DATASCI W210

Synthetic Capstone

Instructors: Stanislav Kelman, Joyce Shen, David Steier, Alberto Todeschini, Puya Vahabi

Course Description

In the capstone class, students combine technical, analytical, interpretive, and social dimensions to design and execute a full data science project, developing their skills as data scientists with a focus on real-world applications and situations. The final project provides an opportunity to integrate all of the core skills and concepts learned throughout the program and prepares students for long-term professional success in the field. It provides experience in formulating and carrying out a sustained, coherent, and impactful course of work resulting in a tangible data science project using real-world data. Students are evaluated on their ability to collaboratively develop and communicate their work in both written and oral form. The capstone is completed as a group project, and each project will focus on open, pre-existing, secondary data. There are relevant readings and case discussions throughout. We provide many real-world examples and perspectives, including through panel discussions with leading data science experts and practitioners.

Course Summary

This course has three main goals:

1. Practice using data science to drive significant impact by solving challenging, valuable problems (i.e., by creating new data-driven capabilities or by influencing decisions that can significantly improve mission-critical results)

2. Reinforce critical problem-solving, communication, influencing, and management skills

3. Engage in a process of teamwork, feedback, and iteration that mirrors the challenges and opportunities of applying data science in a realistic organizational setting

We begin by covering fundamentals for effective execution, including a review of the "top-management perspective," problem structuring, selecting tools/approaches, work planning, and managing ethical and legal considerations. We then broaden our focus to leading others, managing teams, communication/presentation fundamentals, influencing in organizations, and leveraging data visualization. We conclude with advanced topics like institutionalizing data-driven decision-making, building products, and perspectives on the future of data science.

Prerequisites: Students must have completed (or are completing during the same semester) all core courses.

Course Assignments and Final Exam

The primary assignments for this course are three group presentations to the class along with a final web-based deliverable for the group project. In addition to these assignments, students are evaluated based on their participation in class discussions, and by their group-mates based on contributions to the group.

- **55% - Participation** - class discussions and group project participation
- **25% - In-class presentations** - instructor evaluation and peer feedback on presentations
- **20% - Web-based final deliverable**

Course Data Sets, Software, and Tools

This course allows students wide discretion for selecting problems, analyses, data sets, software, and tools for use on their group projects.

Students may select any problem and approach that meets their interests, affords opportunities for truly impactful results, and is feasible within the timeframe of this course. Students may select any open data sets (publicly available, unrestricted) that suit their focal problem/analysis, and may use any software or tools that are appropriate for meeting their goals. Students must present and defend their choices across these dimensions throughout the course.

Students are required to thoroughly document their analyses. They are also strongly encouraged to build an end-to-end analysis pipeline covering data sourcing, cleaning/preparation, transformation, processing, visualization, interpretation, etc.

Readings and Course Pack

There is no textbook for this course. Readings are drawn from various relevant books, articles, reports, and academic papers and are made available either in the course pack or online.

Syllabus

Week 1: Intro to Capstone

Topics
- Why, when, and how data science matters

- What it means to be a data scientist

- Philosophy and mindset for data science work

- Ways in which data science is driving impact over time, along with examples

- Project and group selection (Part 1 of 2)

Required Reading

- McAfee, A. Brynjolfsson, E. Big Data: The Management Revolution. *Harvard Business Review.*

    Retrieved from https://hbr.org/2012/10/big-data-the-management-revolution/ar

- Provost & Fawcett. (2013). Chapter 1-2. In *Data Science for Business*. O'Reilly

- Schutt & O'Neil. (2013). Chapters 1-2, 15-16. In *Doing Data Science*. O'Reilly

- LaValle, S. Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*.

- Press, G. (2013). A very short history of data science. *Forbes*. Retrieved from

    http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/

Optional Reading

- Zumel & Mount. (2014). Chapter 1. In *Practical Data Science with R*. Manning.

