Machine Learning

Dr. Vittorio Zanella Palacios

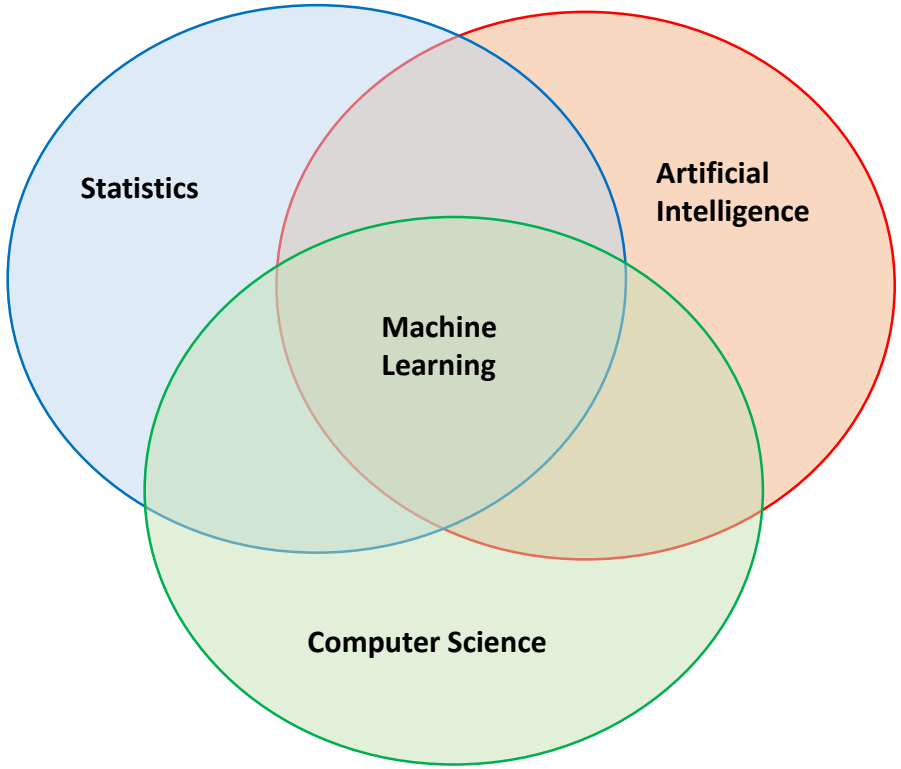
vittorio.zanella@upaep.mx

What is Machine Learning?

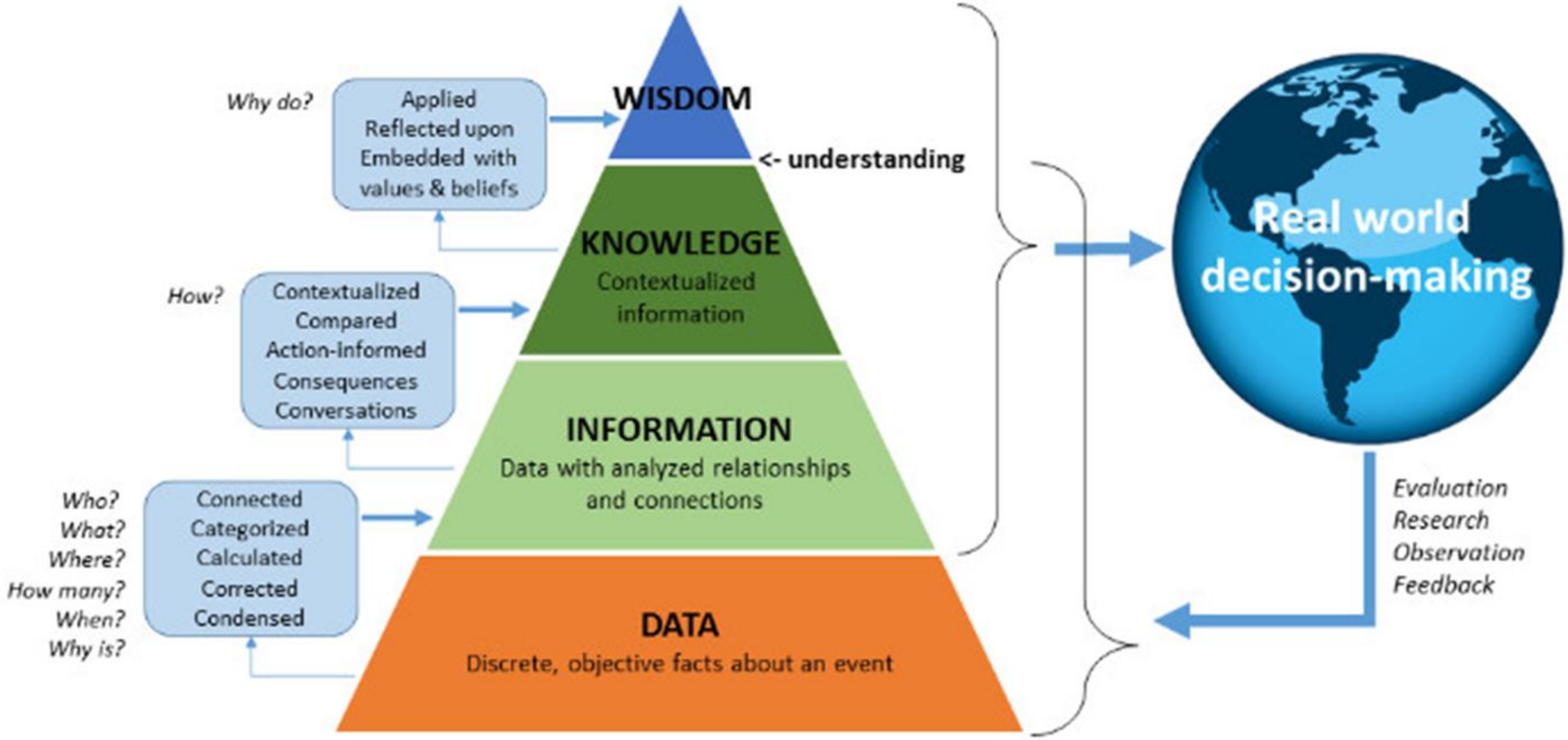
Arthur Lee Samuel
(December 5, 1901 – July 29, 1990)

