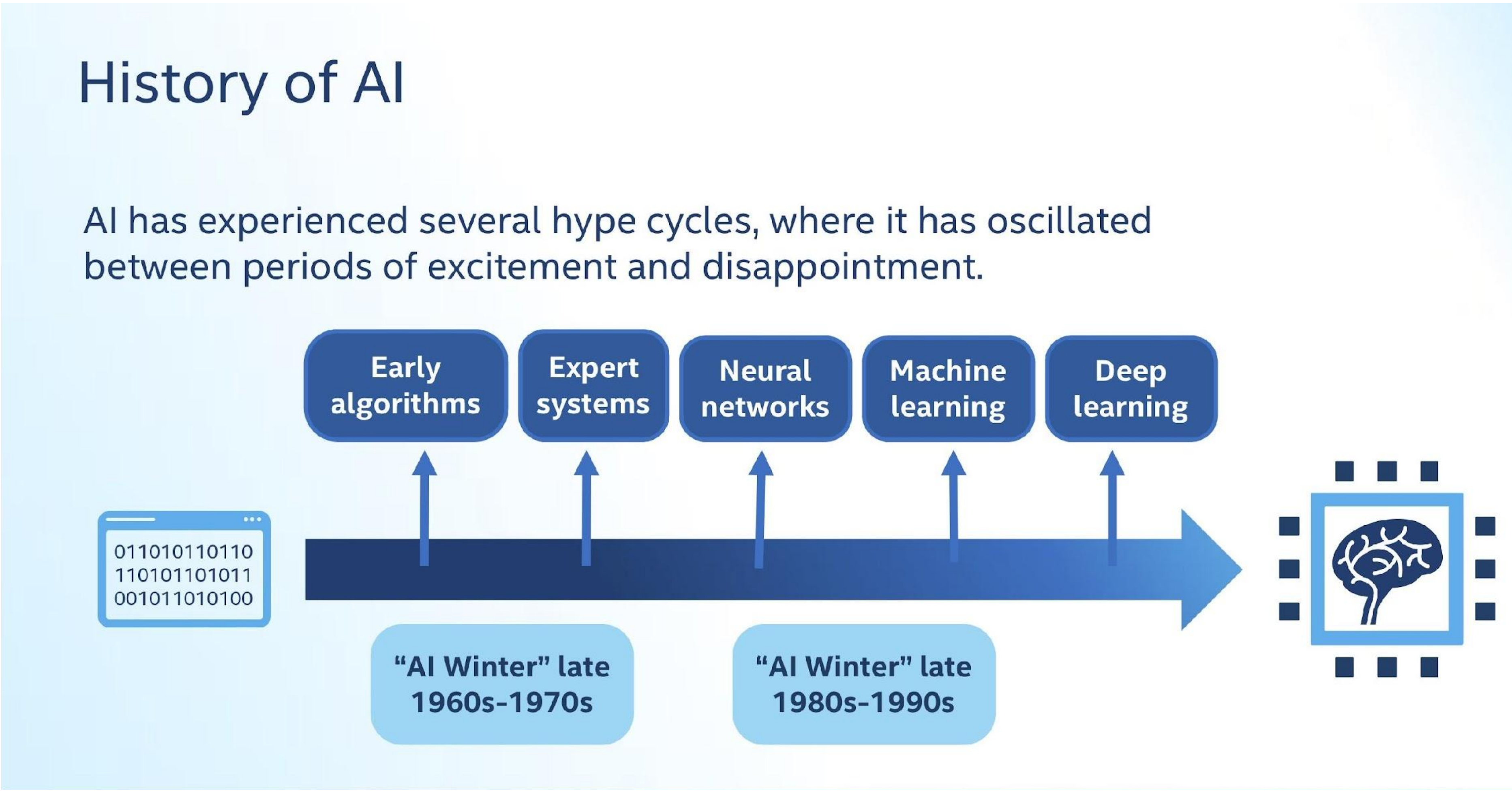
Програмне забезпечення інформаційних систем

What is AI, ML, DL, DS?

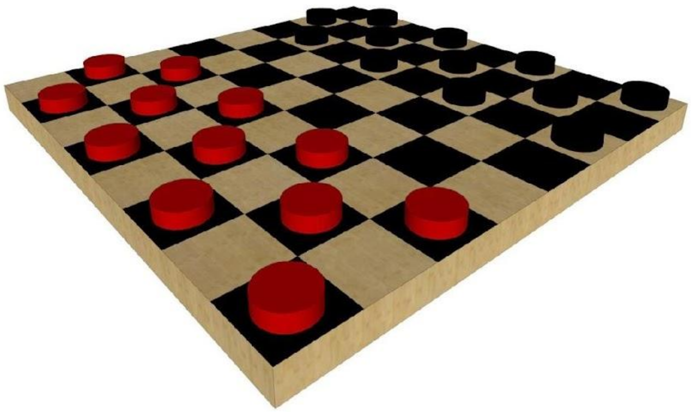
Artificial Intelligence – AI
History



Artificial Intelligence – AI
History

1950s: Early AI

- 1950: Alan Turing developed the Turing test to test a machines ability to exhibit intelligent behavior.
- 1956: Artificial Intelligence was accepted as a field at the Dartmouth Conference.
- 1957: Frank Rosenblatt invented the perceptron algorithm. This was the precursor to modern neural networks.
- 1959: Arthur Samuel published an algorithm for a checkers program using machine learning.



Artificial Intelligence – AI

History

The First "AI Winter"

- 1966: ALPAC committee evaluated AI techniques for machine translation and determined there was little yield from the investment.
- 1969: Marvin Minsky published a book on the limitations of the Perceptron algorithm which slowed research in neural networks.
- 1973: The Lighthill report highlights AI's failure to live up to promises.
- The two reports led to cuts in government funding for AI research leading to the first "AI Winter."

John R. Pierce, head of ALPAC

Artificial Intelligence – AI

History

1980's AI Boom

- Expert Systems - systems with programmed rules designed to mimic human experts.
- Ran on mainframe computers with specialized programming languages (e.g. LISP).
- Were the first widely-used AI technology, with two-thirds of "Fortune 500" companies using them at their peak.
- 1986: The "Backpropogation" algorithm is able to train multi-layer perceptrons leading to new successes and interest in neural network research.

Early expert systems machine

Artificial Intelligence – AI

History

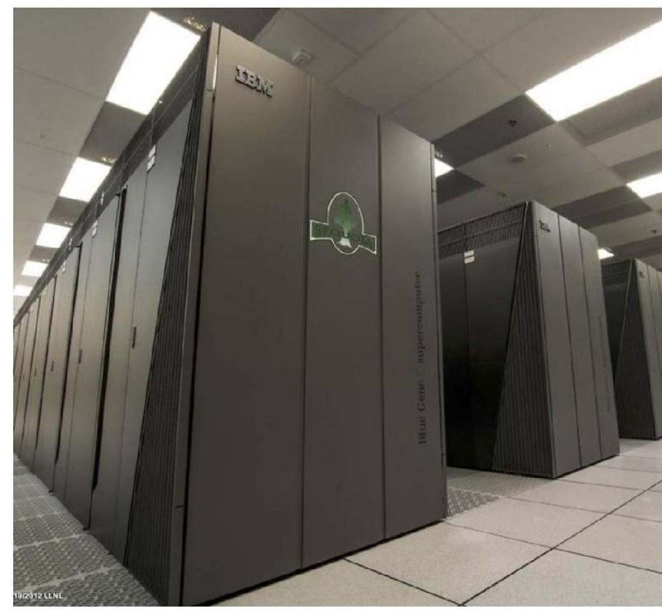
Another AI Winter (late 1980's – early 1990s)

- Expert systems' progress on solving business problems slowed.
- Expert systems began to be melded into software suites of general business applications (e.g. SAP, Oracle) that could run on PCs instead of mainframes.
- Neural networks didn't scale to large problems.
- Interest in AI in business declined.

Artificial Intelligence – AI
History

Late 1990's to early 2000's: Classical Machine Learning

- Advancements in the SVM algorithm led to it becoming the machine learning method of choice.
- AI solutions had successes in speech recognition, medical diagnosis, robotics, and many other areas.
- AI algorithms were integrated into larger systems and became useful throughout industry.
- The Deep Blue chess system beat world chess champion Garry Kasparov.
- Google search engine launched using artificial intelligence technology.



IBM supercomputer

Artificial Intelligence – AI
History

2006: Rise of Deep Learning

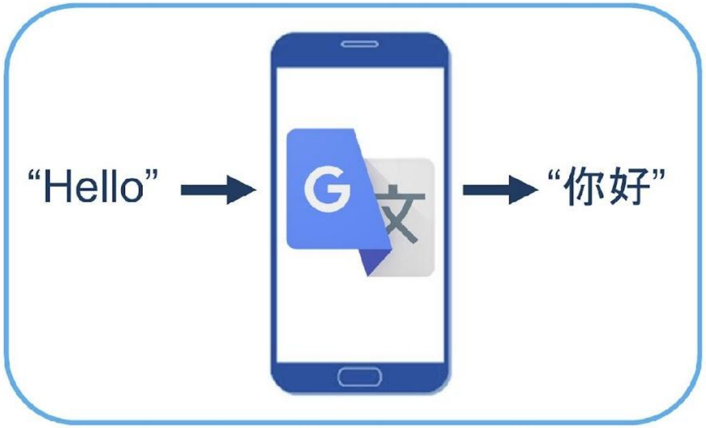
- 2006: Geoffrey Hinton publishes a paper on unsupervised pre-training that allowed deeper neural networks to be trained.
- Neural networks are rebranded to deep learning.
- 2009: The ImageNet database of human-tagged images is presented at the CVPR conference.
- 2010: Algorithms compete on several visual recognition tasks at the first ImageNet competition.

IMAGENET

Artificial Intelligence – AI
Modern Era

Deep Learning Breakthroughs (2012 – Present)

- In 2012, deep learning beats previous benchmark on the ImageNet competition.
- In 2013, deep learning is used to understand "conceptual meaning" of words.
- In 2014, similar breakthroughs appeared in language translation.
- These have led to advancements in Web Search, Document Search, Document Summarization, and Machine Translation.



Google Translate

Artificial Intelligence – AI
Modern Era

Deep Learning Breakthroughs (2012 – Present)

- In 2014, computer vision algorithm can describe photos.
- In 2015, Deep learning platform TensorFlow* is developed.
- In 2016, DeepMind* AlphaGo, developed by Aja Huang, beats Go master Lee Se-dol.

Artificial Intelligence – AI
Modern Era

AI Breakthroughs

Image classification

"Dog" *"Cat"*

As of *2015*, computers can be trained to perform better on this task than humans.

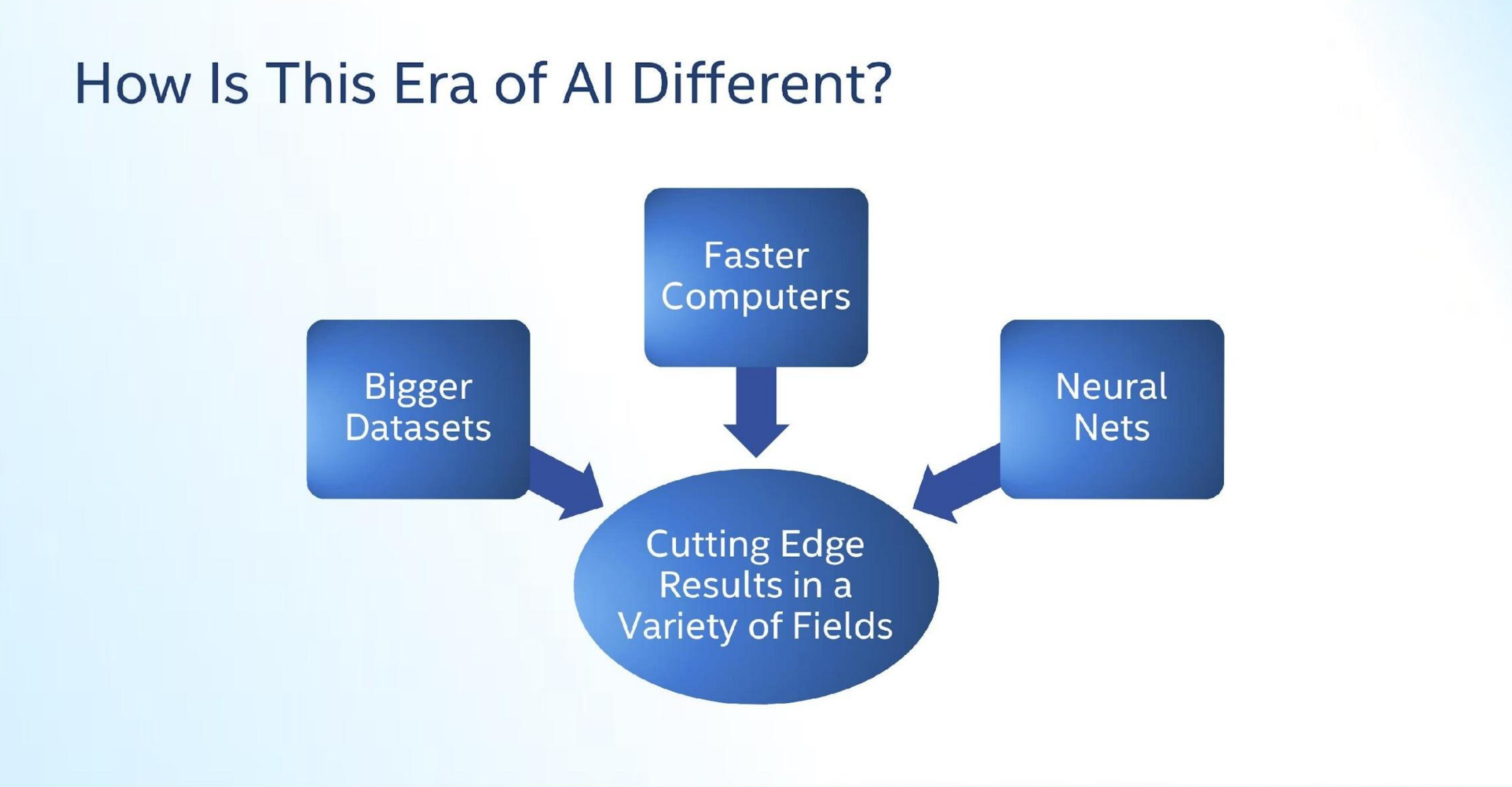
Machine translation

"I am a student"

"Je suis étudiant"

As of *2016*, we have achieved near-human performance using the latest AI techniques.

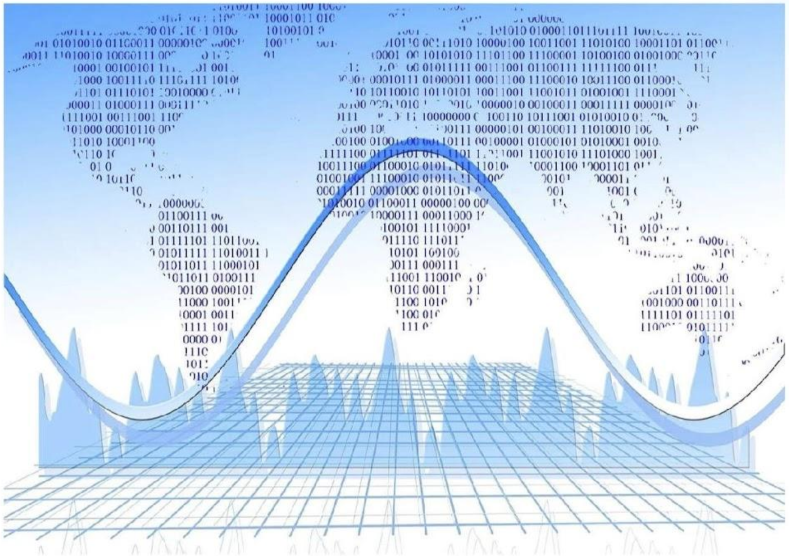
Artificial Intelligence – AI
Modern Era



Artificial Intelligence – AI
Modern Era

Other Modern AI Factors

- Continued expansion of open source AI, especially in Python*, aiding machine learning and big data ecosystems.
- Leading deep learning libraries *open sourced*, allowing further adoption by industry.
- Open sourcing of large datasets of millions of labeled images, text datasets such as Wikipedia has also driven breakthroughs.

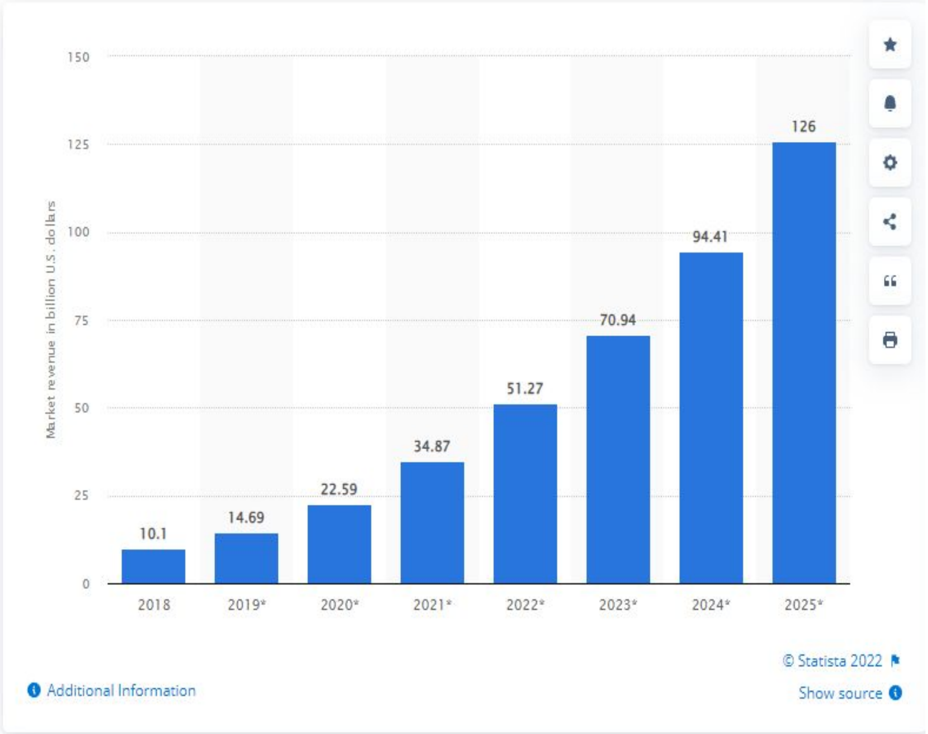


Artificial Intelligence – AI
Modern Era

AI Is The New Electricity

"About 100 years ago, electricity transformed every major industry. AI has advanced to the point where it has the power to transform…every major sector in coming years."

