Introduction to Machine Learning

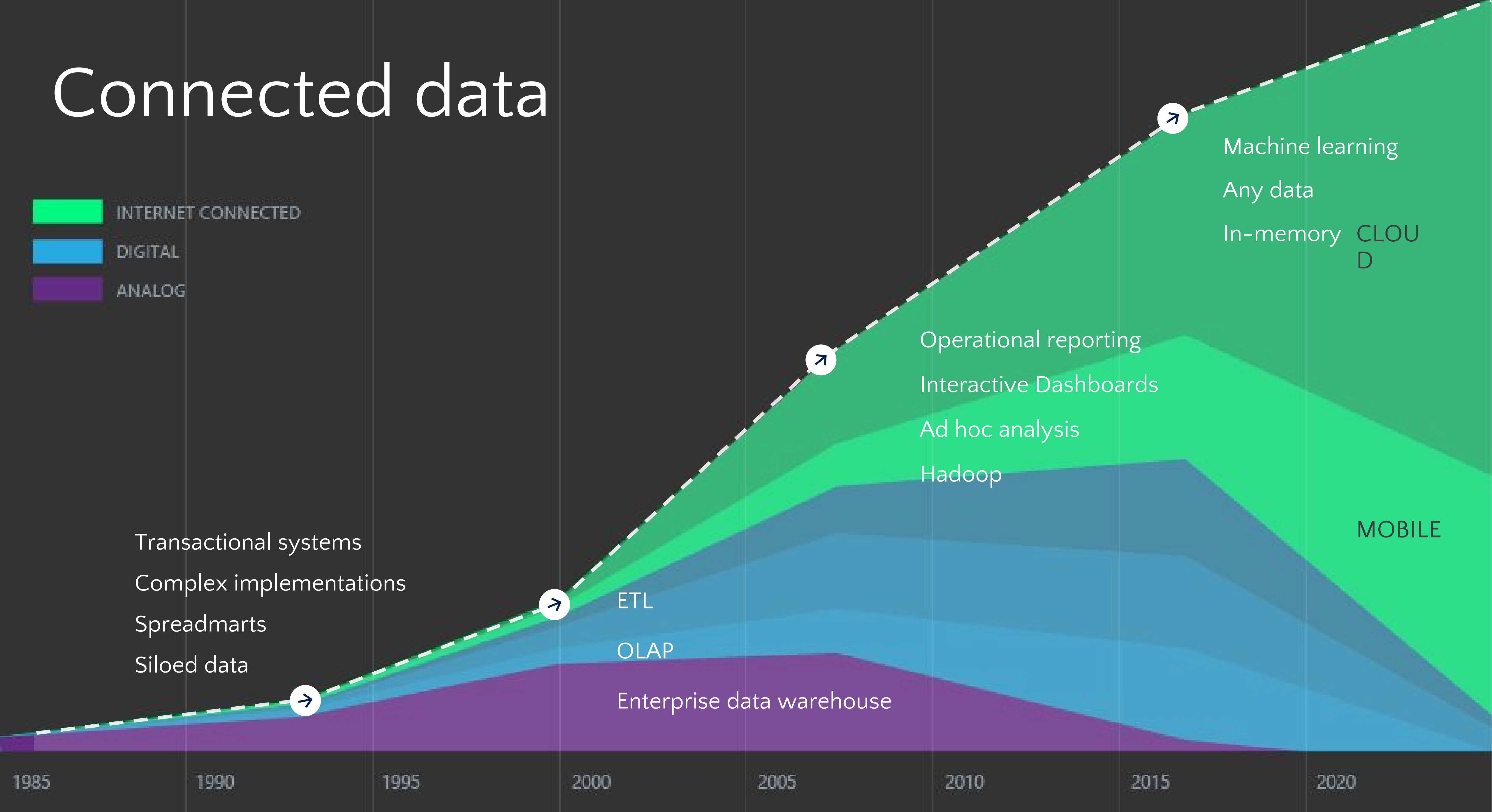
Class Expectations

- Mobile phones are kept in silent mode
- Only ML work, and nothing else.
- No cross talks and parallel discussions
- On time every time
- Hard work is the only shortcut !
- Less theory and more practice

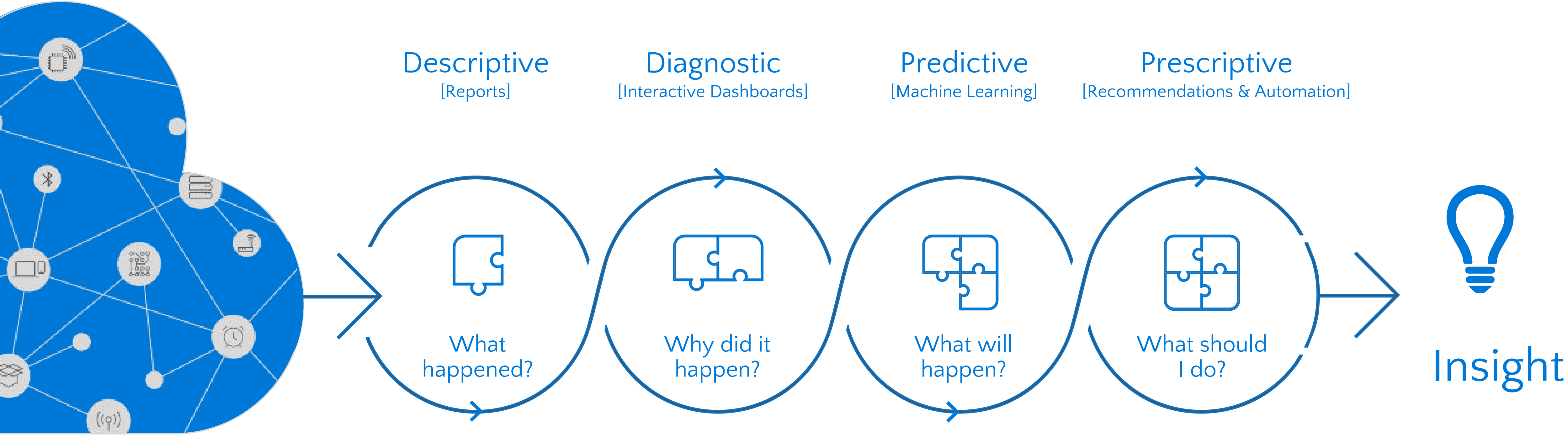
Key expected outcome

- Learn application of ML algorithms in R
- Utilize and demonstrate ML skills in your work
- Clear understanding of Data Science
- Application of Data Visualization in R

Connected data



From data to decisions and actions



What is learning?

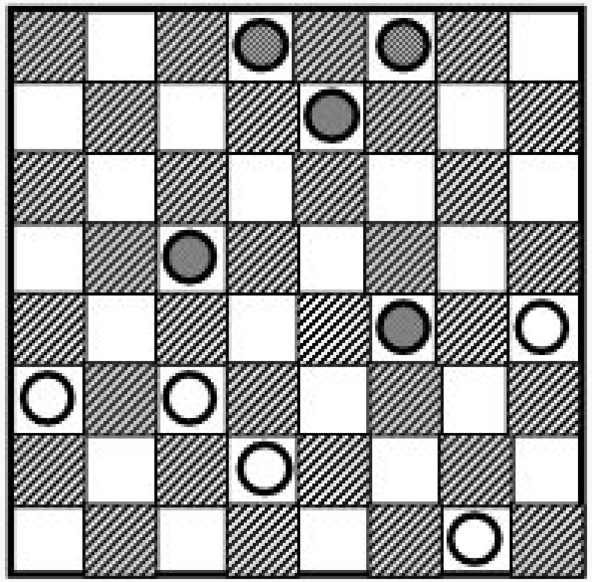
- “Learning denotes changes in a system that ... enable a system to do the same task more efficiently the next time.” –Herbert Simon
- “Learning is any process by which a system improves performance from experience.” –Herbert Simon
- “Learning is constructing or modifying representations of what is being experienced.”
  –Ryszard Michalski
- “Learning is making useful changes in our minds.” –Marvin Minsky
- “The goal of machine learning is to build computer systems that can adapt and learn from their experience.”–Tom Dietterich

So what is Machine Learning ?

Instead of writing programs we collect lots of examples that specify the correct output for a given input. A machine learning algorithm then takes these examples and produces a program that does the job.