In 1959, *Arthur Samuel* defined machine learning as a “Field of study that gives computers the ability to learn with out being explicitly programmed”

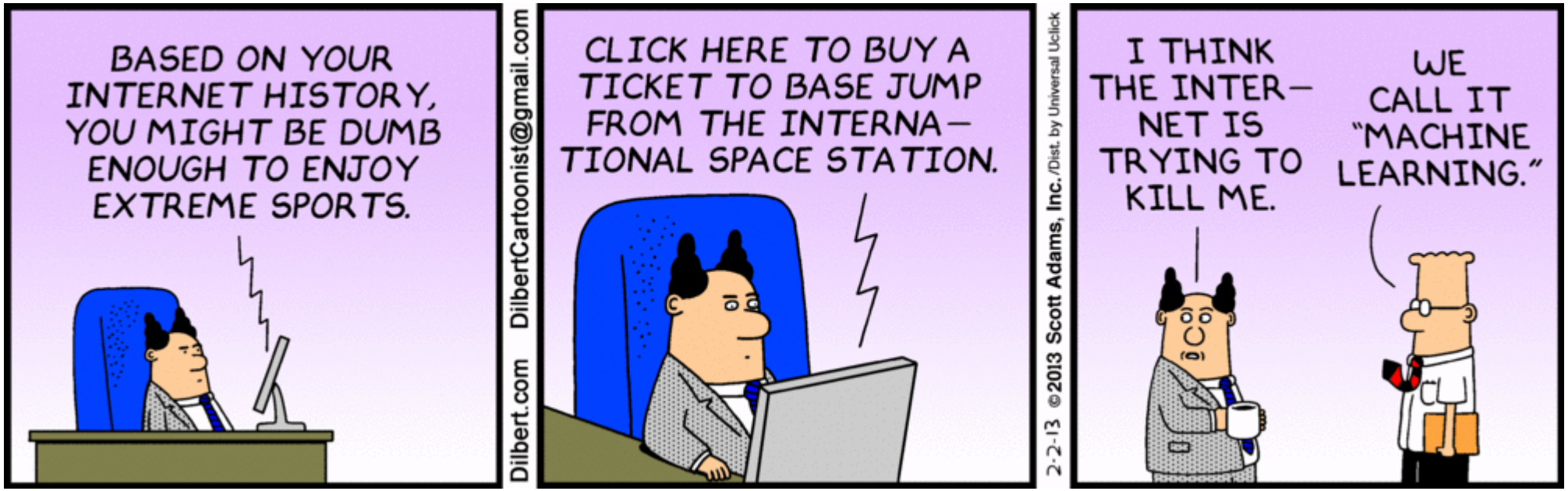
It is a research field at the intersection of statistics, artificial intelligence and computer science



Machine learning is about extracting knowledge from data

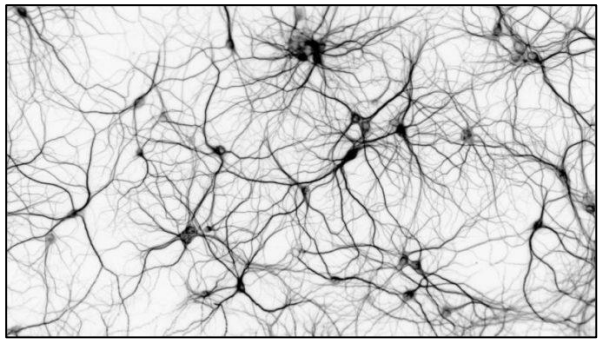

Why do?
Applied
Reflected upon
Embedded with
values & beliefs
WISDOM
<- understanding
How?
Contextualized
Compared
Action-Informed
Consequences
Conversations
KNOWLEDGE
Contextualized
information
INFORMATION
Data with analyzed relationships
and connections
Who?
What?
Where?
How many?
When?
Why is?
Connected
Categorized
Calculated
Corrected
Condensed
DATA
Discrete, objective facts about an event
Real world
decision-making
Evaluation
Research
Observation
Feedback

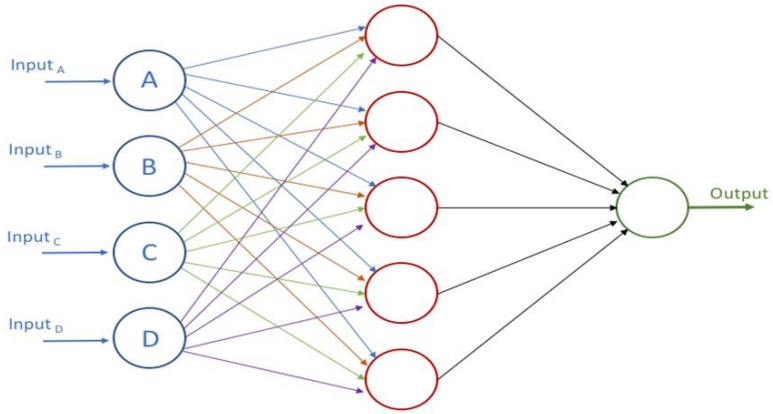
This is what happens in a minute on the internet | World Economic Forum


BASED ON YOUR INTERNET HISTORY, YOU MIGHT BE DUMB ENOUGH TO ENJOY EXTREME SPORTS.
Dilbert.com DilbertCartoonist@gmail.com
CLICK HERE TO BUY A TICKET TO BASE JUMP FROM THE INTERNA— TIONAL SPACE STATION.
2-2-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick
I THINK THE INTER— NET IS TRYING TO KILL ME.
WE CALL IT "MACHINE LEARNING."

