Syllabus for

05-898: Making Products Count:

Data Science for Product Managers

Fall 2017 Mini-course 2 (6 units)

Course Summary

Product managers engage in a variety of complex activities critical to product success including

- Product requirements gathering
- Forecasting customer demand
- Customer segmentation
- Pricing
- Allocating marketing spend
- Identifying buying patterns
- Analyzing and responding to customer feedback

Historically decisions in these areas have often relied on intuition and guesswork, leading to misjudgment of the market and other key factors, and ultimately, product failures. Developments in data science, combining the increasing availability of data from internal and external sources with new algorithms that exploit that data at scale, offer new possibilities for putting product management decisions on a more quantitative and rigorous footing. Students in this course will be introduced to a variety of data science techniques applicable to activities to which product managers typically contribute. These techniques include preference modeling, time series forecasting, regression, clustering, classification, A/B testing, and analytics for unstructured data including clickstreams, text, speech, and images. Along the way, students will learn about practical aspects of applying data science to product management, including:

- Choosing appropriate metrics for product success
- Selecting and assessing data sources
- Instrumenting products for data collection
- Data science infrastructure choices
- Evaluating data science models
- Documenting and communicating data science results
- Planning and delivering data science projects
- Data science pitfalls to watch for

This course is primarily aimed at students with technical backgrounds who wish to apply their skills to product management. Backgrounds in basic statistics, and some programming experience are required, as the course will includes hands-on exercises in Python to illustrate the concepts. In-class exercises and weekly assignments will mainly focus on interpretation of the output of various data science techniques and matching different techniques to product

management needs; each student will also complete a final project that takes a data science need from start to finish (framing the problem, choosing data sources, exploratory data analysis, modeling, communicating results). Currently the final project is planned to be around market sizing and risk mitigation: in the first part of the project, the student selects an idea, identifies customer segments and then draws on publicly available data sources to estimate size of those segments. In the second phase students will identify risks and plans for measuring and mitigating the risks.

Outcomes

The main learning objectives of the course are to:

1. Identify decision points during the product life cycle where data science techniques are applicable
2. Select metrics, product instrumentation, external data sources, modeling techniques and tools for use in product management decision-making
3. Apply selected basic modeling (e.g. classification, clustering, time series analysis, and text analytics) to product management; critique applicability and output from more advanced techniques.
4. Plan and execute a data science project at realistic scale to inform at least one product management decision demonstrating mastery of objectives #1 - #3.

Instructor: David Steier (PhD, CMU SCS '89)

David Steier is Managing Director in the Advanced Analytics and Modeling practice of Deloitte Consulting LLP. At Deloitte, David has helped clients use advanced data analytics and visualization in a variety of industries including healthcare, life sciences, banking, retail, manufacturing, technology, telecommunications, media and the public sector. Prior to Deloitte, David was director in the Center for Advanced Research at PwC, senior director of technology and business development at Kanisa, and managing director at Scient. Beyond his over 25 years of professional experience, David holds a Ph.D. in computer science from Carnegie Mellon and a bachelor's degree in computer science from Purdue University. David has taught analytics project capstone courses and courses on Managing Analytics Projects course at CMU and at the University of California Berkeley.

The teaching assistant for this class will be Prajwal Vasisht, who is enrolled in the School of Computer Science in the MSBIC program.

Course Resources

An online site with the syllabus, selected readings, Python resources has been created in Piazza. . The system is highly catered to getting you help fast and efficiently from classmates, the TA, and myself. Rather than emailing questions, we encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com.

Find our class page at: https://piazza.com/cmu/fall2017/05898/homeCourse Learning. To enroll for access to the site, go to piazza.com/cmu/fall2017/05898

The main text of the course covering the data science material will be *Introduction to Data Science* by Cielen, D., Meysman, A.D. B., and Ali, M. (published by Manning, 2016, https://www.manning.com/books/introducing-data-science). This will be supplemented by material around Python and covering various topics in product management as listed in the course outline below.

