IMT Mines Alès
École Mines-Télécom

HAH811E9 - SCIENCE DES DONNÉES NIVEAU 1

Gérard Dray
Nicolas Sutton-Charani

afaq ISO 9001 Qualité AFNOR CERTIFICATION

EuroMov Digital Health in Motion

Contexte

A la croisée de plusieurs champs disciplinaires : **mathématiques, probabilités, statistiques, informatique, intelligence artificielle, théorie de l'information et visualisation**, la science des données met en œuvre différents outils d'analyse de données afin d'extraire automatiquement des informations utiles, des connaissances, à partir de données potentiellement massives. Le but ultime est de rendre cette information plus facile à exploiter, la protéger et la valoriser. Elle pourra servir de base ensuite à des processus d'évaluation et d'aide à la décision.

Ce cours est organisé en 2 UEs :
Master 1 : Bases mathématiques et informatiques de la science des données.
Master 2 : Notions avancées d'intelligence artificielle et d'apprentissage artificiel.

Objectifs

Renforcer les compétences théoriques avec un approfondissement des statistiques et de la théorie des probabilités.

Introduire les méthodes et techniques d'apprentissage artificiel.

Former à l'utilisation de langages informatiques (Python et R) pour réaliser des projets permettant de mettre en pratique les aspects théoriques de la science des données.

Mots clés : probabilités, statistiques, base de données, apprentissage artificiel, R, python.

Prérequis : notions de base en probabilités, statistiques et programmation.

Moyens

Equipe pédagogique :

- Nicolas Sutton-Charani (NSC)
  Enseignant Chercheur - EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales

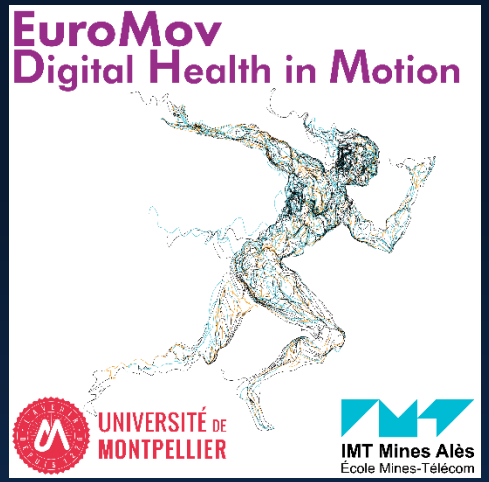
- Gérard Dray (GD)
  Enseignant Chercheur - EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales

Date	Heure début	Heure fin	Intervenant	Salle	Moyens pédagogiques
jeudi 22 janvier 2026	13:15:00	18:15:00	Gérard Dray	TD 36.414	Cours/TD
jeudi 29 janvier 2026	13:15:00	18:15:00	Nicolas Sutton-Charani	TD 36.416	Cours/TD
~~jeudi 5 février 2026~~	~~13:15:00~~	~~18:15:00~~	~~Nicolas Sutton-Charani~~	~~TD 36.412~~	~~Cours/TD~~
jeudi 12 février 2026	13:15:00	18:15:00	Nicolas Sutton-Charani	TD 36.315	Cours/TD
jeudi 26 mars 2026	13:15:00	18:15:00	Equipe pédagogique	TD 36.410	Projet
jeudi 2 avril 2026	13:15:00	18:15:00	Equipe pédagogique	TD 36.417	Projet
jeudi 9 avril 2026	13:15:00	18:15:00	Equipe pédagogique	TD 36.410	Projet
jeudi 16 avril 2026	13:15:00	18:15:00	Equipe pédagogique	TD 36.410	Projet

Gérard Dray
gerard.dray@mines-ales.fr

Chercheur

UMR EuroMov Digital Health in Motion,
Univ Montpellier,
IMT Mines Ales, Ales,
France



Enseignant

IMT Mines Alès
Département d'enseignement
Informatique et Intelligence Artificielle
2IA

Université de Montpellier
IAE
Master 1 E_Mkg
Master 1 MTD
Master 2 E-Mkg
Master 2 MPW

Faculté des Sciences, Faculté de Médecine
Master Sciences et Numérique pour la Santé
Master Santé

UFR STAPS
Masters STAPS

Quick start poll

Scale: 0 = none, 1 = basic awareness, 2 = beginner, 3 = intermediate, 4 = advanced

1. Programming (Python or similar)
2. Math & stats basics for Machine Learning
3. Machine Learning fundamentals (train/test split, overfitting, metrics)
4. Supervised learning (classification & regression)
5. Unsupervised learning (clustering)

What I am / What I am not

What I am:

- A support:
  - for ML concepts, methods, and good practices
  - for problem framing (what to predict, what data to use, how to evaluate)
  - for interpretation and critical thinking (limits, bias, generalization)
  - for clear reporting (results, metrics, conclusions)

What I am not:

- A code debugger or IT support (I won't fix your environment/setup issues)
- A "do-it-for-you" solution: you must run experiments and make choices
- A guarantee of performance: ML results depend on data and validation
- A replacement for scientific reasoning.

Key message:
I help you learn to build and trust ML models,
you remain responsible for implementation and validation.

Course positioning

Data Science for Health Applications

- Turning clinical, physiological, and behavioral data into reliable decision support
- From measurements to models: prediction, classification, and risk stratification
- Focus on methods you can trust: evaluation, interpretability, and limitations

Typical health applications

- Patient monitoring (wearables, vitals, time-series)
- Diagnosis support & screening
- Prognosis and outcome prediction
- Rehabilitation and personalized follow-up
- Medical imaging

Key message:

Build models that are not only accurate, but clinically meaningful and validated.

What is Data Science?

- Data Science is a discipline that :
  - Turns raw data into actionable knowledge
  - Combines statistics, artificial intelligence, programming, and domain expertise
  - Covers the full pipeline: collect, clean, explore, model, communicate

- Why it matters in Health
  - Data are often noisy, heterogeneous, incomplete, and context-dependent
  - Good decisions require reliable data + rigorous evaluation, not only algorithms
  - Most effort is upstream: data quality drives model quality

Key message:

Key message: Better data beats better algorithms

What is Artificial Intelligence?

