

Demystifying Big Data

PRASHANT NATARAJAN | Principal, Analytics & Cognitive | Deloitte Consulting Pty. Ltd., Melbourne, VIC

The Deloitte logo, consisting of the word "Deloitte" in white text on a black rectangular background.

Deloitte.

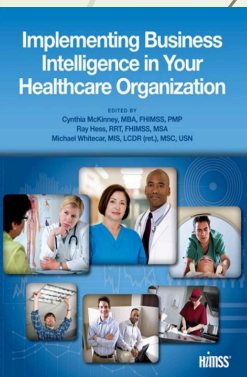
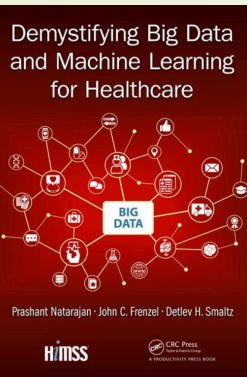
A large orange arrow pointing to the right, located on the left side of the slide.

Prepared for:

ETC 1010 by Prof. Di Cook, Dept of Econometrics & Statistics, Monash University, Clayton, VIC, Australia

All opinions are mine and mine ONLY!

Bio



- Prashant Natarajan
- Principal, Analytics & Cognitive | Deloitte Consulting Pty. Ltd., Melbourne, VIC, Australia
www.deloitte.com.au
- Undergraduate degree in Chemical Engineering; Master's in Technical Communications & Linguistics; PhD courses in Logic & Cognitive Psychology
- AT&T-Yahoo Chancellor's Fellow 2016
- 18+ years in industry – across 5 continents
- Previously, Senior Director of AI Applications at H2O.ai, Mountain View, CA, USA; Director of Product Strategy at Oracle USA
- Lead author or contributor to 5 books on big data analytics, business intelligence, machine learning, AI (best-sellers in 2012, 2017, 2018) – upcoming book on Enterprise AI in 2019
- Co-Faculty Instructor, Applied Data Science & AI, Stanford University, Palo Alto, CA
- Industry Advisor, University of California San Francisco (CIAPM/SF VA)



Why is Data Big or Little?

- ▶ Datafication – capturing data in all its forms and feeds to accomplish new functions
- ▶ Data expands to fill the space available because we don't know what data might turn out to be valuable/usable. All the more so with Machine Learning
- ▶ The 5 V's of Big Data: a primer
- ▶ Each industry/domain determines what makes the data “big” – volume is over-rated thanks to e-commerce
- ▶ Be specific (or) Bring it all in – when to choose what option
- ▶ Veracity and Data Fidelity



Data Fidelity & Veracity

- ▶ Reasons for poor data quality can include
 - ▶ Data entry
 - ▶ Squatting
 - ▶ Data Mgmt
 - ▶ Integration quality
 - ▶ Staleness
 - ▶ Usage
- ▶ 2 schools of thought: 4 C's v "Anything Goes" superseded by NRF Framework for Data Fidelity
- ▶ Discussion: Page 26 from Demystifying Big Data & Machine Learning for Healthcare (Natarajan et. al, CRC Press)

How to Use Big Data

- ▶ Types of Results Produced:

- ▶ Descriptive
- ▶ Discovery
- ▶ Predictive
- ▶ Prescriptive

- ▶ Reasoning used:

- ▶ Deduction
- ▶ Induction
- ▶ Diagnosis
- ▶ Analogy

- ▶ Data types processed:

- ▶ Structured
- ▶ Semi-structured
- ▶ Unstructured

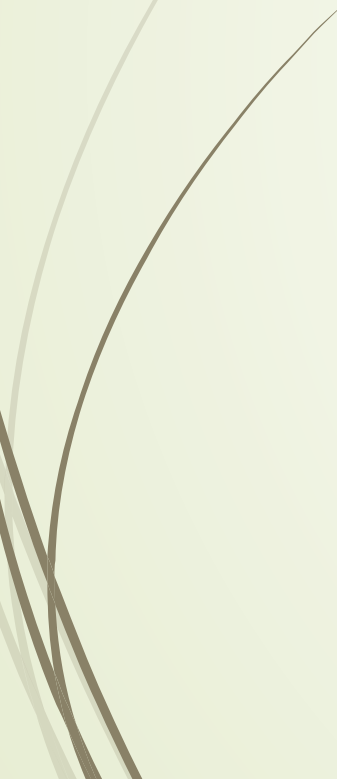
Myths and Best Practices

Practice does not make perfect – only perfect practice makes perfect ~
Vince Lombardi

- ▶ Debunking common myths
 - ▶ Everyone is doing it
 - ▶ Big data is hype
 - ▶ Our organization doesn't have any big data
 - ▶ Data lakes are the universal solution to all big data problems
 - ▶ The DW is “dead”
 - ▶ Big data projects are expensive
 - ▶ We need legions of data scientists
 - ▶ Volume is the most important consideration



Establishing Best Practices

- Determining Purpose and scope – start with small steps and expand progressively
 - Solve a pain point or create a new opportunity for users/businesses
 - Use more data variety – don't focus exclusively on volume
 - Use case definition can be democratic – business users, data scientists, and analysts can all contribute
 - The hammer-nail conundrum
 - Cloud and the hybrid architecture
 - Leadership (witness the creation of CDO & CAO roles in government/industry)
- 



Establishing Best Practices

- Storytelling
- Enable data & info governance
- Ownership, privacy, and cybersecurity
- Don't forget NFRs – load frequency, retention/backup/recovery, versioning, data lineage, impact analysis, performance, TIE (transparency, interpretability, explainability)
- Data fidelity
- Focus on use and action
- Analytics-driven workflows
- Focus on building an Insights-Driven Organization®



Let's continue this conversation

- ▶ Thank you to Prof. Dianne (Di) Cook, Department of Econometrics & Statistics, and Monash University for having me
- ▶ Thank you for your interest and making a great educational choice
- ▶ Deloitte Australia for attracting me to Melbourne from the Silicon Valley and San Francisco Bay Area – a great firm that you **must** consider as your career choice too!
- ▶ Contact me at pranatarajan@deloitte.com.au or @BigDataCXO on Twitter or www.Linkedin.com/in/natarpr