History and relationships to other fields

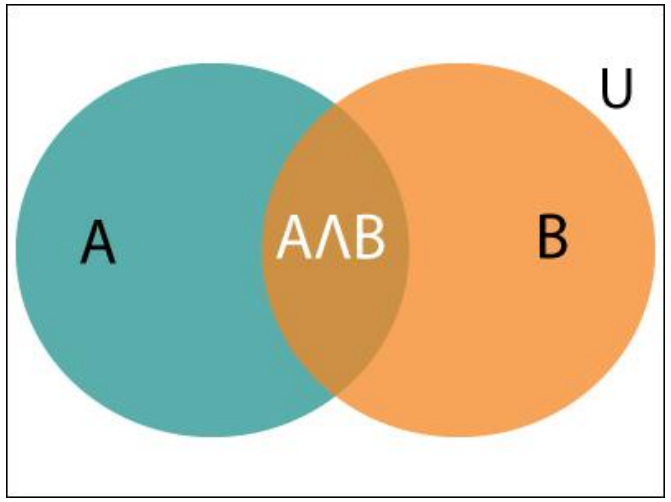
Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data.

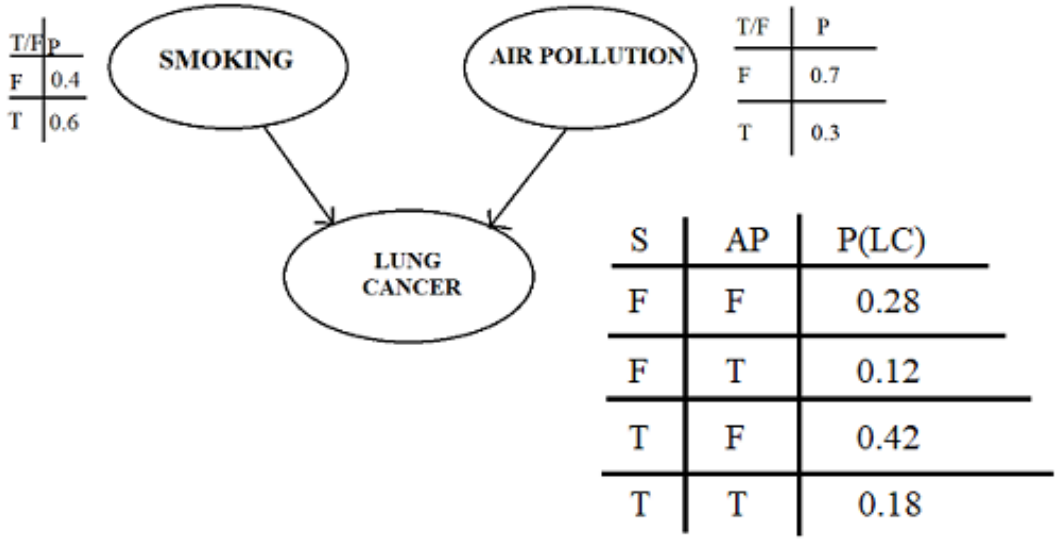




neural networks

Probabilistic reasoning, especially in automated medical diagnosis



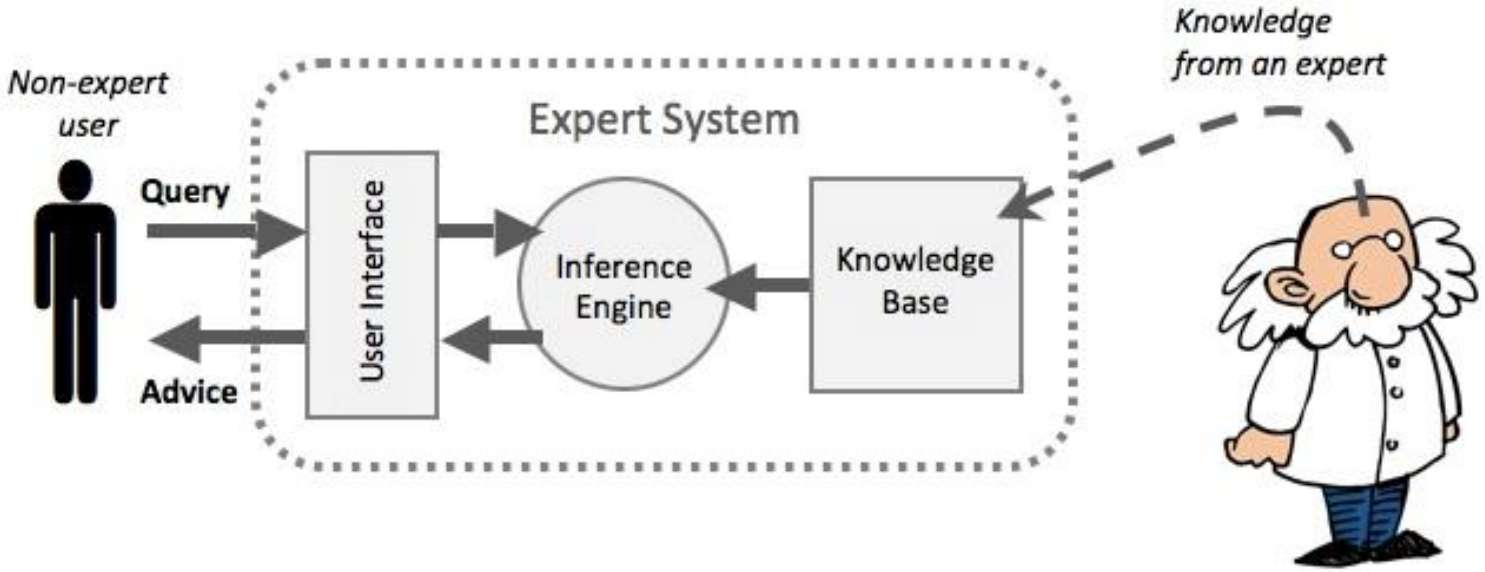


T/F	P
F	0.4
T	0.6

T/F	P
F	0.7
T	0.3

S	AP	P(LC)
F	F	0.28
F	T	0.12
T	F	0.42
T	T	0.18

By 1980, expert systems had come to dominate AI



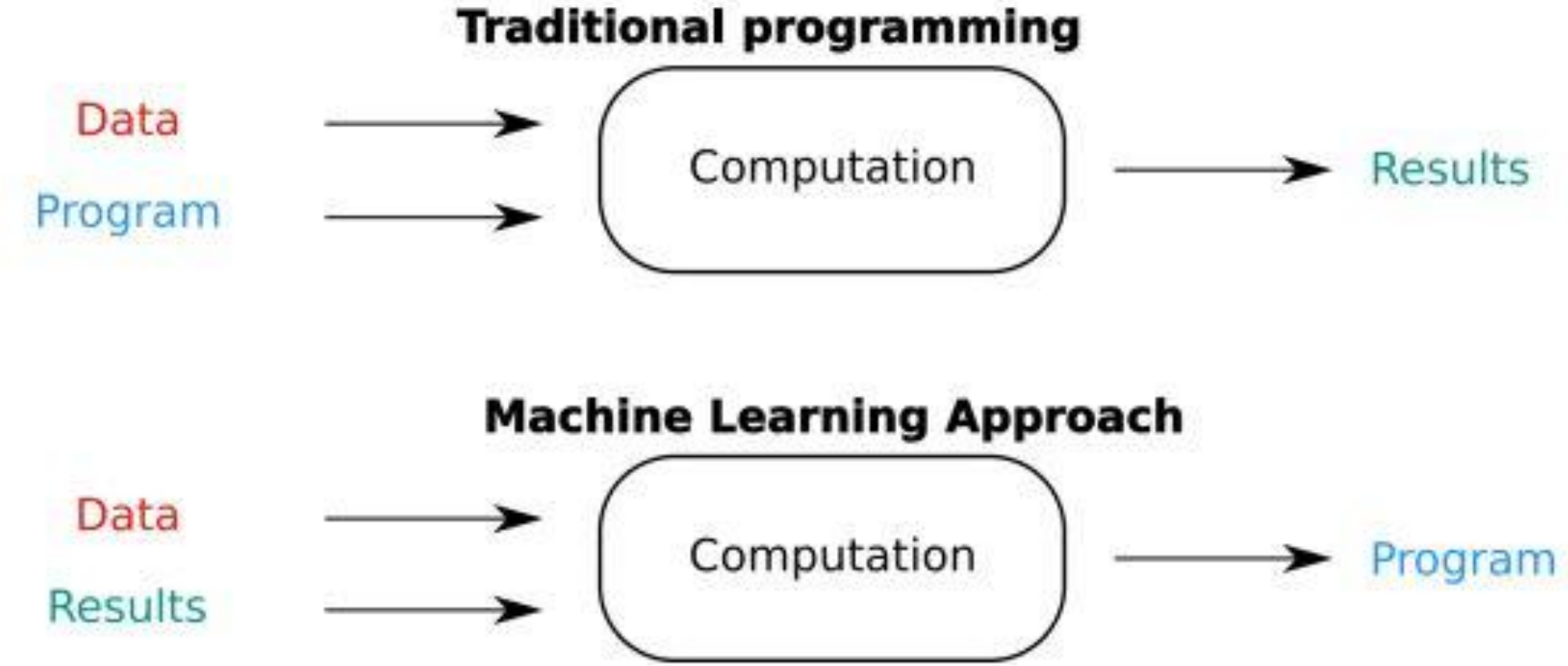
But the more statistical line of research was now outside the field of AI proper, in *pattern recognition.*

Neural networks research had been abandoned by AI around the same time.

Machine learning, reorganized as a separate field, started to flourish in the 1990s

The field changed its goal from achieving artificial intelligence to tackling solvable problems of a practical nature

It also benefited from the increasing availability of digitized information, and the possibility to distribute that via the internet


Traditional programming
Data
Program
Computation
Results
Machine Learning Approach
Data
Results
Computation
Program

BIG
DATA
ALSO
TOOLS
DISK
TARGET
APPLIED
SENSOR
MAY
EXAMPLES
SHARED
DEFINITION
MOVING
EVERY
SET
RELATIONAL
LARGER
CONTINUES
SOCIAL
PRACTITIONERS
CAPTURE
WITHIN
TENS
COMPLEX
CITATION
INTERNET
BUSINESS
ANALYTICS
SETS
GENOMICS
DISTRIBUTED
MANAGEMENT
TERABYTES
CAPACITY
BIOLOGICAL
PROCESSING
PETABYTES
INCLUDE
ABILITY
HUNDREDS
NETWORKS
SYSTEMS
TOLERABLE
DATABASES
STORAGE
SIZE
RECORDS
GARTNER
WORKING
PARALLEL
SAN
SEARCH
MASSIVELY
GROW
DIFFICULTY

