



15.572 ANALYTICS LAB

ACTION LEARNING SEMINAR ON ANALYTICS, MACHINE LEARNING & THE DIGITAL ECONOMY

Instructors	Professor Sinan Aral sinan@mit.edu ; URL: http://web.mit.edu/sinana/www/ Office hours by appointment
Class Times	Thursdays 4:00-5:30pm, E62-276 Special Sessions: 4:00-8:00pm, September 13 , Samberg Conference Center, 7 th Floor (E52) 2:00-8:00pm, December 7 , Bartos Theater (E15)
TA	Prateek Kukreja pkukreja@mit.edu
Administrator	Susan Young susany@mit.edu

Summary and Objectives:

The growth in big data and analytics is transforming management decision-making, operations, marketing, finance, and product innovation. Businesses across the world are wrestling with challenges and opportunities that call for the application of analytics. We are on the cusp of a second machine age – a digital era that holds opportunities and challenges for both individuals and the economy. Workers and professionals in all fields are racing to acquire the skills and capabilities necessary to survive and thrive in this digital revolution.

The purpose of the Analytics Lab (A-Lab) is to match student teams with leading-edge projects involving analytics, machine learning or digital technologies as they apply to business questions and problems. The particular focus of the projects is on the technical and analytical aspects, but business relevance sets the context and nature of the technical problem.

Course Principles and Expectations:

The primary criterion for projects is to provide a learning experience for the students. In addition, the projects should be of high relevance and interest to a particular organization and senior managers and professionals in it.

Project teams of three to four students are expected to work independent of regular class meetings. Project sponsoring organizations will cover costs of travel and lodging, if any. Each project team will have an MIT-associated faculty or research mentor to provide guidance and assistance and a link to outside project sponsors on an as-needed basis.

Two special sessions are scheduled: Pitch Day on September 13 and Final Presentations on December 7. Attendance at both sessions is required.

Notes on Class Activities and Due Dates:

- 9/13: On Pitch Day (4:00-8:00pm, Samberg Conference Center, 7th Floor), we will meet jointly with the representatives from project proposing companies. Each will briefly describe their project as proposed, and students will have an opportunity meet and informally mix with them and fellow students. The session will be followed by a reception. The chief aim of this session is to help inform student team formation and project selection.
- 9/16, 11:59pm: **DUE: Project Ranks.** Each student should complete the survey separately (link to follow). In the following days, faculty, mentors, and the course support team will work out assignments of projects to students/teams, subject to review by the proposing company.
- 9/20: Final team-project pairings will be communicated to students. MIT and every proposing company have executed a jointly signed NDA. Each student team member will be required to review and sign an acknowledgment stating that all will abide by the terms agreed to in the NDA. Additional information will follow from Ellen Baum.
- 9/30, 11:59pm: **DUE: Project plan.** Each team should submit one document to their mentor and Prateek. The project plan is to be developed by the team, reviewed by the team's mentor, and endorsed by the project sponsor before the deadline. It should be thought of as a working document, used by the team and mentor to assess progress and adjust and adapt through the semester. Although there is no hard and fast template, here is a suggested outline:
 - Purpose and Scope: The project purpose and scope should serve as a compass that guides the team throughout the duration of the project. It must reflect the company proposal, but should be more focused. Remember, the project is intended to be a rich learning experience for your team; this project is NOT a consulting engagement with the project sponsor. Bear this in mind when drafting your project purpose and scope
 - Objectives: Break the project down into high-level objectives that you intend to achieve
 - Tasks: For each objective, list one or more granular tasks. For each task, define the following: due date, deliverables, who is responsible, current status, etc.
- 10/11, 10:00am: **DUE: Mid-term presentation slides.** Each team should submit their slides to their mentor and Prateek.
- 10/11 & 10/18, teams will deliver 3-minute presentations on their project work to date and potential lines of future analysis. The chief aim of these sessions is to help illuminate issues common across teams in order to foster discussion and collaboration.
- 12/7, 10:00am: **DUE: Final presentation slides.** Each team should submit their slides to their mentor and Prateek.
- 12/7: During the Final Presentations session (2:00-8:00pm, Bartos Theater), each team will present their project work to an audience of experts, entrepreneurs, and executives, including representatives from project sponsoring organizations, as well as a few "celebrity judges". Teams will have 4 minutes to present their project work, plus 2 minutes for Q&A and judge remarks (6 minutes total per team). See the "Grading" section below for judge evaluation criteria.

