《Data Storage and Data Mining》

⭐ CH00 Machine Learning OverView

➡️ Created by *Wang JingHui*

➡️ Version *2021v* From *2019.12.16* ...

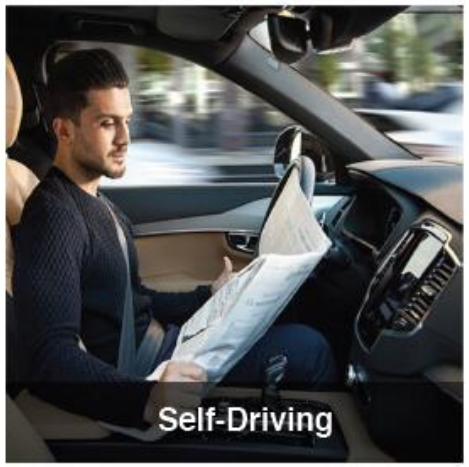
➡️ Last modification time:*2021.02.19*

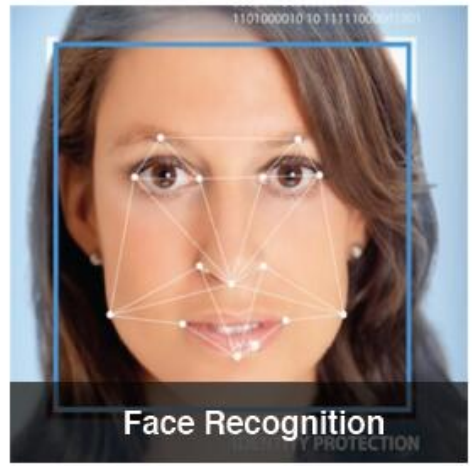
▶ Key points

1. Machine Learning – all around us
2. What's machine learning ?
3. When should you use Machine Learning ?
4. How to create machine learning models ?
5. Typical machine learning tasks