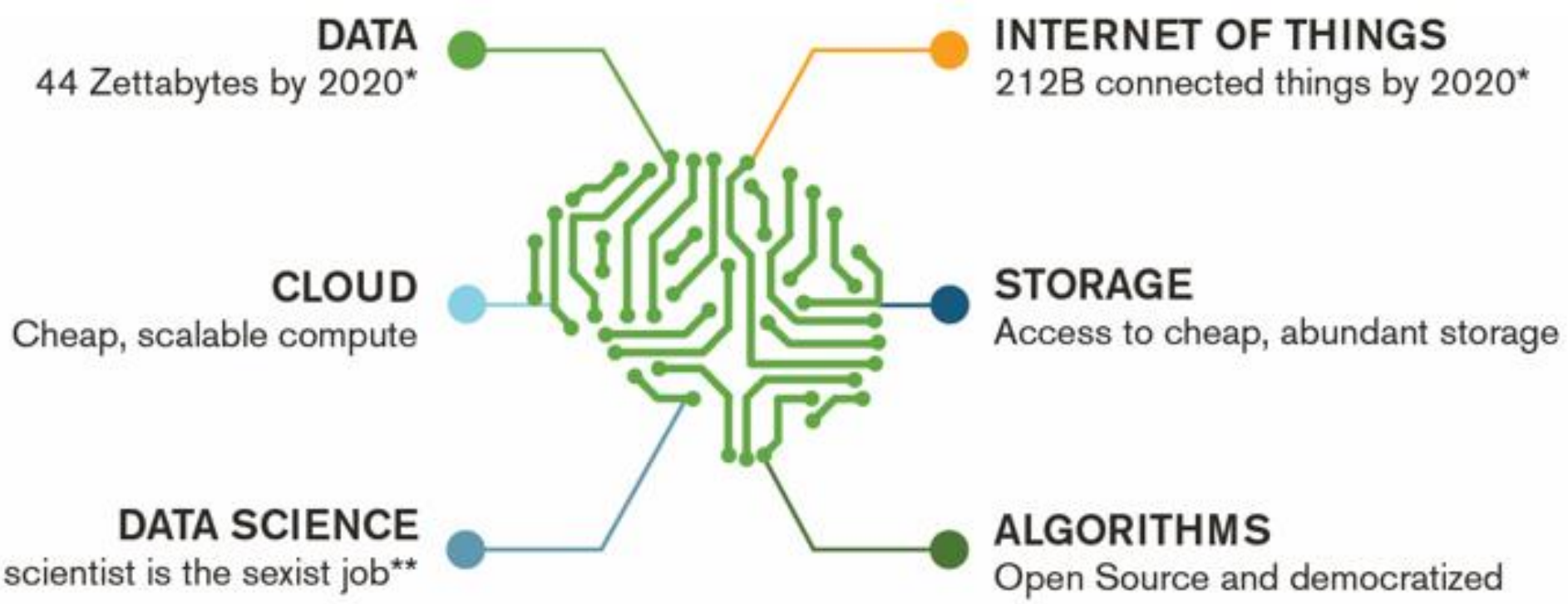
DATA
MINING
PATTERNS
LEARNING
LIKELY
GROUPS
BUSINESS
METHODS
PROCESS
LEARNED
ANALYSIS
DECISION
DATABASE
SET
MIGHT
APPLICATIONS
ALGORITHMS
VISUALIZATION
TEST
LARGE
USING
E-MAILS
WITHIN
OFFERS
DISCOVERY
ASSOCIATION
OFFER
SEND
USE
MACHINE
CUSTOMERS
EXAMPLE
STEP
PEOPLE
KNOWLEDGE
MODELING
TRAINED
INFORMATION
IDENTIFY
INTELLIGENCE
CUSTOMER
RECORDS
DETECTION
ALGORITHM
MODEL
MODELS
FINALLY

Why Machine Learning?

Today there is enough computing power

Huge databases exist on virtually every topic

Why Machine Learning Now?



*IDC Digital Universe report, 2014 http://www.emc.com/infographics/digital-universe-2014.htm

**Data Scientist: The Sexiest Job of the 21st Century, Oct 2012 https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

To summarize, Machine Learning is great for:

Problems for which existing solutions require a lot of hand-tuning or long lists of rules

Complex problems for which there is no good solution at all using a traditional approach

Fluctuating environments

Getting insights about complex problems and large amounts of data

UPAEP

Problems Machine Learning Can Solve

Homework:

To find and describe two problems that can be solved by machine learning

UPAEP

Types of Machine Learning Systems

There are so many different types of Machine Learning systems that it is useful to classify them in broad categories based on:

- Whether or not they are trained with human supervision (supervised, unsupervised, semisupervised, and Reinforcement Learning)

- Whether or not they can learn incrementally on the fly (online versus batch learning)

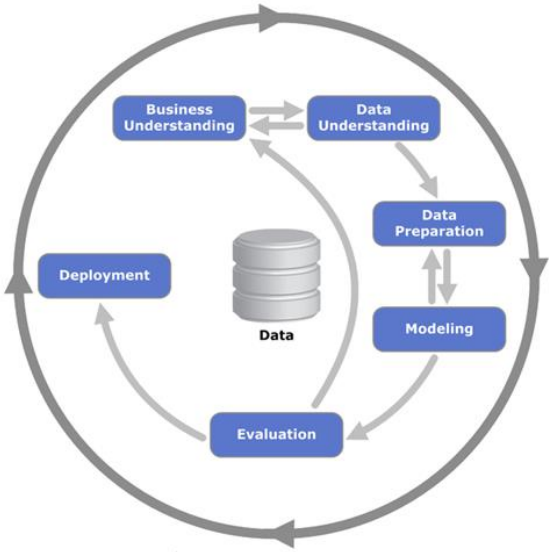
- Whether they work by simply comparing new data points to known data points, or instead detect patterns in the training data and build a predictive model, much like scientists do (instance-based versus model-based learning)

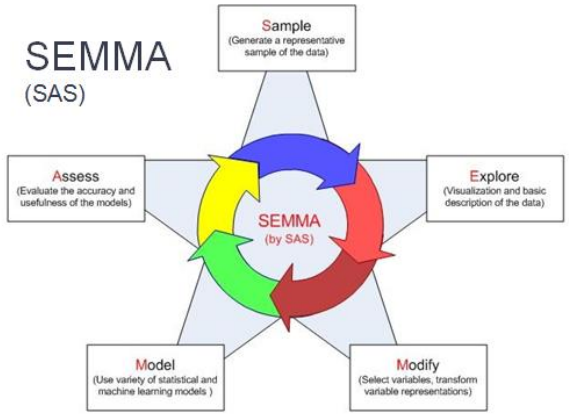
Quite possibly the most important part in the machine learning process is understanding the data you are working with and how it relates to the task you want to solve