- 12/16, 11:59pm: **DUE:**
 - **Final report** (10 pages maximum, 3000 words, not including figures or references); report should consider feedback received during final presentations on 12/7. Each team should submit one document to their mentor and Prateek
 - **Summary of findings** (one page executive summary); summary should include a high-level statement of the challenge posed by the project sponsor and the insights the team generated during the semester.

Please note that teams are required to share your final presentation slides, final report, and summary of findings with project sponsors with enough lead time for them to review for inadvertent disclosure of Confidential Information.

Grading:

- 40% Final presentation content and delivery – team-wide; presentations will be evaluated according to the following criteria:
 - **Technical/Analytical:** How creative or advanced were the techniques used?
 - **Level of Effort:** How much improvement was sought / how many different techniques were attempted before the team selected the one that seemed best?
 - **Business Impact:** Beyond the impressive data analytics described, how clearly did the team convey the bottom-line, real-world impact of their findings?
 - **Presentation:** How informative and interesting was the presentation itself and how well was it delivered?
- 20% Final report – team-wide
- 15% Summary of Findings – team-wide
- 15% Contribution to class discussions and team project enablement – individual
 - Independently evaluated by instructor, mentors, and team members
- 10% Mid-point presentation content and delivery – team-wide

Required Book:

Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking, Foster Provost and Tom Fawcett. 2013. O'Reilly Media Inc. (Online access available at <http://library.mit.edu/item/002221893>)

All other required readings are freely available on the course Stellar site: <https://stellar.mit.edu/S/course/15/fa17/15.572/>

Data Destruction:

The following states the MIT Action Learning Office's policies on data destruction:

Project sponsors share confidential and proprietary information to student teams doing Action Learning projects. MIT Sloan has an obligation to destroy that data at the end of the project so that it does not inadvertently get disclosed to unauthorized people and it is not used for any other purpose than the project.

MIT Sloan depends on the student teams for destroying the data in a timely and appropriate manner. Please note that destruction of data is a requisite step for the completion of course requirements.

What data is required to be destroyed?

Any information supplied by company in any format- emails, notes from a phone meeting, worksheets, records, company documents, any kind of company data. This includes data that is marked confidential and unmarked data. If the company supplied it, it must be destroyed at the end of the project.

What data is NOT required to be destroyed?

Students can keep their final paper and other derivative work that does not include company proprietary or confidential information. If there is any doubt, ask for help to discern what needs to be destroyed.

What are acceptable destruction methods?

- Printed Materials: Documents should be recycled in MIT approved secure recycle bins. Each academic area and many program offices have these bins.
- Digital Data Controlled by Students: If students have the data in Dropbox or on their computer, they must delete the data using appropriate tools.
- Digital Data Controlled by Sloan Technology Services: STS will destroy the data according to MIT Sloan IT policies.

If there are any issues or questions on this issue, please contact Ellen Baum, Contract Administration, at 3-5617 at ebaum@mit.edu or Will Hedglon, STS, at 5-4176 at hedglon@mit.edu.

Class Schedule:

	Date	Time	Session	Lecturer
S1	9/6	4:00-5:30	Welcome – Intro to Analytics	Sinan Aral
S2	9/13	4:00-8:00	Pitch Day (Samberg Conference Center, 7 th Floor, E52)	
S3	9/20	4:00-5:30	NDA & Data Security The Predictive Power of Massive Data about our Fine-Grained Behavior	Ellen Baum Foster Provost
S4	9/27	4:00-5:30	Social Analytics – A Deep Dive	Sinan Aral
S5	10/4	4:00-5:30	Challenges of AI Adoption in Practice	Claudia Perlich
S6	10/11	4:00-5:30	Mid-Term Presentations	
S7	10/18	4:00-5:30	Mid-Term Presentations (cont.)	
S8	11/1	4:00-5:30	Optional Skill Seminar: Deep Learning	Paramveer Dhillon
S9	11/8	4:00-5:30	Optional Skill Seminar: TensorFlow	Daniel Rock
S10	11/15	4:00-5:30	Optional Skill Seminar: Estimation in Networks	Dave Holtz
S11	11/29	4:00-5:30	Optional Skill Seminar: Observational Causal Inference	Michael Zhao
S12	12/7	2:00-8:00	Final Presentations Session	

Reading List:

Session 1: Welcome – Intro to Analytics (Sinan Aral)

1. Carefully review all project proposals and this syllabus.
2. “Chapter 1: Introduction: Data Analytic Thinking” Provost, F. and Fawcett T. 2013. *Data Science for Business*, O’Reilly Media Inc.; <http://library.mit.edu/item/002221893>.
3. “Big Data: The Management Revolution” Brynjolfsson, E. and McAfee, A. 2012. *Harvard Business Review*, 90(10); October: 60-68;
<http://libproxy.mit.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=79996279&site=ehost-live>.