Machine Learning – all around us


Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo
AlphaGo
Lee Sedol


Self-Driving

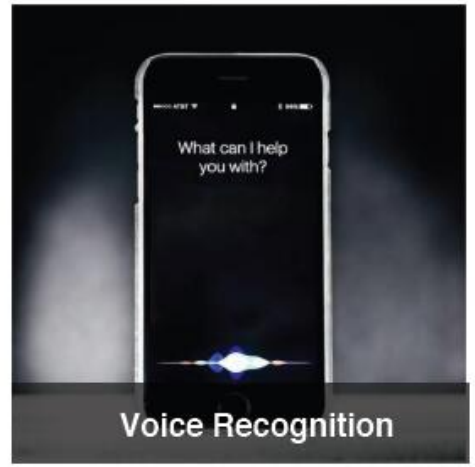

Face Recognition


Sent Mail
Spam (372)
Trash
Spam Detection


Credit card fraud detection


Predictive Maintenance


GROWTH
Sales Forecast

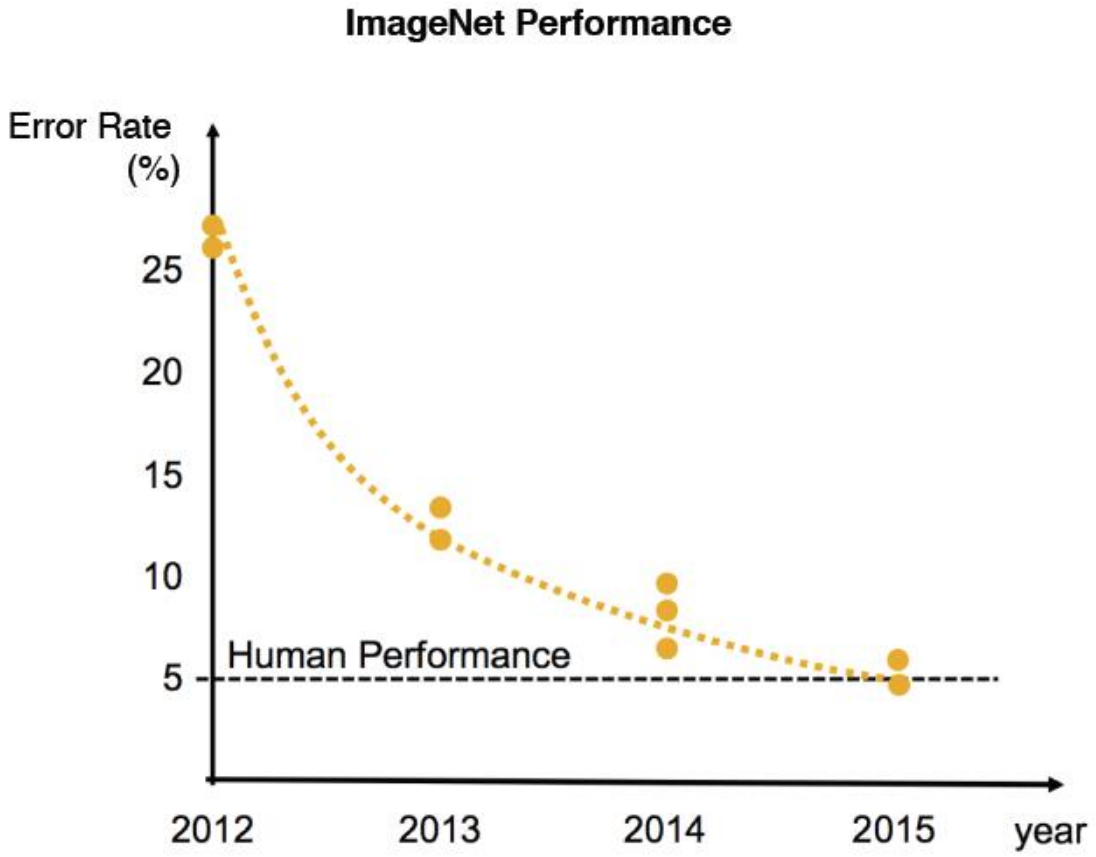

What can I help you with?
Voice Recognition


Robot

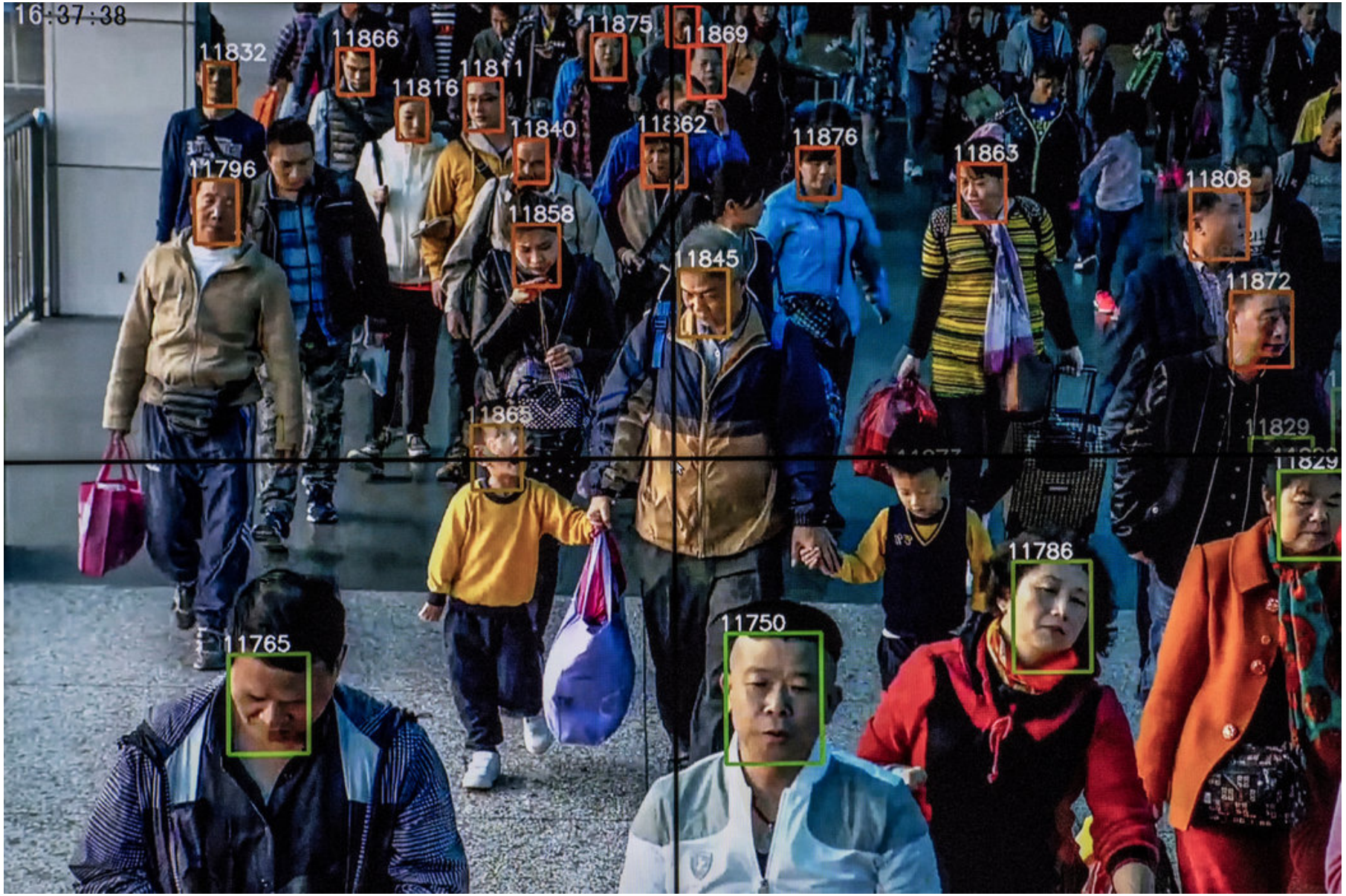
:: Computer vision is surpassing human abilities

- Chair
- Dining Table
- Person

- Dog
- Person
- Leaves



:: Computer vision is surpassing human abilities



:: Computer vision is surpassing human abilities

:: Big Pool of Machine Learning Use Cases

What’s machine learning ?

Arthur Lee Samuel (December 5, 1901 – July 29, 1990) was an American pioneer in the field of computer gaming and artificial intelligence. He popularized the term **"machine learning"** in 1959.The Samuel Checkers-playing Program was among the world's first successful self-learning programs, and as such a very early demonstration of the fundamental concept of artificial intelligence (AI).He was also a senior member in the TeX community who devoted much time giving personal attention to the needs of users and wrote an early TeX manual in 1983.

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, Tom. (1997). Machine Learning, McGraw Hill

Machine Learning is the science of getting computers to act without being explicitly programmed.
-- Andrew Ng

Machine learning is a technique of data science that helps computers learn from existing data in order to forecast future behaviors, outcomes, and trends.
-- Microsoft

:: Machines Learning uses historical data to make predictions

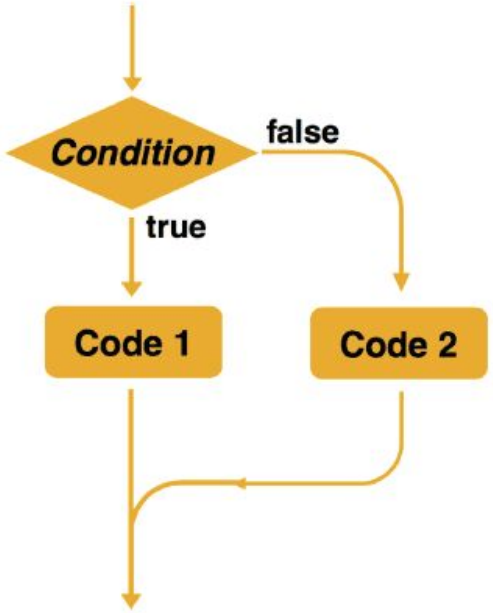
Wrap up

Machines Learning uses historical data to make predictions
Machine Learning is also similar to Data Mining, but whereas data mining is the science of discovering unknown patterns and relationships in data, ML applies previously inferred knowledge to new data to make decisions in real-life applications.

- Computers approximate complex functions from historical data
- Rules are not explicitly programmed but learned from data

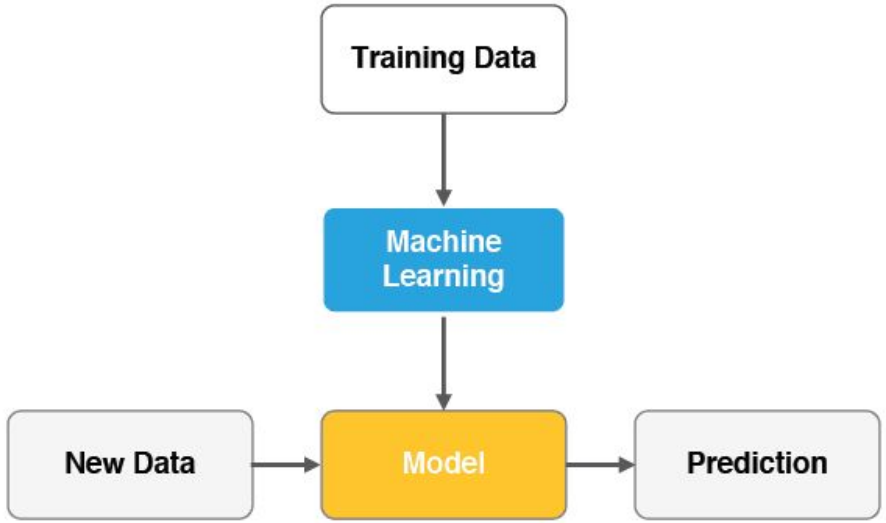
:: The difference between traditional approach and Machine Learning

Rule-based approach



- Explicitly programmed to solve problem
- Decision rules are clearly defined by humans

Machine learning



- Trained from examples
- Decision rules complex or fuzzy
- Rules are not defined by humans but learned by the machine from data

:: AI History

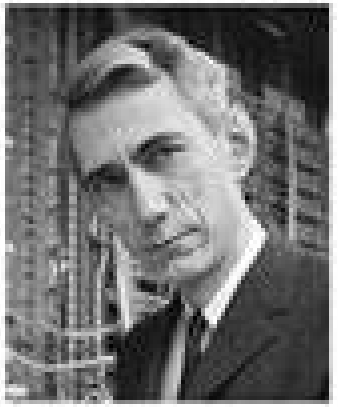
Summer Research Project on Artificial Intelligence。