Data Science project management methodologies

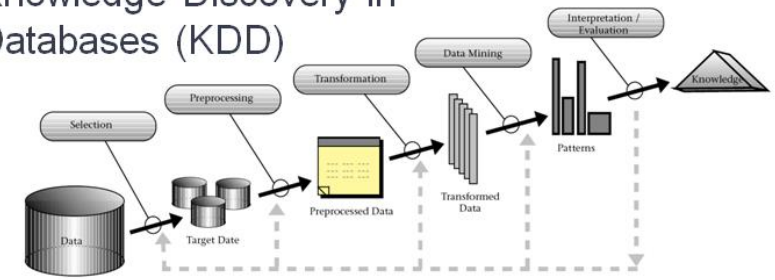
From Data to Insight

Cross Industry Standard Process for Data Mining (CRISP-DM)
(IBM, Teradata, Daimler AG, NCR Corporation and OHRA)





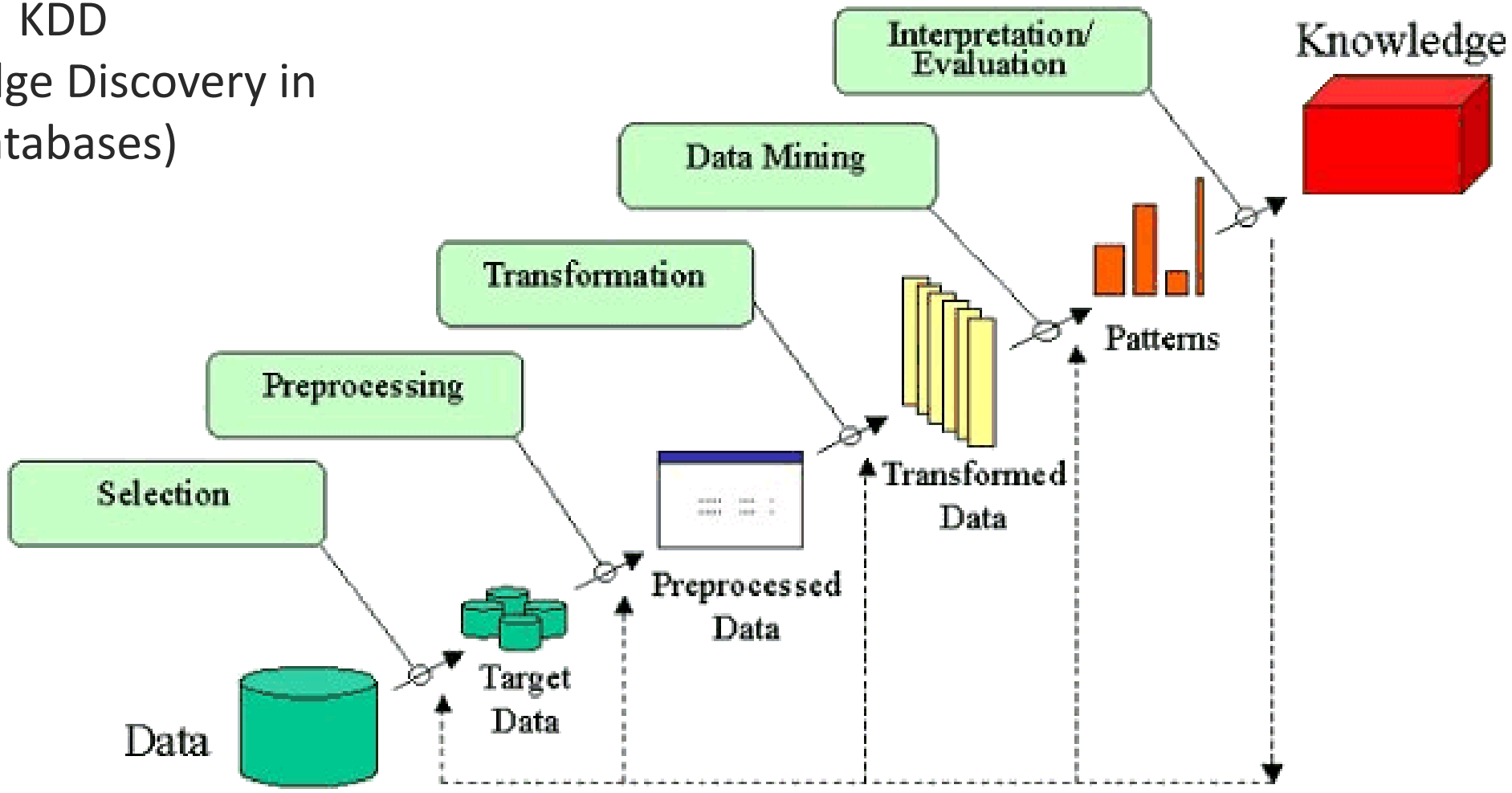
Knowledge Discovery in Databases (KDD)