Optional Reading:

4. “The Business of AI” Brynjolfsson, E. and McAfee, A. 2017. *Harvard Business Review*, July;
<https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence>
5. “The Rapid Adoption of Data-Driven Decision-Making.” Brynjolfsson, E. and McElheran, K. 2016. *American Economic Review*, 106(5): 133-39.
<https://www.aeaweb.org/articles?id=10.1257/aer.p20161016>
6. “Big Data: New Tricks for Econometrics” Varian, H. 2014. *Journal of Economic Perspectives*, 28(2): 3-28;
<https://www.aeaweb.org/articles.php?doi=10.1257/jep.28.2.3>.
7. “Lectures on Machine Learning” Athey, S. and Imbens, G. 2015. NBER;
<http://conference.nber.org/confer/2015/SI2015/ML/syllabus.pdf>.
8. “The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales” Wu, L. and Brynjolfsson, E. 2014. *Economics of Digitization* (A. Goldfarb, S. Greenstein, and C. Tucker, eds.), Univ. of Chicago Press;
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2022293.
9. “Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology” Aral, S., Brynjolfsson, E. and Wu, L. 2012. *Management Science*, 58(5); May: 913-931;
<http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1110.1460>.

Session 2: Pitch Day

Optional Reading:

10. “Chapter 2: Business Problems and Data Science Solutions” Provost, F. and Fawcett T. 2013. *Data Science for Business*, O’Reilly Media Inc.

Session 3: The Predictive Power of Massive Data about our Fine-Grained Behavior (Foster Provost)

11. “Chapter 4: Fitting a Model to Data” Provost, F. and Fawcett T. 2013. *Data Science for Business*, O’Reilly Media Inc.
12. “Chapter 5: Overfitting and Its Avoidance” Provost, F. and Fawcett T. 2013. *Data Science for Business*, O’Reilly Media Inc.

Optional Reading:

13. “The Unreasonable Effectiveness of Data” Halevy, A., Norvig, P. and Pereira, F. 2009. *IEEE Intelligent Systems*;
<http://static.googleusercontent.com/media/research.google.com/en//pubs/archive/35179.pdf>
14. “Machine Science” Evans, J. and Rzhetsky, A. 2010. *Science*;

<http://science.sciencemag.org/content/sci/329/5990/399.full.pdf>

15. "The Machine that Would Predict the Future" Weinberger, D. 2011. *Scientific American*;
[http://www.cs.virginia.edu/~robins/The Machine that would Predict the Future.pdf](http://www.cs.virginia.edu/~robins/The_Machine_that_would_Predict_the_Future.pdf)

Session 4: Social Analytics: A Deep Dive (Sinan Aral)

16. "Chapter 3: Introduction to Predictive Modeling: From Correlation to Supervised Segmentation" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
17. "The Problem with Online Ratings" Aral, S. 2014. *MIT Sloan Management Review*, 55(2); January: 47-52;
<http://libproxy.mit.edu/login?url=http://search.proquest.com/docview/1475566579/6A2DE248EEAB45DDPQ/1?accountid=12492>.
18. "What Would Ashton Do - And Does it Matter?" Aral, S. 2013. *Harvard Business Review*, 91(5); May: 25-27;
<http://libproxy.mit.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=87039770&site=ehost-live>.