A Bit of History

- Arthur Samuel (1959) wrote a program that **learnt** to play draughts (“checkers” if you’re American).

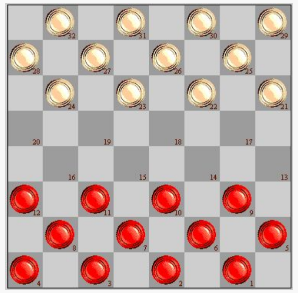


1940s
Human reasoning / logic first studied as a formal subject within mathematics (Claude Shannon, Kurt Godel et al).

1950s
The “Turing Test” is proposed: a test for true machine intelligence, expected to be passed by year 2000. Various game-playing programs built. 1956 “Dartmouth conference” coins the phrase “artificial intelligence”.

1960s
A.I. funding increased (mainly military). Famous quote: “Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved."

Ax. 1. $P(\varphi) \wedge \Box \forall x[\varphi(x) \to \psi(x)] \to P(\psi)$
Ax. 2. $P(\neg\varphi) \leftrightarrow \neg P(\varphi)$
Th. 1. $P(\varphi) \to \Diamond \exists x\, [\varphi(x)]$
Df. 1. $G(x) \iff \forall\varphi[P(\varphi) \to \varphi(x)]$
Ax. 3. $P(G)$
Th. 2. $\Diamond \exists x\, G(x)$
Df. 2. $\varphi \text{ ess } x \iff \varphi(x) \wedge \forall\psi\{\psi(x) \to \Box\forall x[\varphi(x)\to\psi(x)]\}$
Ax. 4. $P(\varphi) \to \Box\, P(\varphi)$
Th. 3. $G(x) \to G \text{ ess } x$
Df. 3. $E(x) \iff \forall\varphi[\varphi \text{ ess } x \to \Box\, \exists x\, \varphi(x)]$
Ax. 5. $P(E)$
Th. 4. $\Box\, \exists x\, G(x)$



Turing Test

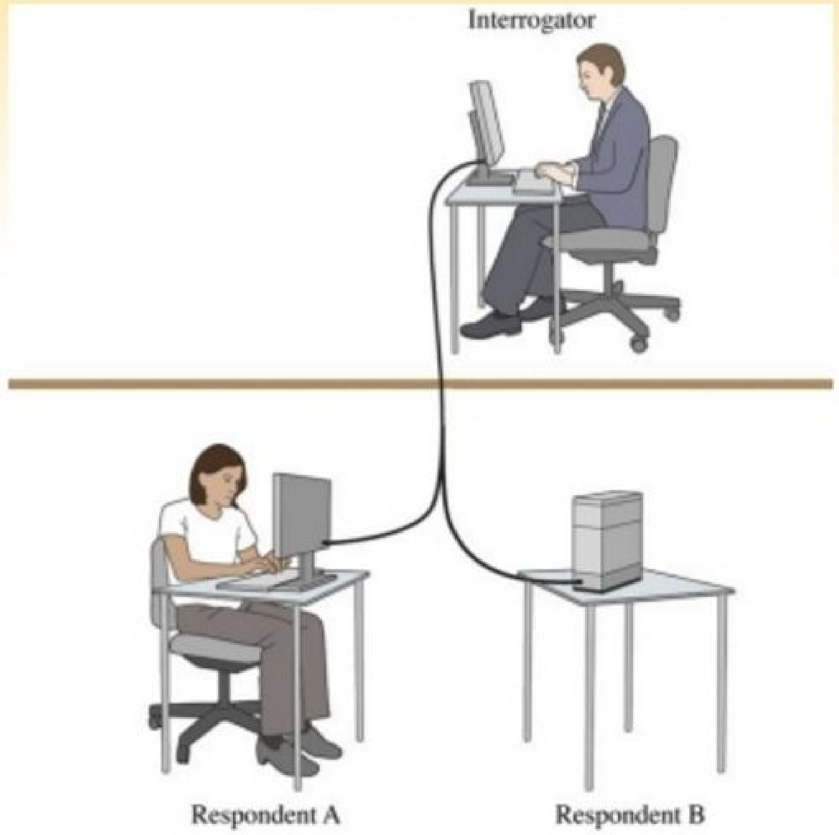
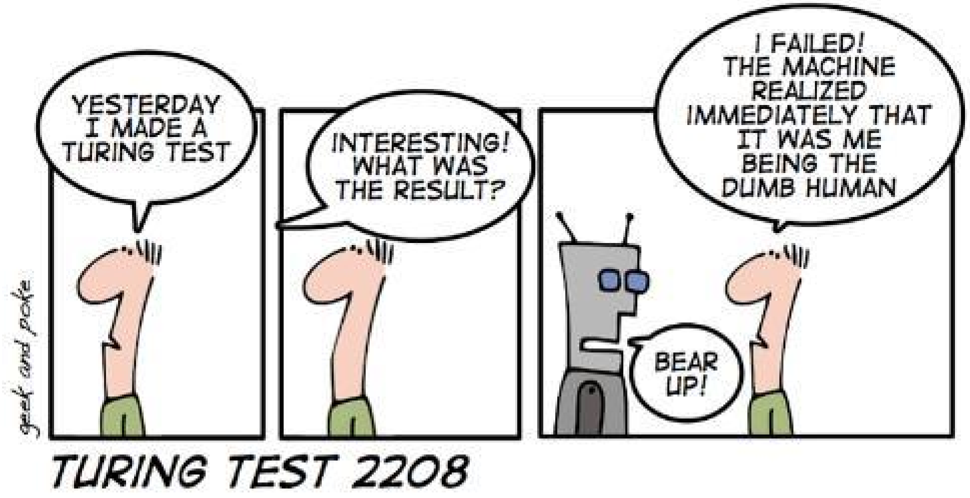


Figure 13.2
In a Turing test, the interrogator must determine which respondent is the computer and which is the human

The “Turing Test” is proposed in 1950s by Alan Turing: a test for true machine intelligence, expected to be passed by year 2000. Various game-playing programs built. 1956 “Dartmouth conference” coins the phrase “artificial intelligence”.

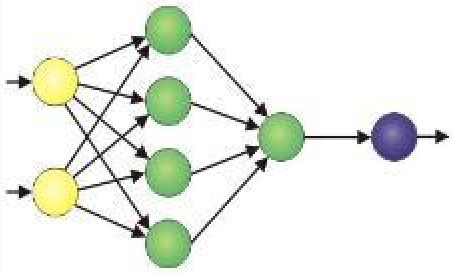


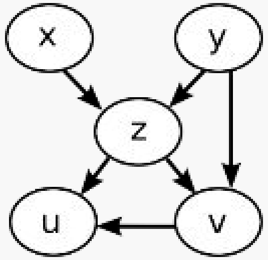
1970s
A.I. “winter”. Funding dries up as people realise it’s hard.
Limited computing power and dead-end frameworks.

1980s
Revival through bio-inspired algorithms: Neural networks, Genetic Algorithms.
A.I. promises the world – lots of commercial investment – mostly fails.
Rule based “expert systems” used in medical / legal professions.

1990s
AI diverges into separate fields: Computer Vision, Automated Reasoning,
Planning systems, Natural Language processing, **Machine Learning**…

…Machine Learning begins to overlap with statistics / probability theory.





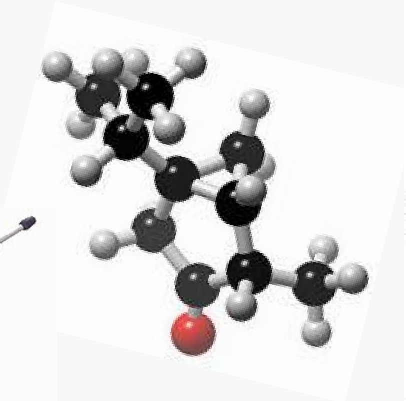
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

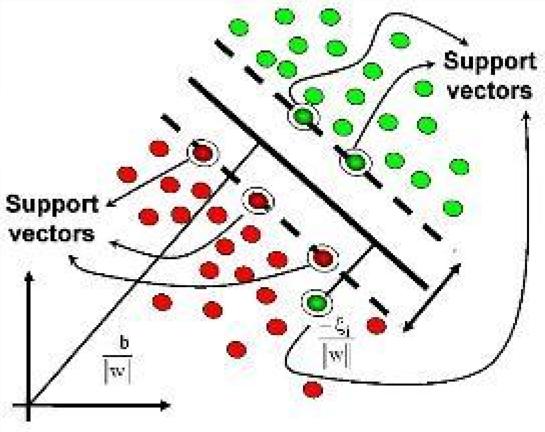
2000s

ML merging with statistics continues. Other subfields continue in parallel.