- Artificial Intelligence (AI): a broad umbrella

- Systems that perform tasks typically requiring human intelligence:
  - pattern recognition
  - decision-making
  - perception (vision, signals)
  - language understanding

- AI ranges from rule-based systems to learning-based approaches

Key message:
In this course, we focus on the AI family that learns from examples: Machine Learning.

What is Machine Learning?

- Machine Learning (ML): learning patterns from examples
- Models learn a mapping from inputs to outputs using data
- Typical goals:
  - prediction (e.g., injury risk, fatigue state)
  - classification (e.g., movement quality categories)
  - pattern discovery (e.g., movement strategies)
- ML is not magic: performance depends on data quality, design choices, and evaluation

Key message:
ML is a methodology to build models that generalize beyond observed examples.

Typical Health use cases

Examples in sport, health, and ergonomics:

- Performance & training: workload monitoring, technique assessment
- Injury prevention: screening, risk indicators, load–response modeling
- Rehabilitation: progress tracking, return-to-activity decision support
- Ergonomics: task recognition, exposure estimation, risk detection
- Clinical & health: activity recognition, functional capacity estimation

Key message:
The same ML tools apply across domains
What changes is the question and constraints.

The ML workflow (big picture)

From measurements to a validated model:

- Define the problem & the target outcome
- Build a dataset (signals, labels, context)
- Preprocess and represent inputs (features or representations)
- Train a model + tune hyperparameters
- Evaluate properly (generalization, bias, robustness)
- Interpret results and consider deployment constraints

Key message:
Most ML failures come from the workflow (problem definition, leakage, evaluation), not from the algorithm.

Course goals & mindset

What you will learn and how we will think

Core ML approaches:

- supervised learning (regression, classification)
- unsupervised learning (clustering, dimensionality reduction)

Evaluation & pitfalls:
overfitting, leakage, cross-validation, metrics

Responsible and critical perspective:
interpretability, bias, reliability, limitations

Practical orientation through examples

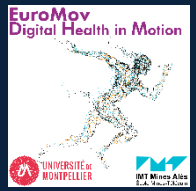
Key message:
The objective is not only to “run models”, but to build trustworthy and useful ones.

IMT Mines Alès
École Mines-Télécom

SCIENCE DES DONNÉES - NIVEAU 1

INTRODUCTION

Gérard Dray



afaq ISO 9001 Qualité AFNOR CERTIFICATION

Objectifs

Être capable de comprendre :

- la définition de la science des données
- les définitions de l'Intelligence Artificielle
- la méthodologie d'Apprentissage Artificiel

Initier une réflexion sur l'usage de la science des données dans le domaine de la santé numérique

Organisation

Programme :

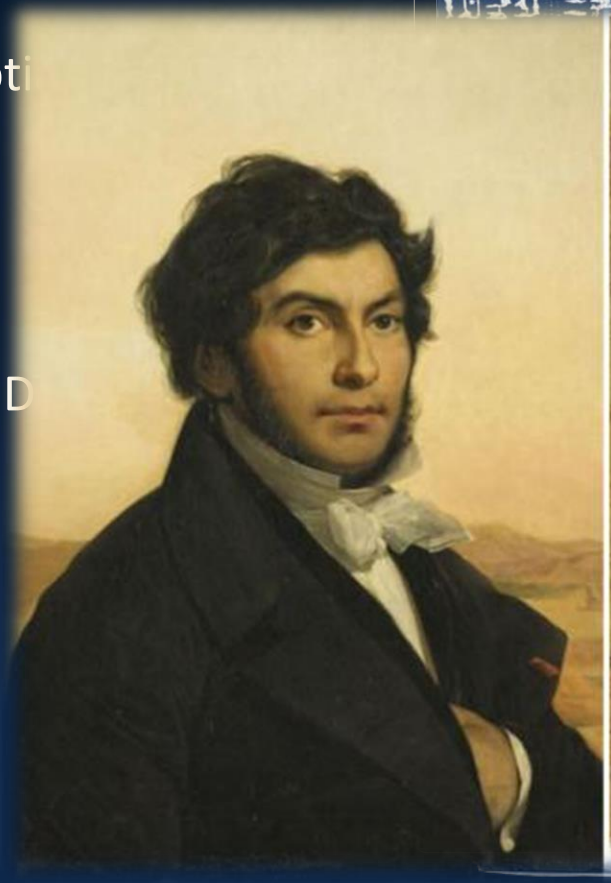
- Une histoire ancienne
- Principales définitions
- Méthode d’apprentissage automatique : une illustration
- Exemples
- Discussion et Conclusion