- A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence J. McKarthy, M.L. Minsky, N. Rochester, C.E. Shannon | 1955 August |
- https://chsasank.github.io/classic_papers/dartmouth-artifical-intelligence-summer-research-proposal.html

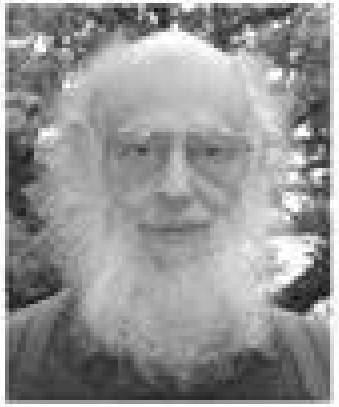
1956 Dartmouth Conference: The Founding Fathers of AI

John MacCarthy

Marvin Minsky



Claude Shannon

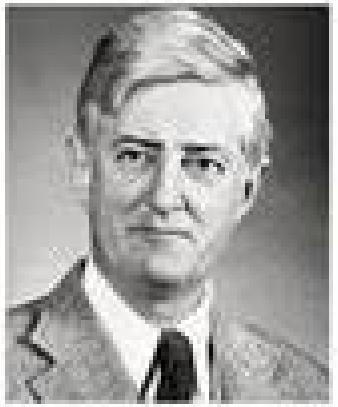


Ray Solomonoff

Alan Newell

Herbert Simon

Arthur Samuel



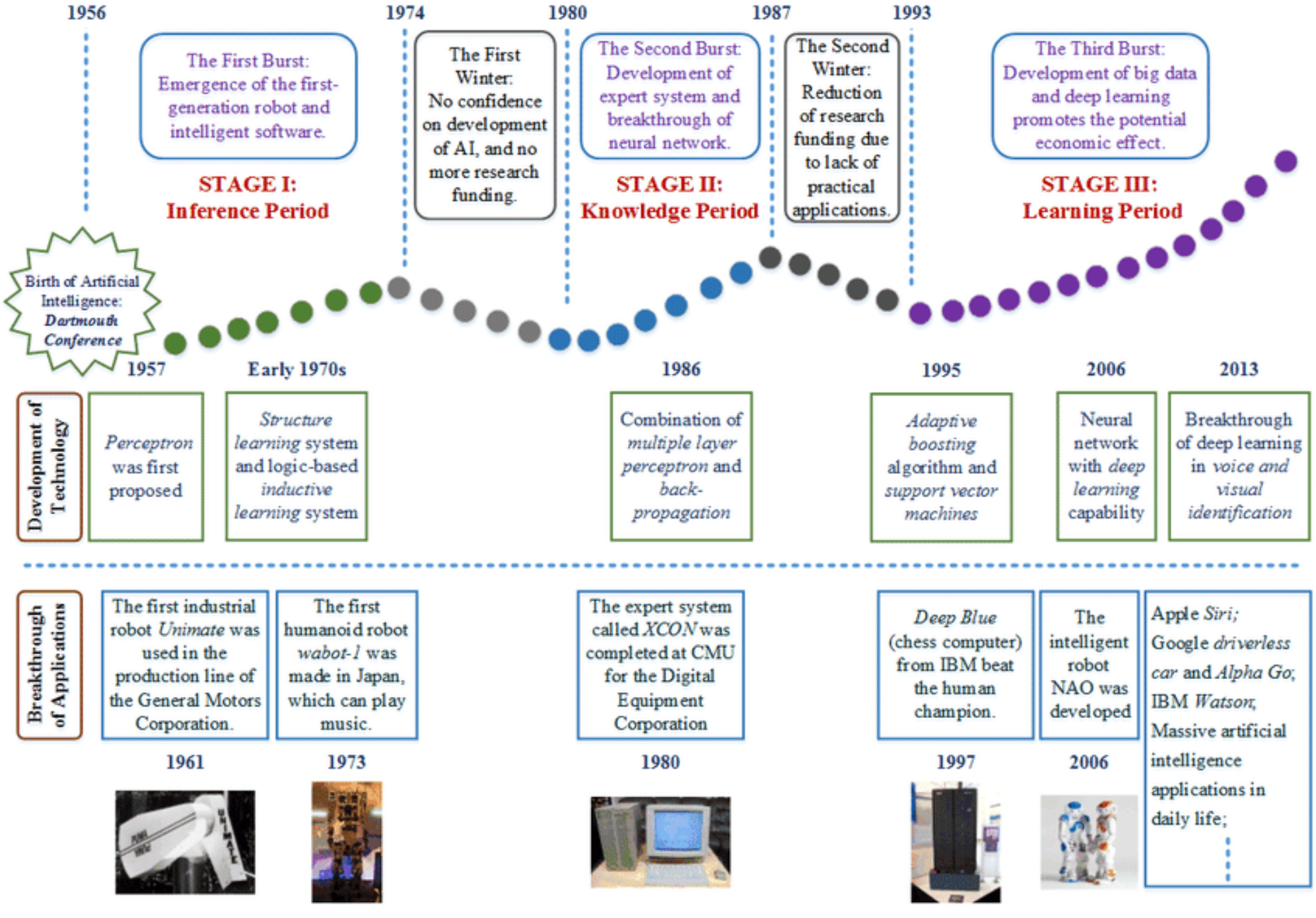
Oliver Selfridge

Nathaniel Rochester

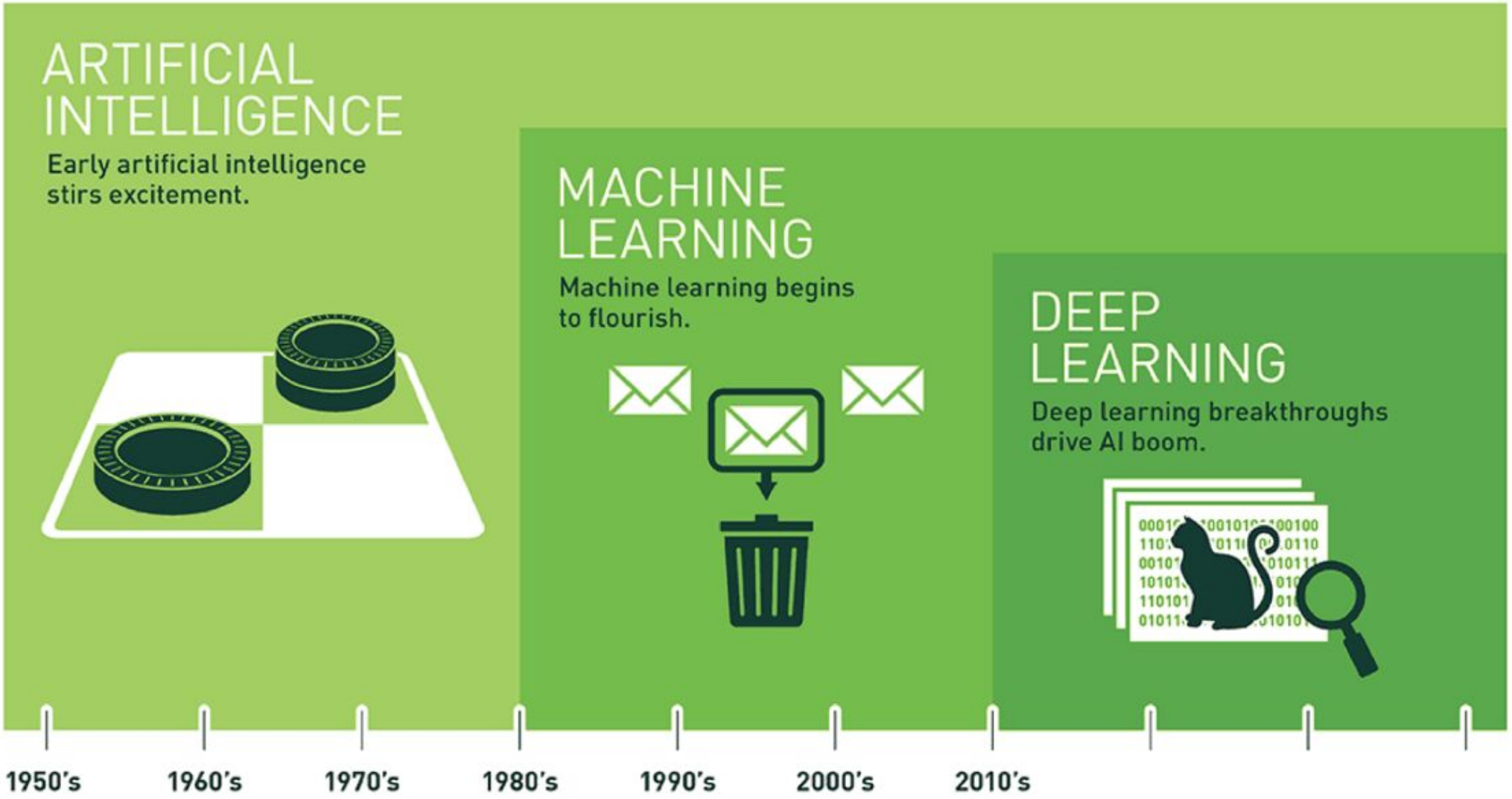
Trenchard More

:: AI History

2006年，会议50年后，当事人重聚达特茅斯（左起：摩尔、麦卡锡、明斯基、塞弗里奇、所罗门诺夫）

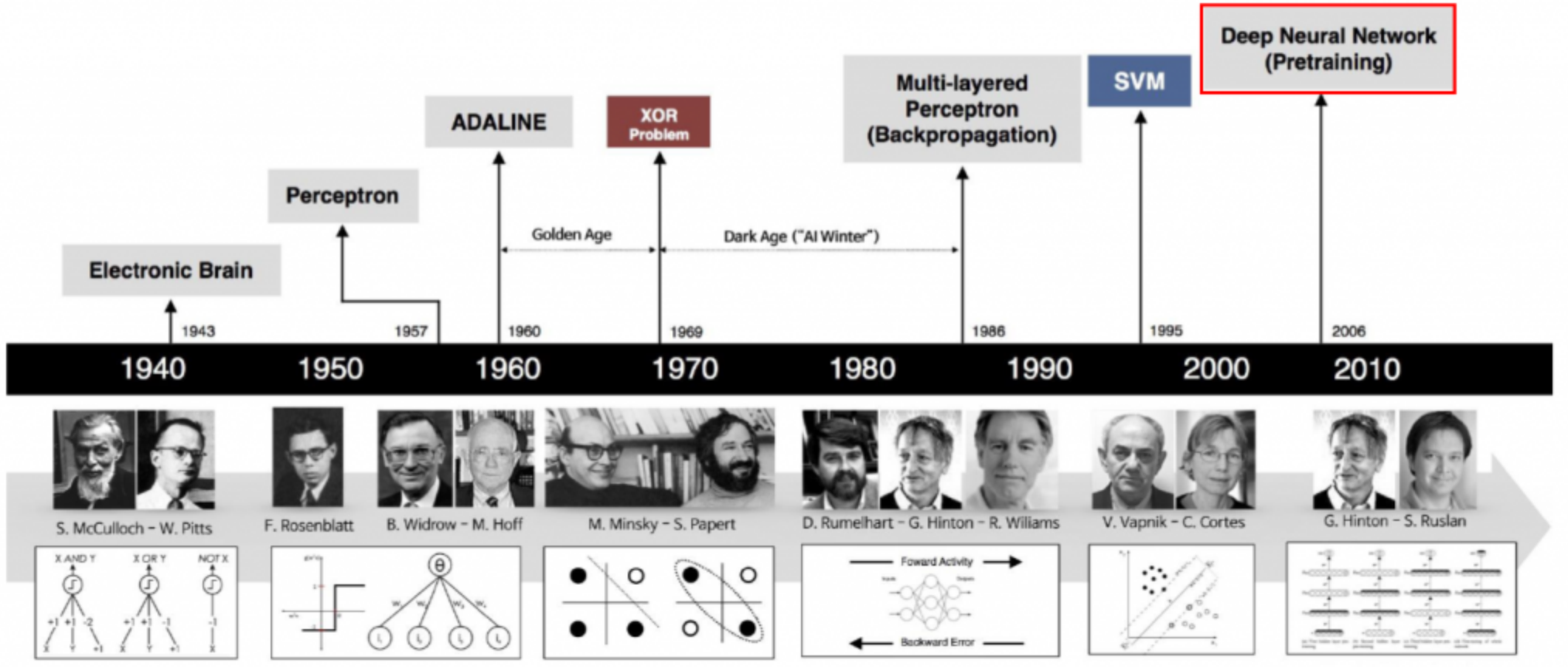

1956
1974
1980
1987
1993
The First Burst: Emergence of the first-generation robot and intelligent software.
The First Winter: No confidence on development of AI, and no more research funding.
The Second Burst: Development of expert system and breakthrough of neural network.
The Second Winter: Reduction of research funding due to lack of practical applications.
The Third Burst: Development of big data and deep learning promotes the potential economic effect.
STAGE I: Inference Period
STAGE II: Knowledge Period
STAGE III: Learning Period
Birth of Artificial Intelligence: Dartmouth Conference
1957
Early 1970s
1986
1995
2006
2013
Development of Technology
Perceptron was first proposed
Structure learning system and logic-based inductive learning system
Combination of multiple layer perceptron and back-propagation
Adaptive boosting algorithm and support vector machines
Neural network with deep learning capability
Breakthrough of deep learning in voice and visual identification
Breakthrough of Applications
The first industrial robot Unimate was used in the production line of the General Motors Corporation.
The first humanoid robot wabot-1 was made in Japan, which can play music.
The expert system called XCON was completed at CMU for the Digital Equipment Corporation
Deep Blue (chess computer) from IBM beat the human champion.
The intelligent robot NAO was developed
Apple Siri; Google driverless car and Alpha Go; IBM Watson; Massive artificial intelligence applications in daily life;
1961
1973
1980
1997
2006