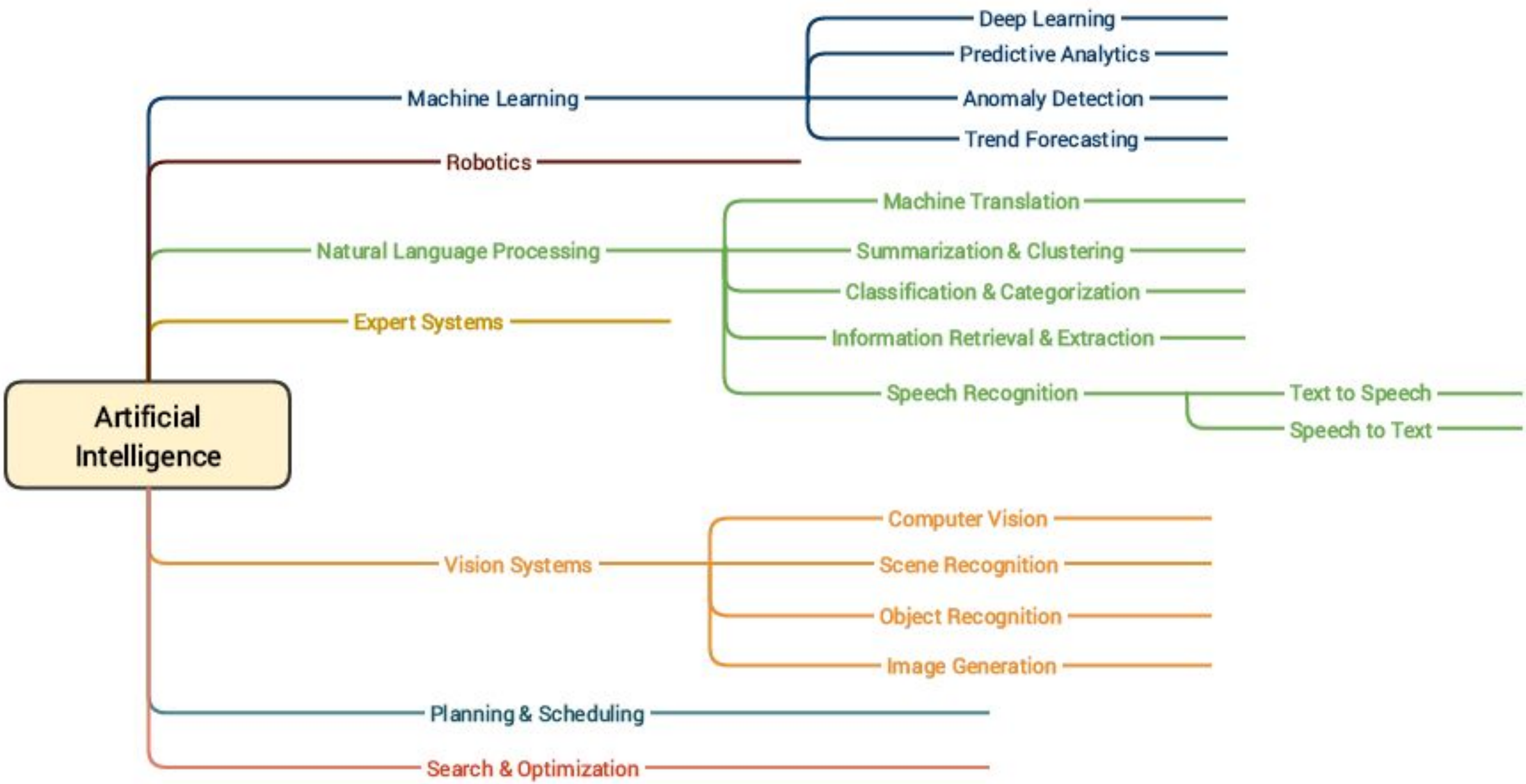
-Andrew Ng, Stanford University



Artificial Intelligence – AI
Definitions

- “Властивість інтелектуальних систем виконувати творчі функції, які традиційно вважаються прерогативою людини. ” (Wikipedia)
- “A branch of computer science dealing with the simulation of intelligent behavior in computers.” (Merriam-Webster)
- “A program that can sense, reason, act, and adapt.” (Intel)
- “Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed.” (Arthur Samuel, 1959)
- “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” (Mitchell)

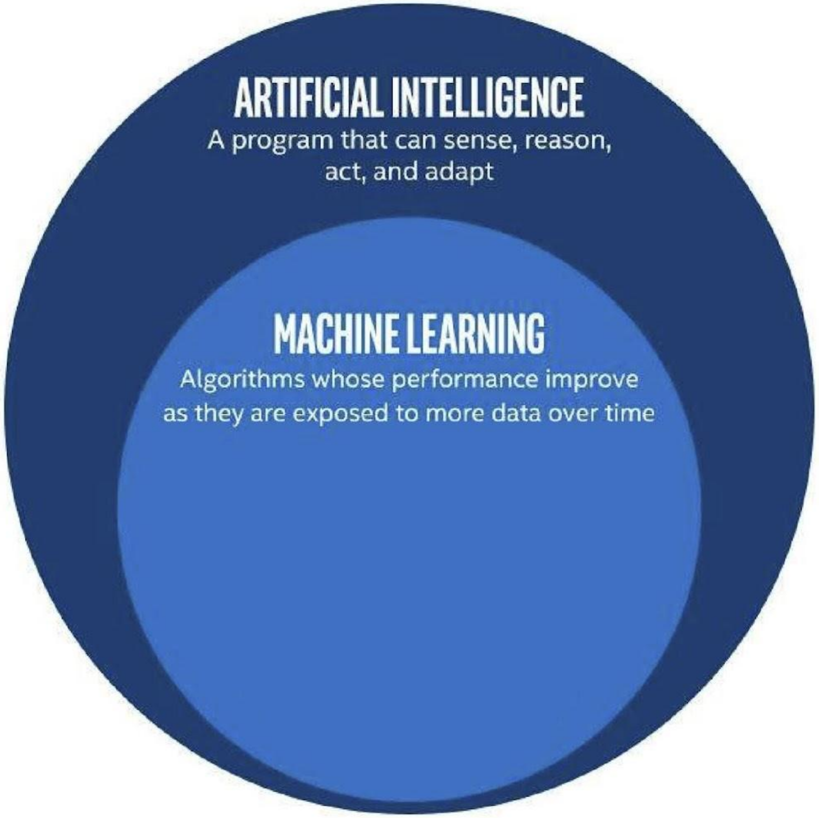
Artificial Intelligence – AI
Problems



Machine Learning – ML
Definitions

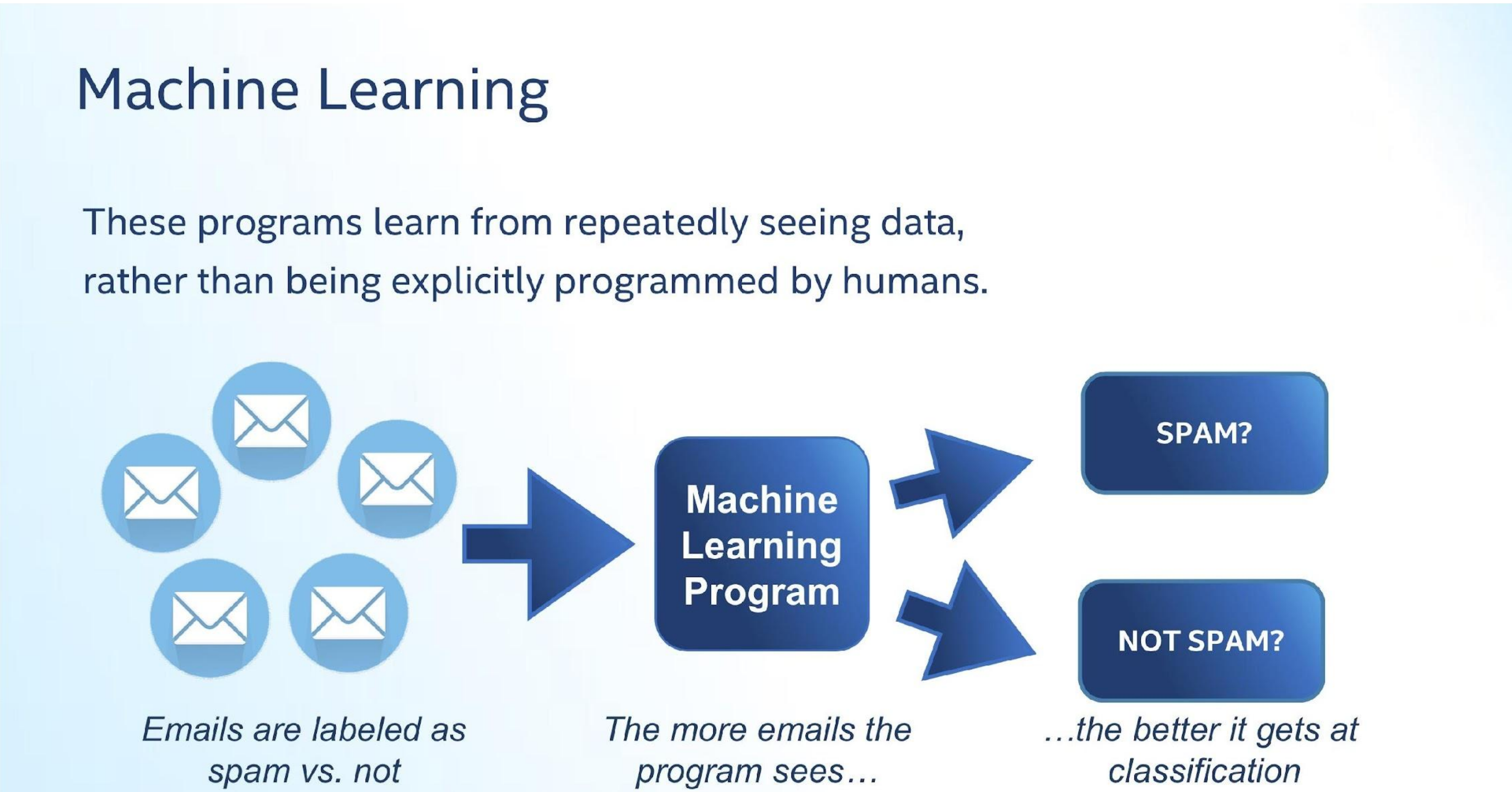
Machine Learning

“The study and construction of programs that are *not explicitly programmed*, but learn patterns as they are exposed to more data over time.” (Intel)



Machine Learning – ML

Learn from data



Machine Learning – ML
Example

Machine Learning Example

- Suppose you wanted to identify fraudulent credit card transactions.
- You could define features to be:
  - Transaction time
  - Transaction amount
  - Transaction location
  - Category of purchase
- The algorithm could learn what feature combinations suggest unusual activity.

Machine Learning – ML
Limitations

Machine Learning Limitations

- Suppose you wanted to determine if an image is of a cat or a dog.
- What features would you use?
- This is where **Deep Learning** can come in.

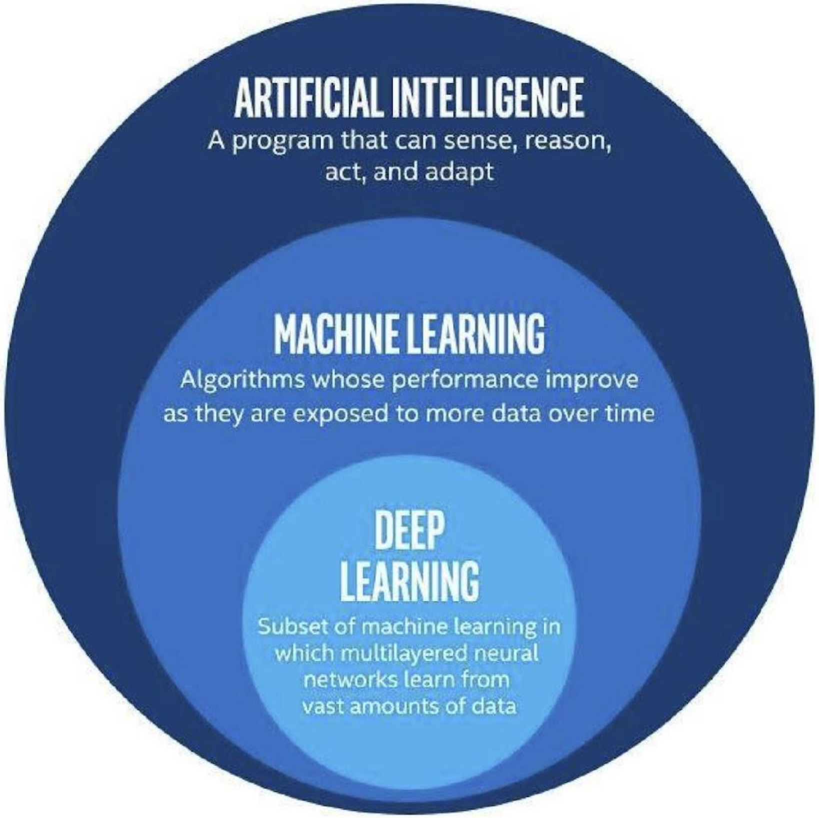
Dog and cat recognition

Deep Learning – DL
Definitions

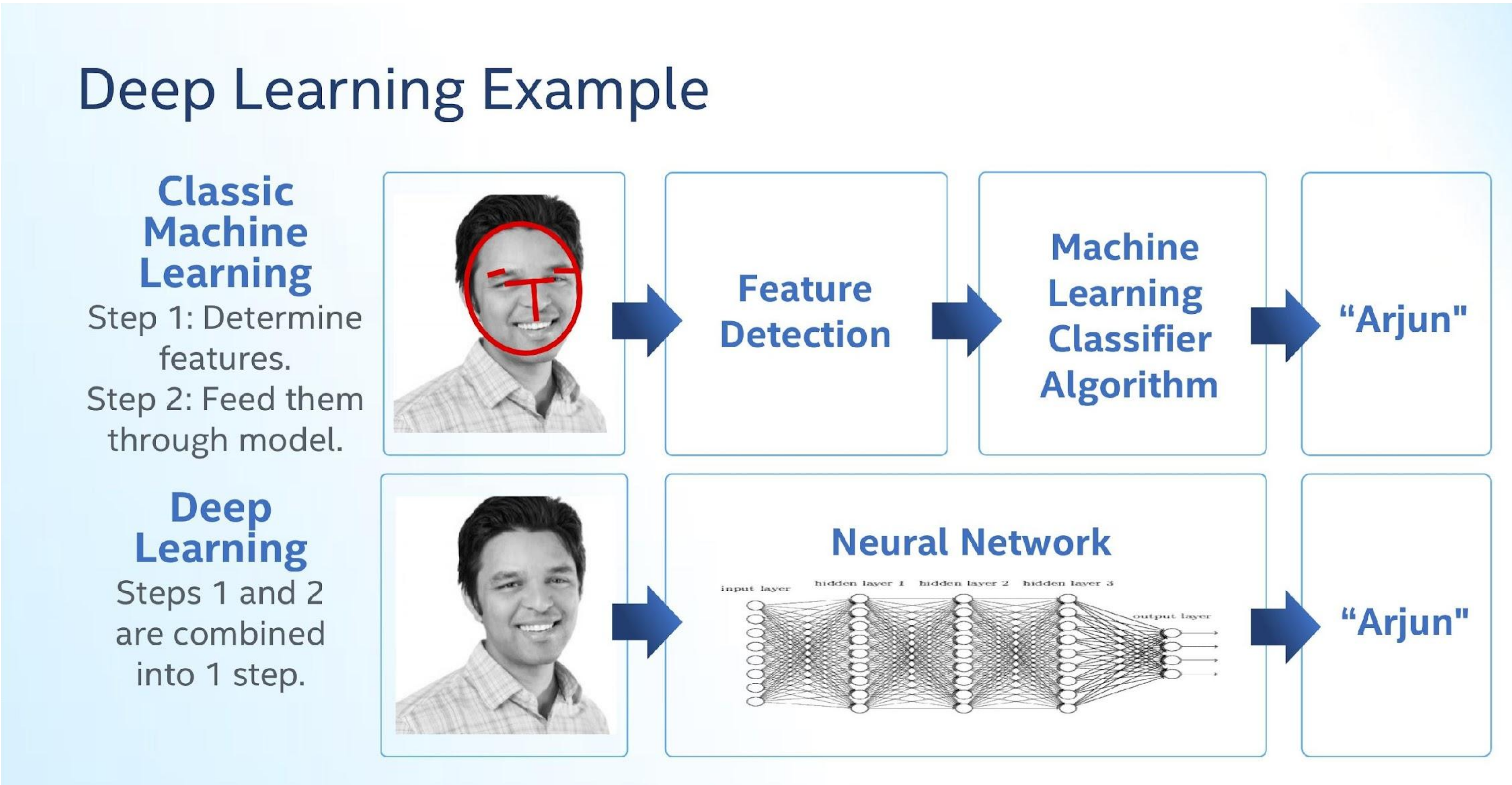
Deep Learning

“Machine learning that involves using very complicated models called “deep neural networks”." (Intel)

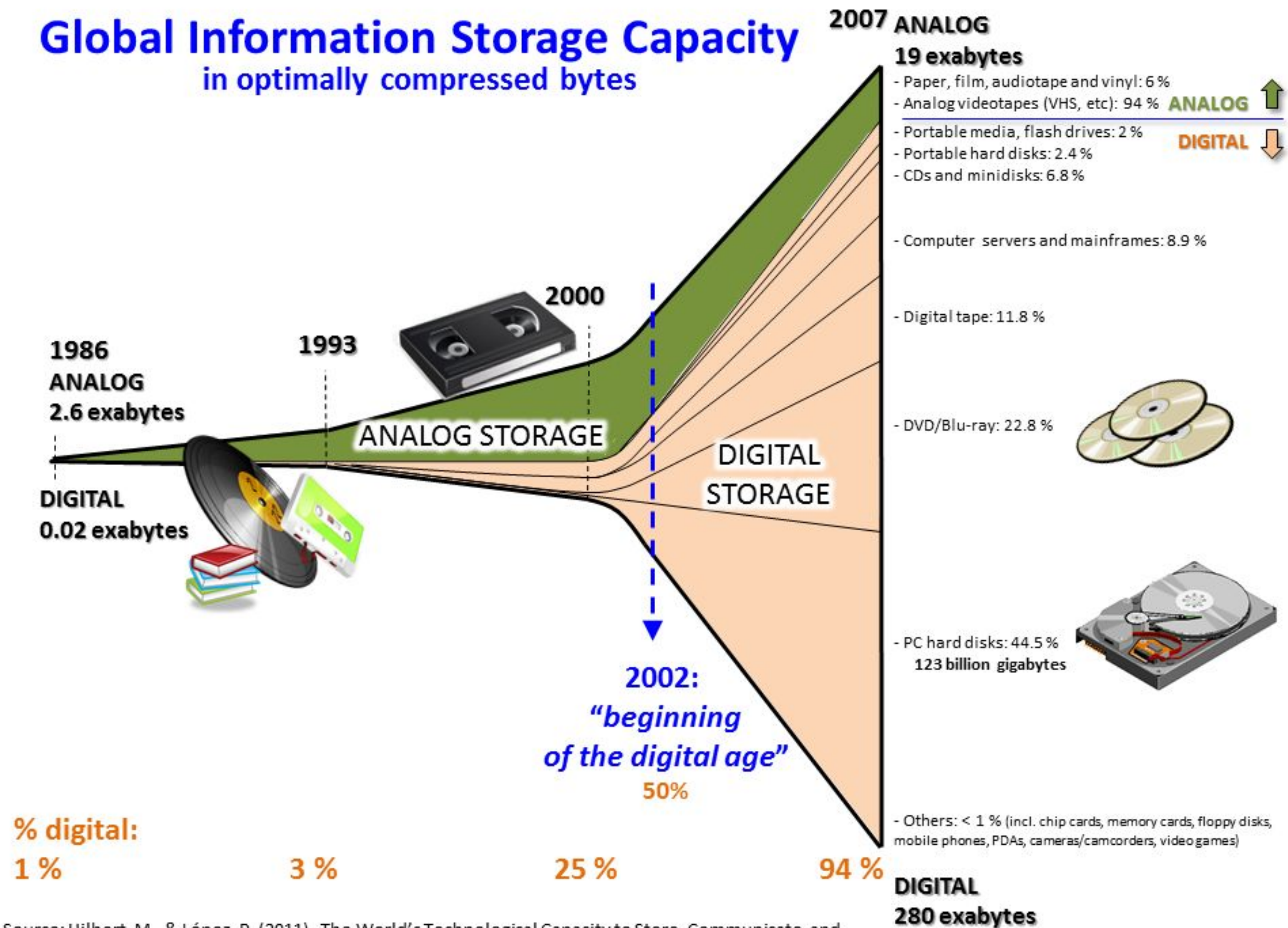
Models determine best representation of original data; in classic machine learning, humans must do this.



