<u>Computational Social Science</u>
<u>Ponnurangam Kumaraguru</u>

Name of the Program **:** Applicable to all Programs on campus including, CSE, CLD, CHD, CND, both at UG & Masters level.

Course Code **: CS9.435**

Credits **: 4**

L - T - P **:** 3-0-1
(L - Lecture hours, T-Tutorial hours, P - Practical hours)

Semester, Year **: Spring, 2022**
(Ex: Spring, 2022)

Pre-Requisites **:** Any UG3, UG4, M.Tech., MS, and Ph.D. student should be able to take it

Course Outcomes **:**

- C0-1: Students will describe the opportunities and challenges that the digital age creates for social sciences research.
- C0-2: Students will evaluate modern social research from the perspectives of both social science and data science.
- C0-3: Students will create research proposals that blend ideas from social science and data science.
- C0-4:Students will be able to summarize and critique research papers in Computational Social Science
- C0-5: Students will conduct, develop, and practice the techniques needed to conduct their proposed research, through course project.

Course Topics **:**
(please list the order in which they will be covered, and preferably arrange these as five to six modules.)

Module 1: Social Research

- Computational Social Science 101
  - What is Computational Social Science?
  - Is Computational Social Science = or ≠ Computer Science + Social Science?
  - Why study Computational Social Science?
  - Challenges with only Computer Science or Social Science
  - Does Social Media data == Computational Social Science? Class debate.
- Social Science vs. Data Science
- Prediction vs. Causality

Read / Listen / Watch:

- Hanna Wallach. 2018. Computational social science ≠ computer science + social data. Commun. ACM 61, 3 (March 2018), 42–44. DOI:https://doi.org/10.1145/3132698 https://cacm.acm.org/magazines/2018/3/225484-computational-social-science-computer-science-social-data/fulltext

- Lazer D, Pentland A, Adamic L, Aral S, Barabasi AL, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M. Social science. Computational social science. Science. 2009 Feb 6;323(5915):721-3. doi: 10.1126/science.1167742. PMID: 19197046; PMCID: PMC2745217. https://pubmed.ncbi.nlm.nih.gov/19197046/
- Coded Bias
    - Trailer https://youtu.be/jZl55PsfZJQ
    - Full documentary https://www.netflix.com/title/81328723

Module 2: Modeling & Causal Inference

- Linear Regression, Model building, Hypothesis testing
- Causal Inference
- Running Experiments – Lab, Real-world
- Read / Listen / Watch:
    - Blumenstock et al. 2015. Predicting Poverty and Wealth from Mobile Phone Metadata. Science. https://www.unhcr.org/innovation/wp-content/uploads/2016/11/blumenstock-science-2015.pdf
    - Lazer, David and Kennedy, Ryan and King, Gary and Vespignani, Alessandro, Google Flu Trends Still Appears Sick: An Evaluation of the 2013-2014 Flu Season (March 13, 2014). Available at SSRN: https://ssrn.com/abstract=2408560 or http://dx.doi.org/10.2139/ssrn.2408560
    - M. R. Khan, J. Manoj, A. Singh and J. Blumenstock, "Behavioral Modeling for Churn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty," 2015 IEEE International Congress on Big Data, 2015, pp. 677-680, doi: 10.1109/BigDataCongress.2015.107.
    - Chapter 3 of Mostly Harmless Econometrics: An Empiricist Companion