:: Artificial Intelligence has been evolving

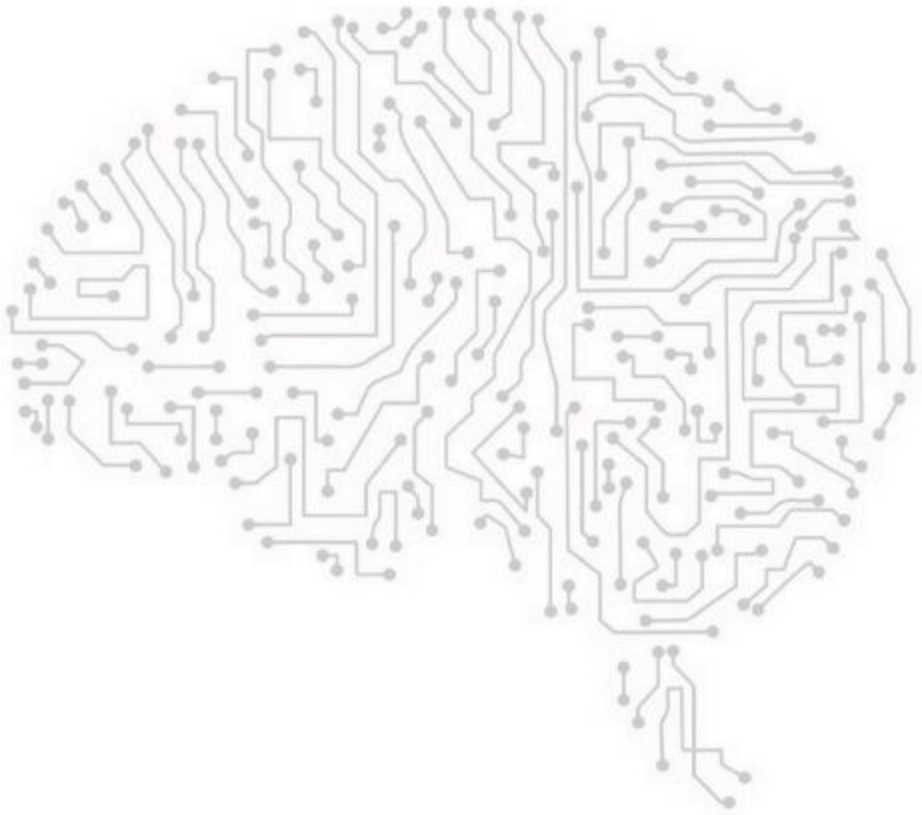


Since an early flush optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions

:: Artificial Intelligence has been evolving



:: Machines can now do things that were not possible before



Why now ?

- **Big Data**
- **Massive Computer power**

  (e.g. GPU, multicore)
- **Better algorithms**

  (e.g. Deep Learning algorithms)

When should you use Machine Learning ?

1. Do you need machine learning?

2. Can you formulate your problem clearly?

3. Do you have sufficient examples?

4. Does your problem have a regular pattern?

5. Can you find meaningful representations of your data?

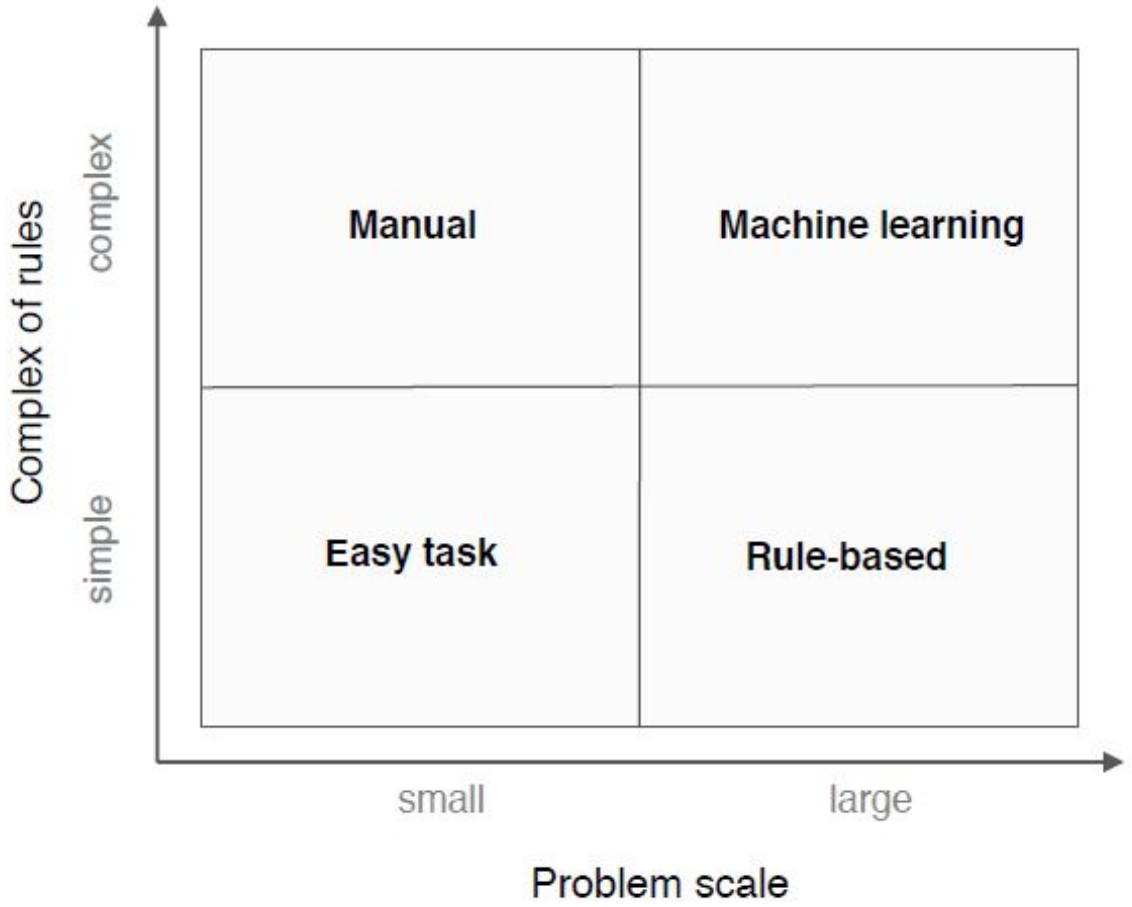
6. How do you define success?

1️⃣ Do you need machine learning?

- Do you need to automate the task?
- High volume tasks with complex rules and unstructured data are good candidates

Example: sentiment analysis

- High volume of reviews on the Web
- Unstructured text
- Human language is complex and ambiguous



2️⃣ Can you formulate your problem clearly?

- What do you want to predict given which input?
- Pattern: “given X, predict Y”
    - What is the input?
    - What is the output?

Example: sentiment analysis

- Given a customer review, predict its sentiment
    - Input: customer review text
    - Output: positive, negative, neutral

3️⃣ Do you have sufficient examples?

- Machine learning always requires data!
- Generally, the more data, the better
- Each example must contain two parts (supervised learning)
    - Features: attributes of the example
    - Label: the answer you want to predict

Example: sentiment analysis

- Thousands of customer reviews and ratings from the Web

4️⃣ Does your problem have a regular pattern?