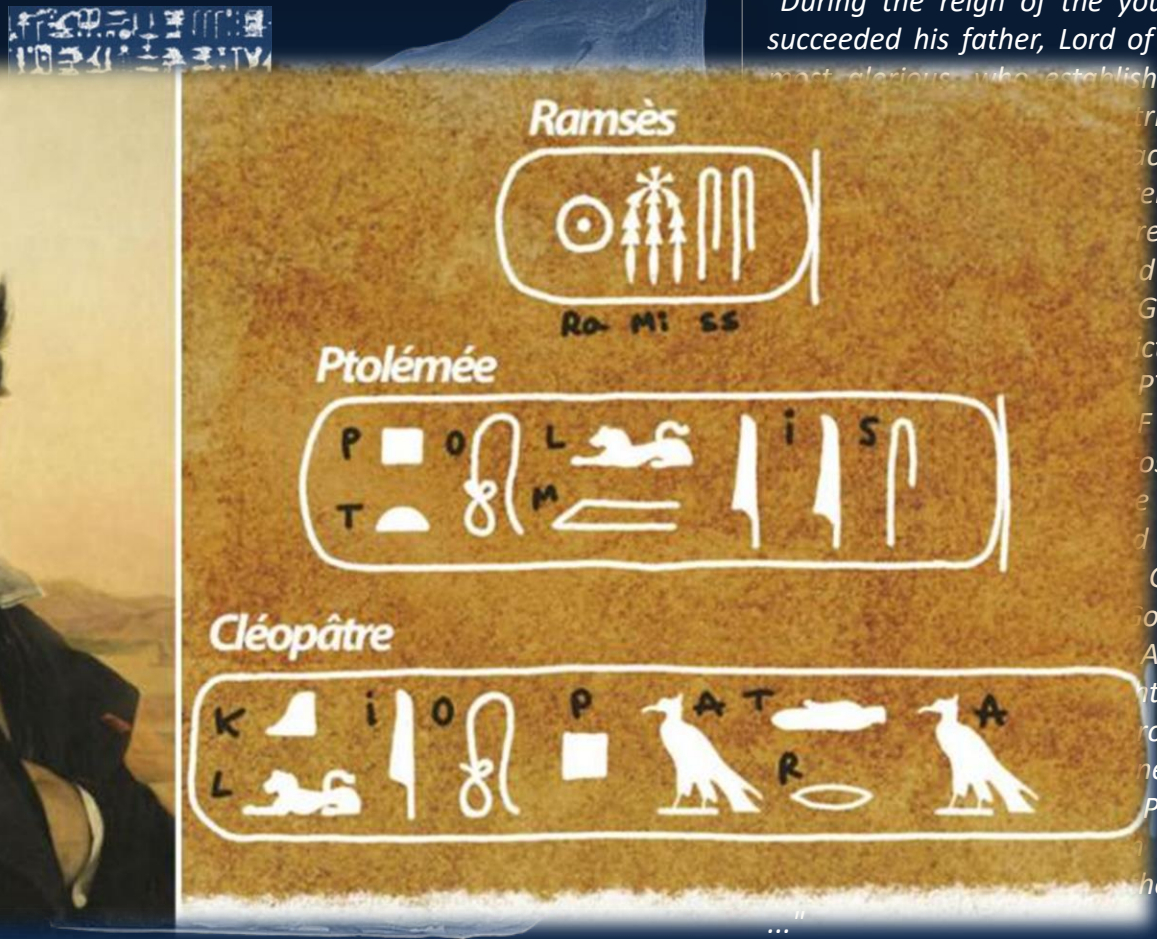
Artificial Intelligence: An ancient story

1822 Jean-François Champollion

Egypti

D





"During the reign of the young man who succeeded his father, Lord of the Diadems, most glorious who established Egypt and triumphing over ace and civilized emonies of the reat, a king like d Lower Egypt, Gods, approved ctory, the living PTOLEMY, WILL F PTAH, in the os of Aetos was e Soteres Gods, d the Euergetai Gods, and the od; daughter Athlophoros of ter Areia of os of Arsinoe ne of Ptolemy Philopator; the of Xandikos, he 18th Mekhir ..."

Pierre de Rosette, 196 av. J.-C

1854 - John Snow
Cholera epidemic, London

Medical Societies.

MEDICAL SOCIETY OF LONDON.

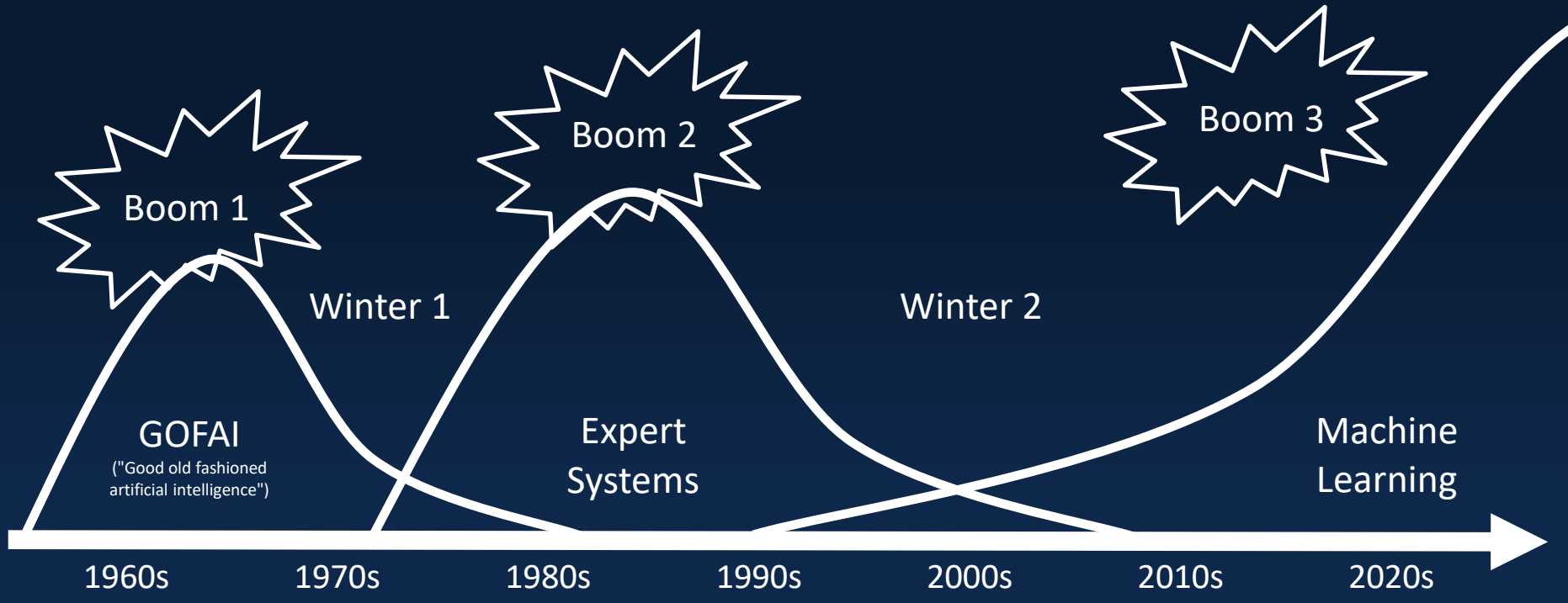
MR. HEADLAND, PRESIDENT.