Deep Learning – DL
Example



Big Data

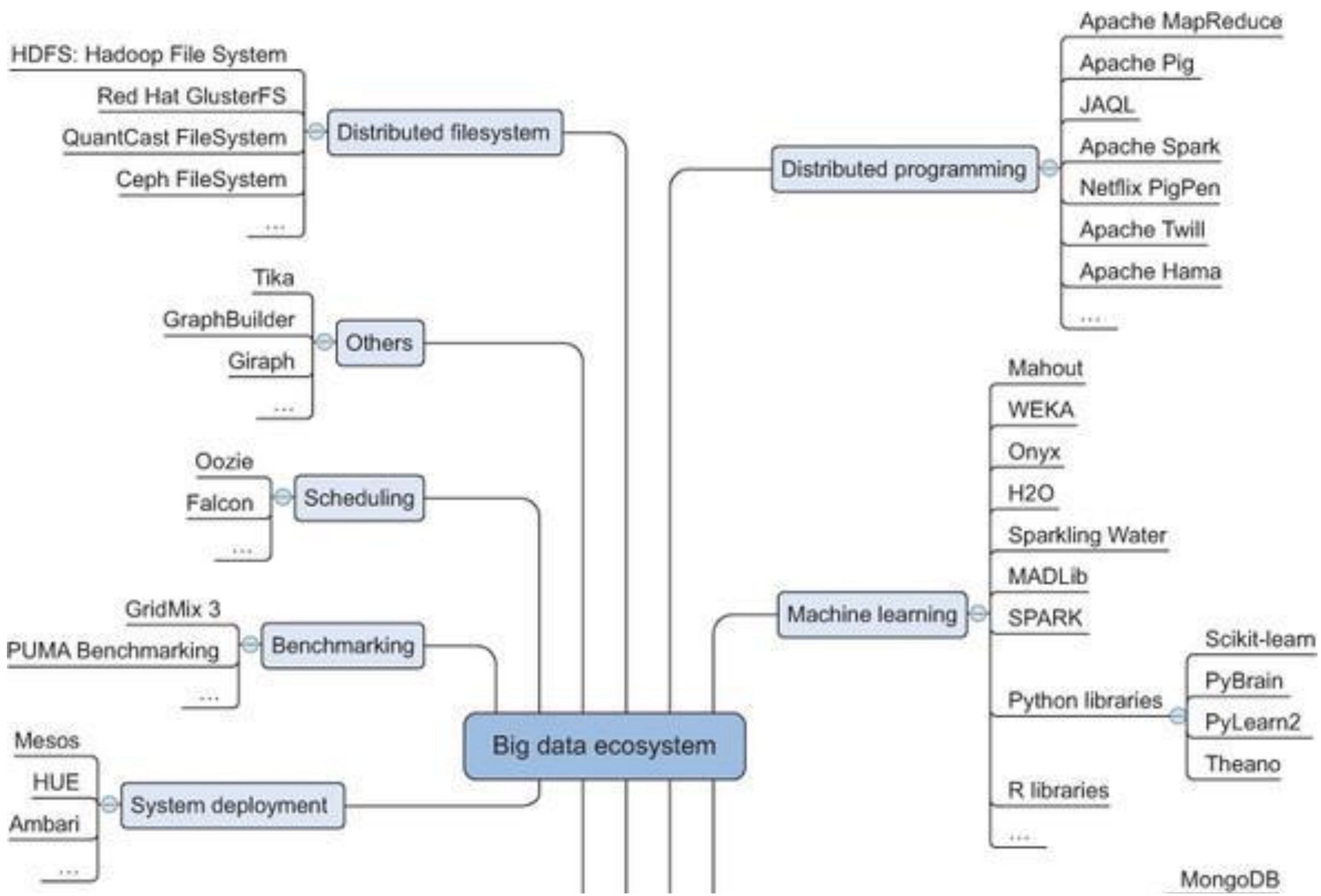


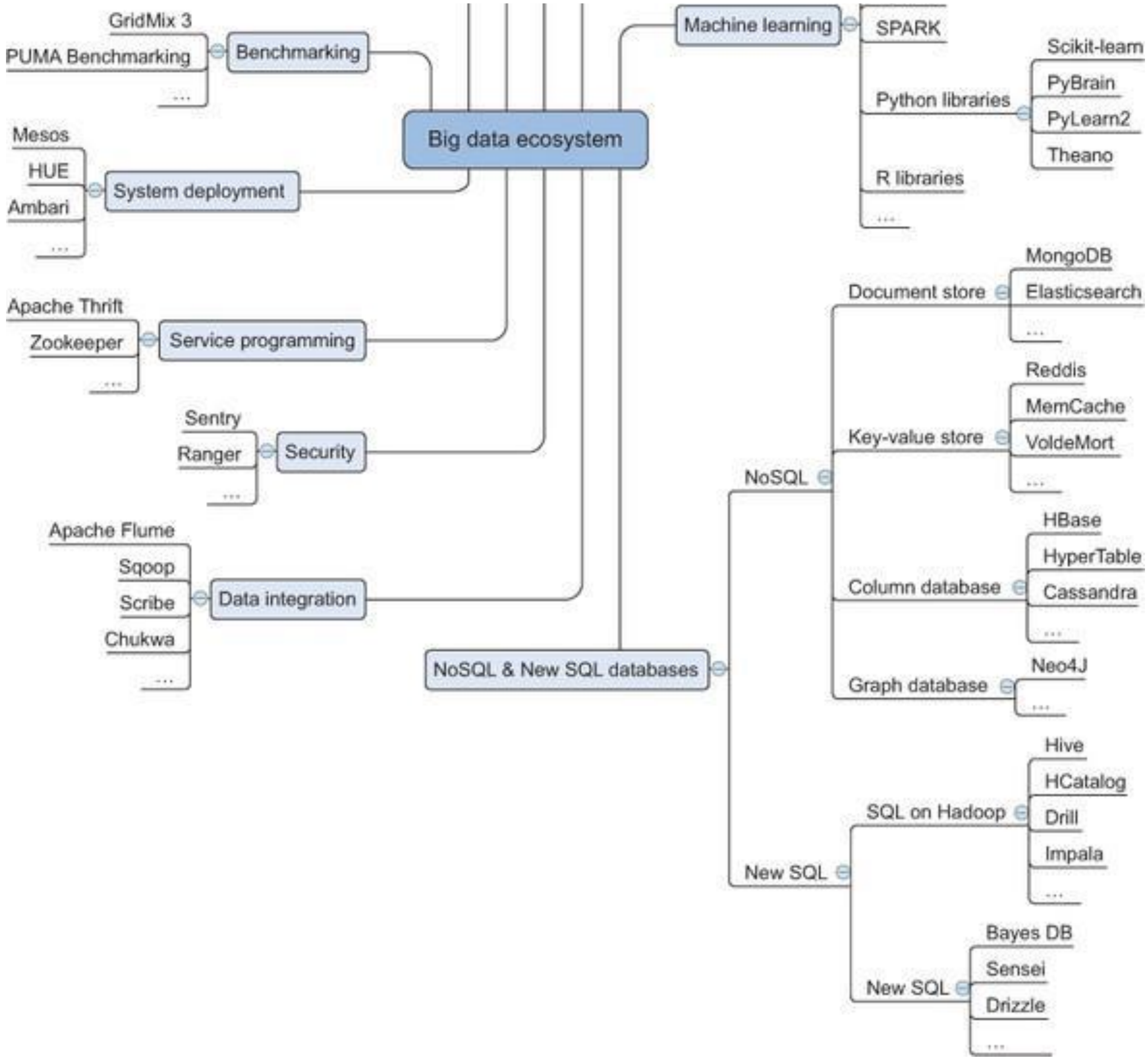
Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://www.martinhilbert.net/WorldInfoCapacity.html

Big Data
Landscape of Technologies

- File system
- Distributed programming frameworks
- Data integration
- Databases
- Machine learning
- Security
- Scheduling
- Benchmarking
- System deployment
- Service programming

Big Data
Ecosystem

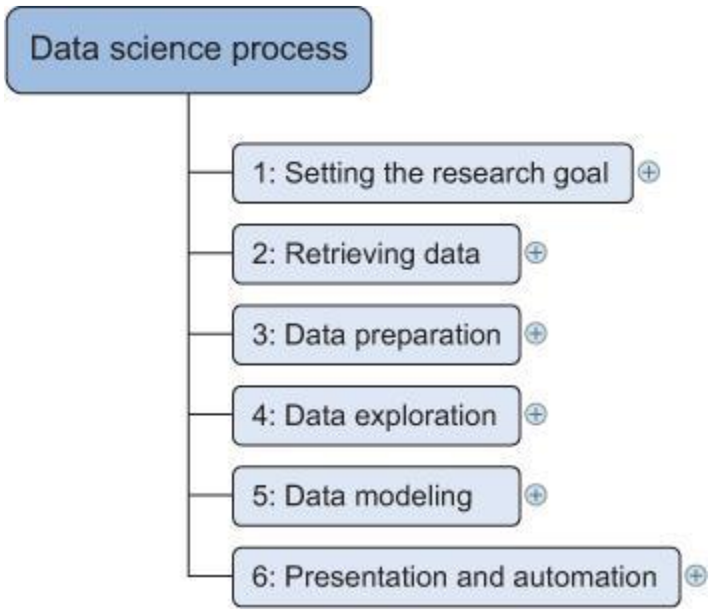



Big data ecosystem
Benchmarking
GridMix 3
PUMA Benchmarking
...
System deployment
Mesos
HUE
Ambari
...
Service programming
Apache Thrift
Zookeeper
...
Security
Sentry
Ranger
...
Data integration
Apache Flume
Sqoop
Scribe
Chukwa
...
Machine learning
SPARK
Python libraries
Scikit-learn
PyBrain
PyLearn2
Theano
R libraries
...
NoSQL & New SQL databases
NoSQL
Document store
MongoDB
Elasticsearch
...
Key-value store
Reddis
MemCache
VoldeMort
...
Column database
HBase
HyperTable
Cassandra
...
Graph database
Neo4J
...
New SQL
SQL on Hadoop
Hive
HCatalog
Drill
Impala
...
New SQL
Bayes DB
Sensei
Drizzle
...

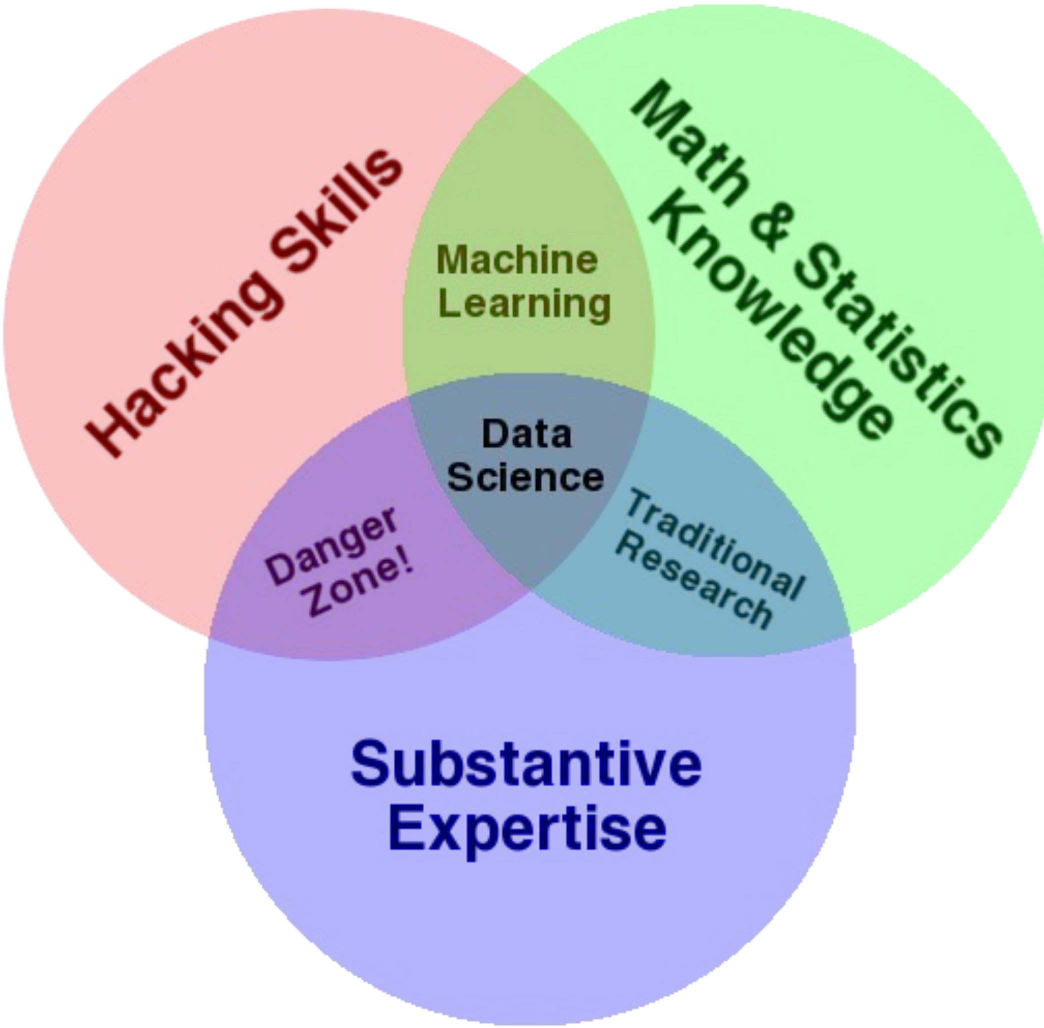
Data Science – DS

- “Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured, similar to data mining.” (Wikipedia)
- Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven)
- "The Sexiest Job of the 21st Century" – Harvard Business Review, 2012