BUSINESS BROADWAY

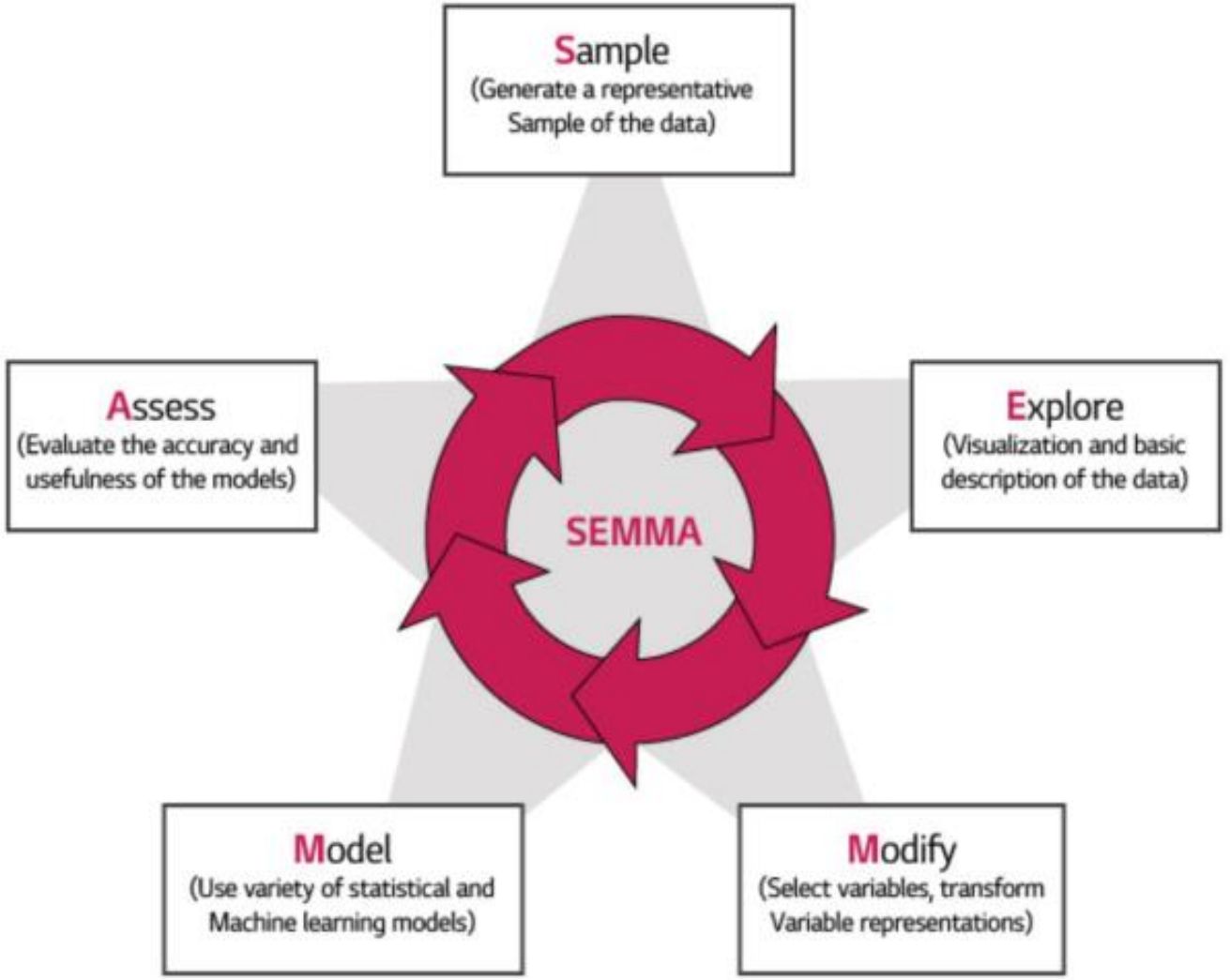
For more information on these methods, see: https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining; https://en.wikipedia.org/wiki/SEMMA; https://en.wikipedia.org/wiki/Data_mining

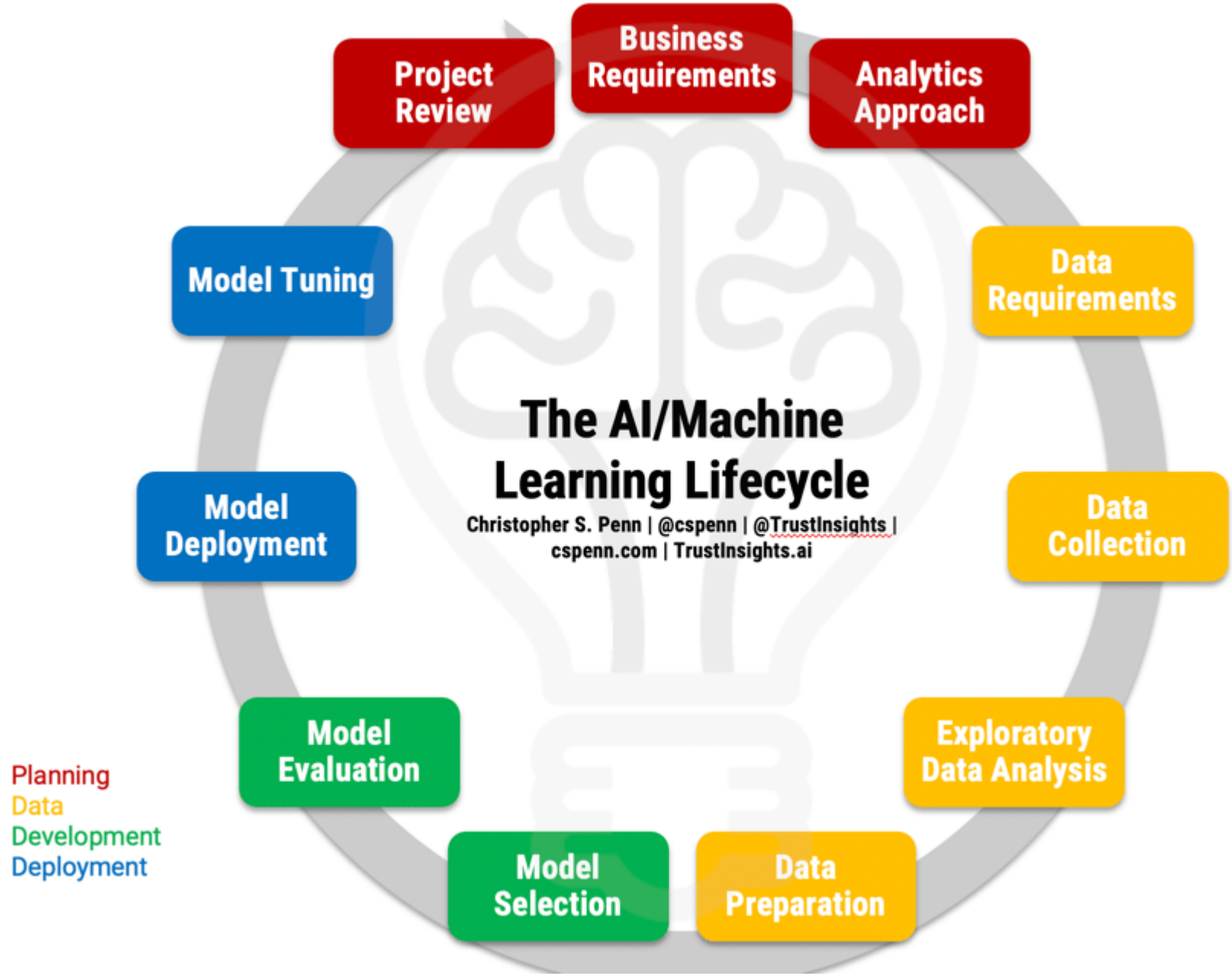
KDD
(Knowledge Discovery in Databases)



