

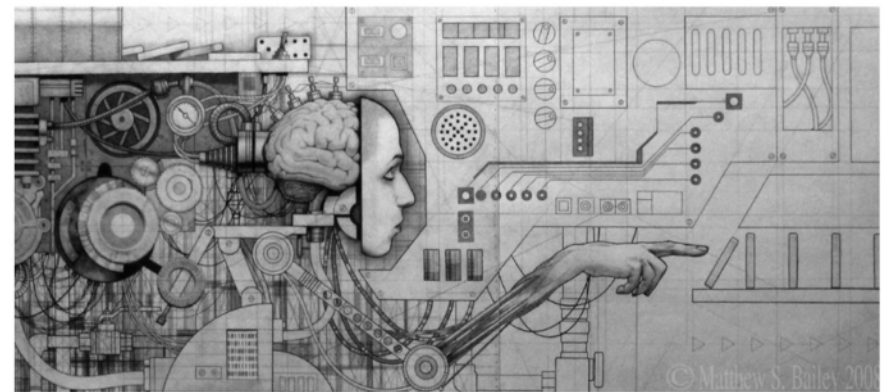
What is Data Science And How You Can Leverage ML to Drive Revenue

A Straightforward, Short, and Non - Academic Approach
for a Non – Technical Audience

By John Thomas Foxworthy

M.S. in Data Science

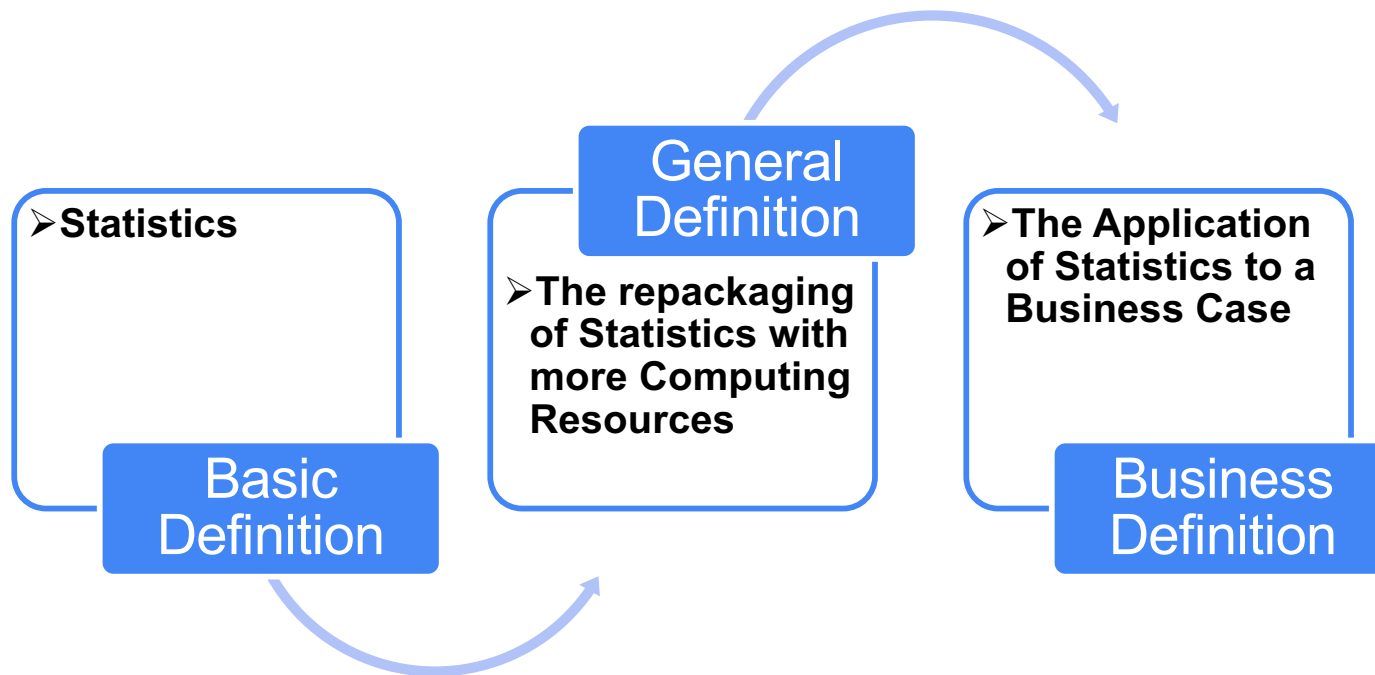
<http://linkedin.com/in/john-t-foxworthy-ms-data-science-1718073>



- ▶ 1. Definitions
- 🏷️ 2. Relabeling
- 💬 3. Motto
- 👥 4. Job Titles
- 🕒 5. Origin
- 📋 6. Some Use Cases for the Business
- 🔍 7. Models and Methodologies
- ❌ 8. Most Common Mistakes
- 🔄 9. Data Science Workflow
- 🎤 10. Data Governance