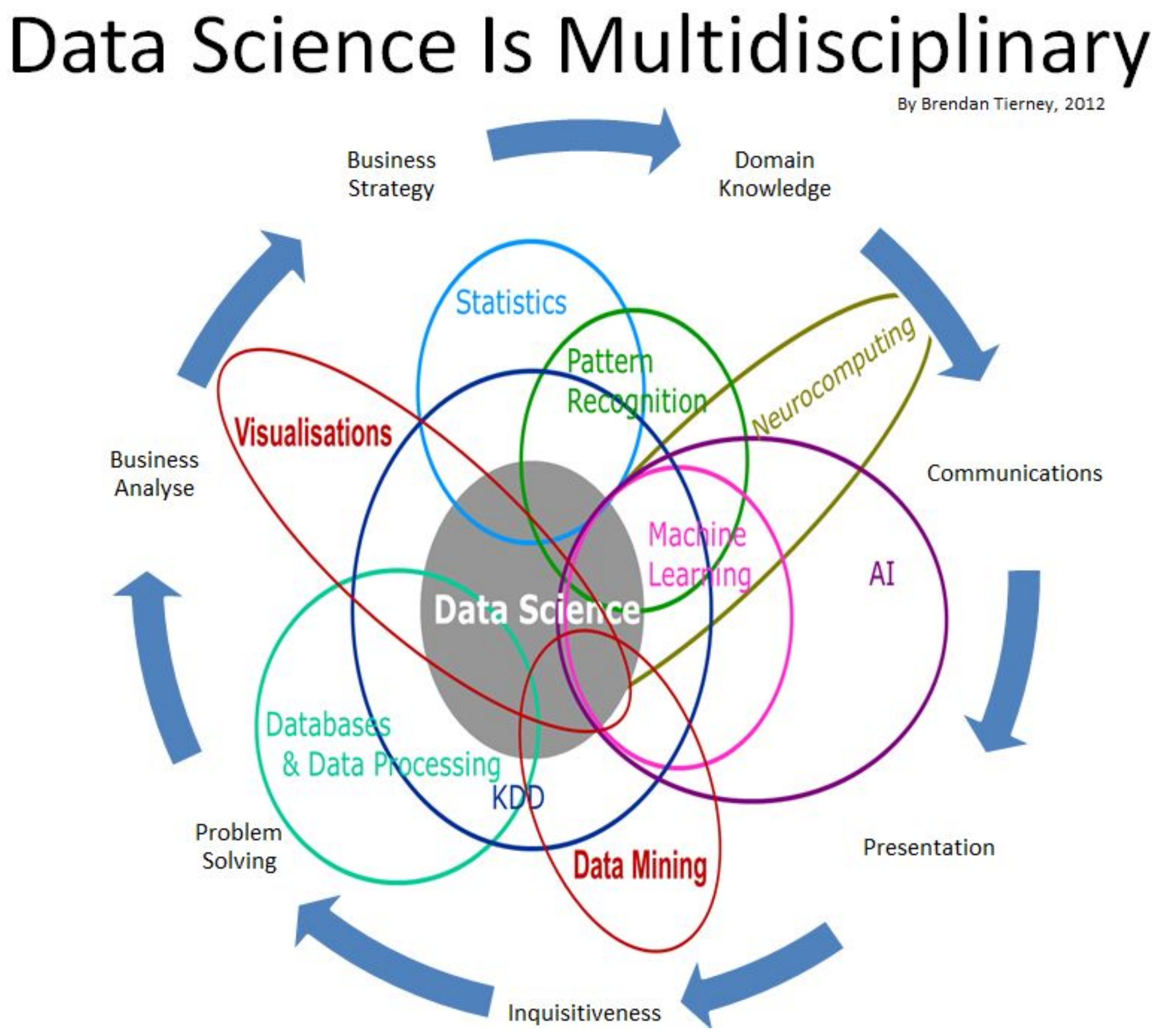
Data Science Process



Data Science – DS
Drew Conway, 2010

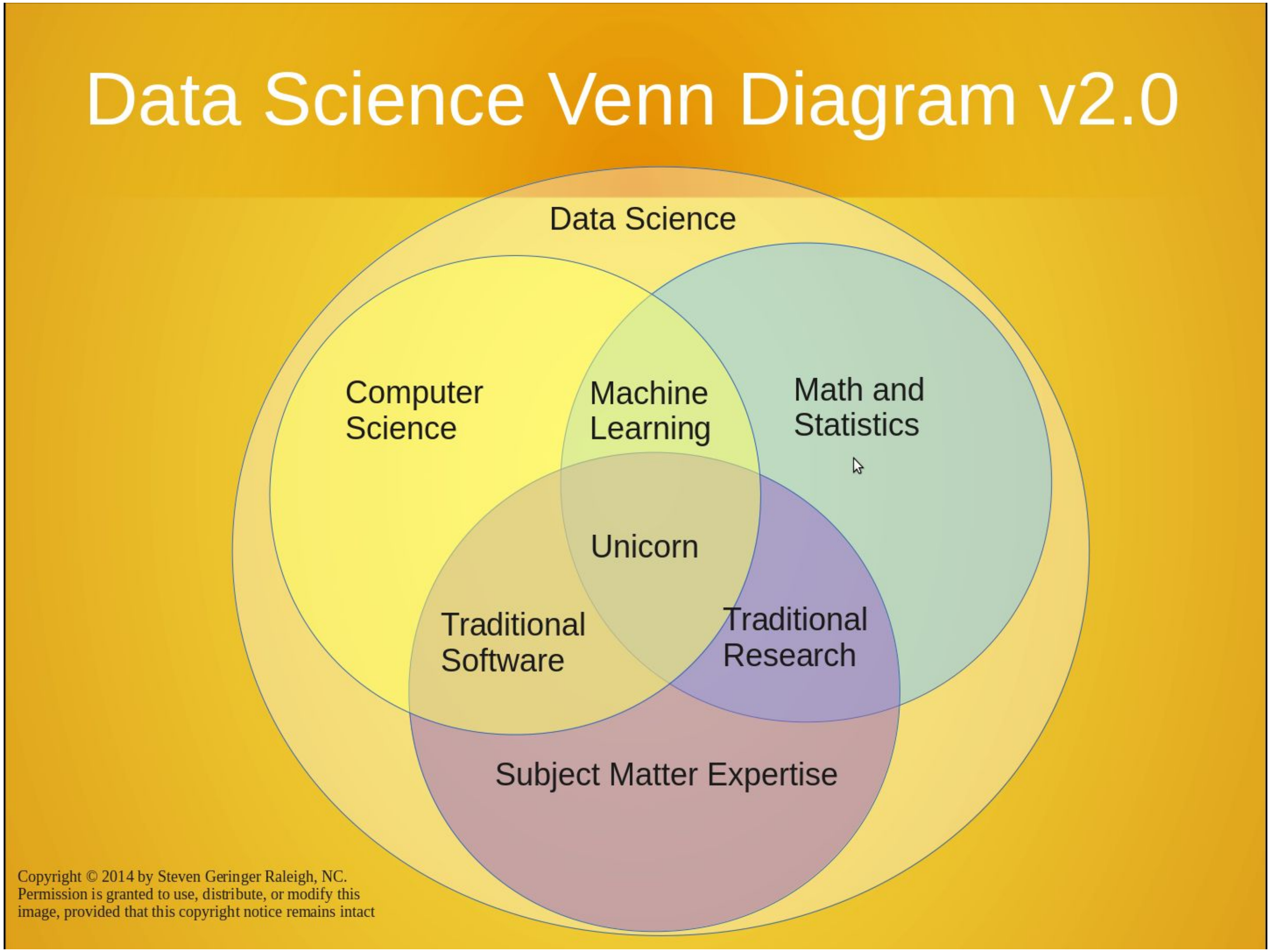


Data Science – DS
Brendan Tierney, 2012

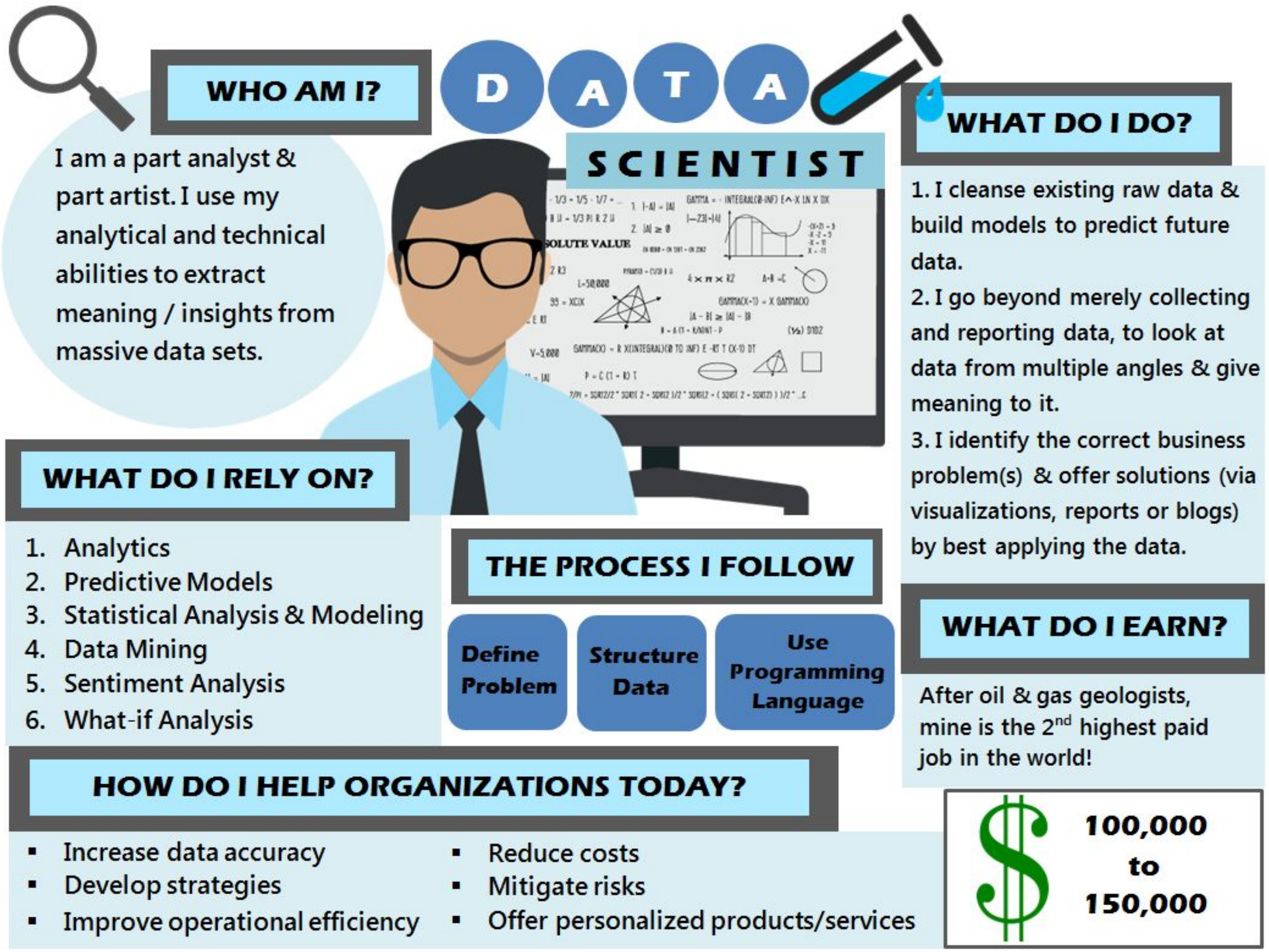


Data Science – DS

Steven Geringe, 2014



Data Science – DS
Data Scientist Portrait



Data Science – DS
Data Scientist – what does it take?

DATA SCIENTIST - WHAT DOES IT TAKE?

Data Scientist is fast becoming the most sought after job of the 21st century, requiring a blend of multidisciplinary skills including, but not limited to, mathematics, statistics, computer science, communication and general commercial acumen.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Bayesian inference
- Graph theory
- NLP
- **Supervised Learning:** decision trees, random forests, logistic regression
- **Unsupervised Learning;** clustering, dimensionality reduction
- **Optimisation:** gradient descent and variants

PROGRAMMING & DATABASE

- Computer science fundamentals
- Python, R, SAS, Scala
- Statistical computing package
- Database SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop, Hive/Pig and Spark
- Customer reducers
- Experience with xaaS like AWS
- Engineering

BI / CONSULTANCY / DOMAIN & SOFT KNOWLEDGE SKILLS

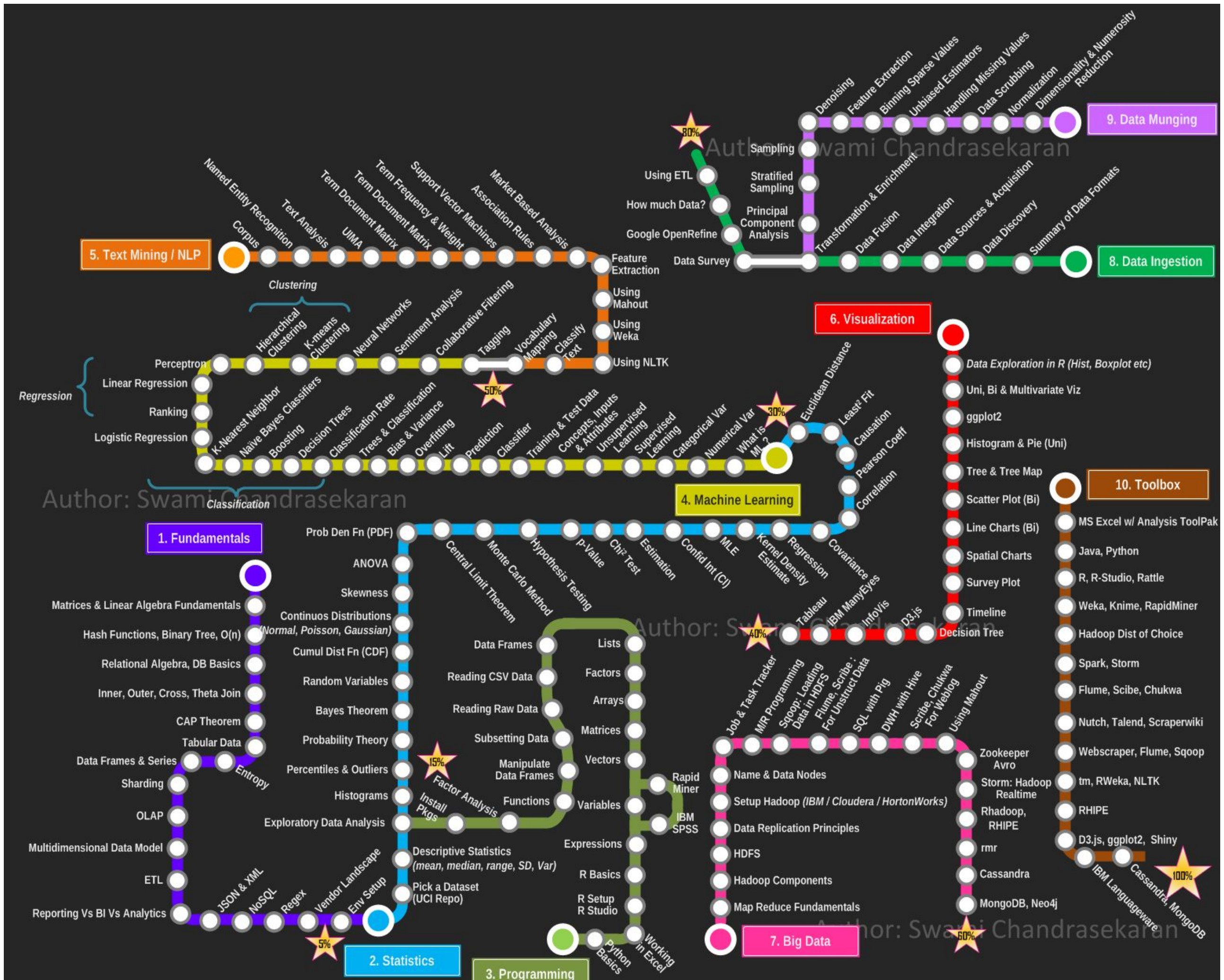
- Passionate about business problems
- Comfortable translating data-driven insights into decisions and actions
- 'Bridge the Gap' between business and technical departments
- Strategic, proactive, creative, innovative and collaborative

- Strong communicator, able to engage with senior stakeholders
- Influence without authority
- Problem solver
- Project management
- Hacker mindset
- Mentoring/Leadership

If you're interested in exploring potential opportunities within this fast paced and dynamic market, our specialist team of consultants are primed and ready to assist you.
Get in touch today!

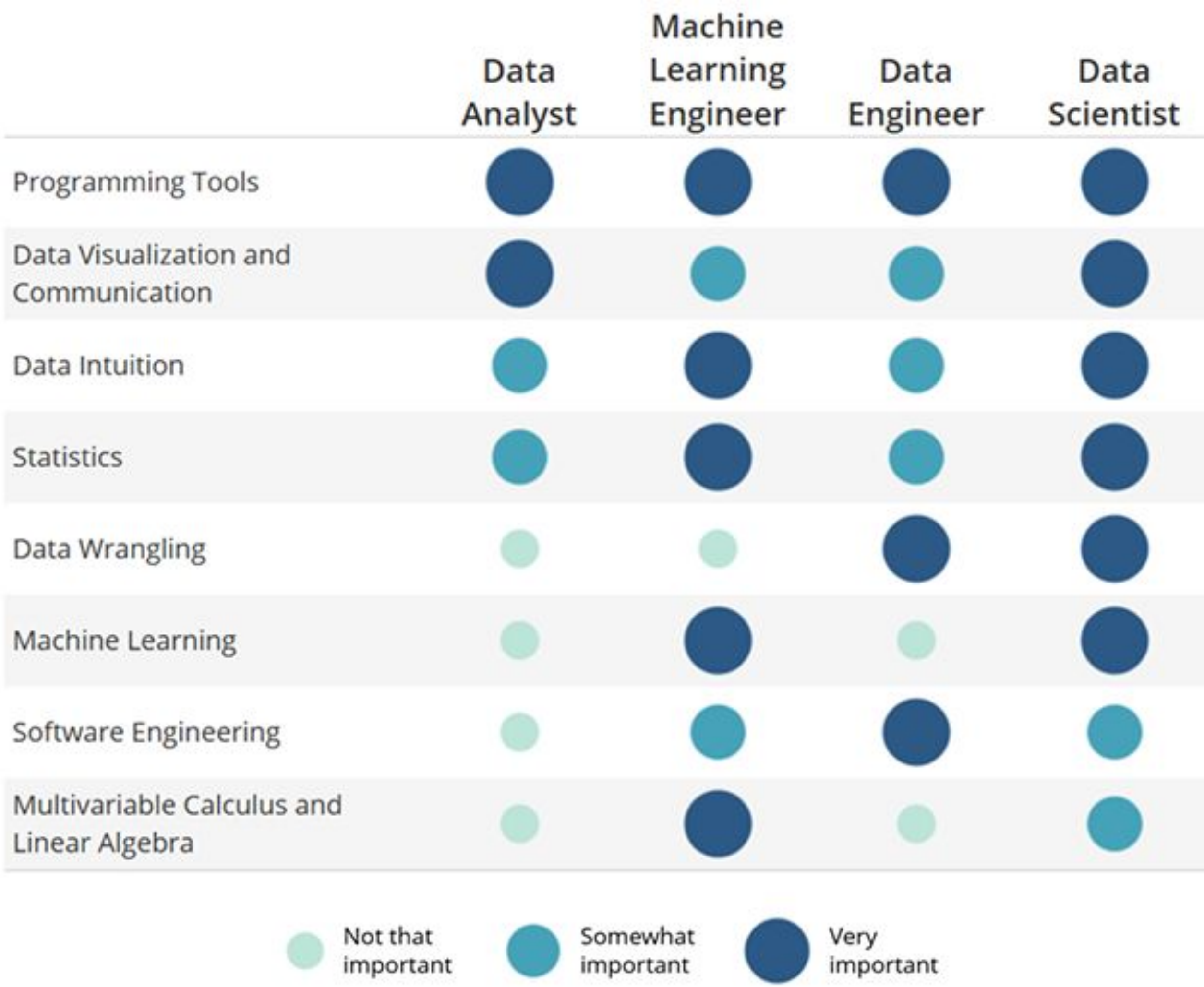
Data Science – DS

Road to Data Scientist



Data Science – DS

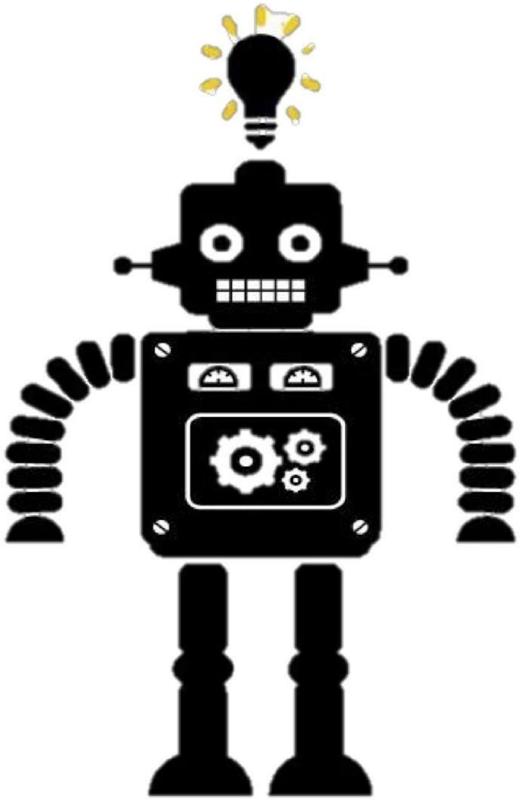
Skills of Data Scientist (Udacity)



Machine Learning – ML
What is Machine Learning?

WHAT IS MACHINE LEARNING?

Machine learning allows computers to learn and infer from data.