- Gozman, D., Currie, W., Seddon, J. (2015). The Role Of Big Data In Governance: A Regulatory And

    Legal Perspective Of Analytics In Global Financial Services. SWIFT Institute. Retrieved from http://www.swiftinstitute.org/wp-content/uploads/2015/12/SIWP-2014-009-The-Role-of-Big-Data-in-Governance-FINAL.pdf

- Phys.org. *Scientist helps move structural biology into 'big data' era*. (2016). Available from

    http://phys.org/news/2016-03-scientist-biology-big-era.html

Week 2: Driving Impact

Topics
- Case example of an organization trying to use "big data" in an effective way

- The top-management perspective and its ramifications on our work

- Techniques for structured problem solving

- Methodologies for tackling data science analysis

- Project and group selection (Part 2 of 2)

Required Reading
- Chandrasekhar, R., & Parker, S. (2014). *Luminar: Leveraging big data using corporate entrepreneurship*. Richard Ivey School of Business Foundation, Version: 2014-05-20.

- Higgins, R., O'Donnell, P., Bhatt, M. (2012). Kyruus: Big Data's Search for the Killer App. Harvard Business School Case 813060-PDF-ENG.

- Davenport, T., Kim, J. (2013). Chapter 2. In *Keeping up with the Quants*. HBR Press.

Optional Reading
- Minto, B. (1996). *The Minto pyramid principle: Logic in writing, thinking, & problem solving*. Norfolk, VA: Minto Publishing, Inc.

- Gabarro, J. J., & Kotter, J. P. (1993). Managing your boss. *Harvard Business Review,* Reprint 93306. Retrieved from http://www.physicianleadership.com/wp/wp-content/uploads/psc/ManagingYourBoss.pdf

Week 3: Product Principles

Topics
- Overview of typical data-enabled products

- Iterative, learning-oriented approach to product development

- Tips for scrappy design, user testing, and roadmapping

- Case examples of iterative product development

Required Reading
- Eisenmann, T., Wagonfeld, A. B., & Barley, L. (2012). Aardvark. Harvard Business School Case 811-064, Revised November 8, 2012.

- Rodden, K. (2015). How to choose the right UX metrics for your product *Google Ventures*. Retrieved from https://library.gv.com/how-to-choose-the-right-ux-metrics-for-your-product-5f46359ab5be

- Croll & Yoskovitz, Analytics Lessons learned, 2013, available at www.leananalyticsbook.com

- Armstrong, G., & Kotler, P. (2015). Chapters 1 & 2. *Marketing: An Introduction*. Essex, England: Pearson Education Limited.

Optional Reading
- Ries, E. (2013). *The Lean Startup*. New York, NY: Crown Business.

- Patil, D. J. (2012). *Data jujitsu: The art of turning data into product*. O'Reilly Radar Report. Available from http://www.oreilly.com/data/free/data-jujitsu.csp

- Kelley, T., & Kelley, D. (2013). Introduction; Chapters 2, 3, 6. *Creative confidence: Unleashing the creative potential within us all*. New York, NY: Crown Business.

- Kagan, M. (2008). In *Inspired: How to create products customers love*. Sunnyvale, CA: SVPG Press.

- Brooks, F. (1975). Chapter 2. In *The mythical man month: Essays on software engineering*. Boston, MA: Addison-Wesley.

Week 4: Ethical Considerations

Topics
- Ethical and legal considerations in data science

- Mitigating the effects of bias (cognitive and algorithmic)

- Recent examples of ethical issues in data science (including privacy and bias)

- Identifying and managing ethical and legal risks in projects

Required Reading
- Kahneman, D. (2011). Chapters 1-3, 21, and 34. In *Thinking, fast and slow* (2nd ed.). New York, NY: Farrar, Straus and Giroux.

- Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016). Machine Bias. *ProPublica.* Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

- Vanian, J. (2018). Solving AI's Bias Problem. *Fortune.* Retrieved from http://fortune.com/longform/ai-bias-problem/

- Nissenbaum, H. (2011). A Contextual Approach to Privacy Online. *Dædalus, the Journal of the American Academy of Arts & Sciences.* Retrieved from https://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf

Optional Reading
- Courtland, R, "Bias detectives: the researchers striving to make algorithms fair", *Nature*, June 20, 2018, https://www.nature.com/articles/d41586-018-05469-3

Week 5: Group Presentation 1

Topics
- Deliver 1st presentation

Required Reading
- None

Optional Reading
- None

Week 6: Selecting and Exploring Data Sources

Topics

- Selecting data sources

- Exploratory data analysis

- Preparing data for modeling

Required Reading

- Mawer, et. al "The Value of Exploratory Data Analysis," March 2017, https://svds.com/value-exploratory-data-analysis/

- Groskopf, "What to Do About Bad Data," 2016, https://github.com/Quartz/bad-data-guide

Optional Reading

- Bourke, "A Gentle Introduction to Exploratory Data Analysis," January 2019, https://towardsdatascience.com/a-gentle-introduction-to-exploratory-data-analysis-f11d843b8184

- Yazdani, "Using Exploratory Data Analysis to Discover Patterns in Image and Document Collections," August 2016, https://pyvideo.org/pydata-chicago-2016/using-exploratory-data-analysis-to-discover-patterns-in-image-and-document-collections.html   (37 min video and slides)

Week 7: Communication and Storytelling

Topics
- Implications of top-management perspective on effective communication

- Power of storytelling and narrative, including how we can leverage it in our work

- Tactics for presenting and sharing information

- Adjustments to project approach/plan following first presentation (as needed)

Required Reading

- Heath, C., & Heath, D. (2007). Introduction; Chapters 2, 5, 6 Epilogue. In *Made to stick: Why some ideas survive and others die*. New York, NY: Random House.

- Zumel, N., Mount, J. (2014). Chapter 11. In *Practical Data Science with R*. 2014. Manning.

- Gilovich, T. (1991). Chapter 6, Believing what we are told. In *How we know what isn't so: The fallibility of human reason in everyday life*. New York, NY: Free Press.

Optional Reading

- Zelazny, G. (2001). *Say it with charts: The executive's guide to visual communication* (4th ed.). New York, NY: McGraw Hill.

- Peruse the following:

  - http://www.visual-literacy.org

  - http://www.informationisbeautiful.net/

  - http://www.economist.com/blogs/graphicdetail

  - http://www.nytimes.com/interactive/2014/12/29/us/year-in-interactive-storytelling.html

Week 8: Influencing as a Data Scientist

Topics

- How people and organizations process information and make decisions

- Examples of common cognitive biases and their implications on decision making

- Tactical guidance for effective influencing and consensus building

Required Reading

- Gilovich, T. (1991). Chapter 2, Something out of nothing. In *How we know what isn't so: The fallibility of human reason in everyday life*. New York, NY: Free Press.

- Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sSequences. *Cognitive Psychology 17*, 295-314. Retrieved from http://psych.cornell.edu/sites/default/files/Gilo.Vallone.Tversky.pdf

Optional Reading

- Ramachandran, V. S. (2012). Chapters 1–3. In *The tell-tale brain: A neuroscientist's quest for what makes us human*. New York: NY: W. W. Norton & Company.

- Bhatia, A. (2012). What does randomness look like? *Wired.* Retrieved from http://www.wired.com/2012/12/what-does-randomness-look-like/

Week 9: The Data-Driven / Data-Informed Organization

Topics
- Ways in which data science can be organized and delivered

- Examples of data-driven organizations and their unique advantages/capabilities

- Challenges to building a data-driven organization, along with potential solutions

- The role of culture, along with important values to establish and reinforce

- Adjustments to project approach/plan following second presentation (as needed)

Required Reading
- Patil, D. J., & Mason, H. (2015). *Data driven*. Available from http://www.oreilly.com/data/free/data-driven.csp .

