm presentations interleaved with paper presentations.
 - ❑ Tentative dates for midterm project presentations are on the schedule, final presentations TBD
-

Component #3 - Project

- ❑ Written reports -- Midterm and Final Report.
 - ❑ Midterm Report - Due week of Nov. 8:
 - ◆ Midterm report should include progress so far, and any obstacles necessary for mid-course correction
 - ❑ Final Report - Due during the final exam week:
 - ◆ Final report should include the results of the analysis, key findings and observations.
 - ◆ Final report should also include a comparison with related work.
 - ◆ In other words, it should be more or less “publication ready”.
 - ◆ No final exam, or midterm for this class.
 - ◆ Most projects from past two year's class (with some extra work) have been submitted/accepted for publication.
 - ❑ Code, data etc:
 - ◆ Submit the code, along with the final report.
-



Course Schedule (Tentative, Subject to Change)

Week #1 (Aug. 31):

1. Aug. 31: Schedule & Logistics.
2. Sept. 2: Paper & Project logistics

Week #2 (Sept. 6):

3. Sept. 7: Introduction & Motivation
4. Sept. 9: Introduction & Motivation (contd..)

Week #3 (Sept. 13):

5. Sept. 14: Data Mining Essentials
6. Sept. 16: Data Mining Essentials

Week #4 (Sept. 20):

7. Sept. 21: Data Mining Essentials (contd..)
8. Sept. 23: Data Mining Essentials (contd..)

Week #5 (Sept. 27):

9. Sept. 28: Student paper presentations
 10. Sept. 30: Student paper presentations
-



Course Schedule (Tentative, Subject to Change)

Week #6 (Oct. 4):

12. Oct. 5: Student paper presentations

12. Oct. 7: Student paper presentations

Week #7 (Oct. 11):

13. Oct. 12: Student paper presentations

14. Oct. 14: Guest lecture

Week #8 (Oct. 18):

15. Oct. 19: Student paper presentations

16. Oct. 21: Student paper presentations

Week #9 (Oct. 25):

17. Oct. 26: Student paper presentations

18. Oct. 28: Student paper presentations

Week #10 (Nov. 1):

19. Nov. 2: Midterm project presentations

20. Nov. 4: Midterm project presentations



Course Schedule (Tentative Subject to Change)

Week #11 (Nov. 8):

21. Nov. 9: Midterm project presentations

22. Nov. 11: Midterm project presentations

(Midterm Report Due)

Week #12 (Nov. 15):

23. Nov. 16: Midterm project presentations

24. Nov. 18: Midterm project presentations

Week #13 (Nov. 22):

Thanksgiving break

Week #14 (Nov. 29):

25. Nov. 30: Student paper presentations

26. Dec. 2: Student paper presentations

Week #15 (Dec. 6):

27. Dec. 7: Student paper presentations

28. Dec. 9: Student paper presentations



Schedule of Assignments (Tentative)

No.	Assignment Topic	Assigned	Due	Weight
0.	Getting to Know You.	08/31	09/01	0%
1.	Selecting Papers & Projects	09/02	09/09	0%
2.	Social Media Engagement	09/07	09/14	3%
3.	Philosophical Issues	09/14	09/21	2%
4.	TF-IDF Example	09/14	09/15	1%
5.	Decision Tree Example	09/16	09/17	1%
6.	Naïve Bayes Example	09/21	09/22	1%
7.	Clustering Example	09/23	09/24	1%
8.	Assessment (Self & Peer)	12/09	12/16	1%
	Total			10%



Grade Rubric

Paper presentation(s)	:20%
Paper synopsis	:10% (5, 2% per synopsis)
Assignment(s)	: 10%
Project	: 60%
-- Midterm Pres.	: 15%
-- Midterm Report	: 15%
-- Final Pres.	: 5%
-- Final Report	: 25%



Past Publications

1. M. Voong, K. Gunda, Predicting the Political Polarity of Tweets Using Supervised Machine Learning, *Proc. of 2020 IEEE 44th Annual Computers, Software, and Applications Conference* , pp. 1707-1712, July 2020.
 2. A. Mondal, S. Gokhale, Mining Emotions on Plutchik's Wheel, *Proc. of 2020 Seventh International Conference on Social Networks Analysis, Management and Security* , December 2020.
 3. H. Khan, F. Yu, A. Sinha, S. Gokhale, *A Practical and Parsimonious Approach to Detecting Offensive Speech*, *Proc. of Intl. Conf. on Computing, Communications, and Intelligent Systems* , February 2021.
 4. A. Mondal, S. Gokhale, Multi-Label Classification of Parrott's Emotions, *Proc. of Intl. Conf. on Software Engineering and Knowledge Engineering*, July 2021.
-

Past Publications (contd..)

5. N. Paul, S. Gokhale, Analysis and Classification of Vaccine Dialogue in the Coronavirus Era, *Proc. of IEEE Big Data Workshop on Smart and Connected Health*, December 2020.
 6. S. Gokhale, Monitoring the Perception of Covid-19 Vaccine using Topic Models, *Proc. of 13th IEEE Intl.Symposium on Social Computing and Networking* , December 2020.
 7. L. Cerbin, J. DeJesus, J. Warnken and S. Gokhale, Unmasking the Mask Debate on Social Media, *Proc. of 2021 IEEE 45th Annual Computers, Software, and Applications Conference*, July 2021.
 8. H. Nguyen, J. Moon, B. Pines, S. Gokhale, Detecting Offensive Content on Social Media During Anti-Lockdown Protests in Michigan, *Proc. of 2021 IEEE 45th Annual Computers, Software, and Applications Conference*, July 2021.
-



Past Publications (contd..)

9. J. Moon, H. Nguyen, S. Gokhale, Sarcasm Detection in Politically Motivated Protests, *Proc. of IEEE Social Computing Conference, October 2021*.
 10. M. Fahim and S. Gokhale, Detecting Offensive Content on Twitter During Proud Boys Riots, *Proc. of IEEE Intl. Conf. on Machine Learning and Applications (under review)*.
 11. L. Cerbin, J. DeJesus, J. Warnken and S. Gokhale, Understanding the Anti-Mask Debate On Social Media, *Intl. Journal of Computers and their Applications (under review)*.
 12. J. Moon, H. Nguyen, B. Pines, S. Gokhale, Identifying Anti-Government Sentiment on Twitter During Michigan Protests, *Proc. of DEVIANCE 2021 (under review)*.
 13. J. Furuness, A. Mendonca, Identifying Myers-Briggs Personality Type by Analyzing Tweets, *(in preparation)*
-