OIDD 898 – Advanced Topics: Understanding Behavior with Big Data

COURSE SYLLABUS

Instructor: Seth Stephens-Davidowitz

Email: seth.stephens@gmail.com

Prerequisites: No advanced mathematical training is required for this course, although
having taken an introductory probability/statistics course may be useful.

1.2 Objectives

This course has two objectives. The first is to enable students to better understand studies that
use data. By the end of this course, students will be able to make sense of data-rich studies,
whether in academic journals, newspapers, or business presentations. Students will learn how
to rigorously critique numerical arguments.

The second objective is to enable students to understand how data might help them answer
questions they have during their careers. How do you determine the right data to collect or
analyze? How do you find it? How do you know if the data can be trusted?

1.3 Course Structure

The course is divided into three parts.

The first part introduces the value of Big Data. This will explain the importance of rigorous data
analysis -- and how the internet is making it easier and easier to collect data to help make
important decisions.

The second part focuses exclusively on prediction. We will discuss how to best use data to
predict what people will do in the future. Ultimately, data analysis is most valuable not when it
can help you make sense of what already happened but if it can offer you new insights into what
might happen. This is how you really improve decision-making.

The third part explores the weaknesses and pitfalls of data analysis. We will discuss ethical
concerns Big Data can raise and the importance of using human intuition to guide data analysis.
We will also focus on the value of incorporating many sources of data to correct blind spots in
any one particular source.

1.4 Instruction Format
The course is primarily lecture based, although it includes some cases and activities. Class discussion is strongly encouraged.

1.5 Final Group Project

Students will think up a difficult decision that a corporation might face. They will discuss how data might be used to help with the decision. They will propose what data they would collect and how they would analyze it. They will present their proposal. The goal of the proposal would be to convince a potentially skeptical manager that their data collection and analysis plan would help make a better decision.

**Course Outline**

<table>
<thead>
<tr>
<th>Number</th>
<th>Topic</th>
<th>Assignment/Readings Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>The Limitations of Human Intuition -- and the Need for Data</td>
<td><strong>Reading</strong>: Ayers, Chapters 1-3, Stephens-Davidowitz, Chapters 1-2; Kahneman, Chapters 1-3;</td>
</tr>
<tr>
<td>3</td>
<td>The Value of New Sources of Data, Part I</td>
<td><strong>Reading</strong>: Stephens-Davidowitz, Chapter 3&lt;br&gt;<strong>Due</strong>: Homework # 1</td>
</tr>
<tr>
<td>4</td>
<td>The Value of New Sources of Data, Part II</td>
<td><strong>Reading</strong>: Dubner and Levitt, Chapters 1-3</td>
</tr>
<tr>
<td>5</td>
<td>Honest Data, Part I</td>
<td><strong>Reading</strong>: Stephens-Davidowitz, Chapter 4</td>
</tr>
<tr>
<td>6</td>
<td>Honest Data, Part II</td>
<td><strong>Reading</strong>: Rudder, Chapters 1-2</td>
</tr>
<tr>
<td>7</td>
<td>Personalization, Part I</td>
<td><strong>Reading</strong>: Stephens-Davidowitz, Chapter 5&lt;br&gt;<strong>Due</strong>: Homework # 2</td>
</tr>
<tr>
<td>8</td>
<td>Personalization, Part II</td>
<td><strong>Reading</strong>: Smith and Telang, Chapter 1</td>
</tr>
<tr>
<td>9</td>
<td>Correlation Versus Causation, Part I</td>
<td><strong>Reading</strong>: Stephens-Davidowitz, Chapter 6&lt;br&gt;Levitt (1994)</td>
</tr>
<tr>
<td>10</td>
<td>Correlation Versus Causation, Part II (A/B Testing)</td>
<td><strong>Reading</strong>: Christian (2012)</td>
</tr>
<tr>
<td>11</td>
<td>The Science of Prediction, Intro</td>
<td><strong>Reading</strong>: Silver, Chapters 1-2&lt;br&gt;<strong>Due</strong>: Homework # 3</td>
</tr>
<tr>
<td></td>
<td>Sports Analytics</td>
<td><strong>Reading</strong>: Silver, Chapter 3; Silver, Chapter 8</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>13</td>
<td>Data and Finance</td>
<td><strong>Reading</strong>: Silver, Chapter 11</td>
</tr>
<tr>
<td>14</td>
<td>Data and Marketing</td>
<td><strong>Reading</strong>: Duhigg (2012)</td>
</tr>
<tr>
<td>15</td>
<td>Data and Economic Output</td>
<td><strong>Reading</strong>: Choi and Varian (2009)</td>
</tr>
</tbody>
</table>
| 16| Data and Politics                | **Reading**: Gabriel and Stephens-Davidowitz (2016)  
**Prepare**: Visit and familiarize yourself with fivethirtyeight.com  
**Due**: Homework # 4 |
| 17| People Analytics                 | **Reading**: Bersin, Chapters 1-4                |
| 18| Midterm Review                   |                                                  |
| 19| Midterm Exam                     |                                                  |
| 20| The Dangers of Data, Intro       | **Reading**: Stephens-Davidowitz, Chapter 6      |
| 21| Ethical Issues in Data Analysis  | **Reading**: Stephens-Davidowitz, Chapter 7      |
| 22| Data Privacy                     | **Reading**: Stephens-Davidowitz, Chapter 8; Schneier, Chapters 1-3  
**Due**: Homework # 5 |
| 24| An Unpredictable World           | **Reading**: Taleb, Chapters 1-3                 |
| 25| In-Class Group Project Preparation|                                                  |
| 26| Group Presentations, 1           |                                                  |
| 27| Group Presentations, 2           |                                                  |
| 28| Conclusion                       |                                                  |

**DETAILED READING SCHEDULE**

**CLASS 1**

Introductory Lecture; No Assigned Reading
CLASS 2

- Ian Ayers, *Supercrunchers*, 2007, Chapters 1-3
- Daniel Kahneman, *Thinking Fast and Slow*, 2011, Chapters 1-3

CLASS 3

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 3

CLASS 4

- Stephen Dubner and Steven Levitt, *Freakonomics*, 2005, Chapters 1-3

CLASS 5

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 4

CLASS 6

- Christian Rudder, *Dataclysm*, 2014, Chapters 1-2

CLASS 7

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 5

CLASS 8

- Michael D. Smith and Rahul Telang, *Streaming Sharing Stealing*, 2016, Chapter 1

CLASS 9

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 6

CLASS 10

CLASS 11

• Nate Silver, *The Signal and the Noise*, 2015, Chapters 1-2

CLASS 12

• Nate Silver, *The Signal and the Noise*, 2015, Chapters 3 and 8

CLASS 13

• Nate Silver, *The Signal and the Noise*, 2015, Chapter 11

CLASS 14


CLASS 15


CLASS 16


CLASS 17

• Josh Bersin, *The Training Measurement Book*, Chapters 1-4

CLASS 18

Midterm Review; No Reading

CLASS 19

Midterm; No Reading
CLASS 20

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 6

CLASS 21

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 7

CLASS 22

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 8
- Bruce Schneier, *Data and Goliath*, 2015, Chapters 1-3

CLASS 23


CLASS 24


CLASS 25

Group Presentation Preparation; No Reading

CLASS 26

Group Presentations; No Reading

CLASS 27

Group Presentations; No Reading

CLASS 28

Concluding Lecture; No Reading