Optional Reading:

19. "Social Influence Bias: A Randomized Experiment" Muchnik, L., Aral, S., and Taylor, S. 2013. *Science*, 341(6146); August 9: 647-651; <http://www.sciencemag.org/content/341/6146/647.full>.
20. "Identifying Influential and Susceptible Members of Social Networks" Aral, S. and Walker, D. 2012. *Science*, 337(6092); July 20: 337-341; <http://www.sciencemag.org/content/337/6092/337.full>.
21. "Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks" Aral, S. and Walker, D. 2011. *Management Science*, 57(9); September: 1623-1639;
<http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1110.1421>.
22. "Distinguishing Influence Based Contagion from Homophily Driven Diffusion in Dynamic Networks" Aral, S., Muchnik, L., and Sundararajan, A. 2009. *Proceedings of the National Academy of Sciences (PNAS)*, 106(51); December 22: 21544-21549; <http://www.pnas.org/content/106/51/21544.full>.

Session 5: Challenges of AI Adoption in Practice (Claudia Perlich)

Optional Reading:

23. "Chapter 6: Similarity, Neighbors and Clusters" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
24. "Chapter 7: Decision Analytic Thinking I: What is a Good Model?" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
25. "Chapter 8: Visualizing Model Performance" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.

Additional Optional Readings

26. "Chapter 9: Evidence and Probabilities" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
27. "Preface: Big Data Is Not About the Data!" King, G. In Press, 2015. *Computational Social Science: Discovery and Prediction*, (Alvarez, M, ed.), Cambridge University Press;
<http://gking.harvard.edu/publications/preface-Big-Data-Not-About-Data>.
28. "Reverse-Engineering Censorship in China: Randomized Experimentation and Participant Observation" King, G., Pan, J., and Roberts, M. 2014. *Science*, 345(6199); August 22: 1-10;
<http://gking.harvard.edu/publications/randomized-Experimental-Study-Censorship-China>.

29. "Systematic Bias and Nontransparency in Us Social Security Administration Forecasts" Kashin, K., King, G and Soneji, S. 2015. *Journal of Economic Perspectives*, 29(2); Spring: 239-258; <http://gking.harvard.edu/publications/systematic-Bias-And-Nontransparency-Us-Social-Securityadministration-Forecasts>.
30. "Explaining Systematic Bias and Nontransparency in Us Social Security Administration Forecasts" Kashin, K., King, G and Soneji, S. 2015 *Political Analysis*, 23(3); May: 336-362; <http://gking.harvard.edu/publications/explaining-Systematic-Bias-And-Nontransparency-Us-Social-Securityadministration>.
31. "How Censorship in China Allows Government Criticism but Silences Collective Expression" King, G., Pan, J., and Roberts, M. 2013. *American Political Science Review*, 107(2); May: 1-18; <http://gking.harvard.edu/publications/how-Censorship-China-Allows-Government-Criticism-Silences-Collective-Expression>.
32. *Data Driven: Creating a Data Culture*, Patil, DJ and Mason, H. 2015. O'Reilly Media Inc.; <http://www.oreilly.com/data/free/data-driven.csp>.
33. "Practical guide to controlled experiments on the web: listen to your customers not to the HiPPO" Kohavi, R, Henne, RM, and Sommerfield, D. 2007. *Proceedings of the 13th ACM*. August: 1-9; <http://www.exp-platform.com/Pages/hippo.aspx>.
34. "Reservoir Sample – Sampling from a stream of elements" Grothaus, G. 2007. *Gregable*. October 8; <http://gregable.com/2007/10/reservoir-sampling.html>.
35. "Seven Rules of Thumb for Web Site Experimenters" Kohavi, R, Deng, A, Longbotham, R, and Xu, Y. 2014. *Proceedings of the 20th*. August: 1-11; <http://www.exp-platform.com/Pages/SevenRulesofThumbforWebSiteExperimenters.aspx>.
36. "Are You Letting a Groundhog Dictate Strategy?" DeFranza, D. 2014. *Brooks Bell*, February 6; <http://www.brooksbell.com/blog/are-you-letting-a-groundhog-dictate-strategy/>.
37. "The Surprising Thing Brand New and Highly Advanced Testing Programs Have in Common" DeFranza, D. 2014. *Brooks Bell*, June 19. <http://www.brooksbell.com/blog/surprising-thing-brand-new-highly-advanced-testing-programs-common/>.
38. "Chapter 13: Data Science and Business Strategy" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
39. Browse: <http://www.enlitic.com>
40. "Competing on Analytics" Davenport, T. 2006. *Harvard Business Review*, 84(1); January: 98-107; <http://libproxy.mit.edu/login?url=http://search.proquest.com/docview/205181007/fulltextPDF/827D3FA2EDFA4F2EPO/2?accountid=12492>.