Machine Learning
in Our Daily Lives

MACHINE LEARNING IN OUR DAILY LIVES

SPAM FILTERING

Machine Learning

in Our Daily Lives

Machine Learning
in Our Daily Lives

MACHINE LEARNING IN OUR DAILY LIVES

- SPAM FILTERING
- WEB SEARCH
- POSTAL MAIL ROUTING

Machine Learning
in Our Daily Lives

MACHINE LEARNING IN OUR DAILY LIVES

SPAM FILTERING	WEB SEARCH	POSTAL MAIL ROUTING
FRAUD DETECTION	MOVIE RECOMMENDATIONS	VEHICLE DRIVER ASSISTANCE
WEB ADVERTISEMENTS	SOCIAL NETWORKS	SPEECH RECOGNITION

Machine Learning – ML

Vocabulary

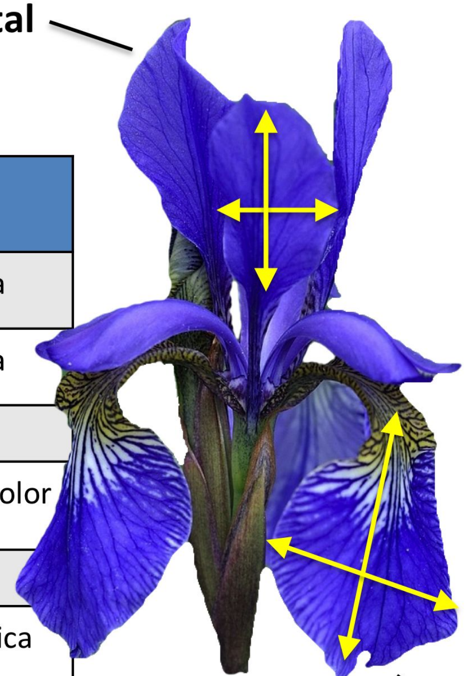
Samples
(instances, observations)

	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
			…		
50	6.4	3.5	4.5	1.2	Versicolor
			…		
150	5.9	3.0	5.0	1.8	Virginica

Features
(attributes, measurements, dimensions)

Class labels
(targets)

Petal



Sepal

Machine Learning – ML
Vocabulary

MACHINE LEARNING VOCABULARY

- **Target: predicted category or value of the data**
  (column to predict)

Machine Learning – ML
Vocabulary

MACHINE LEARNING VOCABULARY

Sepal length	Sepal width	Petal length	Petal width	Species
6.7	3.0	5.2	2.3	Virginica
6.4	2.8	5.6	2.1	Virginica
4.6	3.4	1.4	0.3	Setosa
6.9	3.1	4.9	1.5	Versicolor
4.4	2.9	1.4	0.2	Setosa
4.8	3.0	1.4	0.1	Setosa
5.9	3.0	5.1	1.8	Virginica
5.4	3.9	1.3	0.4	Setosa
4.9	3.0	1.4	0.2	Setosa
5.4	3.4	1.7	0.2	Setosa

Machine Learning – ML

Vocabulary

MACHINE LEARNING VOCABULARY

Sepal length	Sepal width	Petal length	Petal width	Species
6.7	3.0	5.2	2.3	Virginica
6.4	2.8	5.6	2.1	Virginica
4.6	3.4	1.4	0.3	Setosa
6.9	3.1	4.9	1.5	Versicolor
4.4	2.9	1.4	0.2	Setosa
4.8	3.0	1.4	0.1	Setosa
5.9	3.0	5.1	1.8	Virginica
5.4	3.9	1.3	0.4	Setosa
4.9	3.0	1.4	0.2	Setosa
5.4	3.4	1.7	0.2	Setosa

Target

Machine Learning – ML
Vocabulary

MACHINE LEARNING VOCABULARY

- **Target: predicted category or value of the data**
  (column to predict)
- **Features: properties of the data used for prediction**
  (non-target columns)

Machine Learning – ML
Vocabulary

MACHINE LEARNING VOCABULARY

Features

Sepal length	Sepal width	Petal length	Petal width	Species
6.7	3.0	5.2	2.3	Virginica
6.4	2.8	5.6	2.1	Virginica
4.6	3.4	1.4	0.3	Setosa
6.9	3.1	4.9	1.5	Versicolor
4.4	2.9	1.4	0.2	Setosa
4.8	3.0	1.4	0.1	Setosa
5.9	3.0	5.1	1.8	Virginica
5.4	3.9	1.3	0.4	Setosa
4.9	3.0	1.4	0.2	Setosa
5.4	3.4	1.7	0.2	Setosa

Machine Learning – ML
Vocabulary

MACHINE LEARNING VOCABULARY

- **Target: predicted category or value of the data**
  (column to predict)
- **Features: properties of the data used for prediction**
  (non-target columns)
- **Example: a single data point within the data**
  (one row)

Machine Learning – ML
Vocabulary

MACHINE LEARNING VOCABULARY

Examples →

Sepal length	Sepal width	Petal length	Petal width	Species
6.7	3.0	5.2	2.3	Virginica
6.4	2.8	5.6	2.1	Virginica
4.6	3.4	1.4	0.3	Setosa
6.9	3.1	4.9	1.5	Versicolor
4.4	2.9	1.4	0.2	Setosa
4.8	3.0	1.4	0.1	Setosa
5.9	3.0	5.1	1.8	Virginica
5.4	3.9	1.3	0.4	Setosa
4.9	3.0	1.4	0.2	Setosa
5.4	3.4	1.7	0.2	Setosa

Machine Learning – ML

Vocabulary

MACHINE LEARNING VOCABULARY

- **Target: predicted category or value of the data**
  (column to predict)
- **Features: properties of the data used for prediction**
  (non-target columns)
- **Example: a single data point within the data**
  (one row)
- **Label: the target value for a single data point**

Machine Learning – ML
Vocabulary

MACHINE LEARNING VOCABULARY

Sepal length	Sepal width	Petal length	Petal width	Species
6.7	3.0	5.2	2.3	Virginica
6.4	2.8	5.6	2.1	Virginica
4.6	3.4	1.4	0.3	Setosa
6.9	3.1	4.9	1.5	Versicolor
4.4	2.9	1.4	0.2	Setosa
4.8	3.0	1.4	0.1	Setosa
5.9	3.0	5.1	1.8	Virginica
5.4	3.9	1.3	0.4	Setosa
4.9	3.0	1.4	0.2	Setosa
5.4	3.4	1.7	0.2	Setosa

← Label

Machine Learning – ML
Vocabulary (synonyms)

MACHINE LEARNING VOCABULARY (SYNONYMS)

- **Target:** Response, Output, Dependent Variable, Labels
- **Features:** Predictors, Input, Independent Variables, Attributes
- **Example:** Observation, Record, Instance, Datapoint, Row
- **Label:** Answer, y-value, Category

Machine Learning – ML

Two Main Types

Two Main Types of Machine Learning

	Dataset	Goal	Example
Supervised Learning	Has a target column	Make predictions	Fraud detection
Unsupervised Learning	Does not have a target column	Find structure in the data	Customer segmentation

Machine Learning – ML
Types of Machine Learning

TYPES OF MACHINE LEARNING

SUPERVISED	Data points have known outcome

Machine Learning – ML
Types of Machine Learning

TYPES OF MACHINE LEARNING

SUPERVISED	Data points have known outcome
UNSUPERVISED	Data points have unknown outcome

Machine Learning – ML
Types of Machine Learning

TYPES OF MACHINE LEARNING

SUPERVISED — Data points have known outcome

UNSUPERVISED — Data points have unknown outcome

Machine learning – ML

Types of Supervised Learning

TYPES OF SUPERVISED LEARNING

REGRESSION	Outcome is continuous (numerical)
CLASSIFICATION	Outcome is a category

Machine Learning – ML

Types of Machine Learning

TYPES OF MACHINE LEARNING

SUPERVISED	Data points have known outcome
UNSUPERVISED	Data points have unknown outcome

Machine Learning – ML
Types of Unsupervised Learning

TYPES OF UNSUPERVISED LEARNING

CLUSTERING	Identify unknown structure in data
DIMENSIONALITY REDUCTION	Use structural characteristics to simplify data

Machine learning – ML

Types of Supervised Learning

TYPES OF SUPERVISED LEARNING

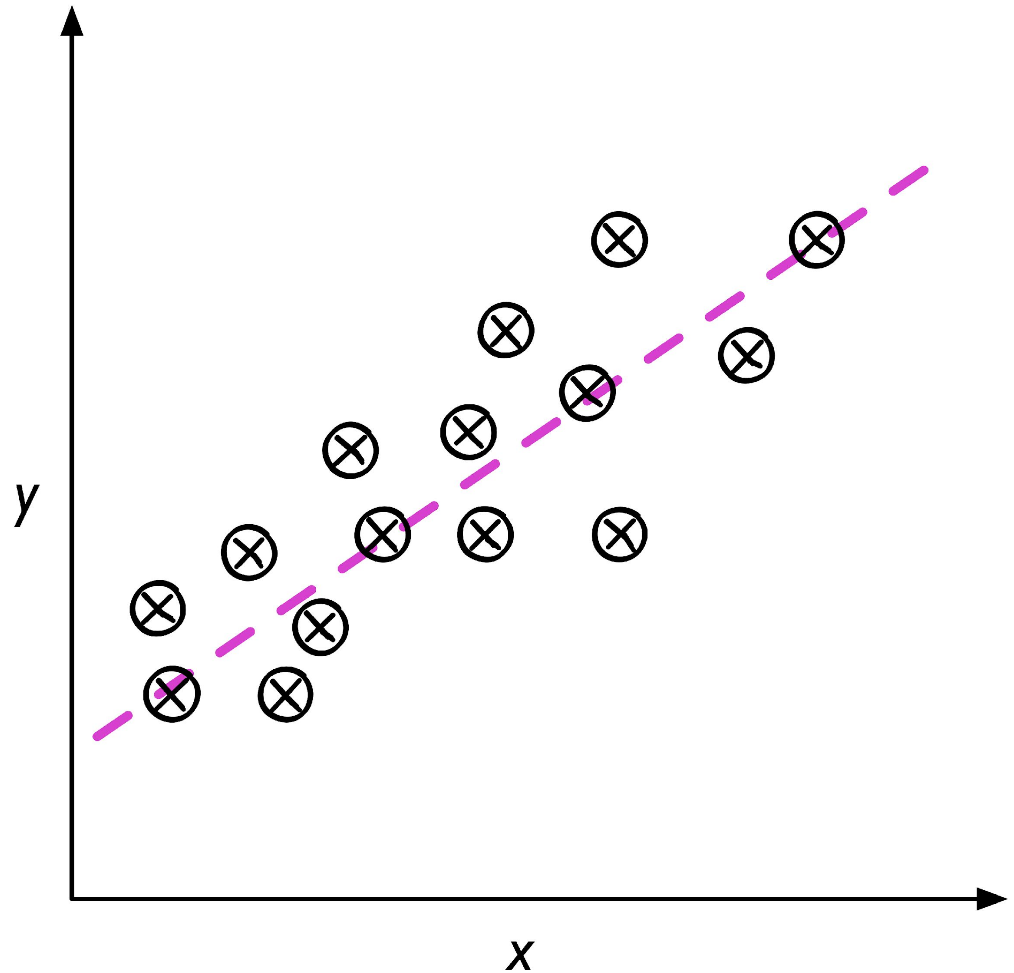
REGRESSION	Outcome is continuous (numerical)
CLASSIFICATION	Outcome is a category

Machine learning – ML
Regression (example)

	вес	рост	ср.дл. волос	пол	возраст (y)
x_1	96	170	короткие	м	42
x_2	60	180	короткие	м	25
x_3	54	165	длинные	ж	30
x_4	83	178	короткие	ж	47
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_{100}	108	193	длинные	ж	32

Задача обучения: определить возраст
$x = (75, 184, \text{“короткие”}, \text{“м”})$, $y = ?$

Machine learning – ML
Regression (graphical representation)



Machine learning – ML

Types of Supervised Learning

TYPES OF SUPERVISED LEARNING

REGRESSION	Outcome is continuous (numerical)
CLASSIFICATION	Outcome is a category

Machine learning – ML
Classification (example 1)

	пульс	гемоглобин	диагноз
x_1	70	140	здоров ($y = -1$)
x_2	60	160	здоров ($y = -1$)
x_3	94	120	миокардит ($y = 1$)
$\dots$	$\dots$	$\dots$	$\dots$
x_{114}	86	98	миокардит ($y = 1$)

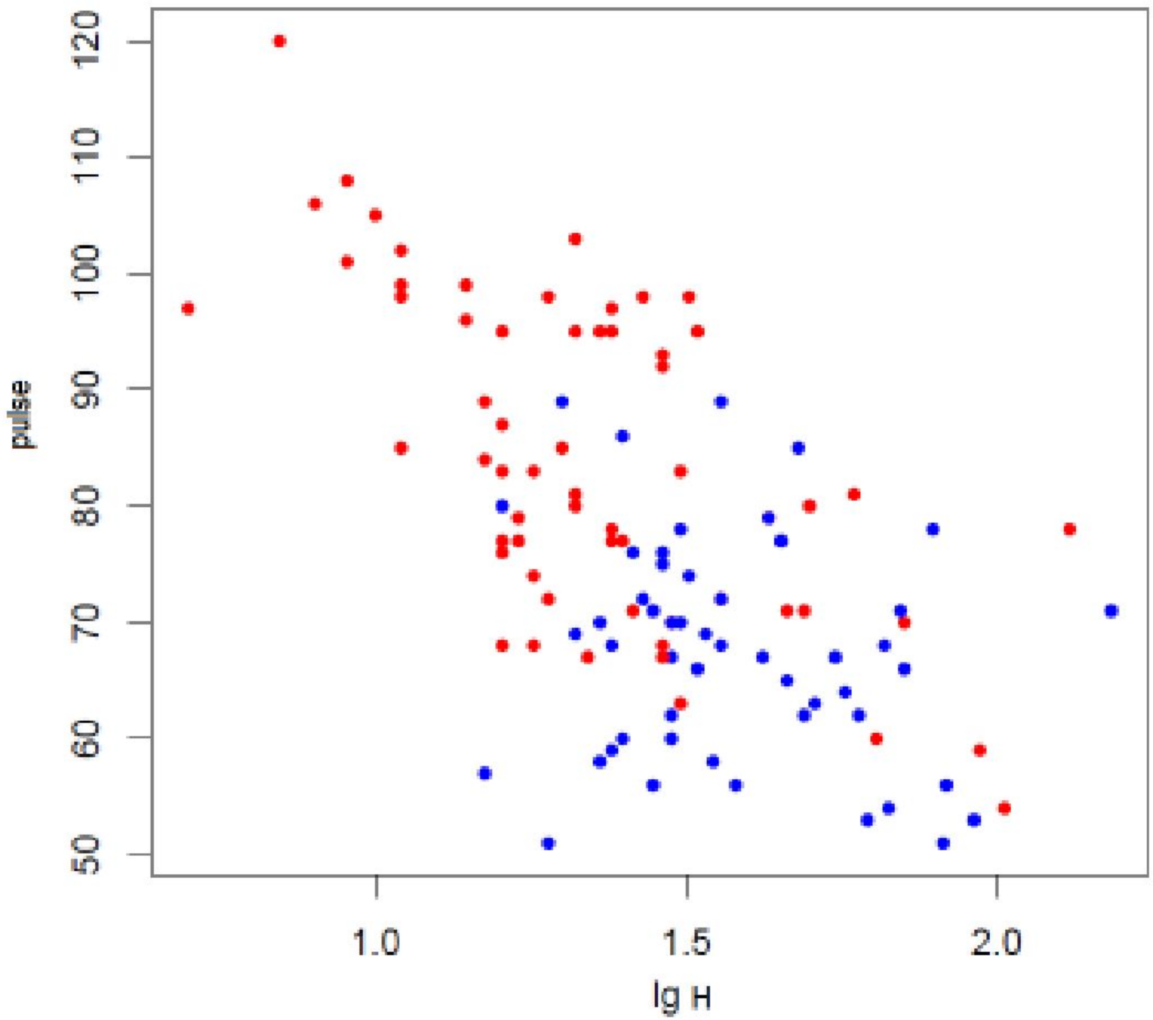
Обучающая выборка:

$$((70, 140), -1), (60, 160), -1), (94, 120), 1)\dots, (86, 98), 1))$$

Задача обучения: новый пациент $x = (75, 128)$, $y =?$

Machine learning – ML

Classification (example 1)

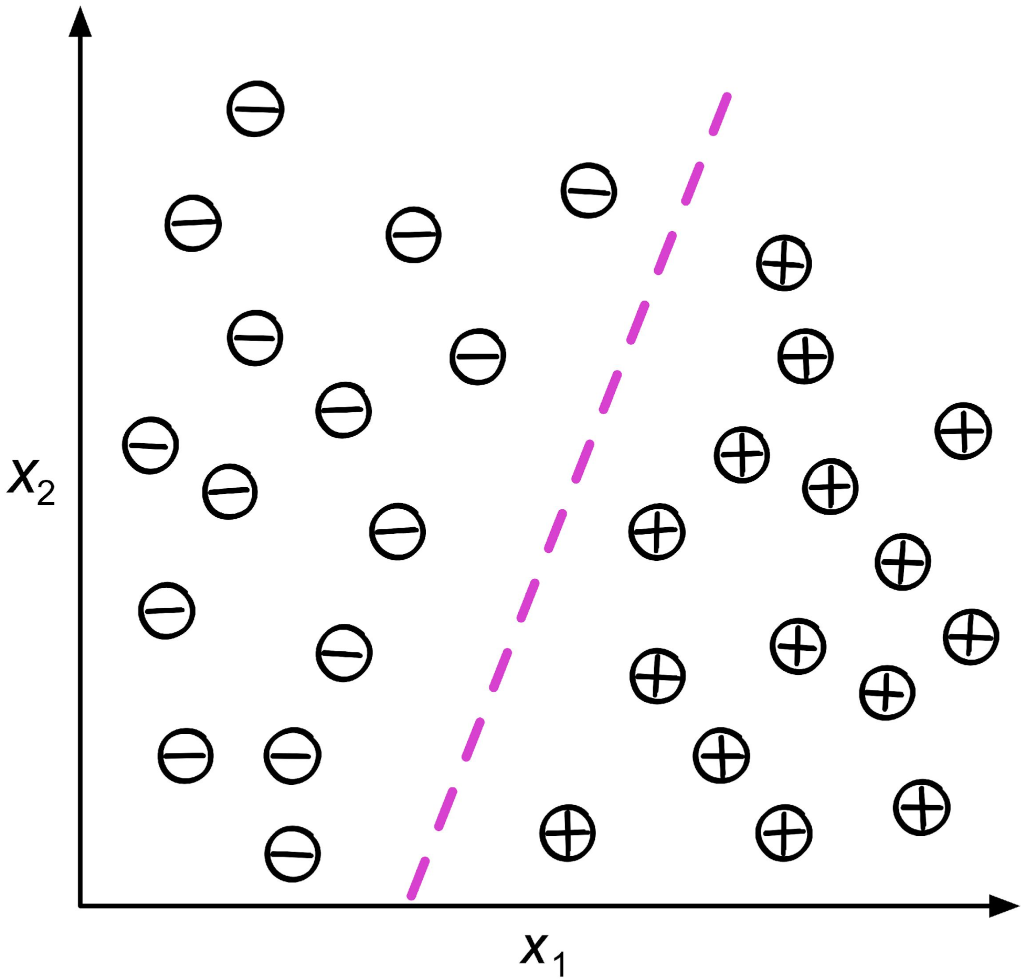


Machine learning – ML
Classification (example 2)