SATURDAY, OCTOBER 14TH, 1854.

Dr. SNOW considered that the cholera poison acted upon the alimentary canal, and not on the blood or nervous system. In every case which he had seen, the evacuations had been sufficient to account for the collapse, without reference to any other cause. There was no poison in the blood in a case of cholera; in the consecutive fever, as it was called, the blood became poisoned from urea getting into the circulation in consequence of the kidneys not acting, but not from any poison having been present from the beginning. There was nothing in the atmosphere to account for the spread of cholera, which he believed was spread from person to person; and that in all cases it could be traced in this manner. If atmospheric, why did it attack one or two persons only in a locality, and these having direct communication with each other? Such cases he had seen at Sydenham, where there had been only two instances of the disease. The first case in the outbreak of 1849 had occurred to a sailor in Bermondsey; the second affected person was the successor to the sailor in the room in which he died. He thought he had collected evidence enough to show that in all cases cholera was propagated by swallowing some portion of the evacuations of an affected person. These, as was well known, flowed into the bed, &c., and persons attending on the sick might easily take the poison unawares. With respect to the class of persons affected by the disease, he believed that the very poor and vagabonds suffered less, in proportion, than decent, respectable persons. He regarded the cholera and diarrhœa, as lately prevalent, to be the same disease in different degrees of intensity. We observed the same difference in scarlatina and other diseases.

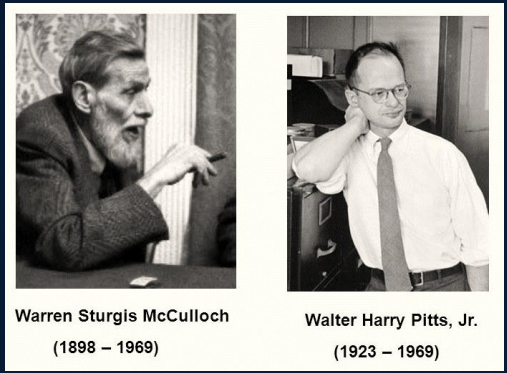
https://fr.wikipedia.org/wiki/%C3%89pid%C3%A9mie_de_chol%C3%A9ra_de_Broad_Street

History of Artificial Intelligence

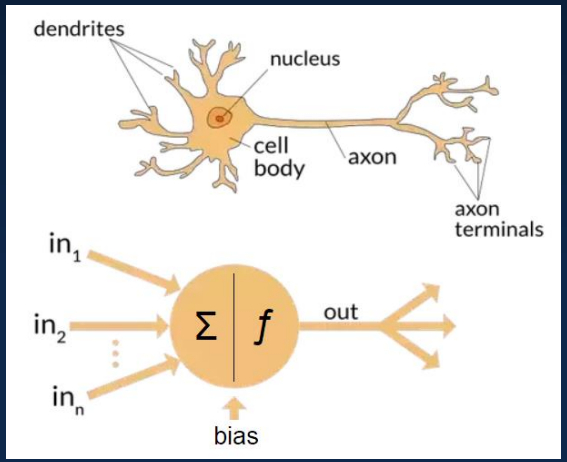


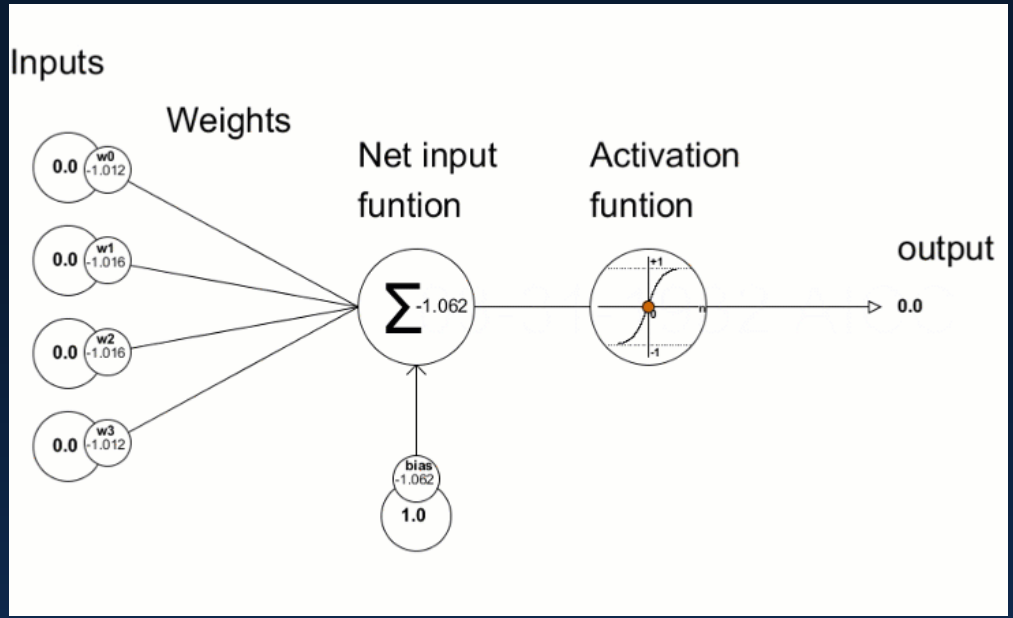
https://www.technologystories.org/ai-evolution/#_ftnref2