Cross Industry Standard Process for Data Mining (CRISP-DM)

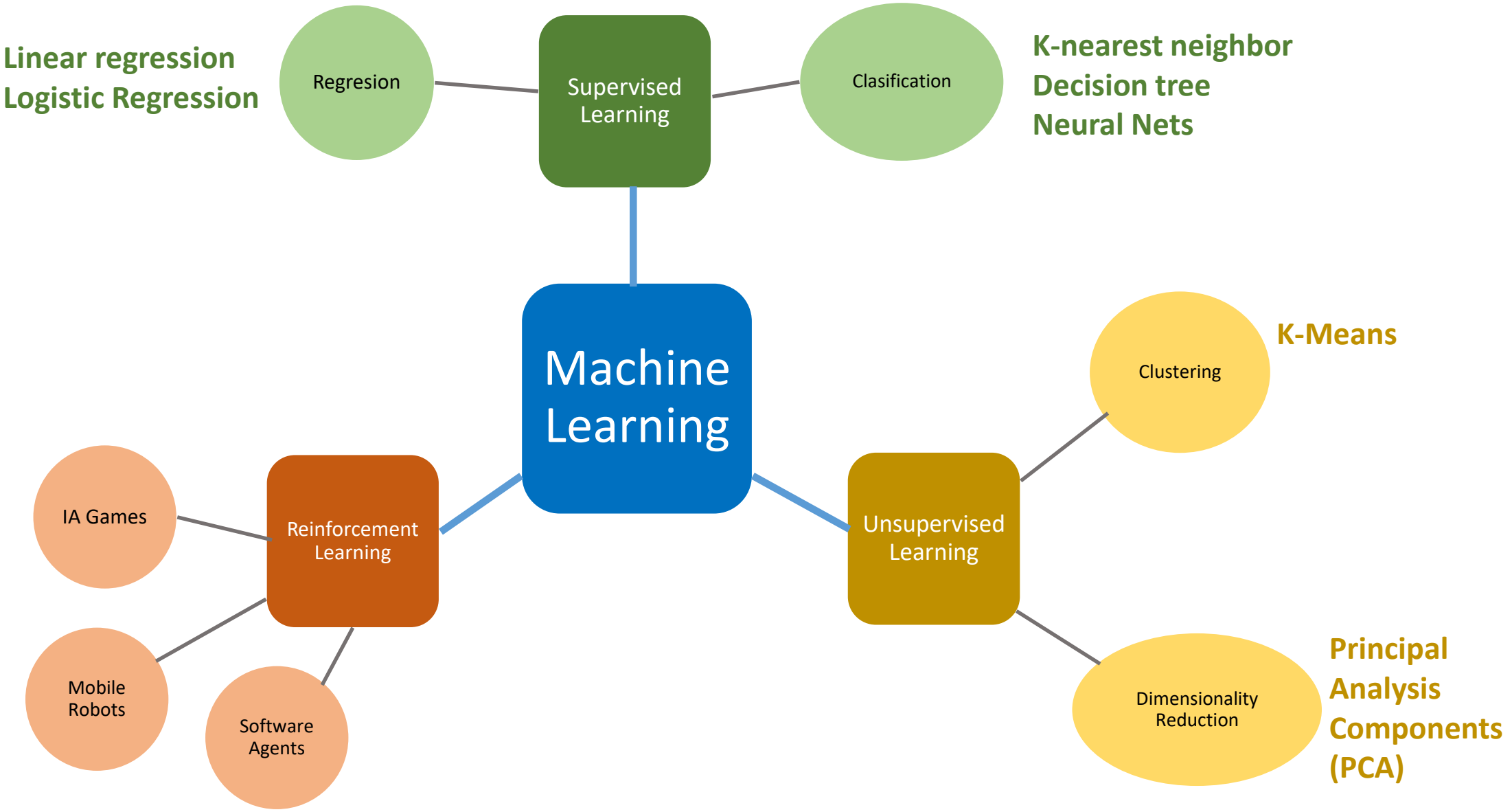
SEMMA
(Sample, Explore, Modify, Model and Access)




Business Requirements
Analytics Approach
Data Requirements
Data Collection
Exploratory Data Analysis
Data Preparation
Model Selection
Model Evaluation
Model Deployment
Model Tuning
Project Review
The AI/Machine Learning Lifecycle
Christopher S. Penn | @cspenn | @TrustInsights |
cspenn.com | TrustInsights.ai
Planning
Data
Development
Deployment

While you are building a machine learning solution, you should answer, or at least keep in mind, the following questions:

- What question(s) am I trying to answer?
- Do I think the data collected can answer that question?
- What is the best way to phrase my question(s) as a machine learning problem?
- Have I collected enough data to represent the problem I want to solve?
- What features of the data did I extract, and will these enable the right predictions?
- How will I measure success in my application?
- How will the machine learning solution interact with other parts of my research or business product?


Linear regression
Logistic Regression
Regresion
Supervised Learning
Clasification
K-nearest neighbor
Decision tree
Neural Nets
Machine Learning
IA Games
Reinforcement Learning
Mobile Robots
Software Agents
Unsupervised Learning
Clustering
K-Means
Dimensionality Reduction
Principal Analysis Components (PCA)