- Machine learning learns regularities and patterns
- Hard to learn patterns that are rare or irregular

Example: sentiment analysis

- Positive words like good, awesome, or love it appear more often in highly-rated reviews
- Negative words like bad, lousy, or disappointed appear more often in poorly-rated reviews

5 Can you find meaningful representations of your data?

- Machine learning algorithms ultimately operate on numbers
- Generally, examples are represented as feature vectors
- Good features often determine the success of machine learning

Example: sentiment analysis

- Represent customer review as vector of word frequencies
- Label is positive (4-5 stars), negative (1-2 stars), neutral (3 stars)
- Picture of data or a mathematical vector

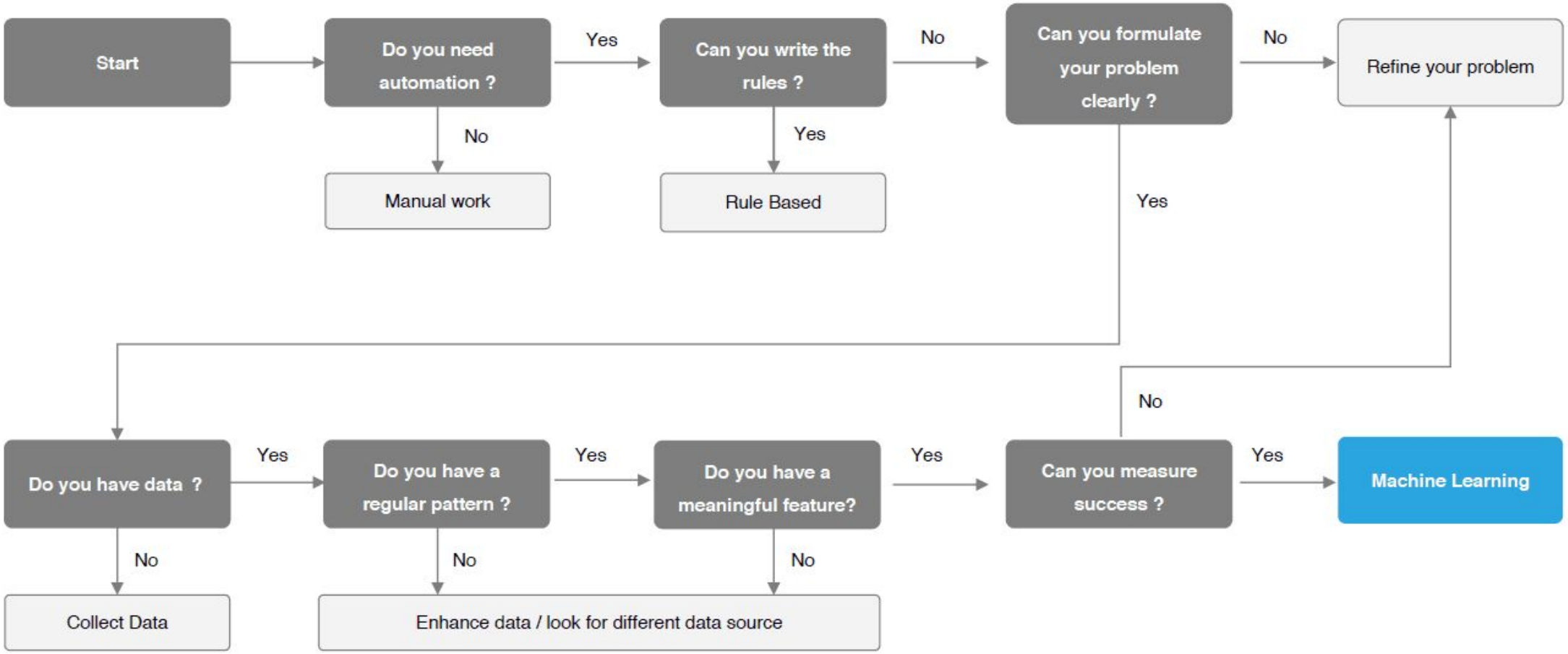
6️⃣ How do you define success?

- Machine learning optimizes a training criteria
- The evaluation function has to support the business goals

Example: sentiment analysis

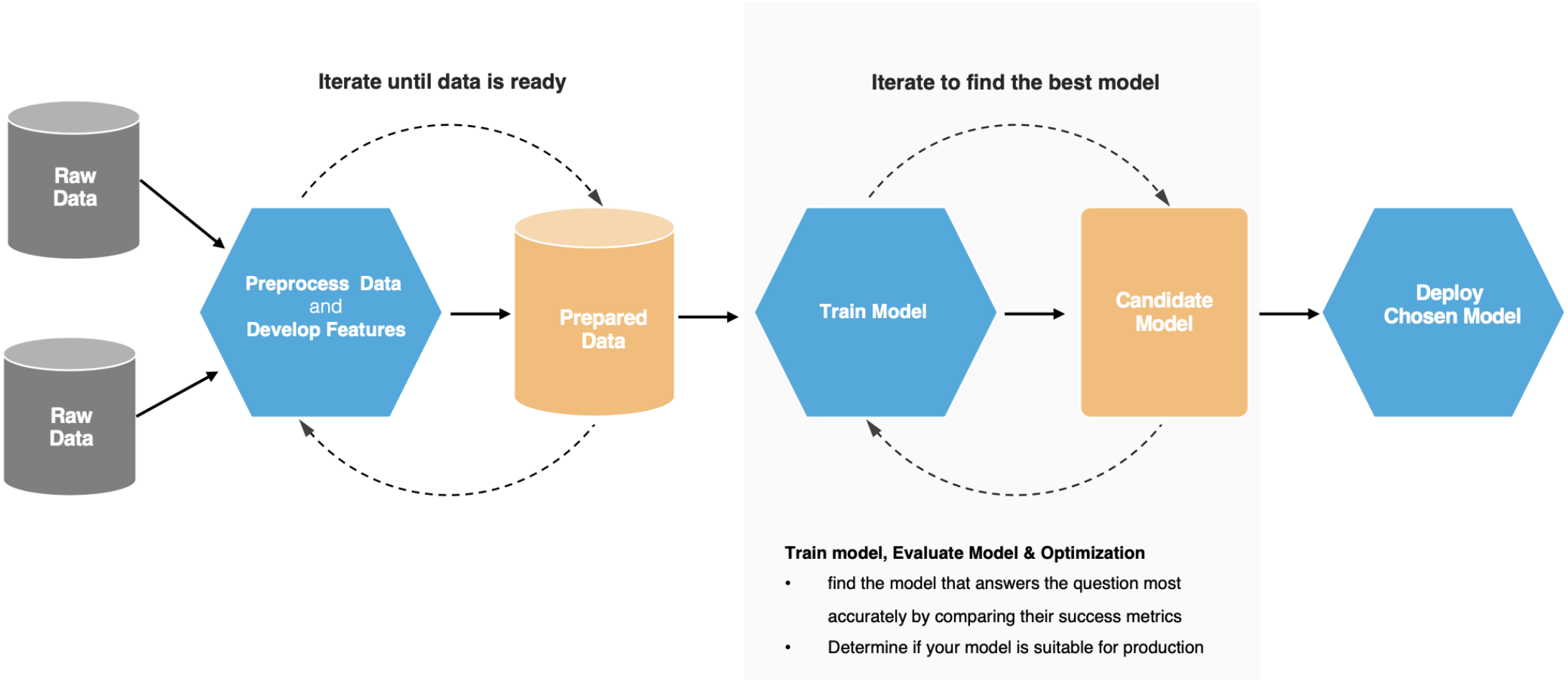
- Accuracy: percentage of correctly predicted labels