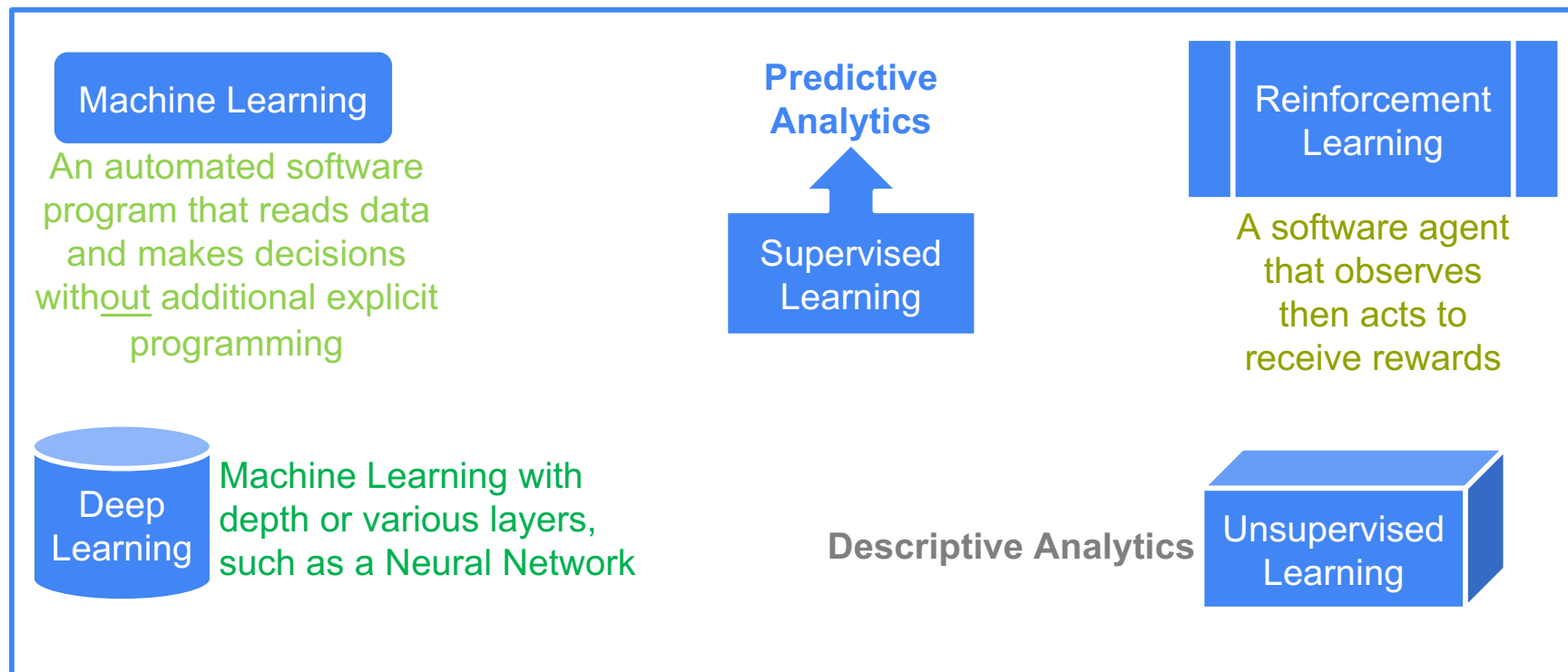


1. Definition: What is Data Science?



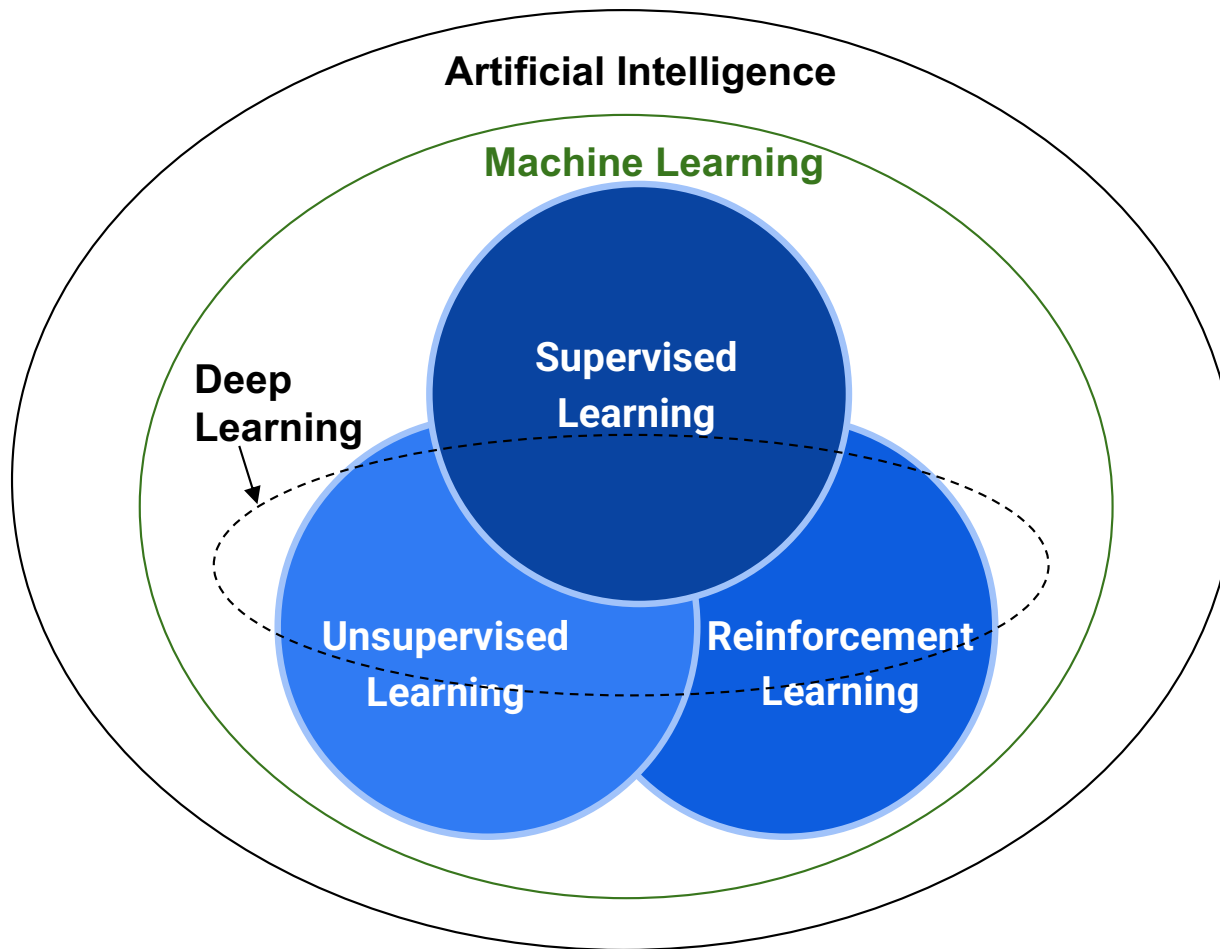
1. Definition: What are types of Data Science?

Artificial Intelligence is anything that is **not** biological that behaves biological



Learning refers to fitting mathematical equations or the estimation process

1. The Current State of Definitions of Data Science



AI encompasses everything in ML

ML encompasses everything in Deep Learning, Supervised Learning, Unsupervised Learning, and Reinforcement Learning

There is either overlap or standalone states for Supervised Learning, Unsupervised Learning, and Reinforcement Learning

Deep Learning is an *extension* of Supervised Learning, Unsupervised Learning, and Reinforcement Learning