First commercial-strength applications: Google, Amazon, computer games, route-finding, credit card fraud detection, etc…

Tools adopted as standard by other fields e.g. biology





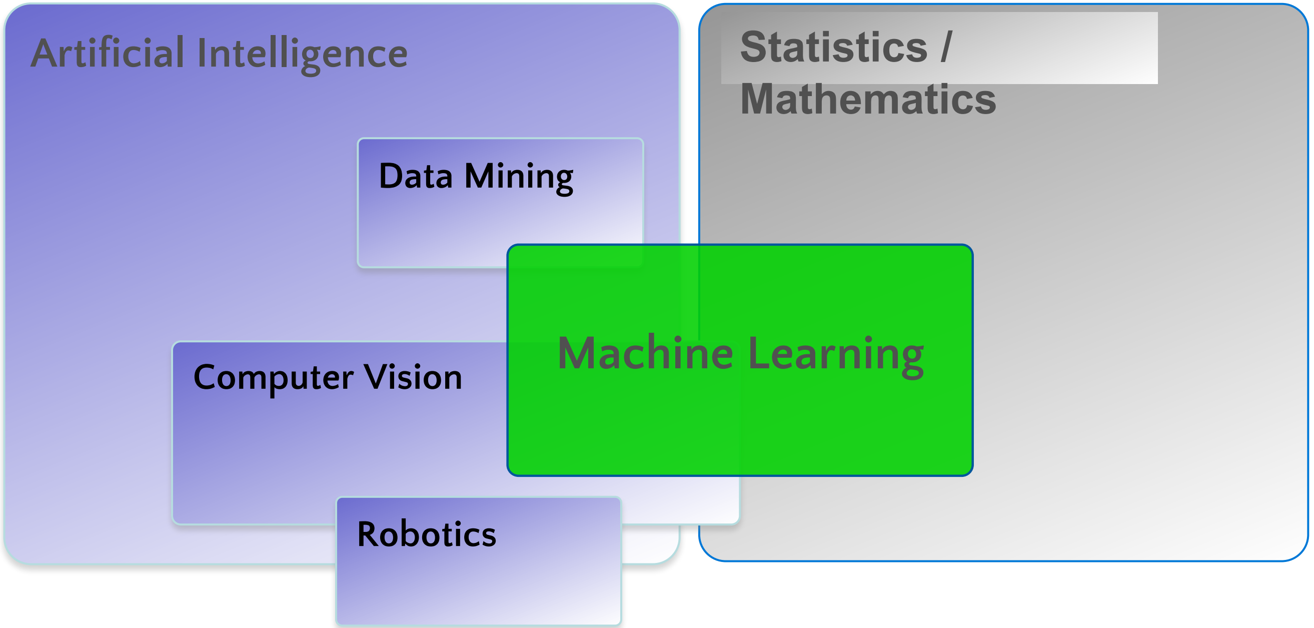
2010s…. IBM Watson, Siri,
Computer Wins GO, Robotics, etc.

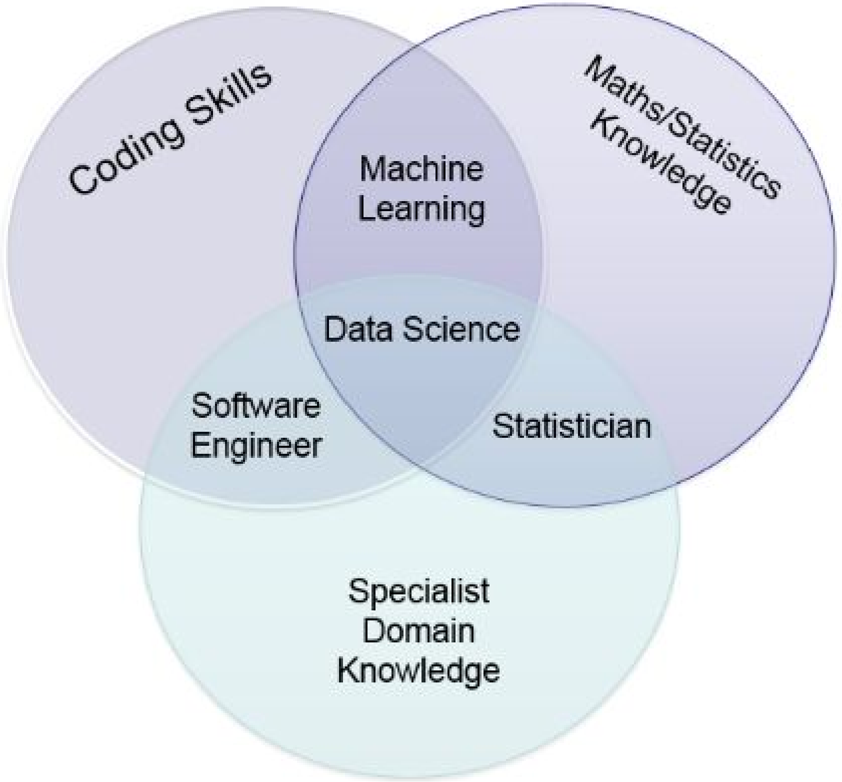
Brief timeline of Machine Learning

- 1950s
  - **Samuel's checker player**
  - Selfridge's Pandemonium
- 1960s:
  - **Neural networks: Perceptron**
  - **Minsky and Papert prove limitations of Perceptron**
- 1970s:
  - Expert systems and the knowledge acquisition bottleneck
  - Mathematical discovery with AM
  - Symbolic concept induction
  - Winston's arch learner
  - Quinlan's ID3
  - Michalski's AQ and soybean diagnosis
  - Scientific discovery with BACON
- 1980s:
  - **Resurgence of neural networks (connectionism, backpropagation)**
  - Advanced decision tree and rule learning
  - Explanation-based Learning (EBL)
  - Learning, planning and problem solving
  - Utility theory
  - Analogy
  - Cognitive architectures
  - Valiant's PAC Learning Theory
- 1990s
  - **Data mining**
  - **Reinforcement learning (RL)**
  - **Inductive Logic Programming (ILP)**
  - **Ensembles: Bagging, Boosting, and Stacking**
- 2000s
  - **Kernel methods**
    - Support vector machines
  - **Graphical models**
  - Statistical relational learning
  - Transfer learning
  - **Deep learning**
- Applications
  - Adaptive software agents and web applications
  - Learning in robotics and vision
  - E-mail management (spam detection)
  - …

Machine Learning

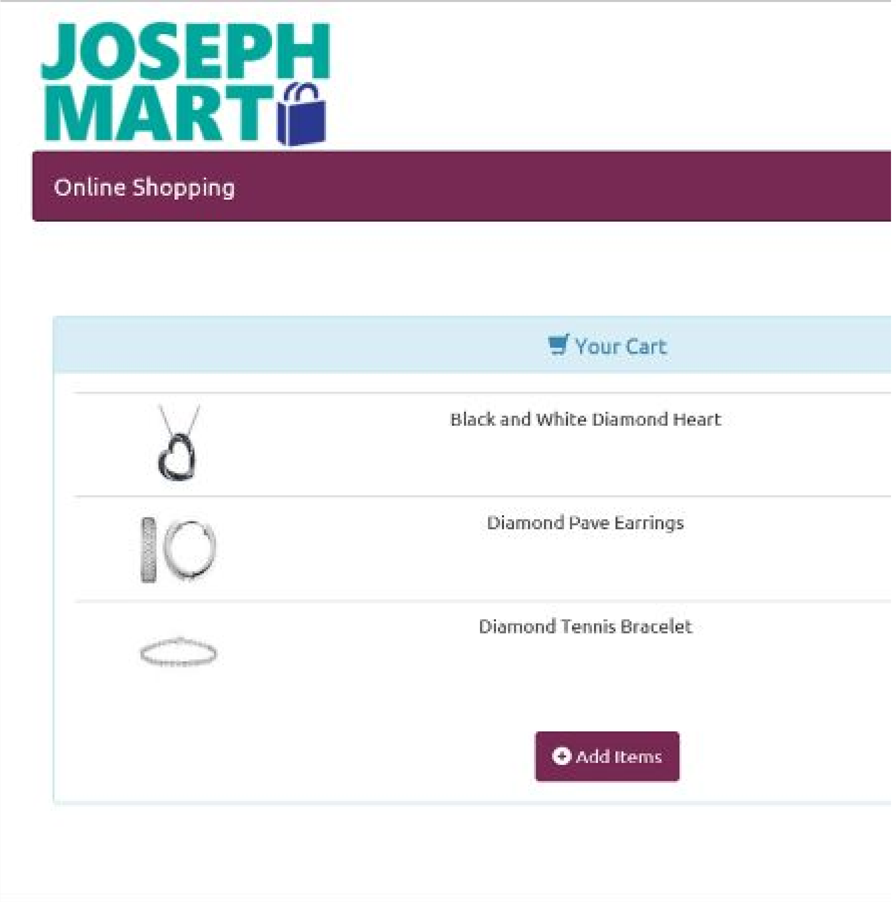
*Where does it fit? What is it **not**?*