Module 3: Mass Collaborations

- Human Computation
    - Galaxy Zoo
        - Lintott, Chris J, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, et al. 2008. “Galaxy Zoo: Morphologies Derived from Visual Inspection of Galaxies from the Sloan Digital Sky Survey.” *Monthly Notices of the Royal Astronomical Society* 389 (3):1179–89. https://doi.org/10.1111/j.1365-2966.2008.13689.x https://watermark.silverchair.com/mnras0389-1179.pdf
        - Kuminski, Evan, Joe George, John Wallin, and Lior Shamir. 2014. “Combining Human and Machine Learning for Morphological Analysis of Galaxy Images.” *Publications of the Astronomical Society of the Pacific* 126 (944):959–67. https://doi.org/10.1086/678977.
    - Crowd-coding of political manifestos
        - Benoit, Kenneth, Drew Conway, Benjamin E. Lauderdale, Michael Laver, and Slava Mikhaylov. 2016. “Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data.” *American Political Science Review* 110 (2):278–95. https://doi.org/10.1017/S0003055416000058 https://kenbenoit.net/pdfs/Crowd_sourced_data_coding_APSR.pdf
- Open Calls
    - Netflix Prize

        - Netflix. 2009. “Netflix Prize: View Leaderboard.” http://www.netflixprize.com/leaderboard.
        - Bell, Robert M., Yehuda Koren, and Chris Volinsky. 2010. “All Together Now: A Perspective on the Netflix Prize.” *Chance* 23 (1):24–24. https://doi.org/10.1007/s00144-010-0005-2.
    - Foldit: Protein-folding game
        - Hand, Eric. 2010. “Citizen Science: People Power.” *Nature News* 466 (7307):685–87. https://doi.org/10.1038/466685a.
- Distributed Data collection
    - eBird: Bird data from birders
        - Kelling, Steve, Daniel Fink, Frank A. La Sorte, Alison Johnston, Nicholas E. Bruns, and Wesley M. Hochachka. 2015. “Taking a Big Data Approach to Data Quality in a Citizen Science Project.” *Ambio* 44 (Suppl 4):601–11. https://doi.org/10.1007/s13280-015-0710-4.
    - Photocity
        - Tuite, Kathleen, Noah Snavely, Dun-yu Hsiao, Nadine Tabing, and Zoran Popovic. 2011. “PhotoCity: Training Experts at Large-Scale Image Acquisition Through a Competitive Game.” In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, 1383–92. CHI ’11. New York: ACM. https://doi.org/10.1145/1978942.1979146 https://dl.acm.org/doi/pdf/10.1145/1978942.1979146
        - Agarwal, Sameer, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, and Richard Szeliski. 2011. “Building Rome in a Day.” *Communication of the ACM* 54 (10):105–12. https://doi.org/10.1145/2001269.2001293.
- How to develop our own (including around course project) Mass Collaborations?
    - Opportunities
    - Methods
    - Challenges

Module 4: Ethics
- Studies of concern
    - Experiment on 700,000 Facebook users
        - Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. “Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks.” *Proceedings of the National Academy of Sciences of the USA* 111 (24):8788–90. https://doi.org/10.1073/pnas.1320040111
    - Tastes, Ties, and Time study on Facebook users
        - Wimmer, Andreas, and Kevin Lewis. 2010. “Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook.” *American Journal of Sociology* 116 (2):583–642. http://www.jstor.org/stable/10.1086/653658.
        - Lewis, Kevin, Marco Gonzalez, and Jason Kaufman. 2012. “Social Selection and Peer Influence in an Online Social Network.” *Proceedings of the National Academy of Sciences of the USA* 109 (1):68–72. https://doi.org/10.1073/pnas.1109739109.
    - Web Censorship

- Burnett, Sam, and Nick Feamster. 2015. "Encore: Lightweight Measurement of Web Censorship with Cross-Origin Requests." In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, 653–67. SIGCOMM '15. London: ACM. https://doi.org/10.1145/2785956.2787485 https://dl.acm.org/doi/pdf/10.1145/2785956.2787485
- Narayanan, Arvind, and Bendert Zevenbergen. 2015. "No Encore for Encore? Ethical Questions for Web-Based Censorship Measurement." *Technology Science*, December. http://techscience.org/a/2015121501/.
- Jones, Ben, and Nick Feamster. 2015. "Can Censorship Measurements Be Safe(R)?" In *Proceedings of the 14th ACM Workshop on Hot Topics in Networks*, 1:1–1:7. HotNets-XIV. New York: ACM. https://doi.org/10.1145/2834050.2834066.