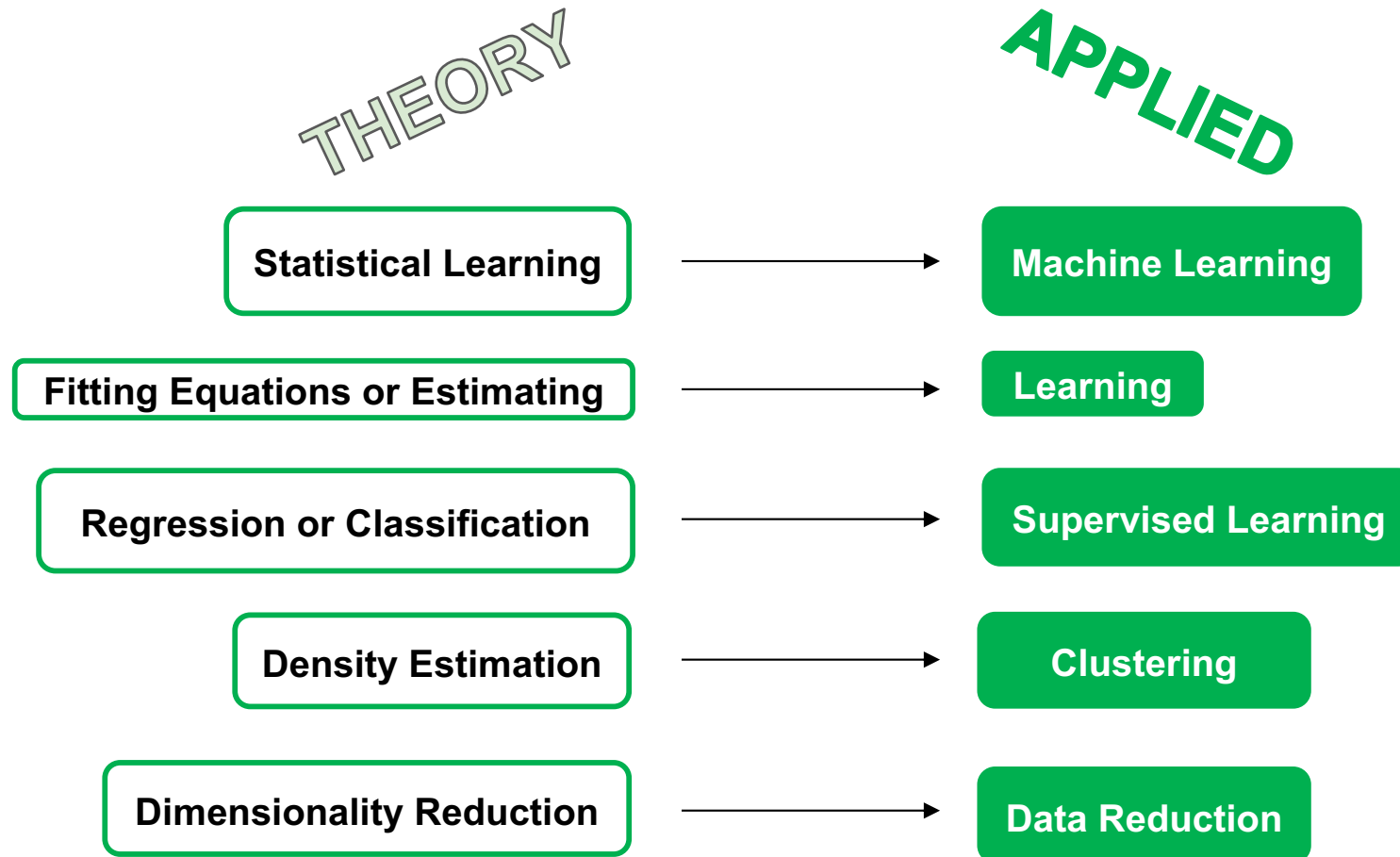
1. Applied (Un)Supervised Learning

Supervised refers to managing the instructions of the algorithm to predict a target

Subject Area	(*) Unsupervised Learning	Supervised Learning
Business	Inputs	Inputs & Outputs
Engineering	Drivers	Drivers & Outcomes
Mathematics	Regressors	Regressors & Regressands
Statistics	Independent Variables	Independent Variables & Dependent Variables
Psychometrics	Predictors	Predictors & Responses
General Science	Explanatory	Explanatories & Focuses
Linguistics	Descriptive	Descriptives & Predictive
Machine Learning	Unlabeled Training Data	Unlabeled Training Data & Labeled Training Data

(*) **Unsupervised** refers to the **lack** of instructions to manage the algorithm to predict a target because it is descriptive analytics, not predictive analytics.

2. Relabeling Defines Data Science and its Purpose



3. Data Science Motto

**All models are wrong,
but some are useful.**

George Box (1976)