Machine learning problem: a Recipe

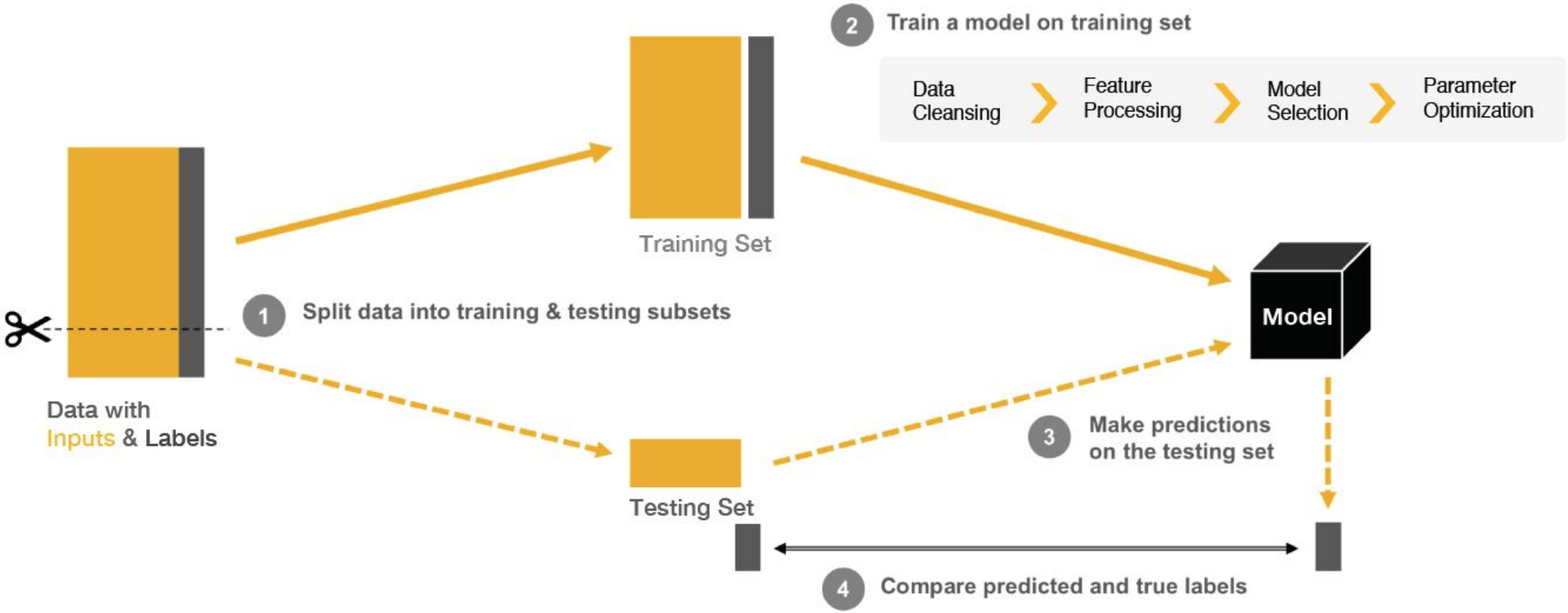


How to create machine learning models ?

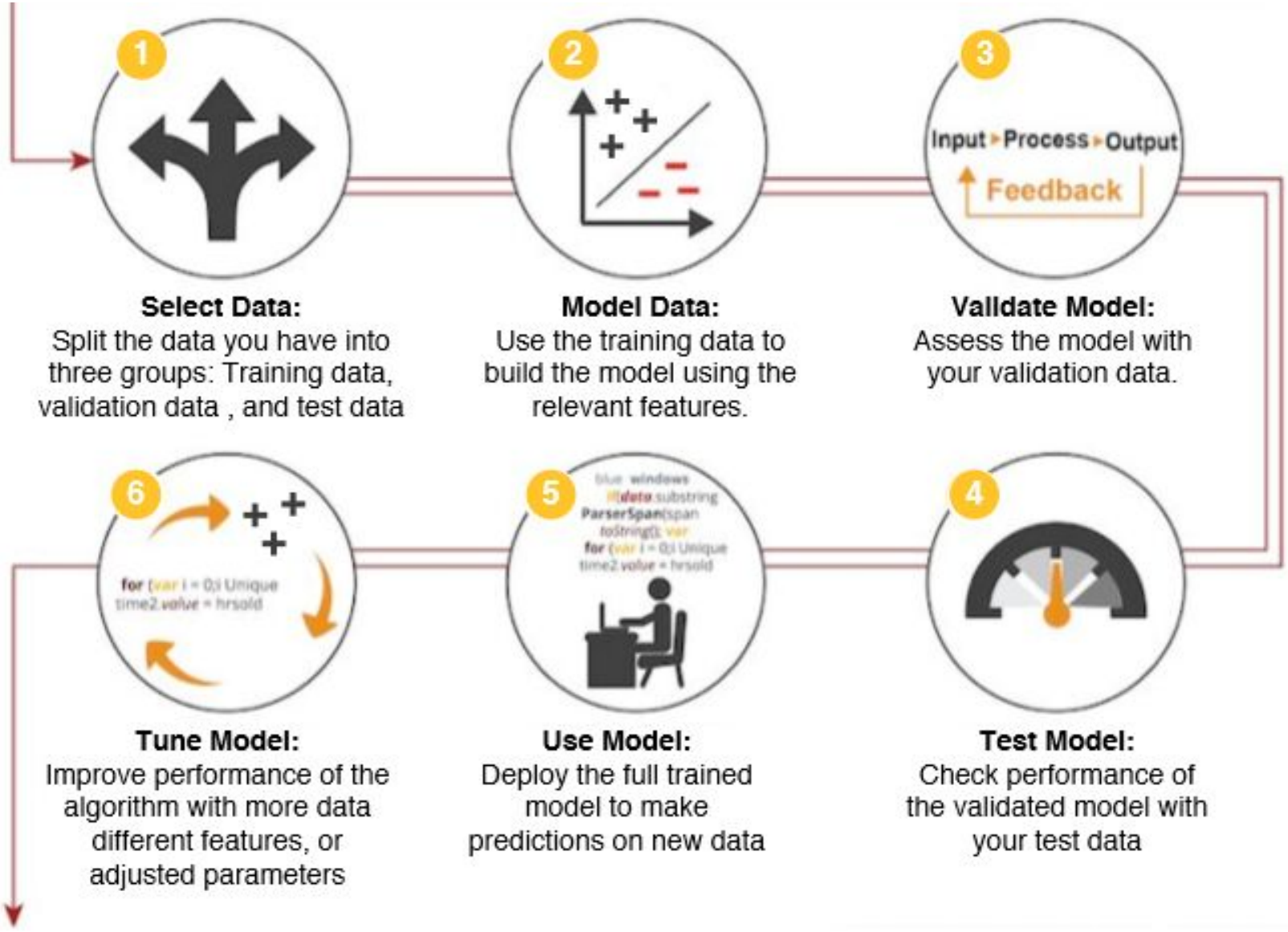
:: The process of Machine Learning



:: How to create machine learning models ?



:: How machine learning works



Select Data:
Split the data you have into three groups: Training data, validation data , and test data

Model Data:
Use the training data to build the model using the relevant features.

