SIADS 501 Reading List

December 20, 2019

1 WEEK 1

Status	Reading
Required	Chapter 2, Business Problems and Data Science Solutions. In Fawcett, Tom. (2013). Data science for business. Sebastopol, CA : O'Reilly.
Required	Chapter 1, Interview with Chris Wiggins. In Gutierrez, Sebastian. (2014). Data Scientists at Work. Berkeley, CA : Apress : Imprint: Apress.
Required	Chapter 4, Interview with Erin Shellman. In Gutierrez, Sebastian. (2014). Data Scientists at Work. Berkeley, CA : Apress : Imprint: Apress.
Required	Chapter 16, Interview with Jake Porway. In Gutierrez, Sebastian. (2014). Data Scientists at Work. Berkeley, CA : Apress : Imprint: Apress.
Optional	Chapter 3, Arms Race: Going to College. In O'Neil, Cathy. (2016). Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy. New York: Broadway Books.
Optional	Rogati, Monica. (2017). How do I become a Data Scientist?. Good Audience blog.
Optional	Kaduk, Taras. (2016). 4 Stages of Data Analytics Maturity: Challenging Gartner's Model. LinkedIn.

2 WEEK 2

Status	Reading
Required	Pages 19-25 and Chapter 10, The Law of Small Numbers (all; pp 109 - 118). In Kahneman, Daniel. (2011). Thinking, fast and slow. New York : Farrar, Straus and Giroux.
Required	Mester, Tomi. (2017). Statistical Bias Types explained (with examples) – part 1. Data36 blog.
Required	Bailey, Brendan. (2017). Data Cleaning 101. Towards Data Science blog.
Required	Tait, Andrew. (2017). 10 Rules for Creating Reproducible Results in Data Science. Dataconomy blog.
Optional	Keng, Brian. (2015). The Gambler's Fallacy and the Law of Small Numbers. Bounded Rationality blog.
Optional	Lee, N.T., Resnick, P., and Barton, G. (2019). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings Institution report.
Optional	Data Cleansing. Wikipedia.org

3 WEEK 3

Status	Reading
Required	Overfitting in Machine Learning: What It Is and How to Avoid It. EliteDataScience.com
Required	Ray, Sunil. (2018). Improve Your Model Performance using Cross Validation (in Python and R). Analytics Vidhya. * Read Introduction section only.
Required	Ranganathan, P., Pramesh, C. S., & Buyse, M. (2016). Common pitfalls in statistical analysis: The perils of multiple testing. Perspectives in clinical research, 7(2), 106–107. doi:10.4103/2229-3485.179436
Required	Anderson, Brian. (N.D.)P-Hacking and the Problem of Multiple Comparisons. Musings, Dr. Brian Anderson's blog.
Required	Spurious Correlations. (N.D.) Tylervigen.org.

Status	Reading
Required	Koehrsen, Will. (2018). Correlation vs. Causation: An Example. Towards Data Science blog.
Required	Wagner, Clifford. (1982). Simpson's Paradox in Real Life. The American Statistician, 36(1), 46-48. doi:10.2307/2684093.
Required	Appleton, D., French, J., & Mark P. J. Vanderpump. (1996). Ignoring a Covariate: An Example of Simpson's Paradox. The American Statistician, 50(4), 340-341. doi:10.2307/2684931
Required	Rohrer, Julia. (2017). That one weird third variable problem nobody ever mentions: Conditioning on a collider. The 100% CI blog.
Optional (Recommended)	Chapter 5, Desperately Seeking Signal. In Silver, Nate. The Signal and the Noise; Why so Many Predictions Fail– But Some Don't. Penguin Press, 2012.
Optional	Section 3.1, Cross-validation: evaluating estimator performance. (N.D.) Scikit-learn.org. * Read Section 3.1 only, no sub-sections.

4 WEEK 4

Status	Reading
Required	Dykes, Brent. (2016). A History Lesson On The Dangers Of Letting Data Speak For Itself. Forbes.com.
Required	Zawadzki, Jan. (2018). Storytelling for Data Scientists. Towards Data Science blog.
Required	Kaynar-Kabul, Ilknur. (2017). Interpretability is crucial for trusting AI and machine learning. The SAS Data Science blog.

Status	Reading
Required	Chapter 2, Are You Smarter than a Television Pundit? (Required: Start with "A Fox-Like Approach to Forecasting" and read through "Principle ;" rest of chapter is optional.). In Silver, Nate. The Signal and the Noise; Why so Many Predictions Fail– But Some Don't. Penguin Press, 2012.
Required	Chapter 6, How to Drown in Three Feet of Water. (Required: Read through Figure 6-2). In Silver, Nate. The Signal and the Noise; Why so Many Predictions Fail– But Some Don't. Penguin Press, 2012.
Required	Irwin, N., & Quealy, K. (2014, May 02). How to avoid being misled by the jobs report. New York Times.
Required	Dudek, Tomasz. (2018). But What Is This "Machine Learning Engineer" Actually Doing? Medium.com.
Required	Newman, Riley. (2015). How We Scaled Data Science to all Sides of Airbnb Over 5 Years of Hypergrowth. VentureBeat.com
Optional	Hypothetical Outcome Plots (HOPs) example. Vega Project.
Optional	UW Interactive Data Lab. (2016). Hypothetical Outcome Plots: Experiencing the Uncertain..