4. Job Titles and the Data Science Profession

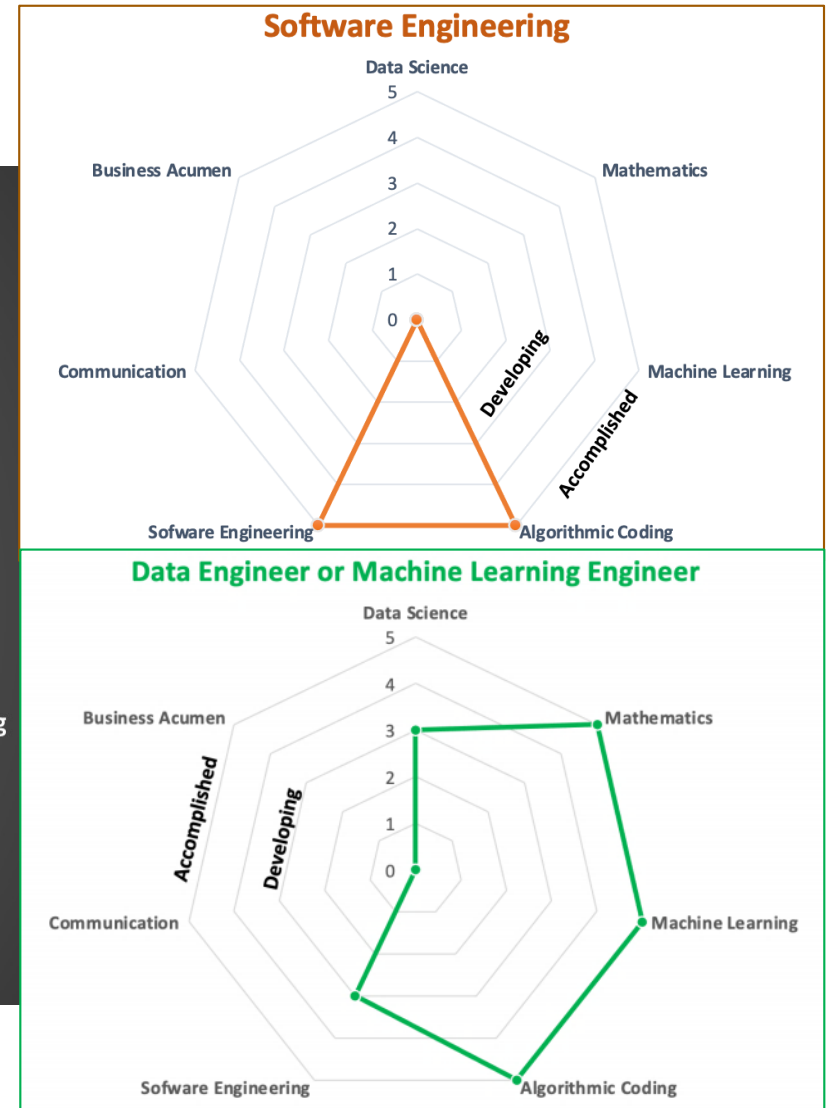
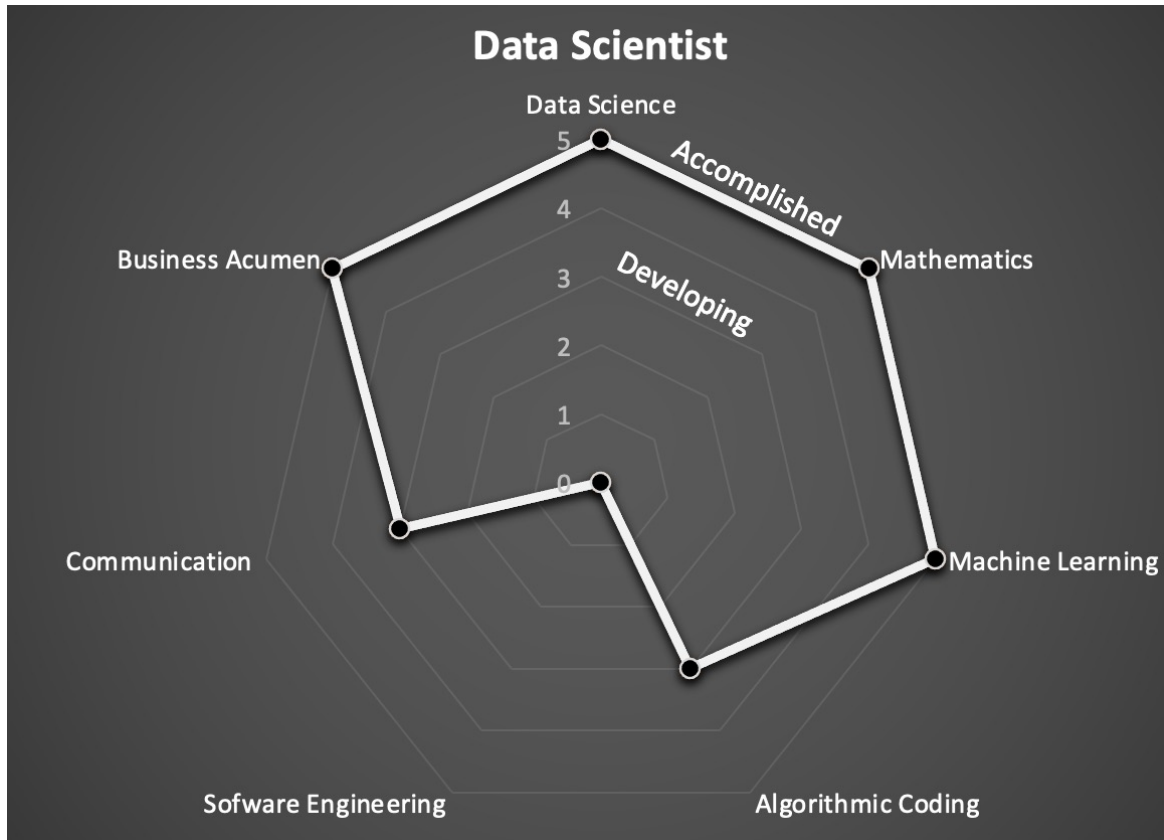
Prior Data Scientist Job Titles
Statistician
Operation Researcher
Econometrician (Statistical Economist)
Computational Linguist (Language Scientist)
Clinical Psychologist (Psychometrician)
Image Processing Engineer
Computer Scientist (*)

(*)
The overwhelming majority of Computer Scientists are **not** Data Scientists, but a small number of Computer Scientists are **exceptional** Data Scientists.