1943 The formal neuron of McCulloch and Pitts



https://images.slideplayer.com/22/6379712/slides/slide_8.jpg

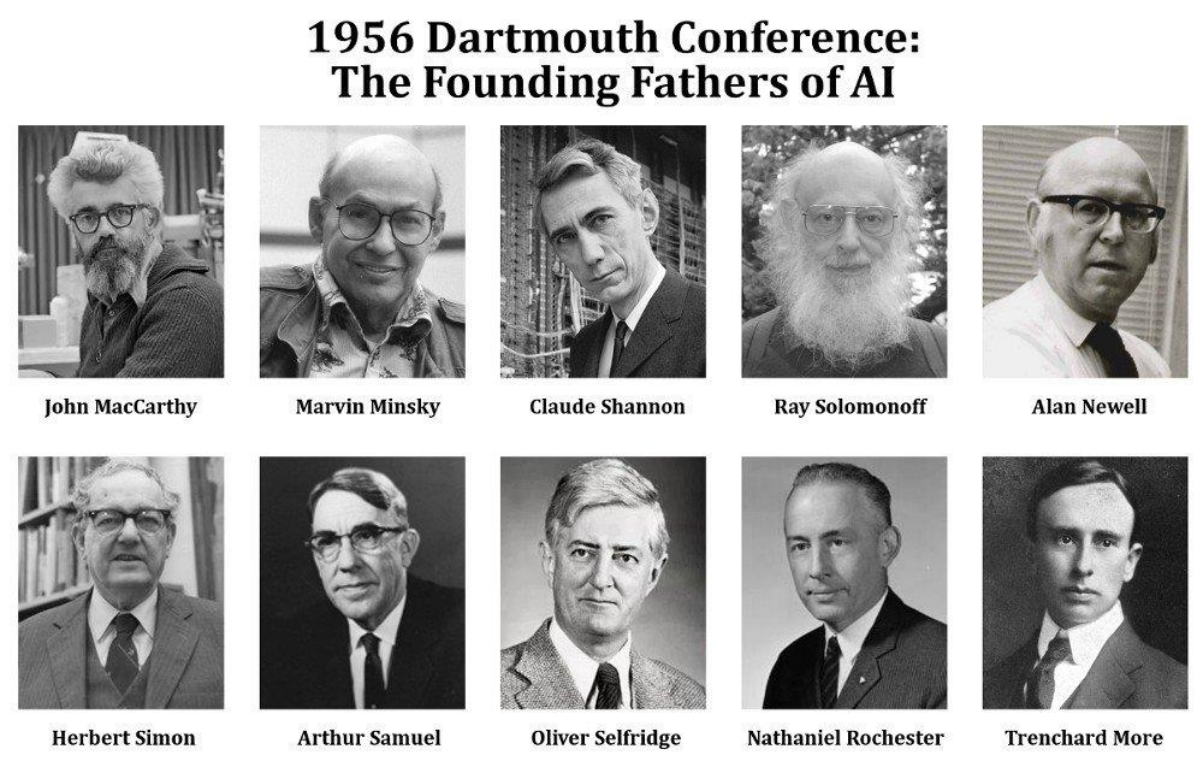




https://www.mql5.com/fr/articles/5486

https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7

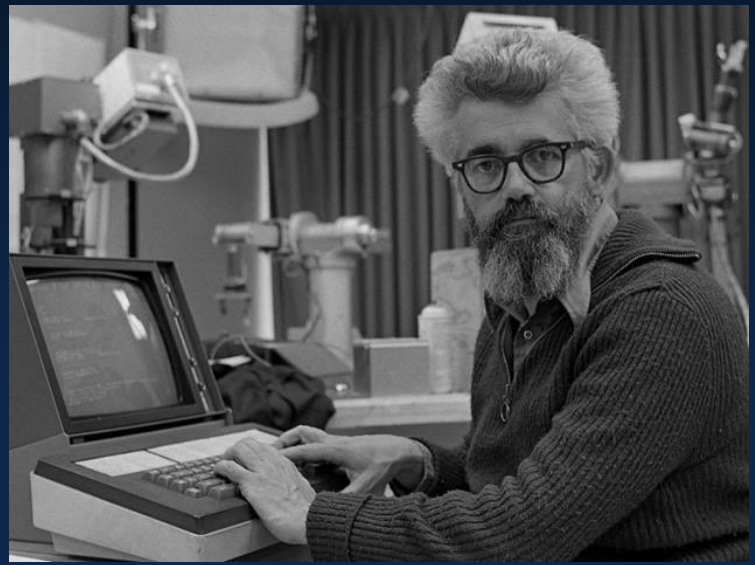
1956 The Dartmouth Conference



The Dartmouth Conference (Dartmouth Summer Research Project on Artificial Intelligence) was a scientific workshop held in the summer of 1956, which is considered to be the birth of artificial intelligence as a field of research.

Artificial Intelligence

“The science and engineering of making intelligent machines, especially intelligent computer programs”

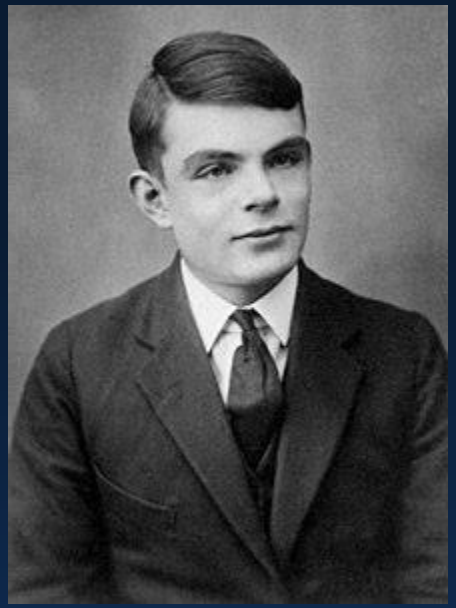


https://becominghuman.ai/introduction-to-artificial-intelligence-5fba0148ec99

“Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed.”

https://www.ibm.com/ibm/history/ibm100/us/en/icons/ibm700series/impacts/

1963 Alan Mathison Turing



“Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's ? If this were then subjected to an appropriate course of education one would obtain the adult brain.”