- Economist Intelligence Unit. (2013). *Fostering a data-driven culture*. Available from http://www.economistinsights.com/technology-innovation/analysis/fostering-data-driven-culture/fullreport

- Khalil, E., & Wood, K. (2014). Aligning data science—Making organizational structure work. Booz Allen Hamilton. Available from https://www.boozallen.com/s/insight/publication/field-guide-to-data-science.html

- Novet, J. (2014). Why LinkedIn's data science reorg actually makes a lot of sense. VentureBeat. Retrieved from http://venturebeat.com/2014/11/03/linkedin-data-scientists-analysis/

Optional Reading
- None

Week 10: Group Presentation 2

Topics
- Deliver 2nd presentation

Required Reading
- None

Optional Reading
- None

Week 11: General Organizational Behavior and Leadership

Topics

- Challenges and opportunities for organizations undergoing major changes

- Ways in which organizations structure behavior (including incentives and culture)

- Approaches for architecting and steering organizational change in a responsible way

Required Reading

- Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management Journal, 18.* Retrieved from http://www.ou.edu/russell/UGcomp/Kerr.pdf

- Singer, E. (2015). Game theory calls cooperation into question. *Scientific American.* Retrieved from http://www.scientificamerican.com/article/game-theory-calls-cooperation-into-question1/

- Collins, J. C., & Porras, J. I. (1994). Cult-like cultures. In *Built to last: Successful habits of visionary companies*. New York, NY: Harper Collins.

Optional Reading

- Senge, P. (1990). The leader's New Work: Building learning organizations. *MIT Sloan Management Review.* Retrieved from http://sloanreview.mit.edu/article/the-leaders-new-work-building-learning-organizations/

- Erickson, T., & Gratton, L. (2007). What it means to work here. *Harvard Business Review.* Retrieved from https://hbr.org/2007/03/what-it-means-to-work-here/ar/1?referral=00134

- Pfeffer, J. (1998). Six dangerous myths about pay. *Harvard Business Review.* Retrieved from https://hbr.org/1998/05/six-dangerous-myths-about-pay

- Terkel, S. (1997). *Working: People talk about what they do all day and how they feel about what they do* (2nd ed.). Fort-Meyers, FL: The News Press.

- Dixit, A. (1993). *Thinking strategically: The competitive edge in business, politics, and everyday life* (2nd ed.). New York, NY: W. W. Norton & Company.

- Bower, J. L., & Christensen, C. M. (1995). Disruptive technologies: Catching the wave. *Harvard Business Review.* Retrieved from https://hbr.org/1995/01/disruptive-technologies-catching-the-wave

Week 12: Possible Futures of Data Science

Topics

- Recent data science–related trends and possible futures

- Perspectives on the future of data science in terms of technology, application areas, and the profession

Required Reading

- Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., & Marrs, A. (2013). Disruptive technologies: Advances that will transform life, business, and the global economy. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/insights/business_technology/disruptive_technologies . (Executive Summary)

- Brynjolfsson, E., & McAfee, A. (2012). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Research Brief. Retrieved from http://ebusiness.mit.edu/research/Briefs/Brynjolfsson_McAfee_Race_Against_the_Machine.pdf

- Simonite, T. (2015). Automating the data scientists. *MIT Technology Review*. Retrieved from

  http://www.technologyreview.com/news/535041/automating-the-data-scientists/

Optional Reading

- Skim: http://www.wolframalpha.com/docs/timeline/

Week 13: Wrap-Up

Topics

- Recap and review

- Goodbye and parting thoughts

Required Reading

- None

Optional Reading

- None

Weeks 14 and 15: Final Group Presentations and Deliverables

Topics

- Prepare for third and final presentation.

- Deliver final presentation and submit project deliverable.

Required Reading

- None

Optional Reading

- None