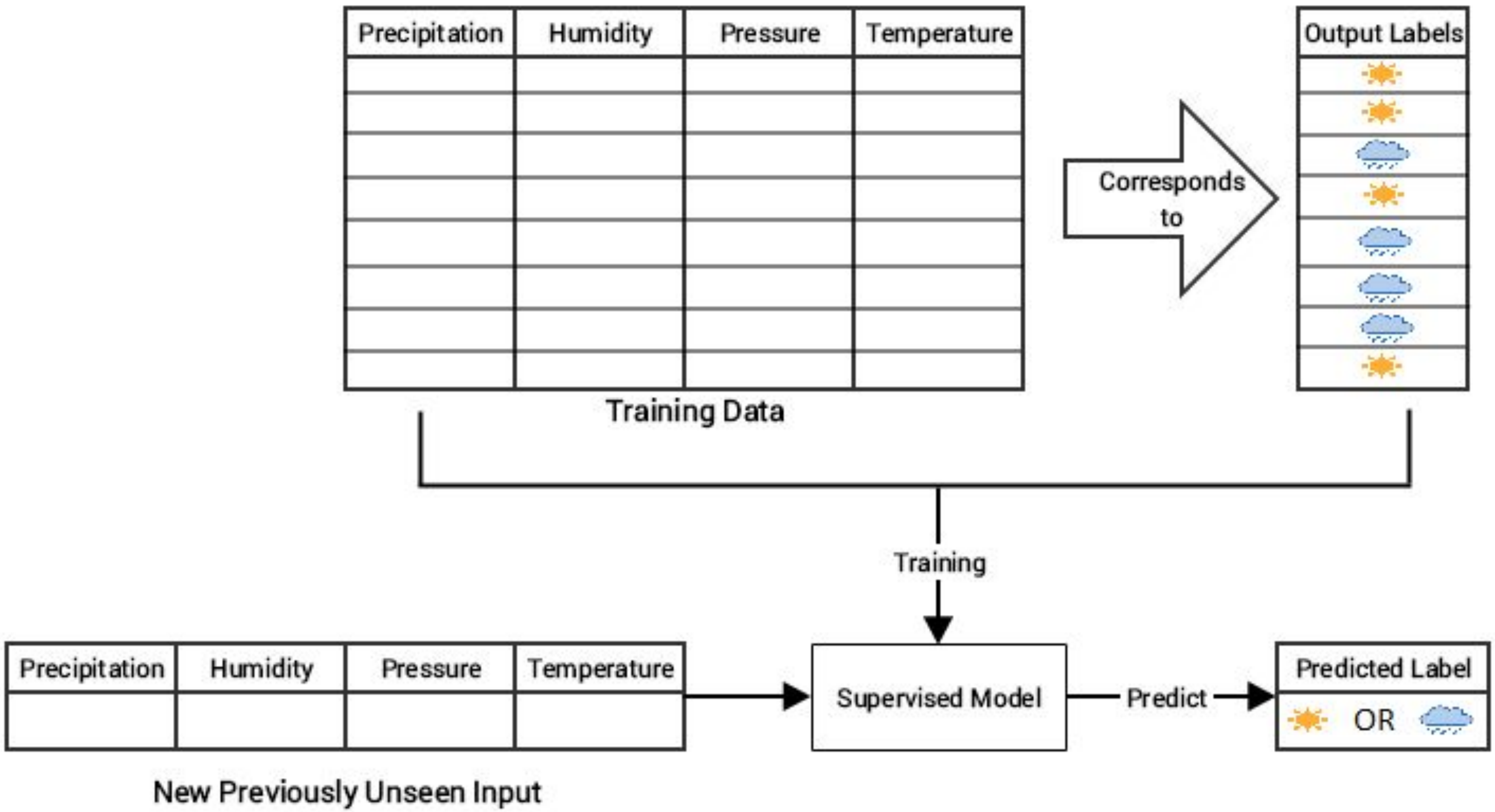
	вес	рост	возраст	ср.дл.волос	пол
x_1	96	170	42	0	м $(y = -1)$
x_2	60	180	25	8	м $(y = -1)$
x_3	54	165	30	21	ж $(y = 1)$
x_4	83	178	47	18	ж $(y = 1)$
...	...	...	...	...	...
x_{100}	108	193	32	40	ж $(y = 1)$

Задача обучения: $x = (75, 184, 28, 10)$, $y = ?$

Machine learning – ML
Classification (graphical representation)

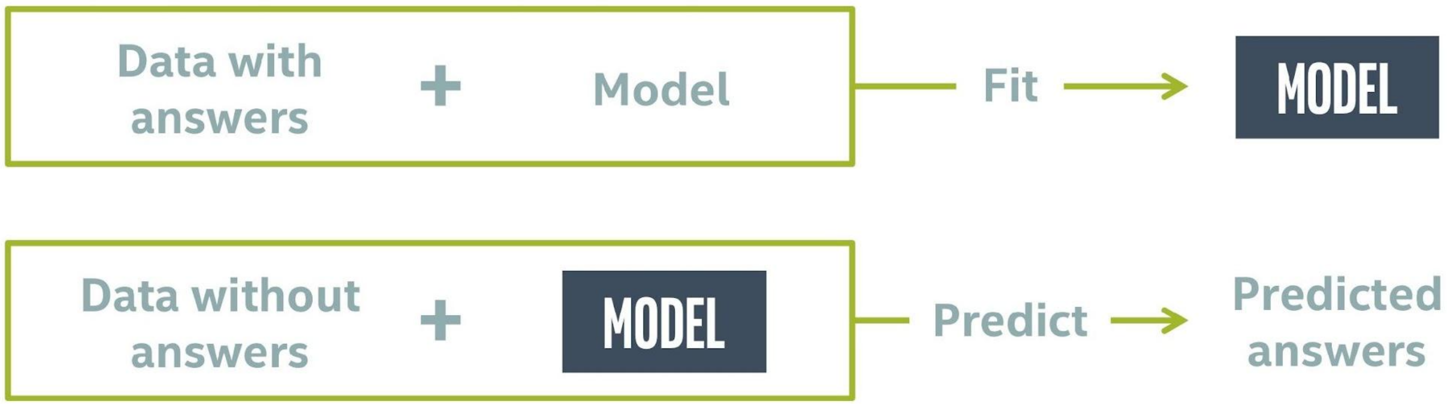


Machine Learning – ML
Supervised Learning Overview



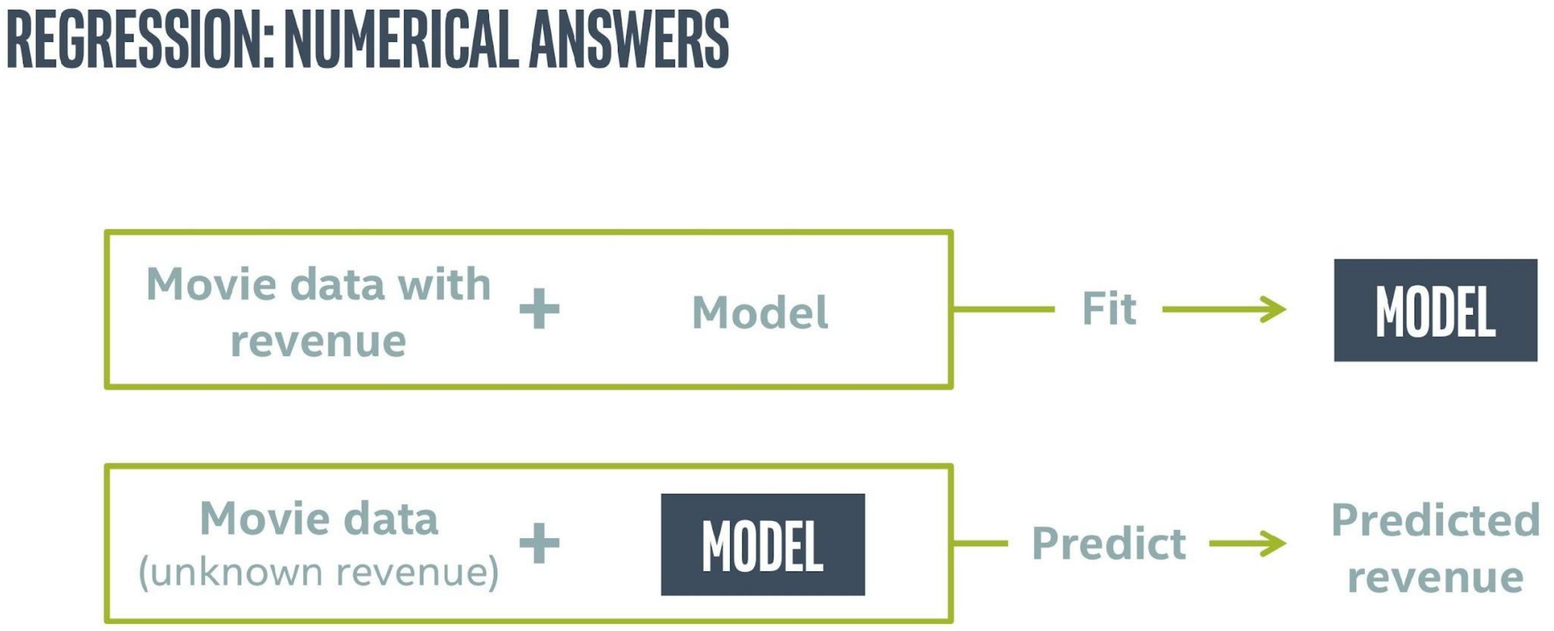
Machine Learning – ML
Supervised Learning Overview

SUPERVISED LEARNING OVERVIEW

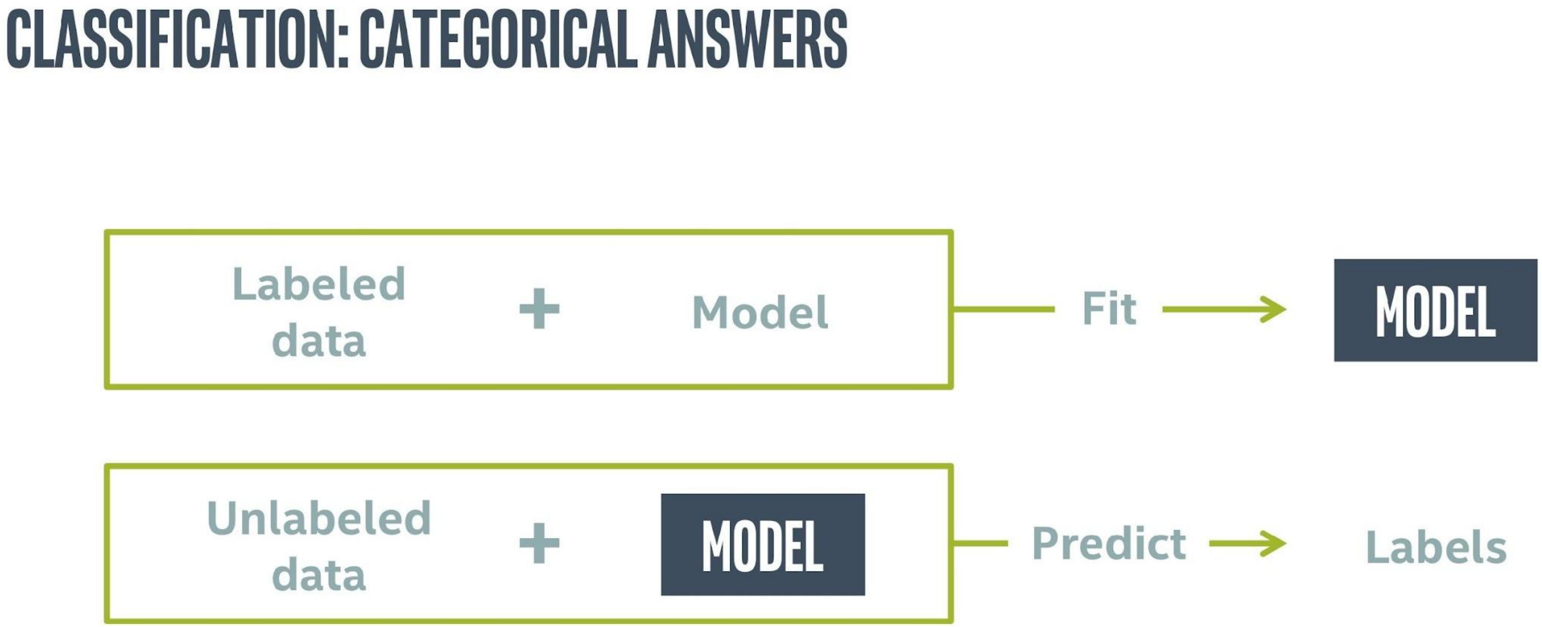


Machine Learning – ML

Regression: Numerical Answers

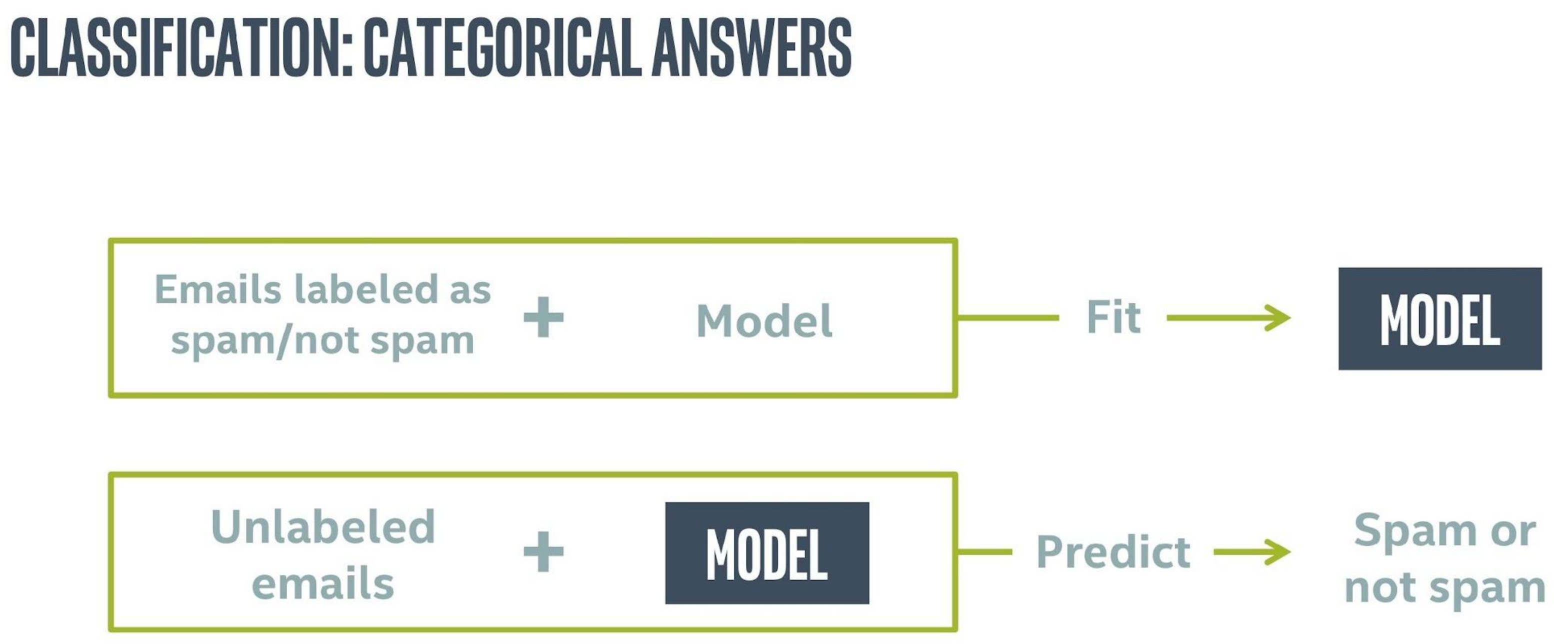


Machine Learning – ML
Classification: Categorical Answers



Machine Learning – ML

Classification: Categorical Answers



Machine Learning – ML
Types of Unsupervised Learning

TYPES OF UNSUPERVISED LEARNING

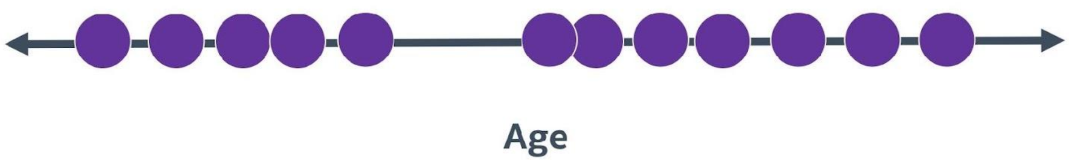
CLUSTERING	Identify unknown structure in data
DIMENSIONALITY REDUCTION	Use structural characteristics to simplify data

Machine Learning – ML
Clustering (example 1)

INTRODUCTION TO UNSUPERVISED LEARNING

Users of a web application:

- One feature (age)

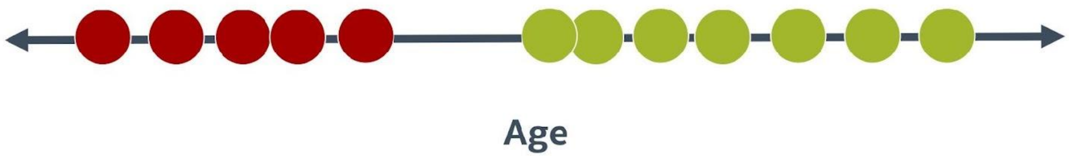


Machine Learning – ML
Clustering (example 1)

INTRODUCTION TO UNSUPERVISED LEARNING

Users of a web application:

- One feature (age)
- Two clusters



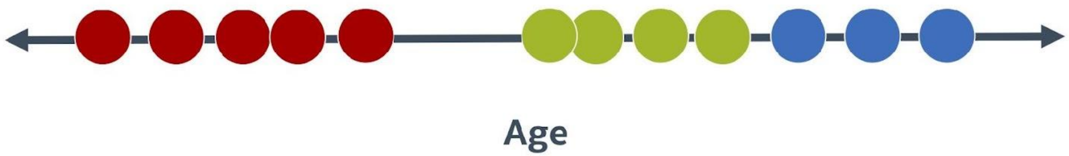
Machine Learning – ML

Clustering (example 1)

INTRODUCTION TO UNSUPERVISED LEARNING

Users of a web application:

- One feature (age)
- Three clusters

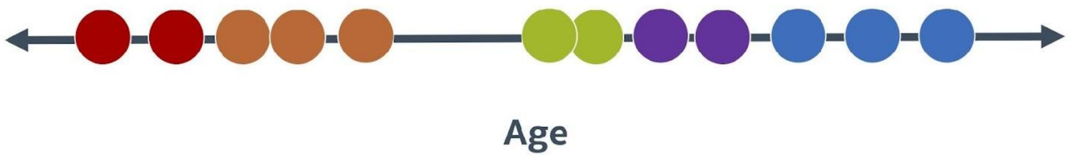


Machine Learning – ML
Clustering (example 1)

INTRODUCTION TO UNSUPERVISED LEARNING

Users of a web application:

- One feature (age)
- Five clusters



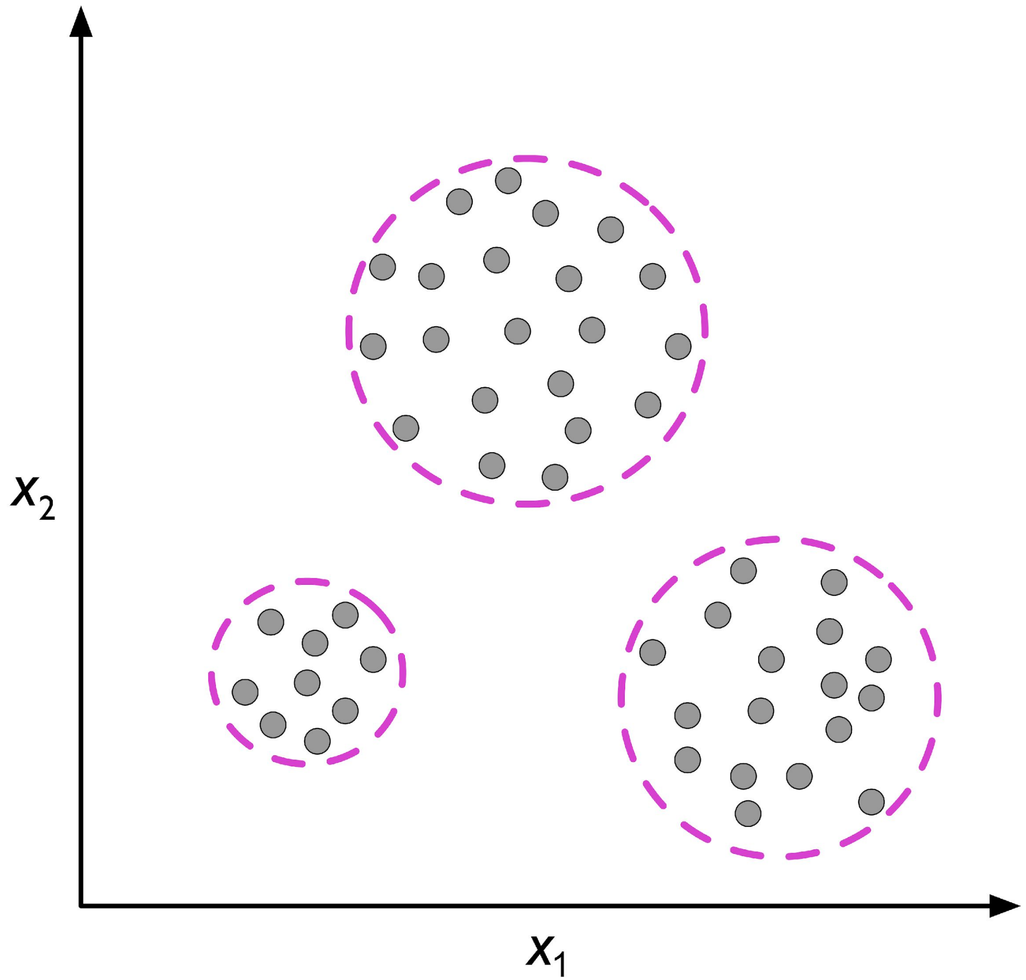
Machine Learning – ML

Clustering (example 2)

	вес	рост	возраст	ср.дл.волос
x_1	96	170	42	короткие
x_2	60	180	25	короткие
x_3	54	165	30	длинные
x_4	83	178	47	короткие
. . .	. . .	. . .	. . .	. . .
x_{100}	108	193	32	длинные

Задача обучения: "отгадать" пол всех людей из обучающей выборки

Machine Learning – ML
Clustering (graphical representation)



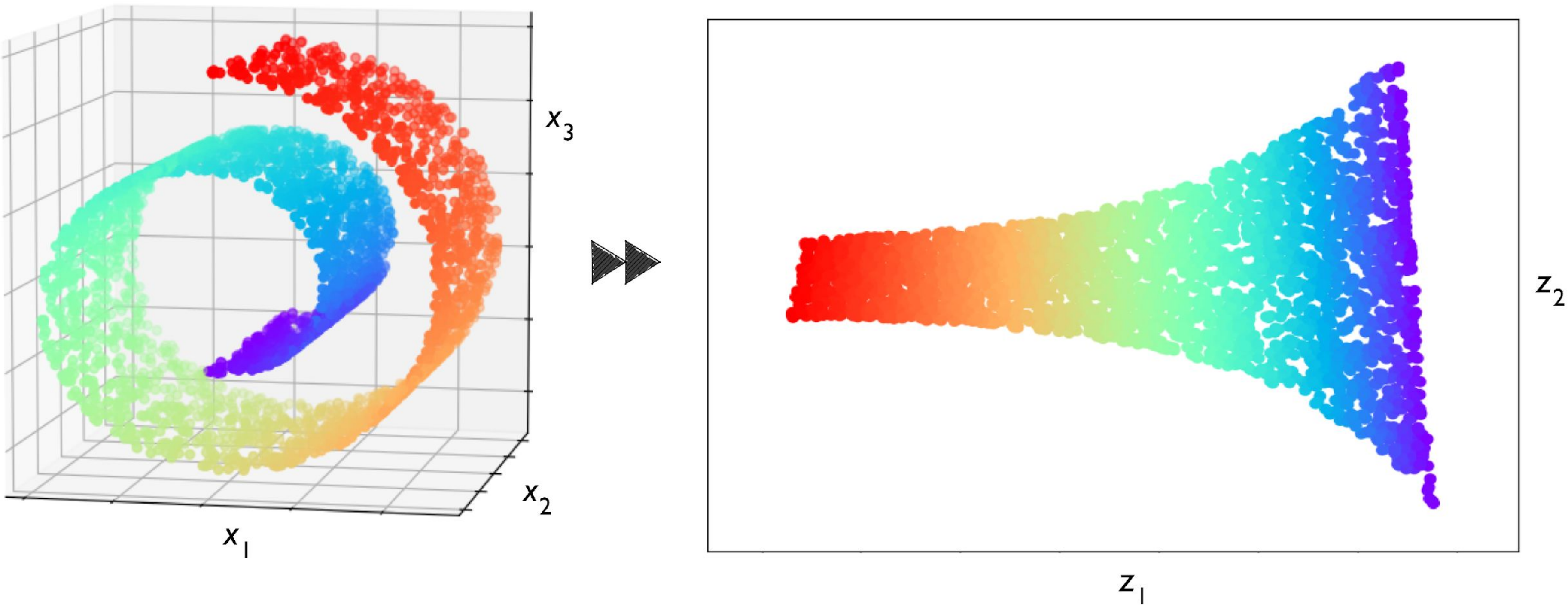
Machine Learning – ML

Types of Unsupervised Learning

TYPES OF UNSUPERVISED LEARNING

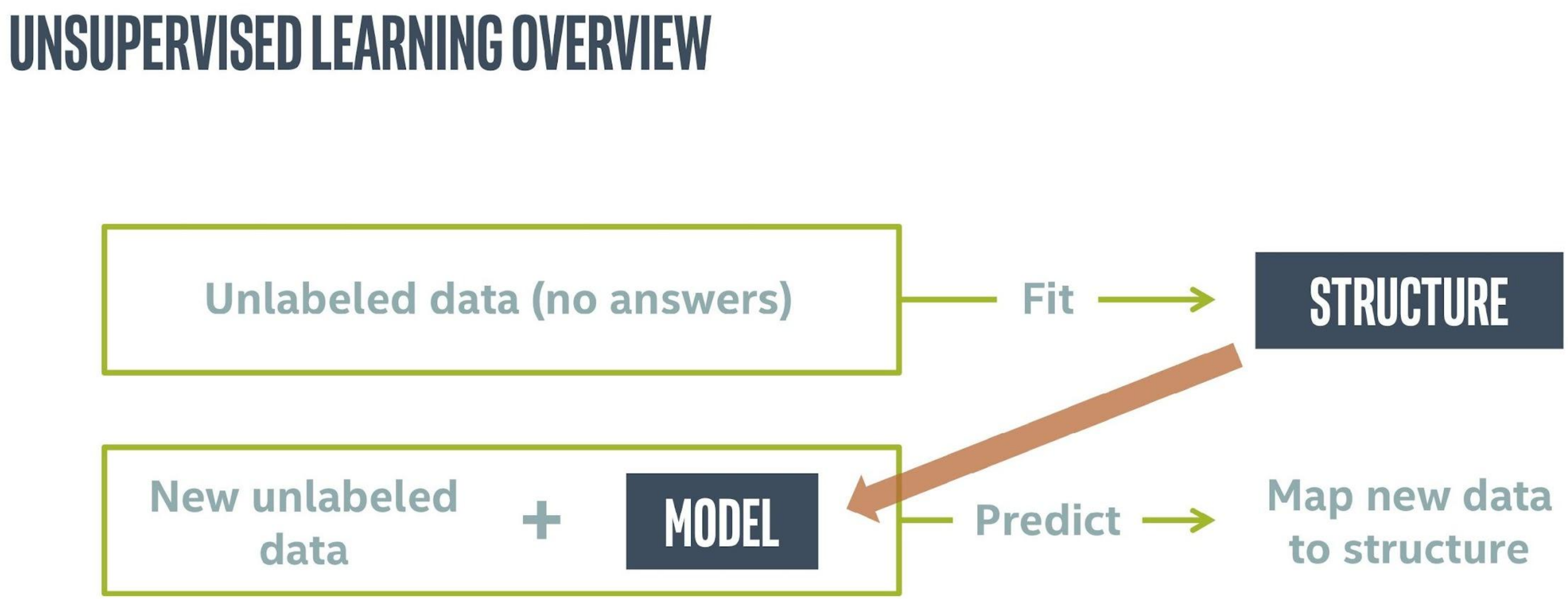
CLUSTERING	Identify unknown structure in data
DIMENSIONALITY REDUCTION	Use structural characteristics to simplify data

Machine Learning – ML
Dimensionality Reduction (graphical representation)