Wrong Data Scientist Job Titles
Software Engineer
Software Developer
Business Systems Analyst
Data Analyst
Data Engineer
IT Project Manager
Chief Technology Officer (CTO)
Computer Scientist (*)

(*) Short Sample Write-Up to Clarify the Difference: <https://medium.com/p/a04497543052>

4. Job Titles and their Skill Sets



4. Job Titles and their Skill Sets



The Labor Market for Data Scientists has more risk for organizations

- ❑ There are **more** people with a quantitative background who want to be Data Scientists than **real** Data Scientists by a wide margin. **The risk of the blind leading the blind is always present.**

Examples

- A Software Engineer **claims** that he or she can explain anything if they can code it up.
- A Physics Ph.D. **may oversimplify** Data Science after reading a book on Data Science.
- A Data Analyst **claims to be** a Data Scientist after reading and interpreting a dataset.

Hearsay

Oversimplification

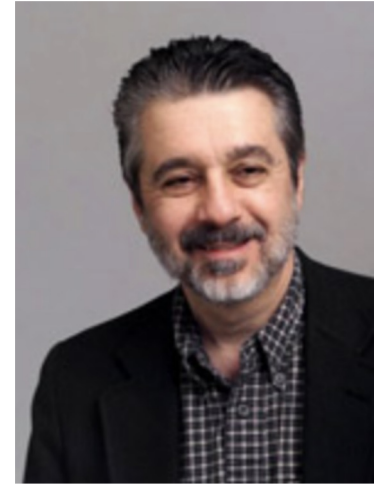
Bias

Underestimation

5. Where did Data Science come from?

William S. Cleveland

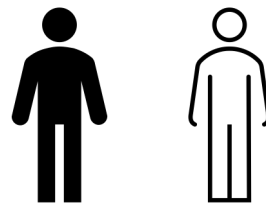
Professor of Statistics at Purdue University, Indiana






- ❖ Vote at the Statistical Symposium in 2001
- ❖ Reasons provided for the establishment of Data Science.
 - ✓ The processing power of computing
 - ✓ The quantity and quality of data
 - ✓ The first source of Data Science: <https://www.jstor.org/stable/1403527> (The International Statistical Institute)
 - The first official sentence of Data Science: *“This document describes a plan to enlarge the major areas of technical work of the field of statistics. Because the plan is ambitious and implies substantial change, the altered field will be called “data science””.*
- ❖ The term “Data Science” did **not** become popular, according to Google Search Analytics, until 2010
 - ✓ From 2010 to the present time, Data Science projects were slowly funded first, then exponentially

Question

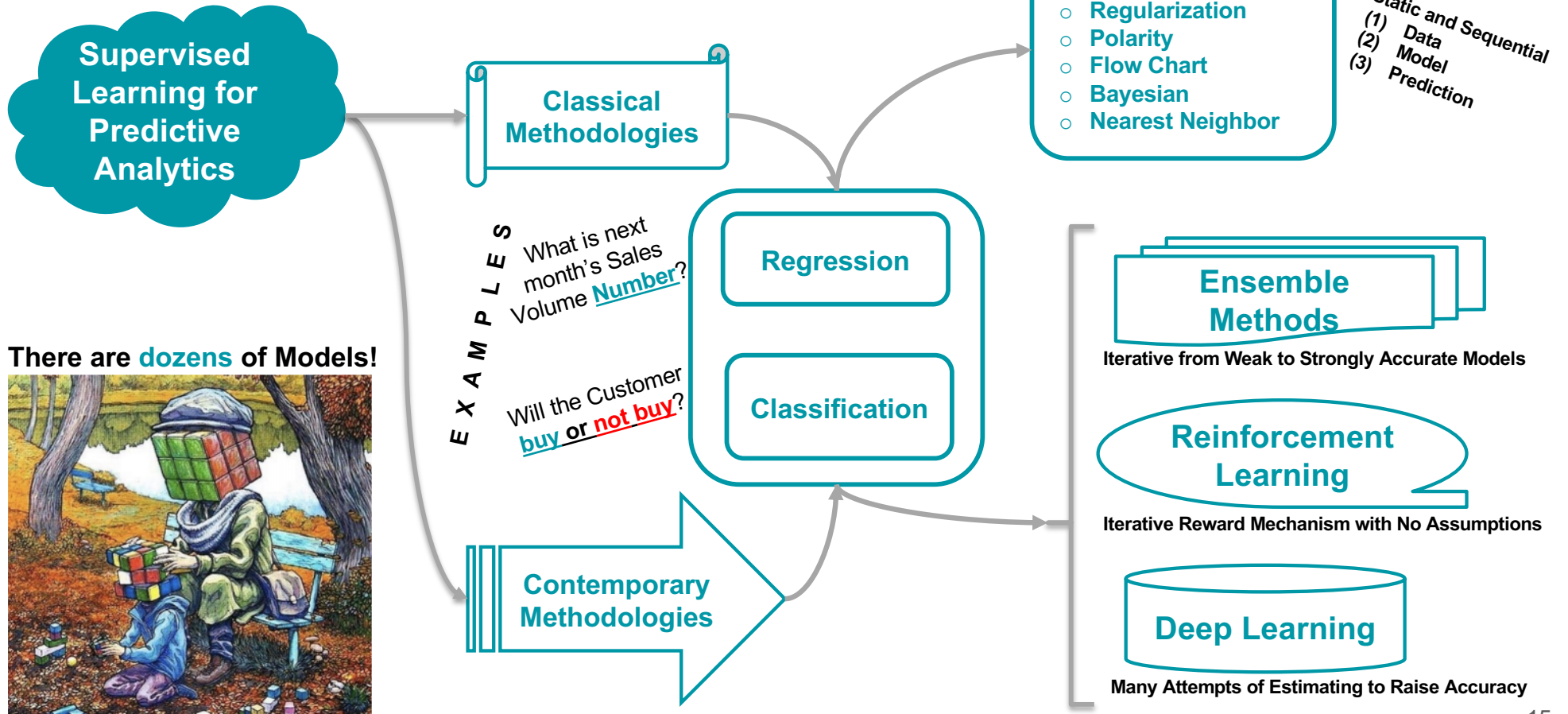
**If you clone a human being, then
does the clone have artificial intelligence?**