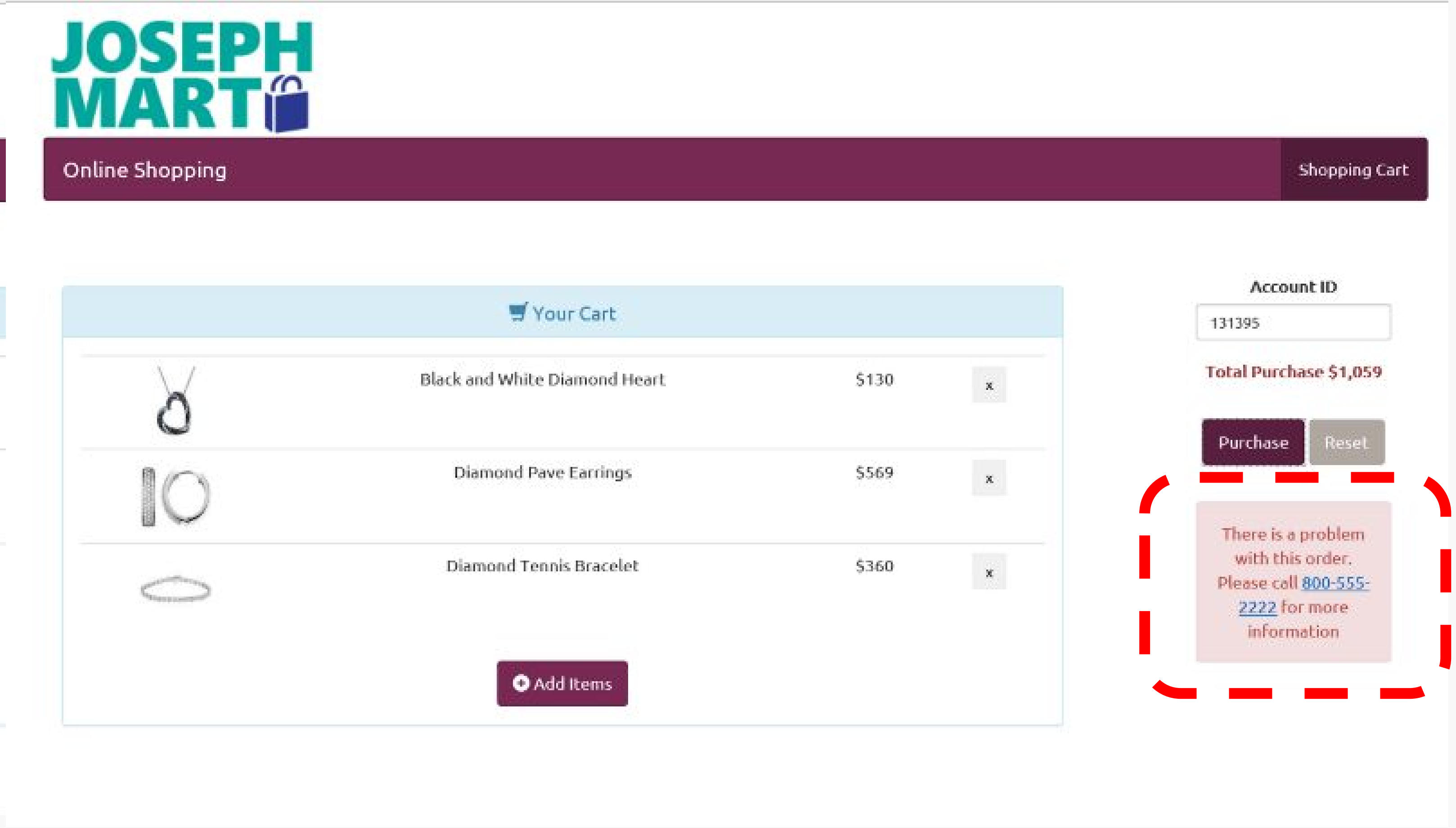


Application before and after

Before



With Intelligence, handle suspect



Real-world Examples of Intelligent Apps

10% conversion uplift by recommending high propensity to buy

Large computers and related products/services

Delight Loyal customers through personalized offers

Pier1 imports

Save $300,000 a day through predictive maintenance

Rockwell Automation

Understand six years of patient data in a few hours

Leeds General Infirmary

ImagineCare using ML & Dynamics for improved patient health care

Dartmouth-Hitchcock

Offer personalized service across 1,400 locations

Ziosk

Prevent electrical blackouts across thousands of buildings

eSmart SYSTEMS

Reduce cost: Predict insurance ETA and need for follow-up

GAFFEY HEALTHCARE

Virtually, Every Business App can be made Intelligent

Sales and marketing

- Customer Acquisition
- Cross-sell and upsell
- Loyalty programs
- Marketing mix optimization

