15.572: ANALYTICS LAB: ANALYTICS, MACHINE LEARNING & THE DIGITAL ECONOMY

SYLLABUS AS OF NOVEMBER 13, 2015 – CONSULT STELLAR FOR MOST UP-TO-DATE INFORMATION AND READINGS

Instructors	**Professors Sinan Aral & Erik Brynjolfsson**
	Email: sinan@mit.edu; URL: http://web.mit.edu/sinana/www/
	Email: erikb@mit.edu; URL: http://digital.mit.edu/erik/
Class times	Tuesdays 4-5:30pm, E51-335
	Workshops 4-8pm, September 18, Bartos Theater (E15)
	4-9pm, December 8, Bartos Theater (E15)
Project Mentors	Sagit Bar-Gill sbargill@mit.edu
	Renee Richardson Gosline rgosline@mit.edu
	Jeff Lee jeff_lee@mit.edu
	William Li wli@csail.mit.edu
	Adel Malek a_malek@mit.edu
	Aparna Ramesh aparna76@mit.edu
	Daniel Rock drock@mit.edu
	Guillaume Saint-Jacques gsaintja@mit.edu
	Deborah Soule dsoule@mit.edu
	Thomas Stephens tpsjr@mit.edu
	George Westerman georgew@mit.edu
	Irving Wladawsky-Berger irving@irvingwb.com
Teaching Assistant	Fei Gao feigao@mit.edu
Course Support Team	Susan Young susany@mit.edu
	Chuck Gibson cgibson@mit.edu
Office Hours	By Appointment

Summary and Objectives:

The growth in big data and analytics is transforming in 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 aspects, but business relevance sets the context and nature of the technical problem.

Teams of three will form and select projects in the first two weeks of classes. Project proposals have been received from the following organizations: **Accenture, Akamai, Amazon, BASF, Boston Public Schools, BuyerZone, Christian Science Monitor, Dell, Graduate Management Admission Council, Intursa, Kairos, MasterCard, Micronotes, Nasdaq, Raise Marketplace, Schindler, Telemetry Sports, and Zensar.**

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 are expected to work independent of regular class meetings, possibly including travel. 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.

Students will have the opportunity to attend skill seminars designed around issues they are confronting as they work to complete their project work. Attendance is optional but strongly encouraged.

Two special sessions are scheduled: Match Day on September 18 (4-8pm) and Final Presentations Workshop on December 8 (4-9pm). Attendance at both sessions is required.

Notes on Class Activities and Due Dates:

- Prior to 9/15: Students will given project proposals from all project proposing companies and are expected to be familiar with them before the first session.
- 9/15: The first session will be followed by an informal session to facilitate team formation. Pizza will be provided.
- 9/18: On Match Day (4-8pm, Bartos Theater, http://whereis.mit.edu/?go=E15), 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.
- 9/22: The third session will be followed by another informal pizza session to further facilitate the team formation and project selection process ahead of the 9:00pm deadline (see below).
- 9/22, 9:00pm: **DUE**: team formation and project selection preferences; each team should submit one document (guidelines posted on Stellar, "Team Formation for Projects") to Fei Gao (feigao@mit.edu). In the following days, faculty, mentors, and the course support team will work out assignments of projects to teams, subject to review by the proposing company.
- 9/28: Final team-project pairings will be communicated to students.
- 10/6, 9:00pm: **DUE**: project plan; each team should submit one document to their mentor and Fei Gao.
- 10/26, 9:00pm: **DUE:** mid-point presentation slides; each team should submit their slides to their mentor and Fei Gao.
- 10/27 and 11/3: Student teams will deliver 5-7 minute updates on their project work at the mid-point of the term. The chief aim of these sessions is to help illuminate issues common across teams in order to foster collaboration and set the agenda for the optional skill seminars to follow.
- 12/7, 9:00pm: **DUE**: final presentation slides; each team should submit their slides to their mentor and Fei Gao.
- 12/8: During the Final Presentations Workshop (4-9pm, Bartos Theater), each team will present their project work to an audience of experts, entrepreneurs, and executives, including representatives from

project sponsoring organizations. Teams will have 8 minutes to present their project work and 4 minutes for Q&A.

- 12/11, 9:00pm: **DUE**: final report (10 pages maximum, 3000 words, not including figures); report should consider feedback received during final presentations on 12/8. Each team should submit one document to their mentor and Fei Gao.