6. Some Data Science Use Cases for the Business

 Business Use Case	 Data Science Model	 Improvement
Fraud Detection	Anomaly Detection as an Unsupervised Learning Model	Credit Risk Reduction
Propensity to Buy New Product Launch or Existing Product with New Feature Launch	Predictive Analytics with a Classification Model	Revenue Generator for Product Management
Sentiment Analysis How people outside the company view your organization and its reputation	Natural Language Processing (NLP) on all news reports on the company	Executives have an outside view or attitude of the company whether positive or negative
Customized Recommendations For individual customers with no assumptions about any customer	Reinforcement Learning to suggest what customers want in the company's product line	Increases the Return on Investment (ROI) and Profit Margins for the company
Explain Driving Factors What are the inputs to explain a \$1 Billion Valuation as a Unicorn for Venture Capital Investment?	Explanatory AI Regression Model with Automated Factor Selection to reveal what drives a successful Start-Up from a set of various Start-Ups	Ensures the the Return on Investment (ROI) and Maximizes Profit for investors

7. Models and Methodologies



6. Unsupervised Learning and Descriptive Analytics

(Use Cases)

Clustering

Describe Structures in Datasets

Data Reduction Techniques

Pre – Model Preparation

Anomaly Detection

Fraud Detection



Feature Selection Automation

Inputs for Explainable AI and Predictive Analytics

Autoencoder

File Compression to Reduce Data Noise

Group Segmentation

Customer Segmentation

Should you ask your Supervisor about Unsupervised Learning? Yes!