Course Outline

This mini-course is planned around two sessions of 1 hour 20 minutes apiece per week for 7 weeks. For the Fall 2017 offering of the minicourse course, there will be an additional class on Friday Oct 27, and there will be no class the Wednesday before Thanksgiving (Nov 22). Note that the class venue will change - see room assignments below. The first six sessions and last two (1-6 and 13-14) are in person on campus (PH = Porter Hall, NSH = Newell-Simon Hall), while sessions 7-12 will be conducted online using Zoom.

- **Session 1 (Oct 23, PH 125C): Introduction; Metrics for Success**
  - **Topics**
    - Introductions
    - Data science and product management
    - Structure of the class
    - Metrics for success in product management
  - **Readings**:
    - Cielan et al, pp. 1-14
    - (optional) Croll and Yoskovitz book, pp 45-72, then skim pp 71-151, or Croll and Yoskovitz e-book
  - **Assignment out**: Complete python proficiency exercise, Datacamp tutorials as necessary
- **Session 2 (Oct 25, PH 125C): Statistics and Regression**
  - **Topics**
    - Descriptive statistics
    - Probability distributions
    - Linear regression
    - Logistic regression
  - **Readings/preparation**
    - DataCamp online course, Statistical Thinking in Python, part 1
    - Datacamp online course, Statistical thinking in Python, Part 2
  - **Assignment out:** HW #1
- **Session 3 (Oct 27, NSH 3305): Data Sources; Demand Forecasting**
  - **Topics**

        - The data science process
        - Finding and evaluating data sources
        - Forecasting
        - Time series analysis
    - **Readings**
        - Cielan et. al,2016, pp 1-10, 22-56.
        - Quartz guide to bad data, 2016
        - Jain, Time series 2016
    - **Assignment due:** Python proficiency demonstration
    - **Assignment out**: HW1
- **Session 4 (Oct 30, PH 125C): Decision Trees**
    - **Topics**
        - The space of data science techniques
        - Supervised and unsupervised learning
        - Decision trees for classification
        - Selecting modeling techniques
        - Decision tree analysis
    - **Readings**:
        - Cielan, et al 2016, Chapter 3 (Machine learning)p. 228-230
        - Magee 1964: Decision trees for decision making
        - Markham, 2015, Comparing supervised learning algorithms http://www.dataschool.io/comparing-supervised-learning-algorithms/
        - (optional) Provost and Fawcett, Data Science for Business,, Chapter 3-4
- **Session 5 (Nov 1, PH 125C): Market Segmentation and Clustering**
    - Topics
        - Unsupervised learning
        - Clustering (*k*-means)
        - Clustering for market segmentation
        - Customer segmentation
        - Selecting marketing techniques
    - **Readings**
        - Dancho, 2016
        - Cast, 2013;
        - Wind and Bell, 2008 “Market Segmentation”
    - **Assignment due:**
        - Prerequisite exercise demonstrating proficiency in Python
        - HW1
    - **Assignment out**: HW2
- **Session 6 (Nov 6, PH 125C): Conjoint Analysis and Customer Feedback**

- **Topics**
    - Conjoint analysis to gather customer preferences
    - Analyzing customer feedback – text analytics, topic modeling
    - Social media listening to gather customer feedback
- **Readings**
    - Cielan, et.al, 2016: Chapter 8

● Session 7 (Nov 8, Online): Evaluating Data Science Models

- **Topics**
    - Feature selection: information gain, PCA, lasso
    - Model evaluation – MSE, lift, AUC, Type 1 vs 2 errors
    - Survey and experimental design
    - A/B testing
- **Readings:** TBD
- **Assignment due**: HW2
- **Assignment out:** HW3

● Session 8 (Nov 13, Online): Data Science Infrastructure

- **Topics**
    - Instrumentation
    - Clickstream analytics
    - Parallel and distributed computing for scalability:
    - Resolving technical tradeoffs
- **Readings**
    - Cielan: et. al 2016, Chapter 4