Grading:

- 40% final presentation content and delivery – team-wide
- 30% final report – team-wide
- 10% mid-point presentation content and delivery – team-wide
- 20% contribution to class discussions and team project enablement – individual

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/fa15/15.572/

Optional:

SQL in 10 Minutes, Sams Teach Yourself, Ben Forta. 2012. Sams Publishing.

Class Schedule:

	Date	Time	Session	Lecturer
S1	9/15	4:00-5:30	Welcome Lecture: The Economic Payoff From Analytics	Erik Brynjolfsson
S2	9/18	4:00-8:00	Match Day	-
S3	9/22	4:00-5:30	Lecture: Social Analytics – A Deep Dive	Sinan Aral
S4	9/29	4:00-5:30	Lecture: Big Data Is Not About the Data!	Gary King
S5	10/6	4:00-5:30	Lecture: Data Driven – Creating a Data Culture	Hilary Mason
No Class For Two Weeks: Monday Schedule on Tuesday (10/13) and SIP (10/20)				
S6	10/27	4:00-5:30	Mid-Point Presentations	
S7	11/3	4:00-5:30	Mid-Point Presentations (cont.)	
S8	11/10	4:00-5:30	No Class – Team-Mentor Meetings	
S9	11/17	4:00-5:30	*Optional: Skill Seminar – Handling large datasets/building databases*	*Paramveer Dhillon*
S10	11/24	4:00-5:30	*Optional: Skill Seminars – Data visualization and "big data as a UI problem"*	*Cesar Hidalgo*
S11	12/1	4:00-5:30	*Optional: Skill Seminars – Deep learning*	*Daniel Rock*
S12	12/8	4:00-9:00	Final Presentations Workshop/Wrap Up	Guests, Erik/Sinan

Reading List:

Session 1: *The Economic Payoff from Analytics* (Erik Brynjolfsson) (9/15):

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. "*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.
5. "*Lectures on Machine Learning*" Athey, S. and Imbens, G. 2015. NBER; http://conference.nber.org/confer/2015/SI2015/ML/syllabus.pdf.
6. "*Strength in Numbers: How Does Data-Driven Decision Making Affect Firm Performance?*" Brynjolfsson, E., Hitt, L., and Kim, H. 2011; http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486.
7. "*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.
8. "*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: *Match Day: Meet Project Proposers* (9/18):

Optional Reading:

9. "*Chapter 2: Business Problems and Data Science Solutions*" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.

Session 3: *Social Analytics: A Deep Dive* (Sinan Aral) (9/22):

10. "*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.
11. "*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.
12. *"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:

13. "*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.

14. "*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.
15. "*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.
16. "*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 4: *Big Data Is Not About the Data!* (Gary King) (9/29):

17. "*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.
18. "*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.

Optional Reading:

19. "*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.
20. "*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.
21. "*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.
22. "*Chapter 4: Fitting a Model to Data*" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
23. "*Chapter 5: Overfitting and Its Avoidance*" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.

Session 5: *Data Driven: Creating a Data Culture* (Hilary Mason) (10/6):

24. *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.
25. "*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.
26. "*Chapter 8: Visualizing Model Performance*" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.

Optional Reading:

27. "*Reservoir Sample – Sampling from a stream of elements*" Grothaus, G. 2007. *Gregable*. October 8; http://gregable.com/2007/10/reservoir-sampling.html.
28. "*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.

Session 6: Mid-Point Presentations (10/27):

Session 7: Mid-Point Presentations (11/3):

Session 8: Discussion of Issues (11/10):

29. "*Chapter 6: Similarity, Neighbors and Clusters*" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
30. "*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.

Sessions 9-11: Optional Skill Seminars (11/17, 11/24, 12/1):

Optional Reading:

31. "*Chapter 9: Evidence and Probabilities*" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
32. *"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/.
33. *"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/.
34. "*Chapter 13: Data Science and Business Strategy*" Provost, F. and Fawcett T. 2013. *Data Science for Business*, O'Reilly Media Inc.
35. Browse: http://www.enlitic.com
36. "*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/827D3FA2EDFA4F2EPQ/2?accountid=12492.

Session 12: Final Presentations Workshop/Wrap Up (12/8):