Finance and risk

- Fraud detection
- Credit risk management

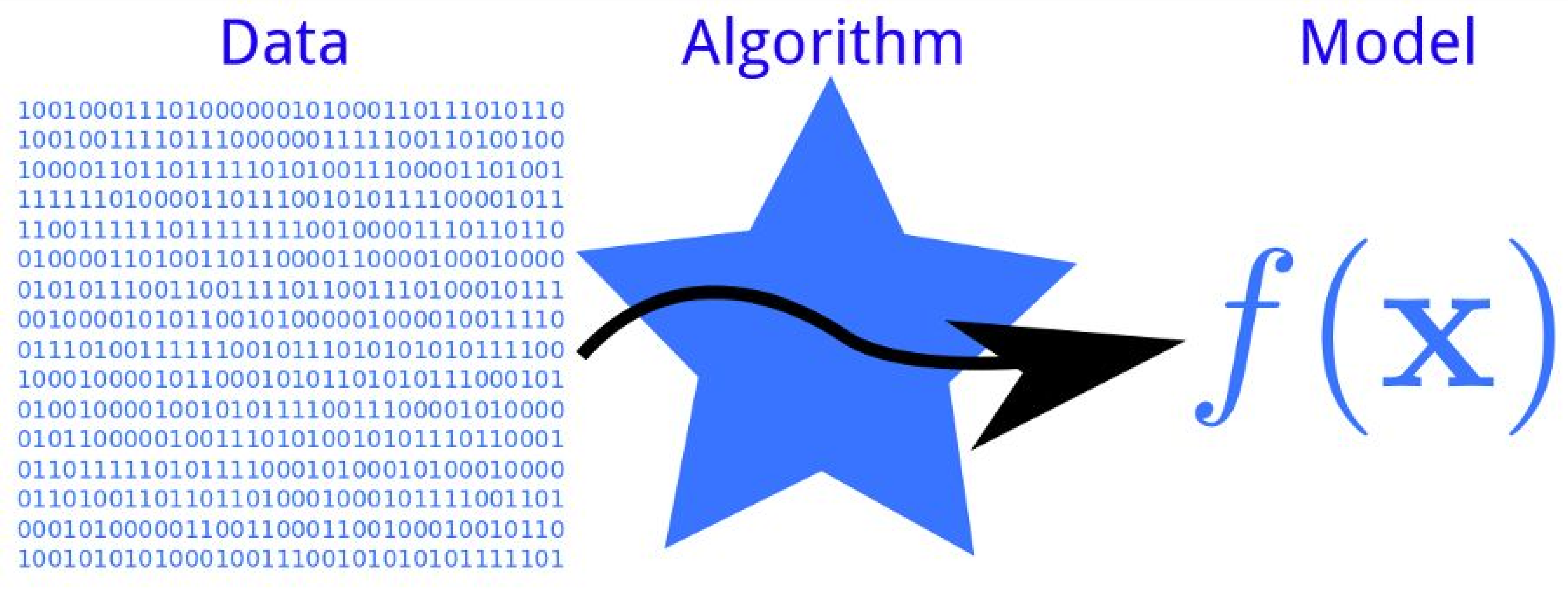
Customer and channel

- Lifetime customer value
- Personalized offers
- Product recommendation

Operations and workforce

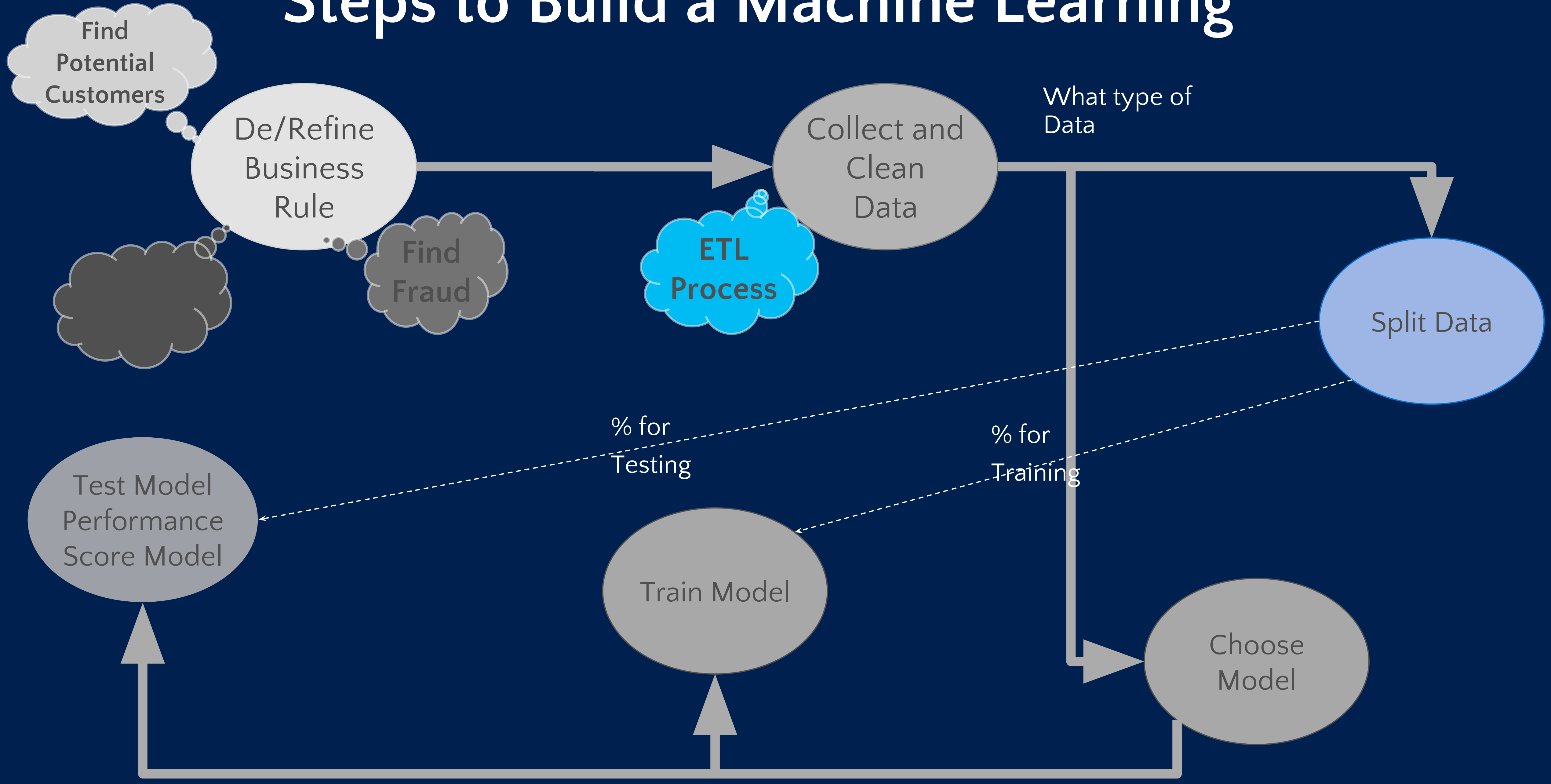
- Pay for performance
- Operational efficiency
- Smart buildings
- Predictive maintenance
- Supply chain management

Machine learning

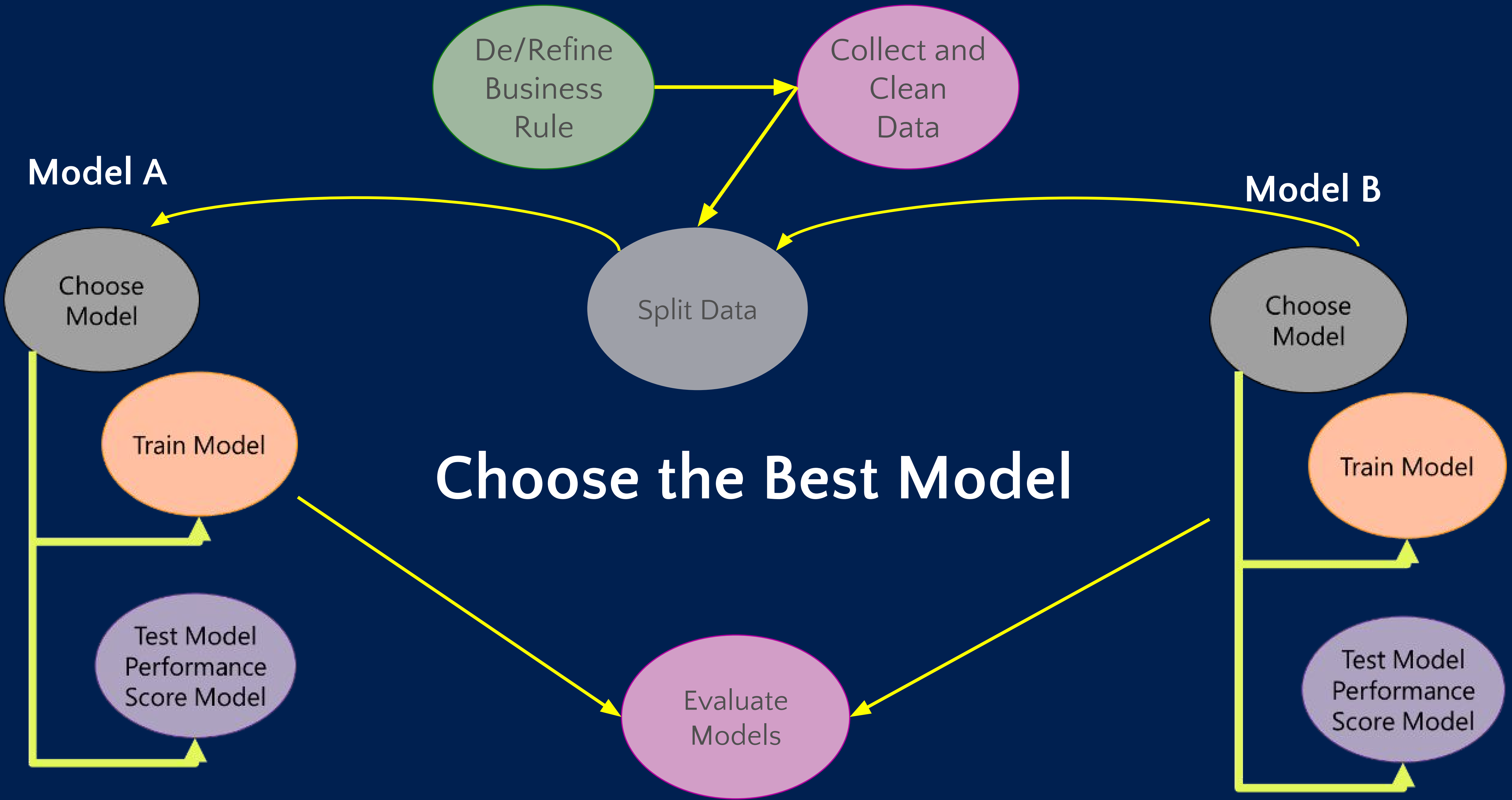


http://phdp.github.io/posts/2013-07-05-dtl.html

Steps to Build a Machine Learning



Choose the Best Model



Data History

Innovation	Business Question	Technologies
Data Collection (1960's)	"What was total revenue in the past 5 years?"	Mainframe computers, tape backup
Data Access (1980's)	"What were unit sales in New England last March?"	RDBMS, SQL, ODBC
Data Warehousing (1990's)	"""What were unit sales in New England last March? Drill down to Boston"	OLAP, multi-dimensional databases, data warehouses
Data Science (Today)	"What's likely to happen to Boston sales next month and why?"	Advanced algorithms, massively parallel databases, Big Data