● Session 9 (Nov 15, Online): Data Science for Marketing

- **Topics**
    - Marketing to different customer segments: Uplift modeling
    - Optimizing marketing spend: market mix modeling
- **Readings**
    - Lo, 2015
    - Gutierrez, 2016
- **Assignment due:** HW3
- **Assignment out**: HW4

● Session 10 (Nov 20, Online): Data Science for Pricing

- **Topics**
    - Conjoint analysis for pricing
    - Price elasticity and optimization
- **Readings**: TBD

[Note: No class Nov 22]

● Session 11 (Nov 27, Online): Deep Learning

- **Topics**
    - Introduction to neural networks

        - Convolutional Neural Networks
        - Recurrent Neural Networks
        - Long Short Term Memories
    - **Readings**
        - Jackar, 2016
        - (Optional,) Ali, 2016
    - **Assignment due**: HW4
    - **Assignment out**: Final project
- **Session 12 (Nov 29, Online): Internet of Things**
    - **Topics**
        - Real-time personalization
        - Internet of things
    - **Readings**: TBD
- **Session 13 (Dec 4, PH 125C): Sustaining Data-Driven Product Management**
    - **Topics**
        - Communicating data science results
        - Data visualization
        - Turning data into product
        - Organizing, draining, recruiting, and retaining data science talent
    - **Readings**
        - Patil, 2012
        - Zumel & Mount, 2013
- **Session 14 (Dec 6, PH 125C): Final Project Presentation and Wrap-Up**

Course References

1. Ali, S. “Guide to Deep Learning,” http://www.datasciencecentral.com/profiles/blogs/guide-to-deep-learning
2. Cielen, D., Meysman, A.D. B., and Ali, M ; *Introduction to Data Science* , Manning, 2016
3. Croll, Adam and Yoskovitz, Alistair, *Lean Analytics*, O’Reilly, 2013. https://www.amazon.com/Lean-Analytics-Better-Startup-OReilly/dp/1449335675
4. Croll and Yoskovitz e-book http://leananalyticsbook.com/wp-content/uploads/2013/01/Analytics-Lessons-Learned.pdf f
5. Cast, 2013 http://www.mindtheproduct.com/2013/02/everything-a-product-manager-needs-to-know-about-analytics/
6. Dancho, 2016, “Customer Segmentation, Part 1: K-means clustering “http://www.business-science.io/business/2016/08/07/CustomerSegmentationPt1.html

7. DataCamp online course, Statistical Thinking Python, Part 1, https://www.datacamp.com/courses/statistical-thinking-in-python-part-1
8. DataCamp online course, Statistical Thinking in Python, Part 2 https://www.datacamp.com/courses/statistical-thinking-in-python-part-2
9. Gutierrez, 2016 https://www.slideshare.net/PierreGutierrez2/introduction-to-uplift-modellig
10. Jackar, 2016, “Introduction to deep learning and its role for IoT future, http://www.datasciencecentral.com/profiles/blogs/an-introduction-to-deep-learning-and-it-s-role-for-iot-future
11. Jain, 2016, https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/
12. Lo, 2015, Uplift Modeling workshop https://www.slideshare.net/odsc/victor-lomachinelearningpresentation?next_slideshow=1
13. Magee 1964: Decision trees for decision making (https://hbr.org/1964/07/decision-trees-for-decision-making)
14. Markham, 2015, Comparing supervised learning algorithms http://www.dataschool.io/comparing-supervised-learning-algorithms/
15. Patil, 2012, “Data Jujitsu: The Art of Turning Data into Product,” 2012, http://www.oreilly.com/data/free/data-jujitsu.csp
16. Provost & Fawcett. *Data Science for Business,* 2014. O’Reilly
17. Quartz, Guide to Bad data, 2016 https://github.com/Quartz/bad-data-guide
18. Wind and Bell, Chapter 11 Marketing segmentation https://marketing.wharton.upenn.edu/files/?whdmsaction=public:main.file&fileID=566 in Baker and Hart, The Marketing Book, 2008 (sixth edition).
19. Zumel & Mount:, Chapter 11, “Producing Effective Presentations”, in P*ractical Data Science Using* R, Manning, 2013,