8. Common Data Science Mistakes

Always Inaccurate

Reselect Inputs
Change Model

Confusing Cause and Effect

Endogeneity or Simultaneity

The Old Way does not work anymore

Regime Switching

Too Many Similar Features

Multicollinearity or try Deep Learning

Same Model, Different Results

High Variance

Overgeneralizing

Overfitting

Missing Data causes No Prediction

Wrong Model

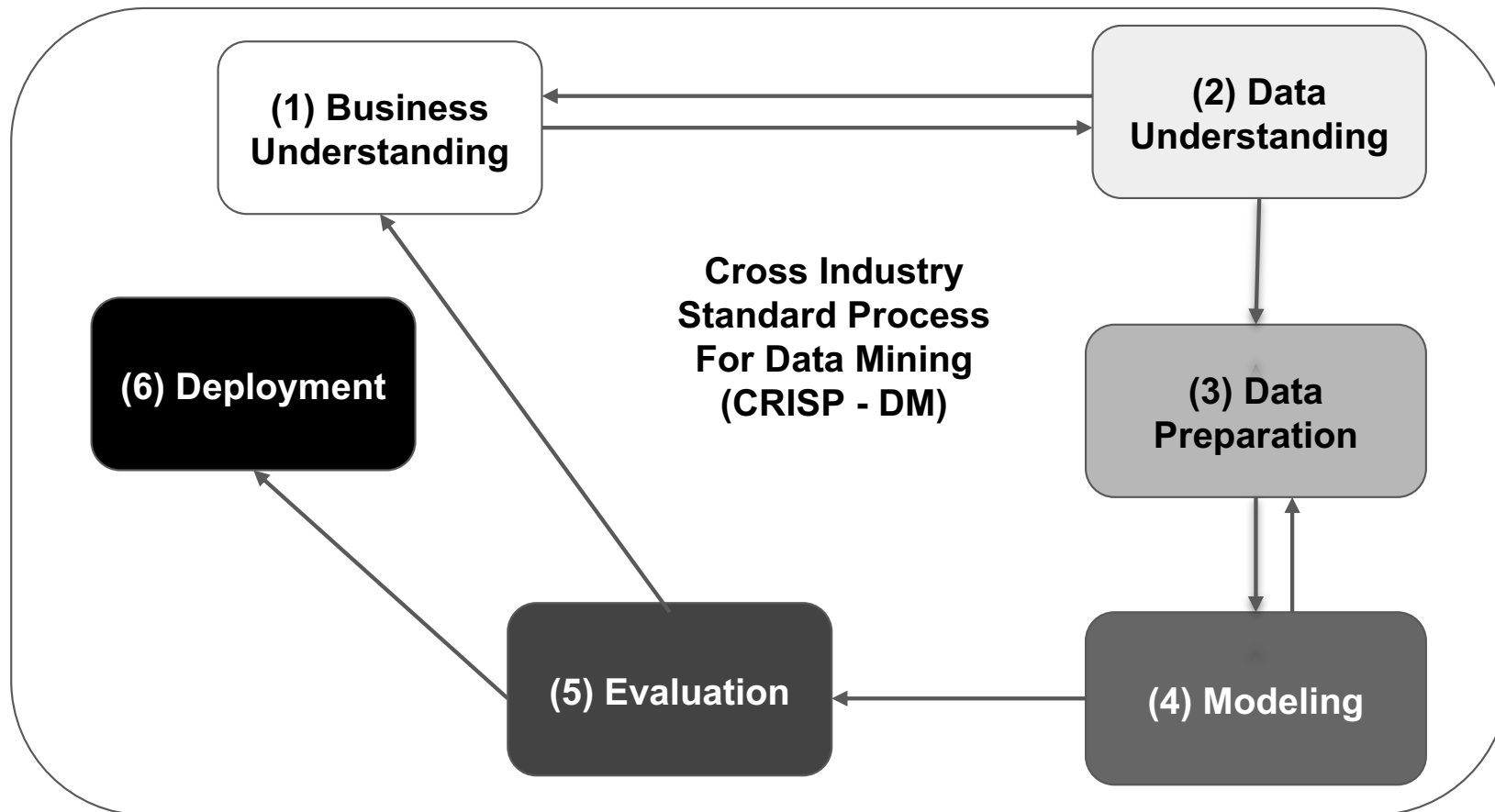
More Data, Less Usefulness

High Bias

Misclassification

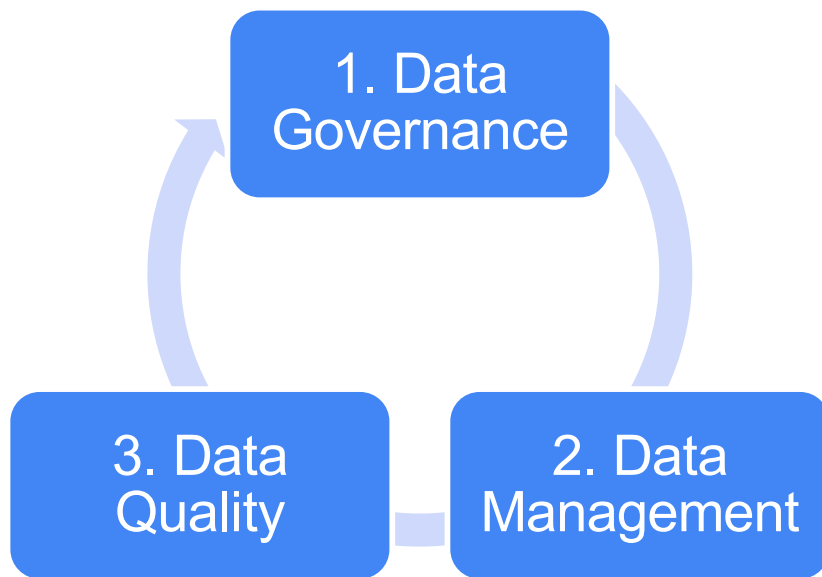
Type I, False Positive

9. Generalized Data Science Workflow



Similar to Agile Software Development, but more Business Interaction and more Business Impact

10. Data Science Governance



All the practices and processes that ensure the management of data assets in an organization is **Data Governance**.

The execution of **Data Governance** is **Data Management**.

The highest priority of **Data Management** is **Data Quality**.

The definition of **Data Quality** is **Usefulness**.

Short Sample Write Up: <https://medium.com/codex/data-governance-from-a-data-scientist-dab4a80cd551>

**Thank you, and I hope I have raised your
knowledge base and interest in Data Science.**

By John Thomas Foxworthy

M.S. in Data Science

<http://linkedin.com/in/john-t-foxworthy-ms-data-science-1718073>