Data science life-cycle (DSLC)

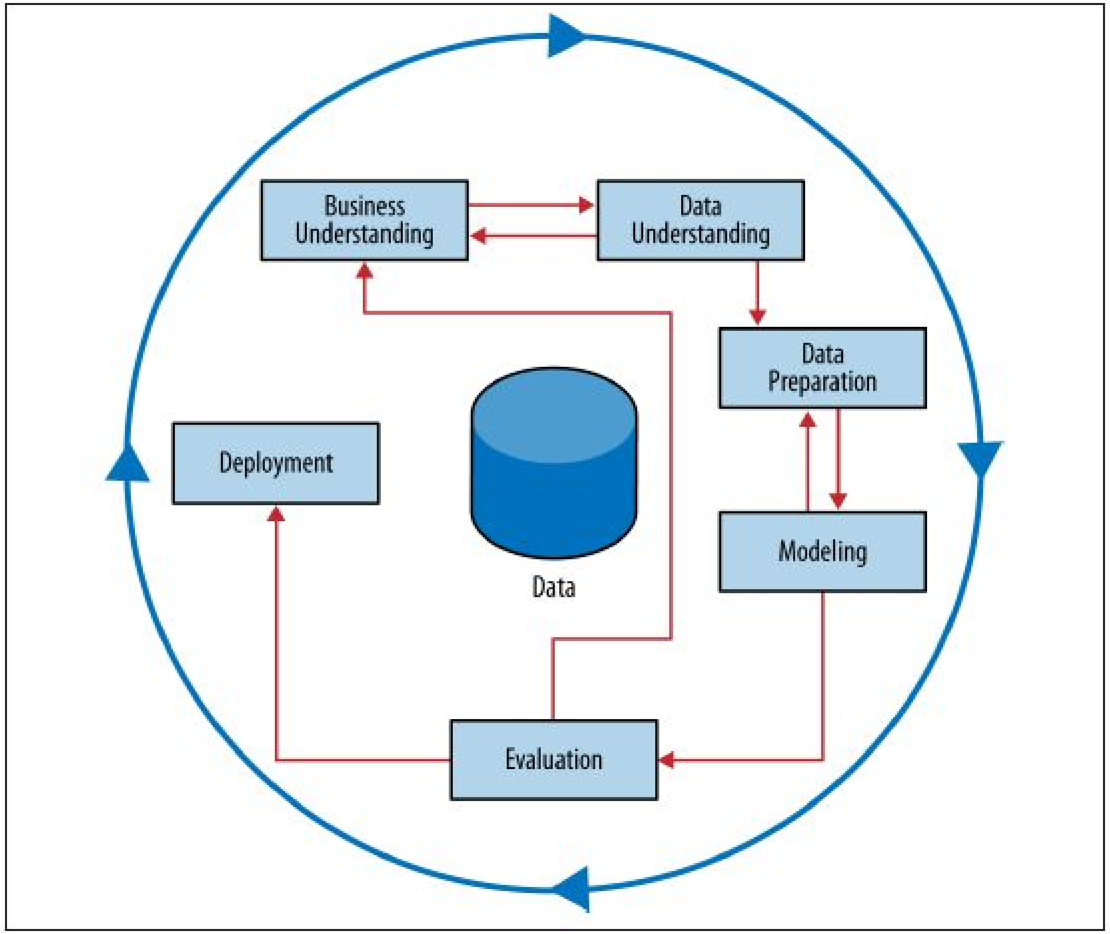
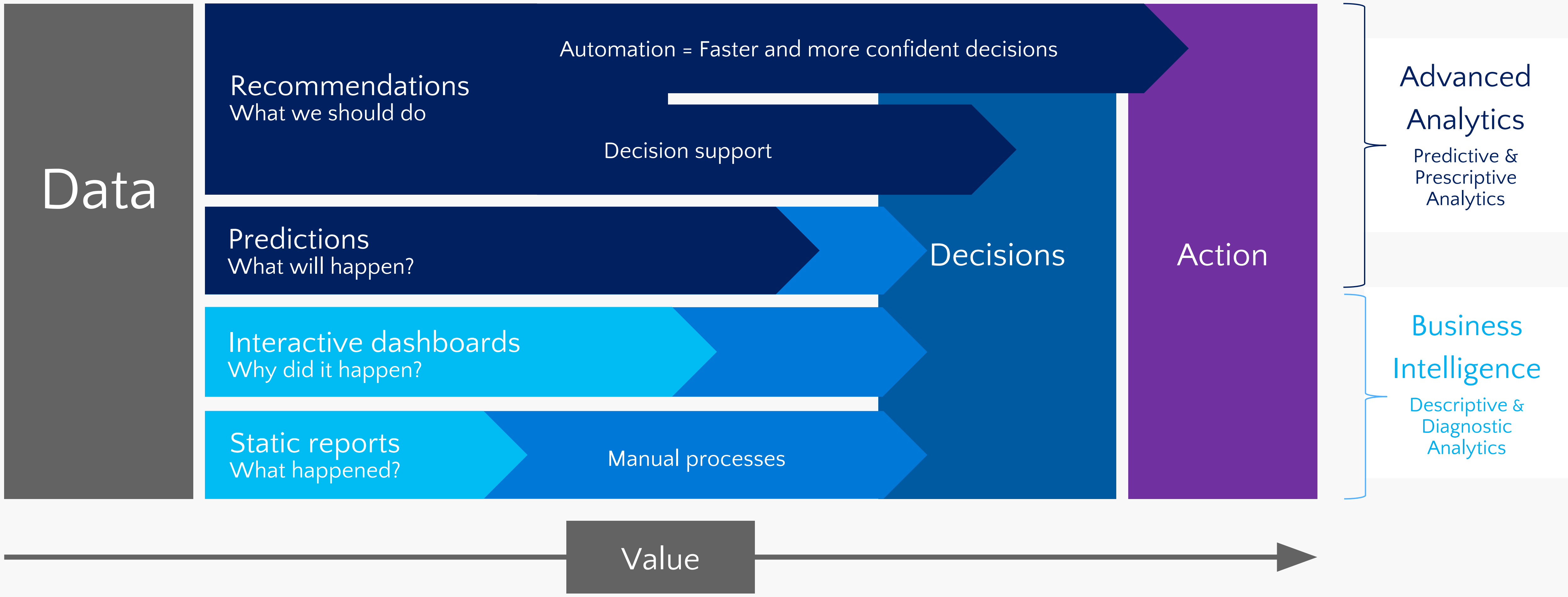


Figure 2-2. The CRISP data mining process.