Validate Model:
Assess the model with your validation data.

Tune Model:
Improve performance of the algorithm with more data different features, or adjusted parameters

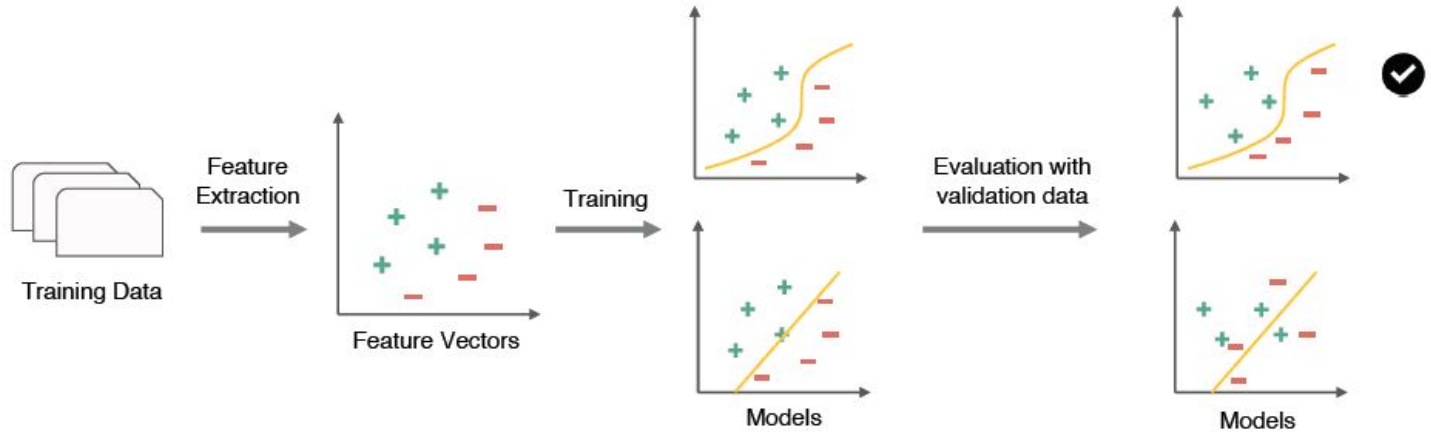
Use Model:
Deploy the full trained model to make predictions on new data

Test Model:
Check performance of the validated model with your test data

Typical machine learning tasks

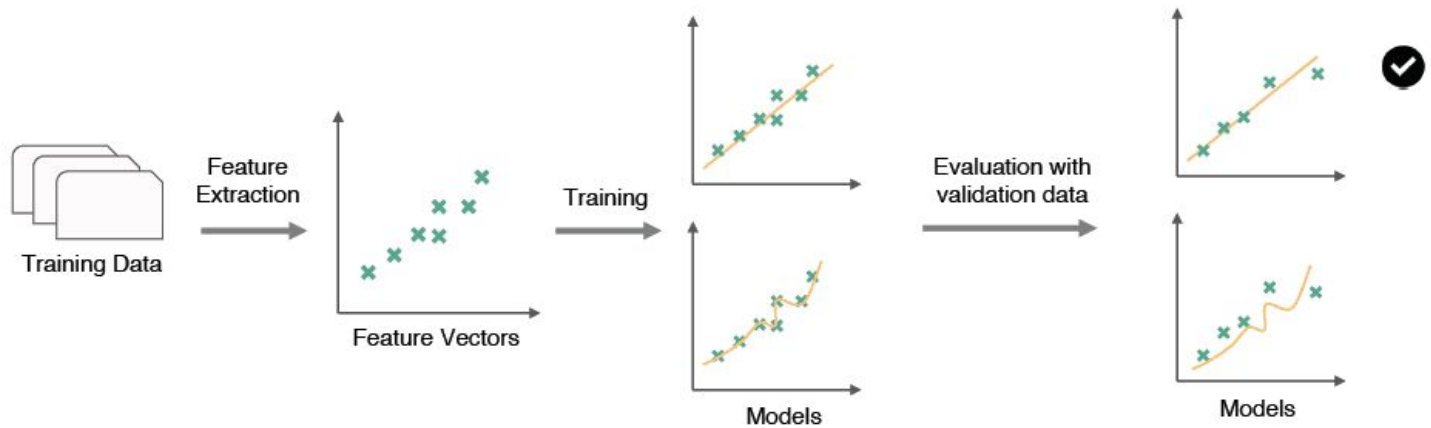
Classification (Predict the category)

Identifying to which category an object belongs to



Regression (Predict the value)

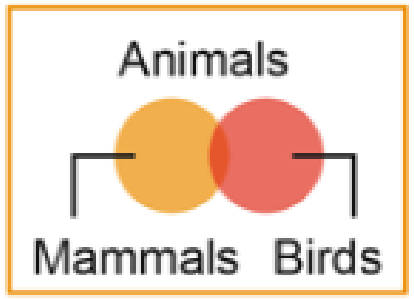
Predicting a continuous-valued attribute associated with an object




Is this tweet positive?
Which service will this customer choose?
Which of two coupons draws more customers?
A/B
Predict Categories
Classification
Identify what category new information belongs in.
Regression
Forecast the future by estimating the relationship between variables.
Predict Values
Estimate product demand
Predict sales figures
Analyze marketing returns
Predict credit risk
Detect fraud
Catch abnormal equipment readings
Find Unusual Occurrences
Anomaly Detection
Identify and predict rare or unusual data points.
Clustering
Separate similar data points into intuitive groups.
Discover Structure
Perform customer segmentation
Predict customer tastes
Determine market price

What are the five tribes?

Symbolists



Use symbols, rules, and logic to represent knowledge and draw logical inference

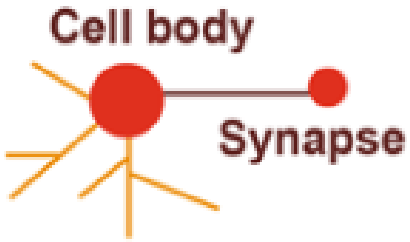
Favored algorithm
Rules and decision trees

Bayesians

Assess the likelihood of occurrence for probabilistic inference

Favored algorithm
Naive Bayes or Markov

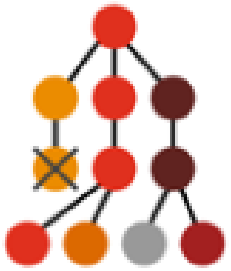
Connectionists



Recognize and generalize patterns dynamically with matrices of probabilistic, weighted neurons

Favored algorithm
Neural networks

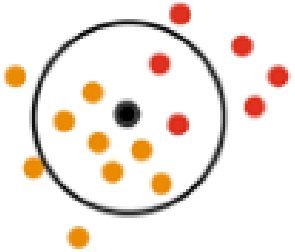
Evolutionaries



Generate variations and then assess the fitness of each for a given purpose

Favored algorithm
Genetic programs

Analogizers



Optimize a function in light of constraints ("going as high as you can while staying on the road")

Favored algorithm
Support vectors

Source: Pedro Domingos, *The Master Algorithm*, 2015

Enjoy your machine learning!

https://github.com/wjssx/

E-mail: csr_dsp@sina.com

Copyright © 2021 Yjssx

This software released under the BSD License.