

# Walker's Data Science Laws



Presentation by:

Michael Walker

Rose Business Technologies

720.373.2200

m@rosebt.com

<http://www.rosebt.com>



# Walker's Data Science Laws

A "Data Scientist" is a professional who uses scientific methods to liberate and create meaning from raw data - somebody who can play with data, spot trends and learn truths few others know.

Data scientists are inquisitive: exploring, asking questions, doing "what if" analysis, questioning existing assumptions and processes.

# Walker's Data Science Laws

**I. Organizations with the best quality and variety of data, machine learning and algorithms will win in the long term.**

# Walker's Data Science Laws

## Corollaries:

- A. Decisions based on solid, unbiased data science evidence from machine learning and algorithms will be superior - over time - to decisions based on human intuition and reasoning alone.
  
- B. Data science evidence in conjunction with human reasoning and creativity is superior to either data evidence alone or human reasoning and logic alone.
  
- C. Data quality and veracity can make or break data science results.

# Walker's Data Science Laws

## Corollaries:

D. Bad or mediocre data engineers can make great data scientists produce sub-optimal results.

E. Human bias, stupidity, unethical behavior and bad laws can impede or eliminate any and all data science advantages.

F. Data variety usually trumps large volume - big data is hyped, dumb and overrated.

G. Organizations with the best data engineers and data scientists will produce superior results most of the time.

# Walker's Data Science Laws

**II. The life of data science is experience and not models, statistics, math, logic or elegant computer code.**

# Walker's Data Science Laws

## Corollaries:

A. Data science is about creating value, finding new truths and helping optimal decision-making - it is not about creating models, statistics, math or coding.

B. Models, statistics, math, coding and tech are simply tools that may or may not be useful to add value.

C. The best data scientists are creative, skeptical, question assumptions and great communicators.

D. Data scientists highest value may be designing and executing rigorous empirical experiments - especially ethical online human experiments and computer simulations.



# Walker's Data Science Laws

## Corollaries:

E. Data scientists highest value may be designing and executing machine learning systems and algorithms.

F. Data scientists highest value may be collecting the right “Smart Data” and designing processes to collect needed data not currently available.

G. Data and business analysts are not data scientists: nurses are not physicians and paralegals not lawyers.

H. Human behavior and other complex phenomena may be high causal density environments where causality is difficult if not impossible to understand.

# Walker's Data Science Laws

## Corollaries:

I. Big data is stupid data.

J. Data science is a team sport: unicorns are fantasy and data scientists specialize and add value in collaboration with colleagues and other stakeholders.

K. Data engineers and computer programmers are not data scientists.

# **Walker's Data Science Laws**

**II. Data science shall honor the scientific method or lose credibility.**

# Walker's Data Science Laws

## Corollaries:

- A. The scientific method in addition to deep analytical and creative thinking is what separates data scientists from garden variety data and business analysts.
- B. Forecasting and predictions are very difficult if not impossible in high causal density environments.
- C. Data scientists should help people think in terms of probabilities and not absolute predictions.
- D. Detecting signal from noise in data is very difficult and “big data” is meaningless without data science to interpret meaning using scientific methods as well as deep analytical and creative thinking.

# Walker's Data Science Laws

## Corollaries:

E. The current infatuation with "big data" - where detecting patterns and correlations is in vogue - is dangerous and unsustainable in the long run - although it may be useful at appropriate times.

F. Well designed and executed randomized, controlled experiments is still the gold standard in any venue or domain.

G. All data scientists are biased - do not fool yourself.

H. Data science results must be replicable or clearly labeled not replicable for decision-makers.

# Walker's Data Science Laws

## Corollaries:

I. Cherry picking data to support a desired result is not acceptable data science practice: if a data scientist has sufficient freedom to select the data she can support almost any result.

J. Data science that produces interesting results explained by complex reasoning should be greeted with skepticism and may or may not be true or useful.

K. Simple algorithms with lots of good quality data usually outperforms complex algorithms with small amounts of near perfect quality data (See Peter Norvig, "The Unreasonable Effectiveness of Data").

# Walker's Data Science Laws

**IV. Organizations that run the most low-cost, low-risk experiments create durable competitive advantage.**

# Walker's Data Science Laws

## Corollaries:

- A. Ability to consistently innovate and optimize new products, services and operational processes depends on continual trial and error.
- B. High value data science designs and executes rigorous empirical experiments.
- C. Detecting patterns and correlations is usually unreliable and may or may not be useful.
- D. Well designed and executed randomized, controlled experiments is the gold standard.



# Walker's Data Science Laws

**V. Models should be judged by reasonable, dubious or untestable assumptions - not only predictive results.**

# Walker's Data Science Laws

## Corollaries:

- A. Models are always an illusion of reality - yet may or may not be useful.
- B. Searching for assumptions that produces a desired result is not acceptable data science practice.
- C. Bad assumptions have consequences: freedom to select any assumptions allows the creation of a model to support any result.
- D. Data scientists should always ask: are there other more reasonable assumptions that explain observations?

# Walker's Data Science Laws

## Corollaries:

E. Simplifying assumptions usually makes them unrealistic and disconnected from the real world.

F. All models need to be subjected to rigorous empirical tests to avoid creating an illusion of reality.

G. Data scientists should use models properly: to gain understanding when no real alternatives available - build simplified models with reasonable and testable assumptions that capture what is important and subject both assumptions and results to rigorous empirical tests.

# Thank You

Presentation by:

Michael Walker

Rose Business Technologies

720.373.2200

m@rosebt.com

<http://www.rosebt.com>