https://fr.wikipedia.org/wiki/Alan_Turing

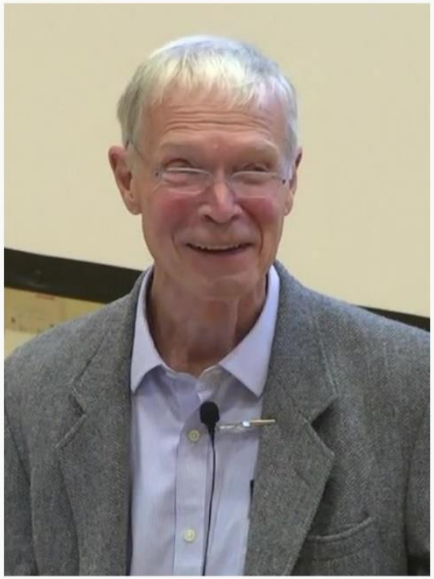
DATA MINING

1997 The day Deep Blue defeated Garry Kasparov at chess

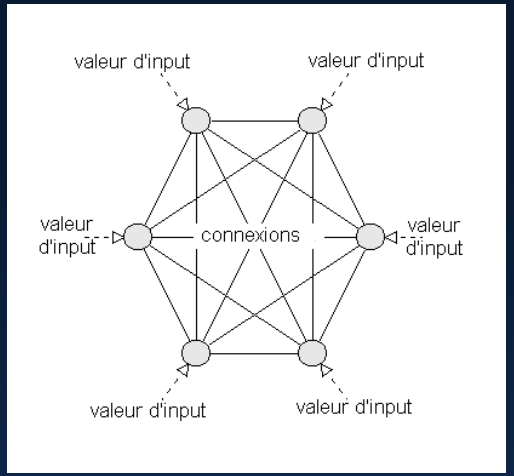
May 11, 1997, chess champion Garry Kasparov loses the sixth game of a historic match.

https://www.cscience.ca/2020/08/07/le-jour-ou-deep-blue-a-battu-garry-kasparov-aux-echecs/

1982 John Joseph Hopfield



John Hopfield en 2016



Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2554–2558, April 1982
Biophysics

Neural networks and physical systems with emergent collective computational abilities

(associative memory/parallel processing/categorization/content-addressable memory/fail-soft devices)

J. J. HOPFIELD

Division of Chemistry and Biology, California Institute of Technology, Pasadena, California 91125; and Bell Laboratories, Murray Hill, New Jersey 07974

Contributed by John J. Hopfield, January 15, 1982

ABSTRACT **Computational properties of use to biological organisms or to the construction of computers can emerge as collective properties of systems having a large number of simple equivalent components (or neurons). The physical meaning of content-addressable memory is described by an appropriate phase space flow of the state of a system. A model of such a system is given, based on aspects of neurobiology but readily adapted to integrated circuits. The collective properties of this model produce a content-addressable memory which correctly yields an entire memory from any subpart of sufficient size. The algorithm for the time evolution of the state of the system is based on asynchronous parallel processing. Additional emergent collective properties include some capacity for generalization, familiarity recognition, categorization, error correction, and time sequence retention. The collective properties are only weakly sensitive to details of the modeling or the failure of individual devices.**

Given the dynamical electrochemical properties of neurons and their interconnections (synapses), we readily understand schemes that use a few neurons to obtain elementary useful biological behavior (1–3). Our understanding of such simple circuits in electronics allows us to plan larger and more complex circuits which are essential to large computers. Because evolution has no such plan, it becomes relevant to ask whether the ability of large collections of neurons to perform "computational" tasks may in part be a spontaneous collective consequence of having a large number of interacting simple neurons.

In physical systems made from a large number of simple elements, interactions among large numbers of elementary components yield collective phenomena such as the stable magnetic orientations and domains in a magnetic system or the vortex patterns in fluid flow. Do analogous collective phenomena in a system of simple interacting neurons have useful "computational" correlates? For example, are the stability of memories, the construction of categories of generalization, or time-sequential memory also emergent properties and collective in origin? This paper examines a new modeling of this old and fundamental question (4–8) and shows that important computational properties spontaneously arise.

All modeling is based on details, and the details of neuroanatomy and neural function are both myriad and incompletely known (9). In many physical systems, the nature of the emergent collective properties is insensitive to the details inserted in the model (e.g., collisions are essential to generate sound waves, but any reasonable interatomic force law will yield appropriate collisions). In the same spirit, I will seek collective properties that are robust against change in the model details.

The model could be readily implemented by integrated circuit hardware. The conclusions suggest the design of a delocalized content-addressable memory or categorizer using extensive asynchronous parallel processing.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

The general content-addressable memory of a physical system

Suppose that an item stored in memory is "H. A. Kramers & G. H. Wannier *Phys. Rev.* **60**, 252 (1941)." A general content-addressable memory would be capable of retrieving this entire memory item on the basis of sufficient partial information. The input "& Wannier, (1941)" might suffice. An ideal memory could deal with errors and retrieve this reference even from the input "Vannier, (1941)". In computers, only relatively simple forms of content-addressable memory have been made in hardware (10, 11). Sophisticated ideas like error correction in accessing information are usually introduced as software (10).

There are classes of physical systems whose spontaneous behavior can be used as a form of general (and error-correcting) content-addressable memory. Consider the time evolution of a physical system that can be described by a set of general coordinates. A point in state space then represents the instantaneous condition of the system. This state space may be either continuous or discrete (as in the case of N Ising spins).

The equations of motion of the system describe a flow in state space. Various classes of flow patterns are possible, but the systems of use for memory particularly include those that flow toward locally stable points from anywhere within regions around those points. A particle with frictional damping moving in a potential well with two minima exemplifies such a dynamics.

If the flow is not completely deterministic, the description is more complicated. In the two-well problems above, if the frictional force is characterized by a temperature, it must also produce a random driving force. The limit points become small limiting regions, and the stability becomes not absolute. But as long as the stochastic effects are small, the essence of local stable points remains.