- Crime prediction using Social data, Tracking immigrants through their phone apps
- Institutional Review Board / Ethics Committee – Expectations, Why is it necessary?
- Informed consent, Privacy, Risk

Module 5: Biases in CSS Research

- Biases & inaccuracies at the source of the data
- Biases & inaccuracies during processing
- Biases in social data
- Inferences from biased data
- Read / Listen / Watch:
  - danah boyd & Kate Crawford (2012) CRITICAL QUESTIONS FOR BIG DATA, Information, Communication & Society, 15:5, 662-679, DOI: 10.1080/1369118X.2012.678878
  - Daniel Gayo-Avello. 2011. Don't turn social media into another 'Literary Digest' poll. Commun. ACM 54, 10 (October 2011), 121–128. DOI:https://doi.org/10.1145/2001269.2001297
  - McFarland DA, McFarland HR. Big Data and the danger of being precisely inaccurate. Big Data & Society. December 2015. doi:10.1177/2053951715602495
  - https://en.wikipedia.org/wiki/Weapons_of_Math_Destruction
  - Ricardo Baeza-Yates. 2018. Bias on the web. Commun. ACM 61, 6 (June 2018), 54–61. DOI:https://doi.org/10.1145/3209581
  - Hargittai E. Potential Biases in Big Data: Omitted Voices on Social Media. Social Science Computer Review. 2020;38(1):10-24. doi:10.1177/0894439318788322
  - Olteanu Alexandra, Castillo Carlos, Diaz Fernando, Kıcıman Emre. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Frontiers in Big Data. VOLUME 2, 2019 DOI=10.3389/fdata.2019.00013 https://www.frontiersin.org/article/10.3389/fdata.2019.00013

Preferred Text Books :

1. Salganik, Matthew J., Bit by Bit: Social Research in the Digital Age, Princeton University Press, 2018. Free Online Version: https://www.bitbybitbook.com/en/1st-ed/preface/

Reference Books :

E-book Links :

https://www.bitbybitbook.com/en/1st-ed/preface/

https://www.nature.com/collections/cadaddgige/
https://www.researchgate.net/profile/Joshua-Angrist/publication/51992844_Mostly_Harmless_Econometrics_An_Empiricist%27s_Companion/links/00b4953344a9a0cb13000000/Mostly-Harmless-Econometrics-An-Empiricists-Companion.pdf

Grading Plan :
(The table is only indicative)

Type of Evaluation	Weightage (in %)
Quizzes	20.0
Assignments + Activities	30.0
Project report + Blog + Video	20.0 [15 + 2.5 + 2.5]
Project	30.0
Other Evaluation __________	

Mapping of Course Outcomes to Program Objectives: (1 – Lowest, 2—Medium, 3 – Highest, or a '-' dash mark if not at all relevant).

	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO10	PO11	PO12	PSO 1	PSO 2	PSO 3	PSO 4
CO 1	3	1	-	-	-	3	-	-	-	3	-	1	3	-	-	3
CO 2	3	1	-	-	3	3	-	-	3	3	-	1	3	-	-	3
CO 3	3	3	-	-	-	3	1	3	3	3	-	1	3	3	-	3
CO 4	3	1	-	-	-	-	-	-	3	3	-	1	3	1	-	3
CO 5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Teaching-Learning Strategies in brief (4-5 sentences) :

Learning
- Lectures
- Reading research papers
- Class participation: questions, discussions

- Online discussion: Teams

Learning by doing

- Course project
- Real world issues
- Interdisciplinary approach
- Real world implementation

POTENTIAL GUEST LECTURES:

1. Prof. Mathew Salganik, Princeton University
2. (Soon to be Dr.) Ashwin Rajadesingan, University of Michigan
3. Dr. Hemank Lamba, Dataminr

==============

Note: This course description format comes into effect from Spring 2022.