[Wikipedia page on the CRISP-DM process model](#)

From Data to Decisions and Actions



DSLC vs SDLC

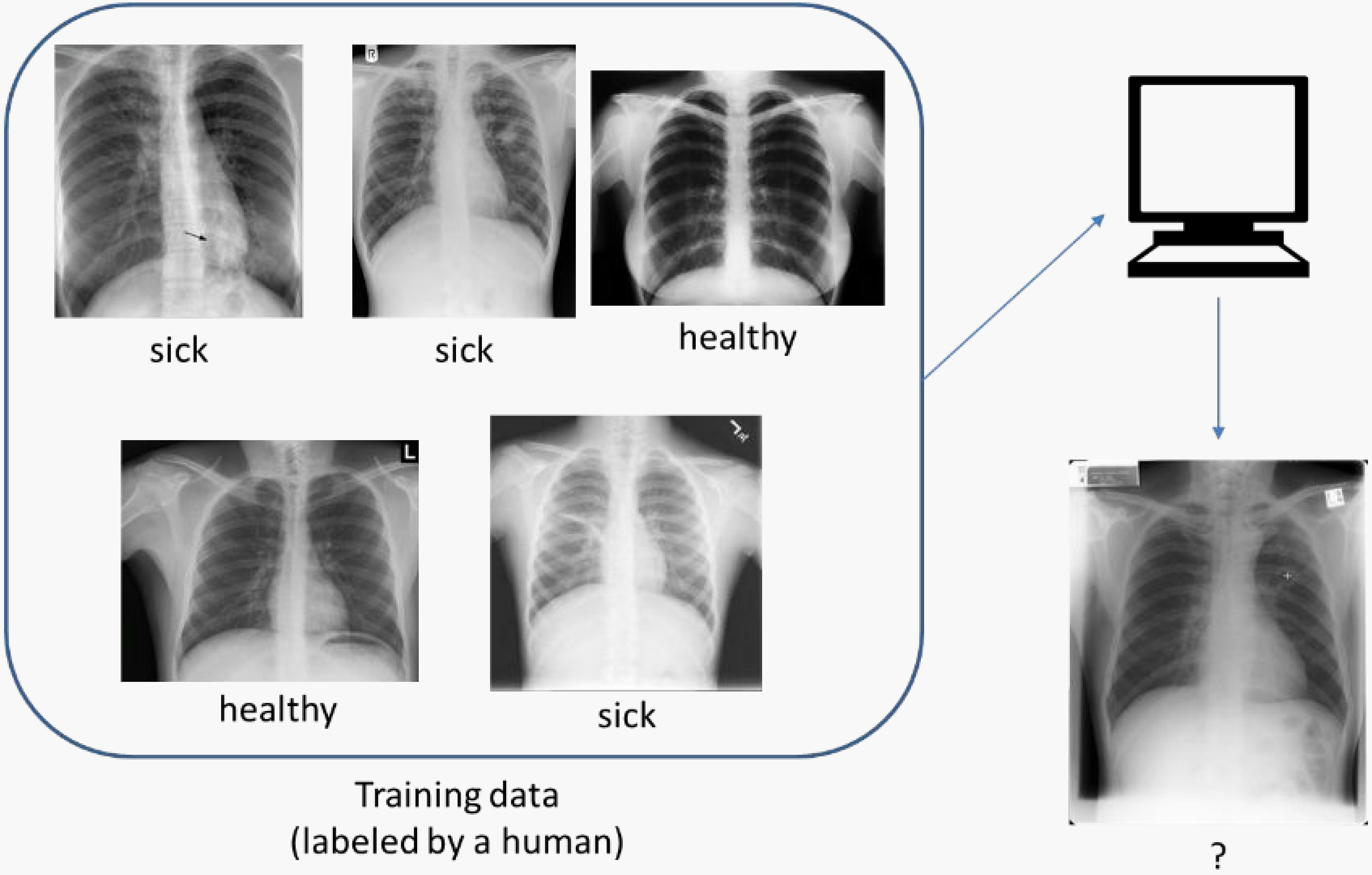
Characteristic	SDLC	DSLC
Begins with	Specs	A vague request
Ends with	Successful UAT	Value realization*
Key components are	Deterministic	Probabilistic
Progress can be measured at	Always (milestones reached)	Typically after 75% effort
Knowledge of science & math	Incidental	Fundamental
Design or build first?	Design first	Build first
Cost of poor design	Cost overrun	Catastrophic
Research & Innovation	In niche areas	Diverse and Applied

Curtsey: Sandeep Rajput

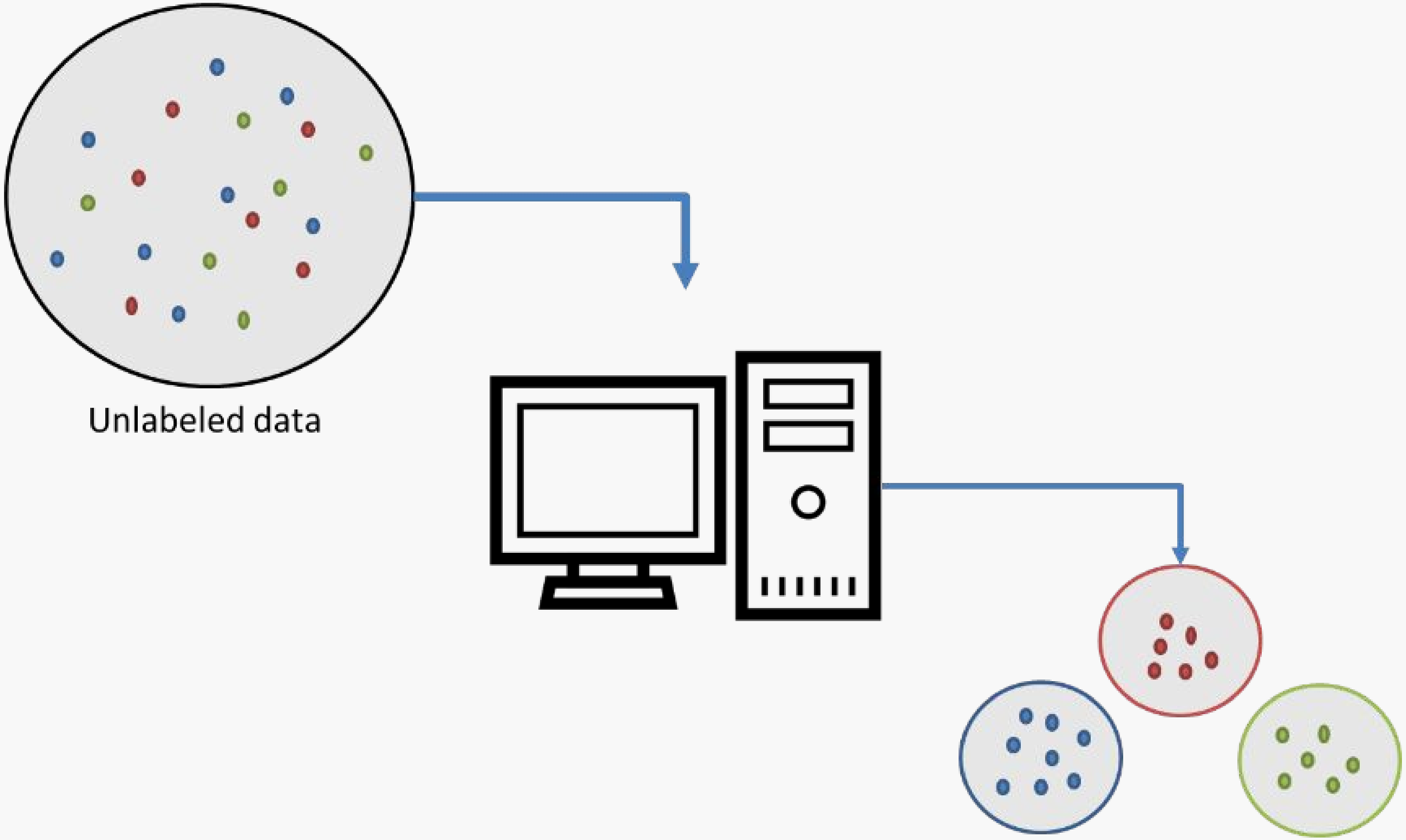
Major paradigms of machine learning

- **Rote learning** – "Learning by memorization."
  - Employed by first machine learning systems, in 1950s
    - Samuel's Checkers program
- **Supervised learning** – Use specific examples to reach general conclusions or extract general rules
    - Classification (Concept learning)
    - Regression
- **Unsupervised learning (Clustering)** – Unsupervised identification of natural groups in data
- **Reinforcement learning**– Feedback (positive or negative reward) given at the end of a sequence of steps
- **Analogy** – Determine correspondence between two different representations
- **Discovery** – Unsupervised, specific goal not given

Supervised learning (classification)



Unsupervised learning (clustering)