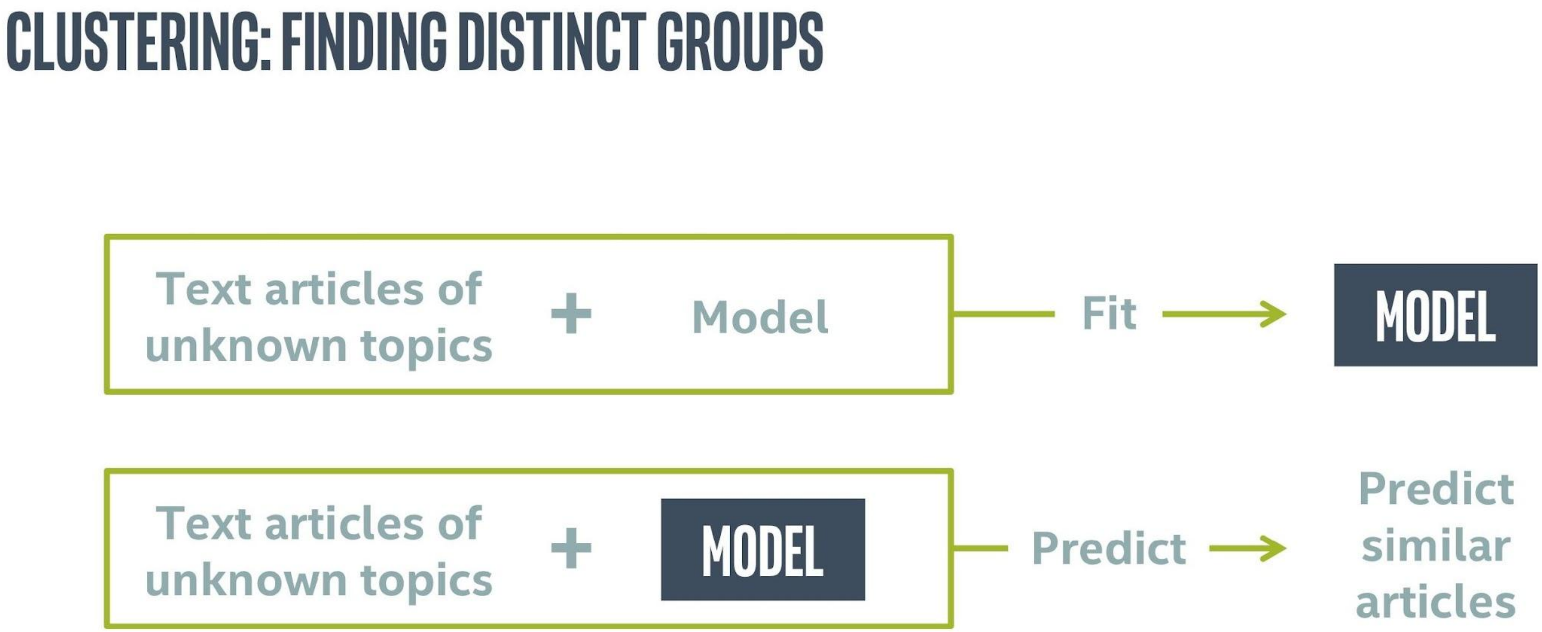


Machine Learning – ML

Unsupervised Learning Overview

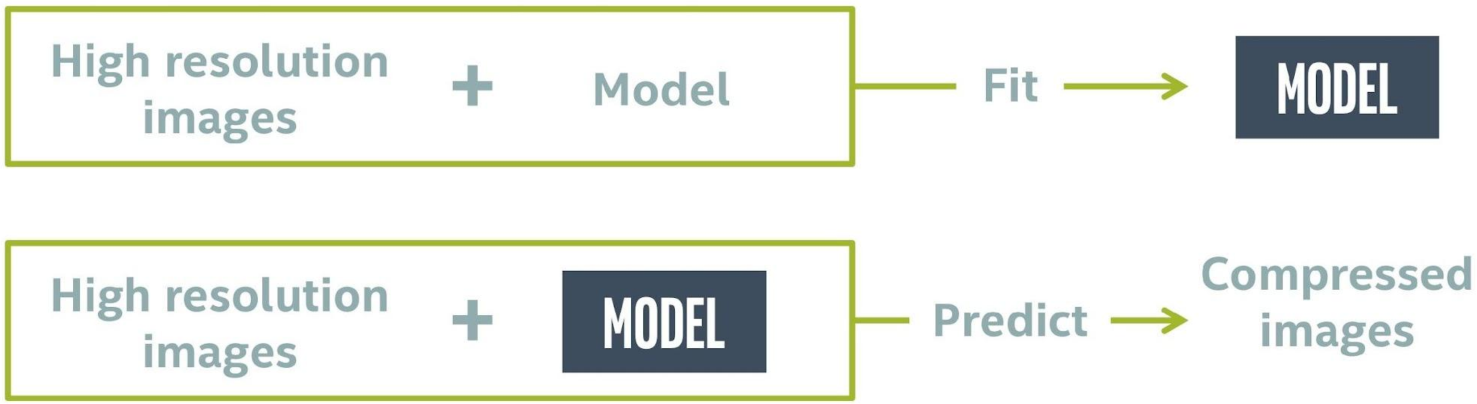


Machine Learning – ML
Clustering: Finding Distinct Groups

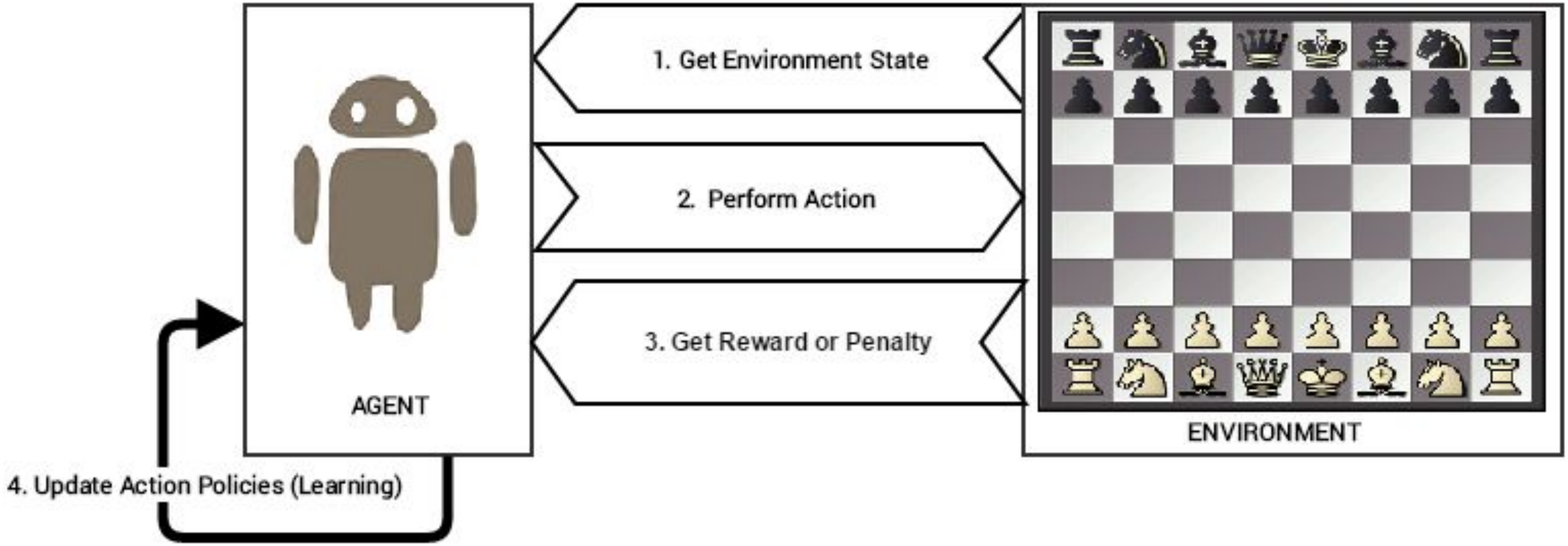


Machine Learning – ML
Dimensionality Reduction: Simplifying Structure

DIMENSIONALITY REDUCTION: SIMPLIFYING STRUCTURE

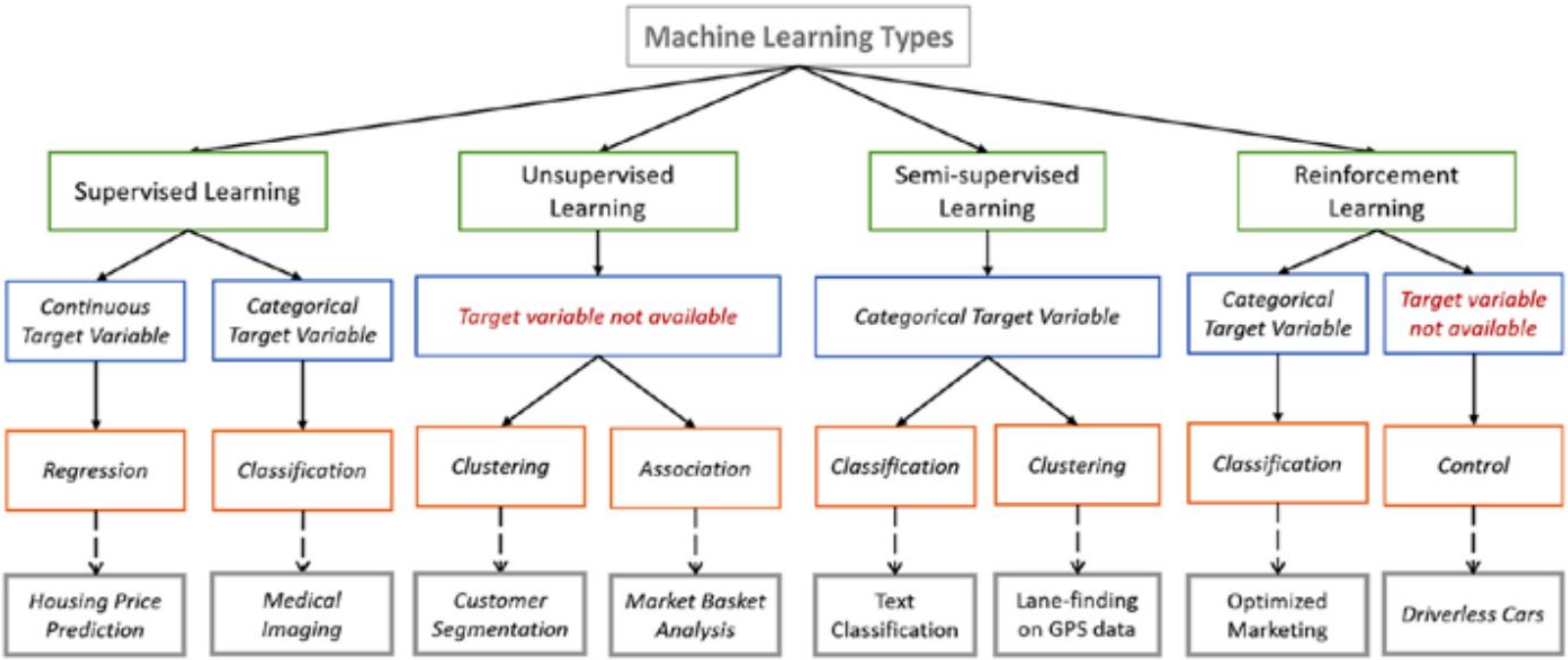


Machine Learning – ML
Reinforcement Learning



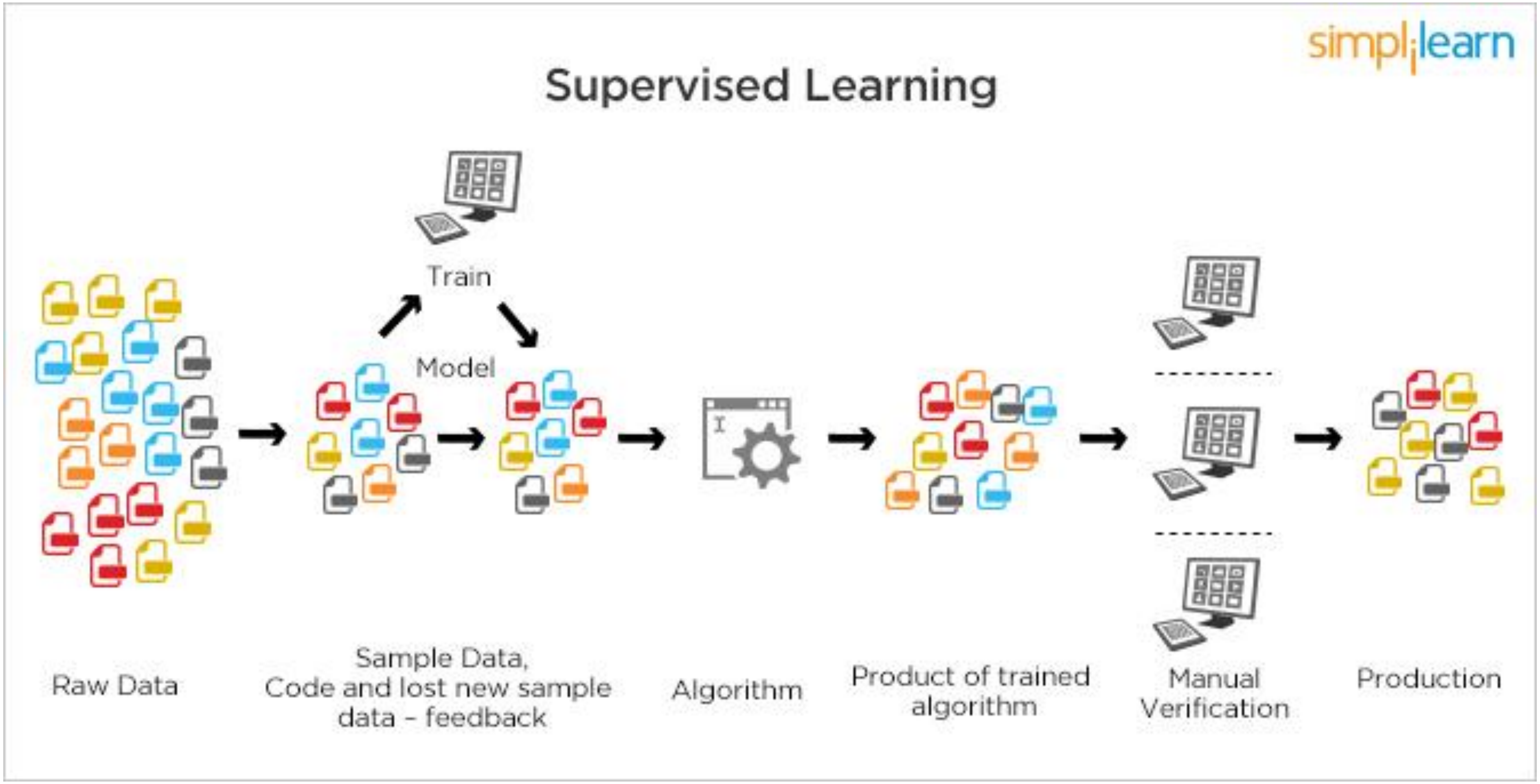
Machine Learning – ML

Types of Machine Learning

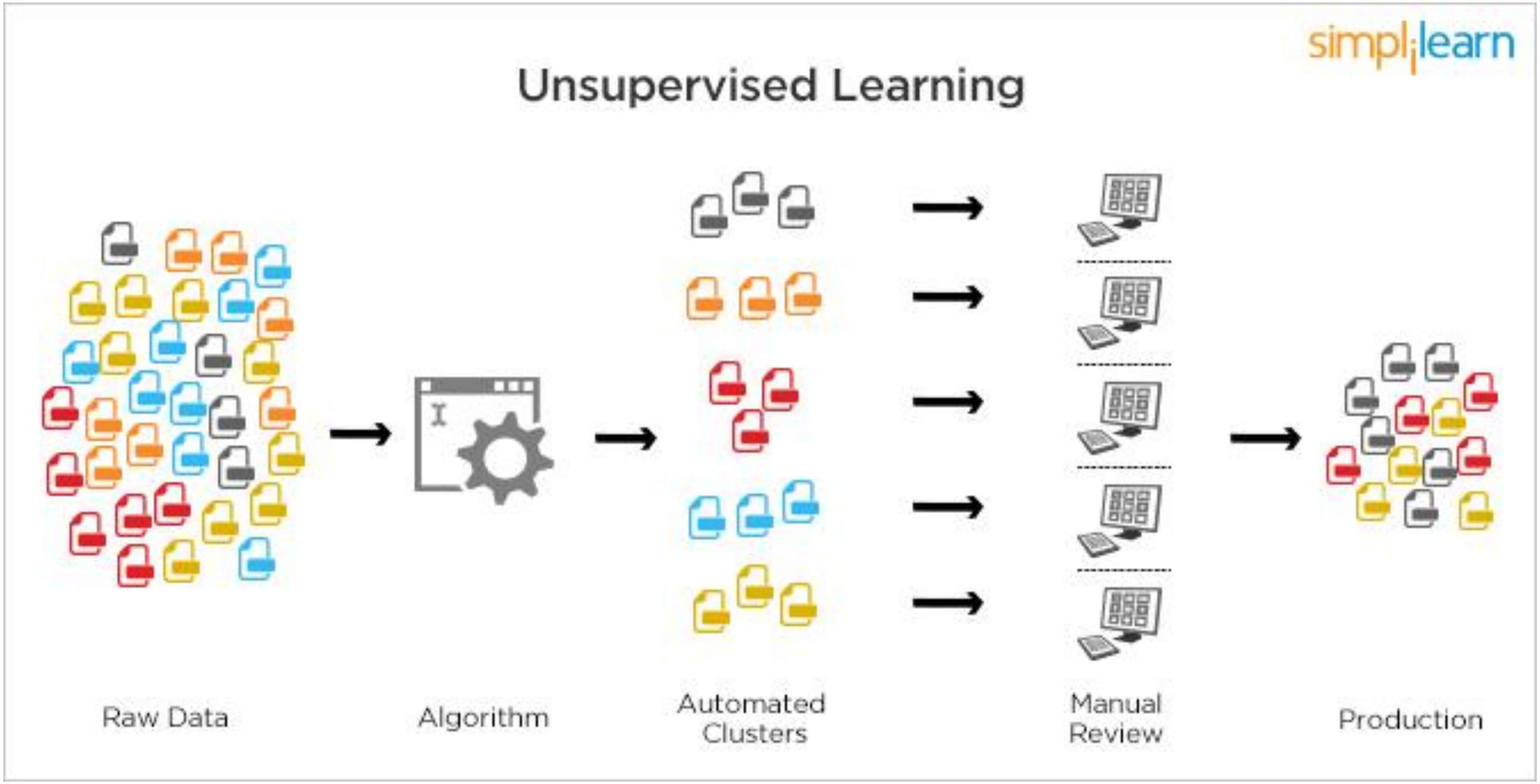


Machine Learning – ML

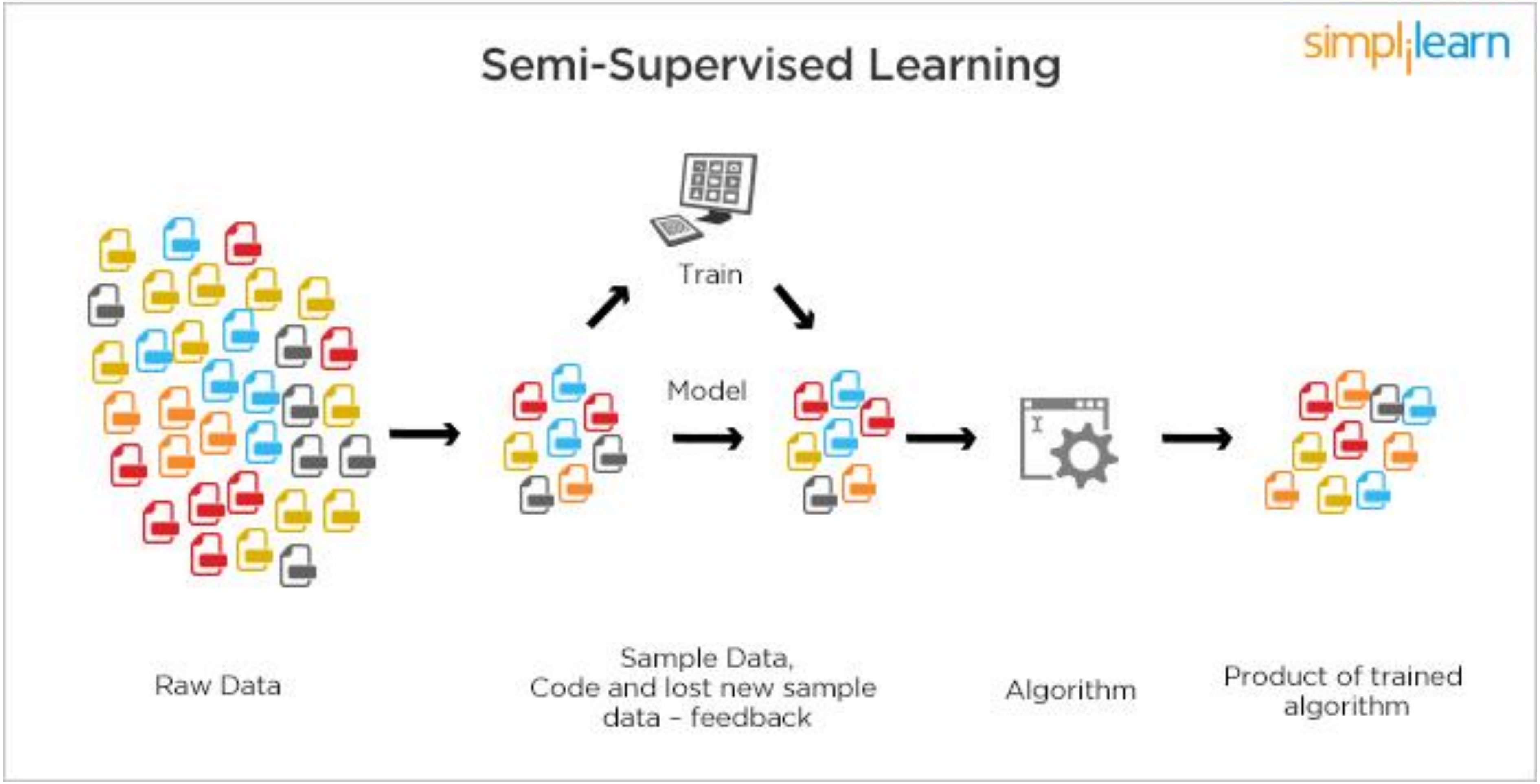
Supervised Learning



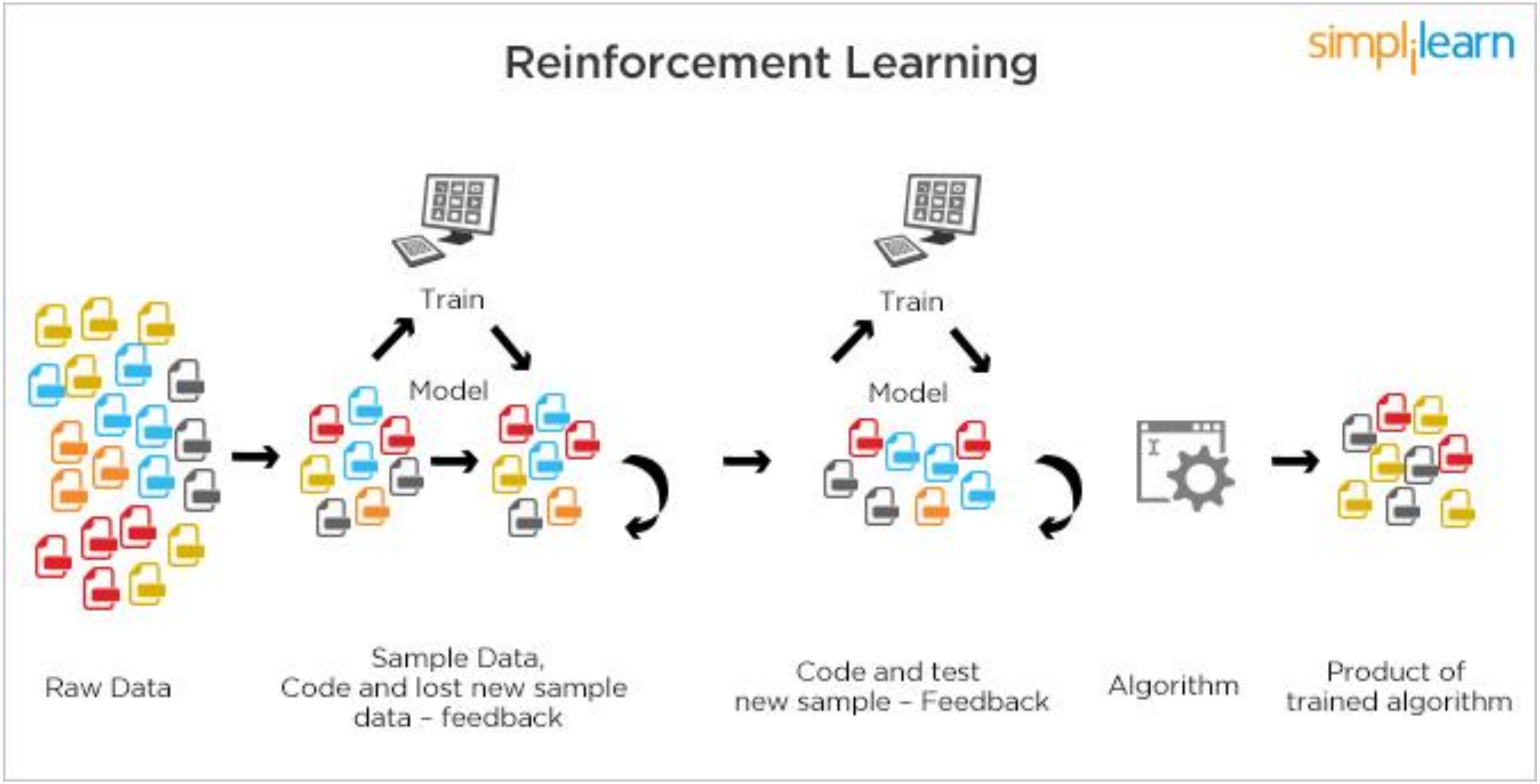
Machine Learning – ML
Unsupervised Learning



Machine Learning – ML
Semi-Supervised Learning



Machine Learning – ML
Reinforcement Learning



References:

- Battle of the Data Science Venn Diagrams
  - https://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html
- Becoming a Data Scientist – Curriculum via Metromap
  - http://nirvacana.com/thoughts/2013/07/08/becoming-a-data-scientist/
- 8 Skills You Need to Be a Data Scientist
  - https://blog.udacity.com/2014/11/data-science-job-skills.html
- Machine Learning: What it is and Why it Matters
  - https://www.simplilearn.com/what-is-machine-learning-and-why-it-matters-article


Thank You