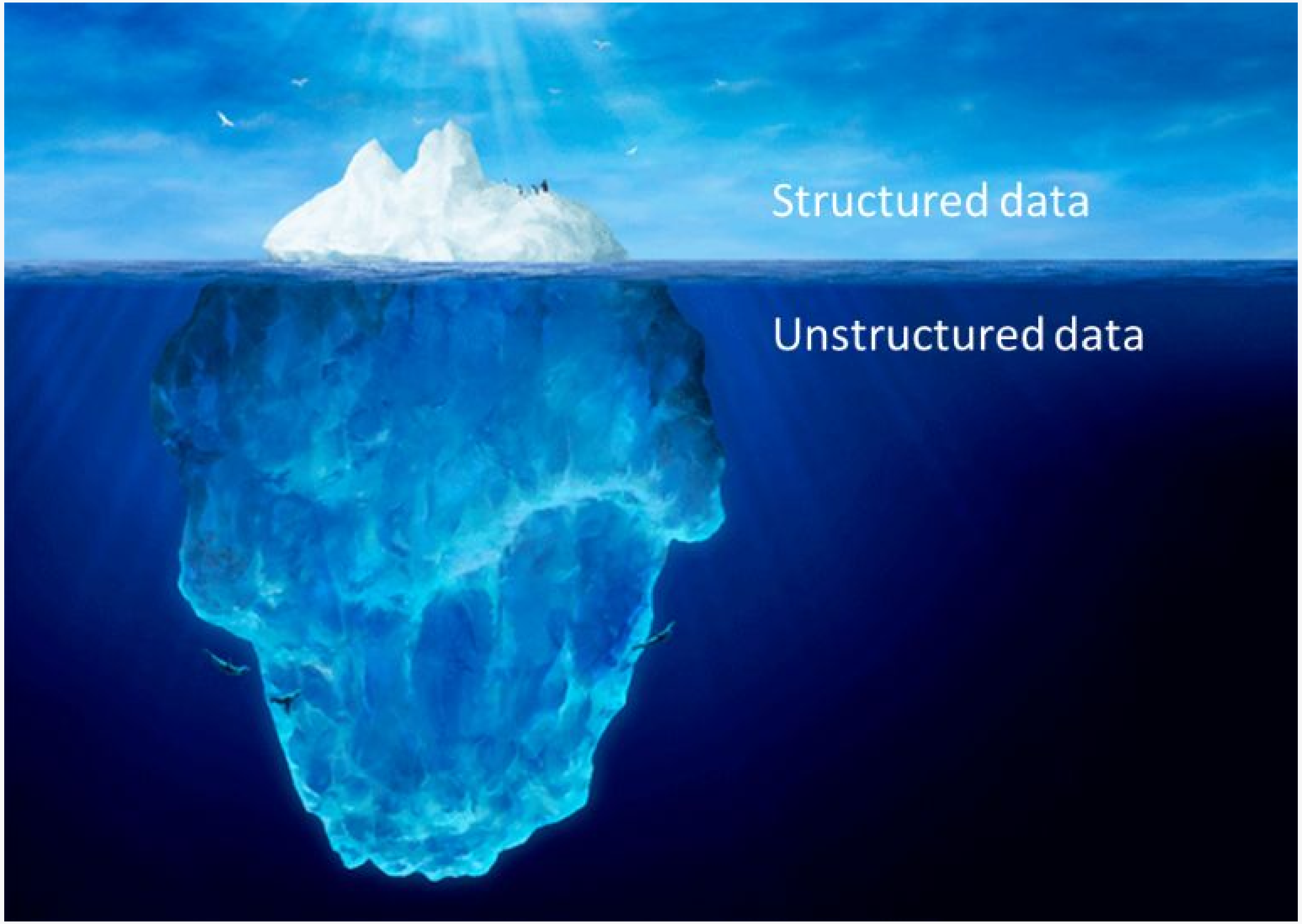
Tools of the trade

- $$$
- Proprietary (ie closed data format)

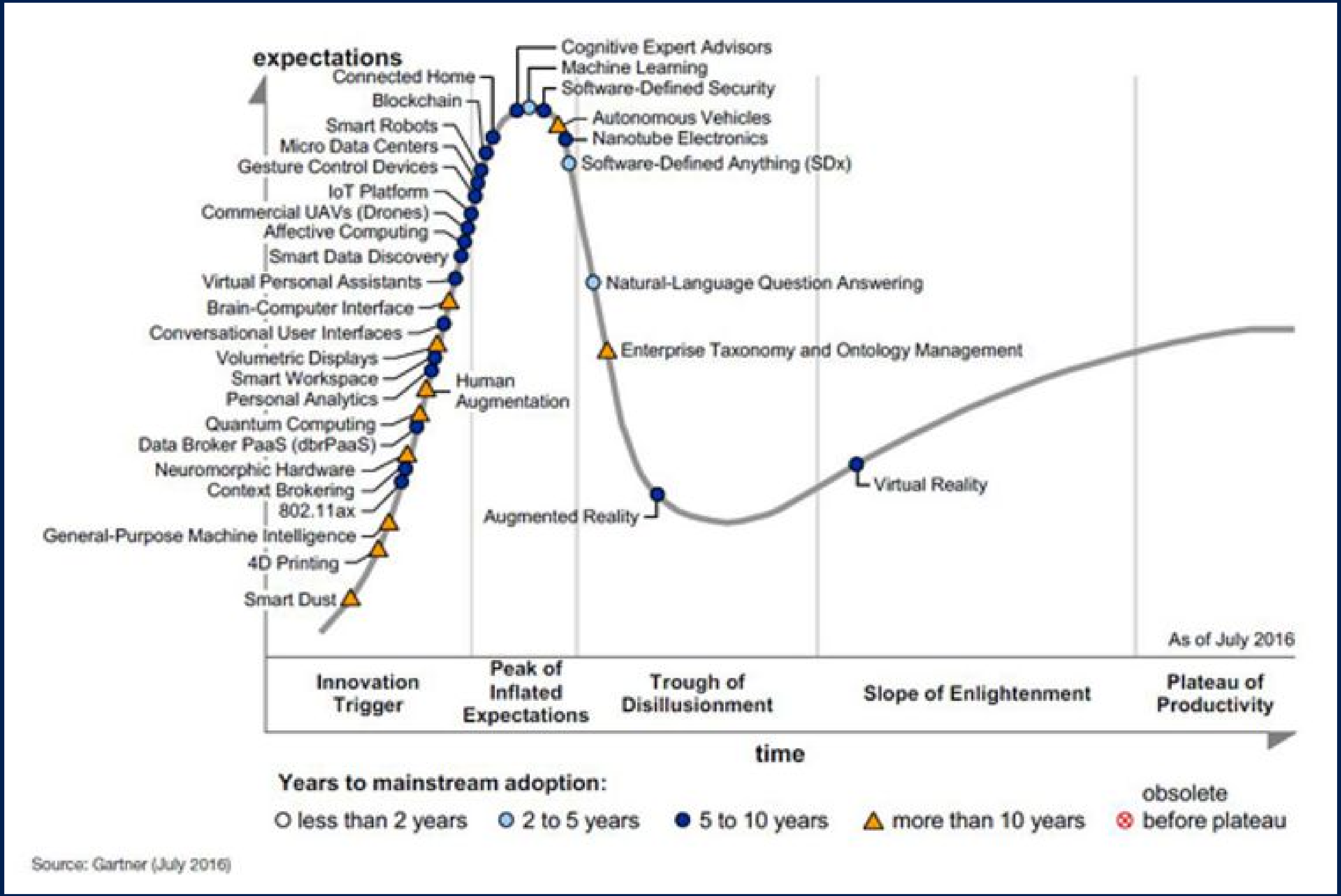
- Free
- Open source (ie open data format)

http://nihlibrary.campusguides.com/dataworkstation

Leveraging unstructured data



Don't forget the hype curve



Summary

Problem Detection?

- What are you top 10 challenges that you want data to solve?
- What are you trying to solve?
- What data do you have?

- Can you group it? –> Clustering
- Can you categorize it? –> Classification
- How much / How many? –> Regression
- Is it weird? –> Anomaly detection

Summary

The Problem to Solve	The Category of Techniques	Covered in this Course
I want to group items by similarity. I want to find structure (commonalities) in the data	Clustering	K-means clustering
I want to discover relationships between actions or items	Association Rules	Apriori
I want to determine the relationship between the outcome and the input variables	Regression	Linear Regression Logistic Regression
I want to assign (known) labels to objects	Classification	Naïve Bayes Decision Trees
I want to find the structure in a temporal process I want to forecast the behavior of a temporal process	Time Series Analysis	ACF, PACF, ARIMA
I want to analyze my text data	Text Analysis	Regular expressions, Document representation (Bag of Words), TF-IDF

Thank you