Consider a physical system described by many coordinates $X_1 \cdots X_N$, the components of a state vector X. Let the system have locally stable limit points X_a, X_b, $\cdots$. Then, if the system is started sufficiently near any X_a, as at $X = X_a + \Delta$, it will proceed in time until $X \simeq X_a$. We can regard the information stored in the system as the vectors X_a, X_b, $\cdots$. The starting point $X = X_a + \Delta$ represents a partial knowledge of the item X_a, and the system then generates the total information X_a.

Any physical system whose dynamics in phase space is dominated by a substantial number of locally stable states to which it is attracted can therefore be regarded as a general content-addressable memory. The physical system will be a potentially useful memory if, in addition, any prescribed set of states can readily be made the stable states of the system.

The model system

The processing devices will be called neurons. Each neuron i has two states like those of McCullough and Pitts (12): $V_i = 0$

https://fr.wikipedia.org/wiki/John_Joseph_Hopfield

https://tpebrodinserise.wordpress.com/hebb-modele-hopfield/

https://www.pnas.org/doi/abs/10.1073/pnas.79.8.2554

2000 Machine Learning

Chat

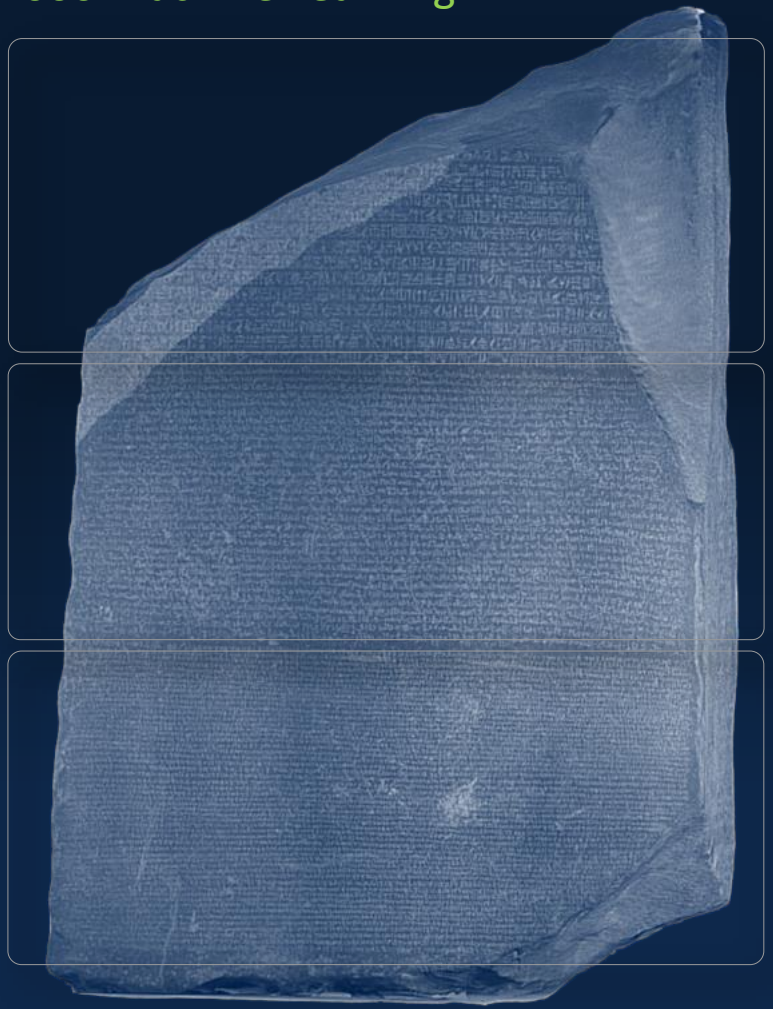
Cat

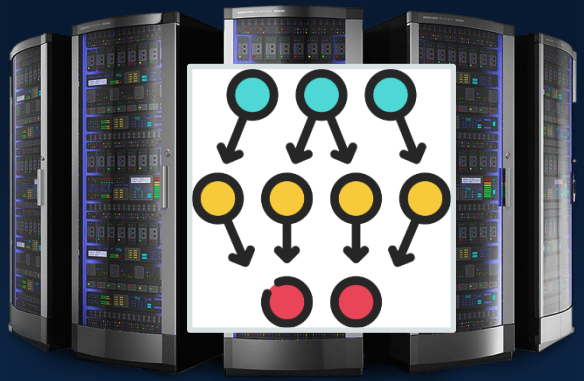
2000 Machine Learning

Chien

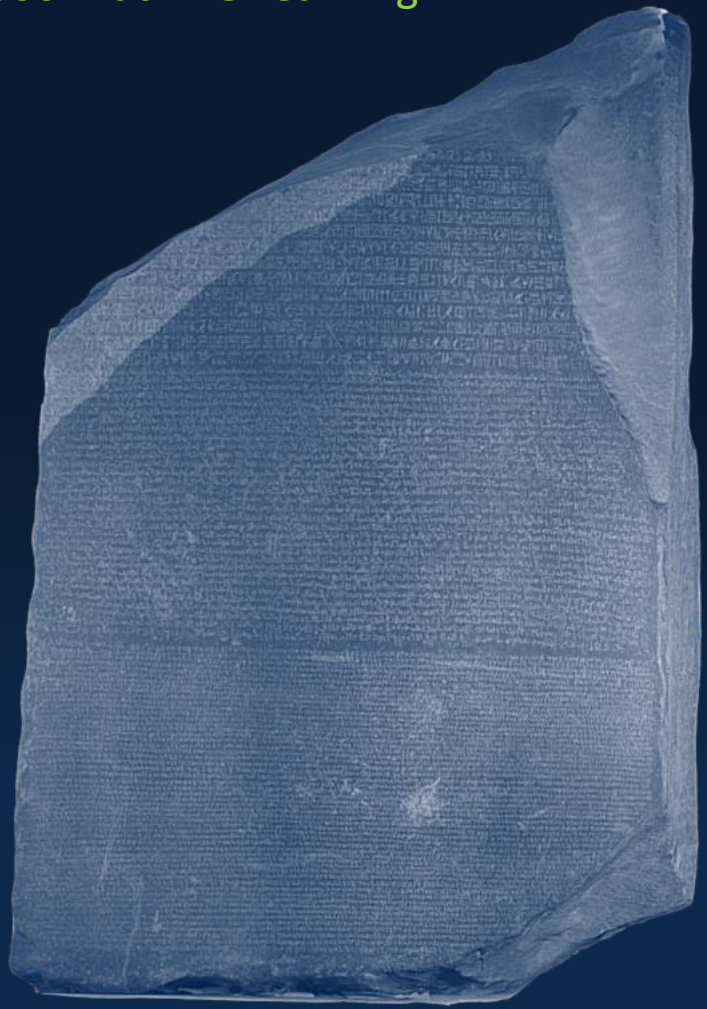
Dog

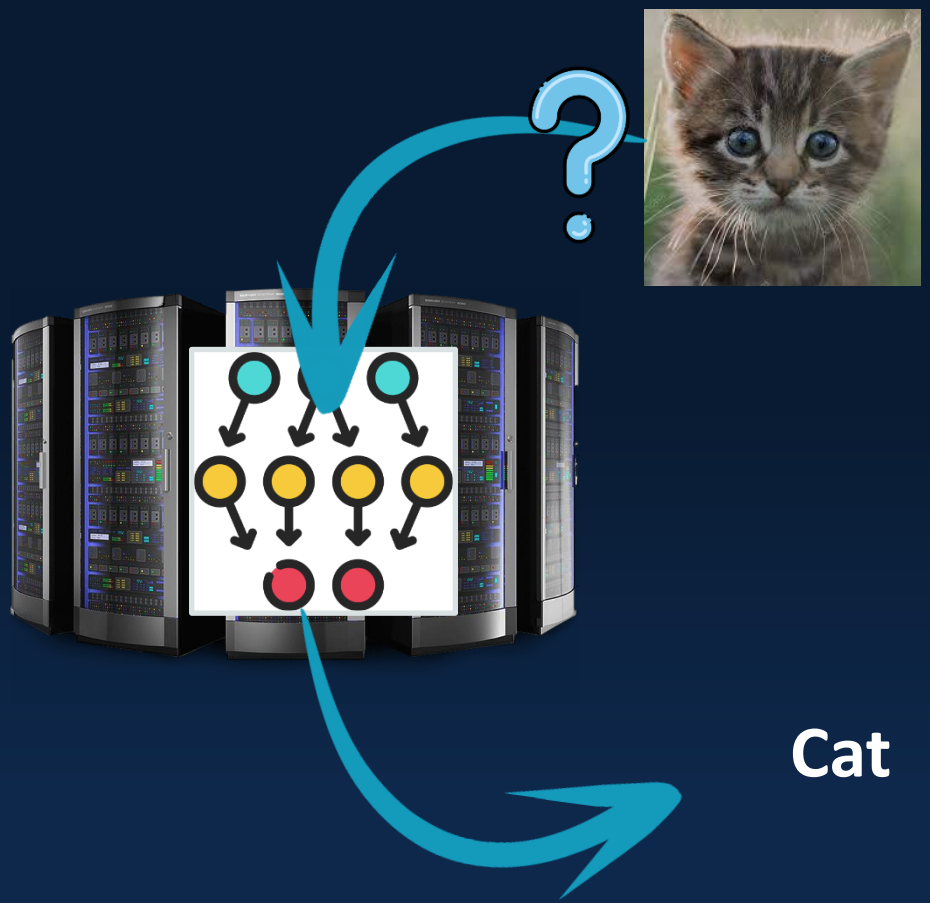
2000 Machine Learning



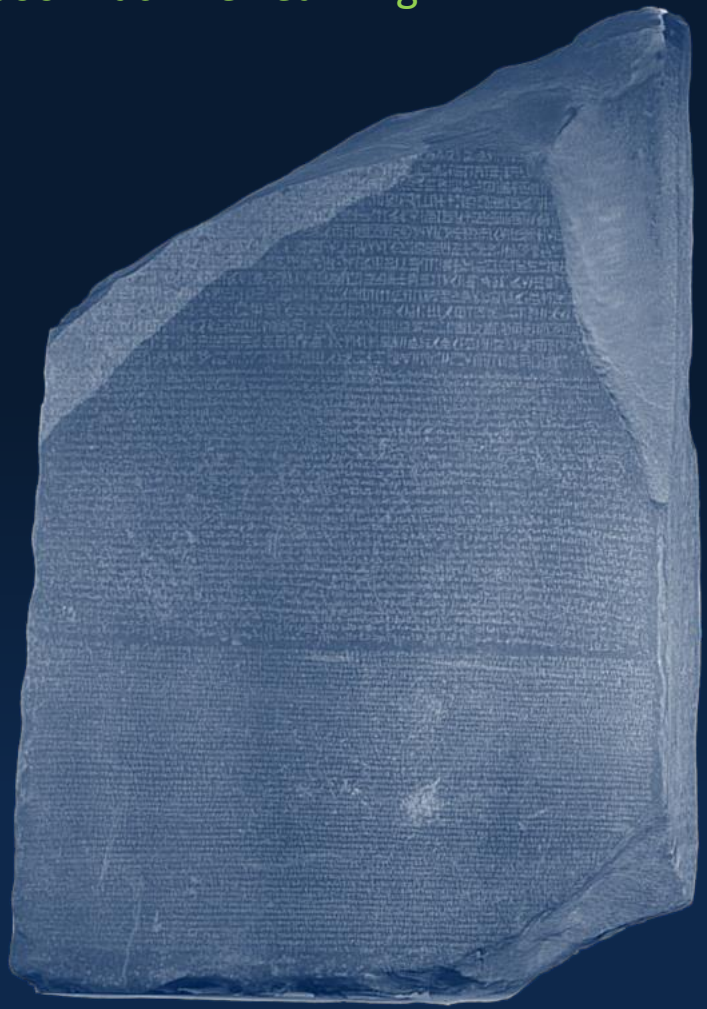


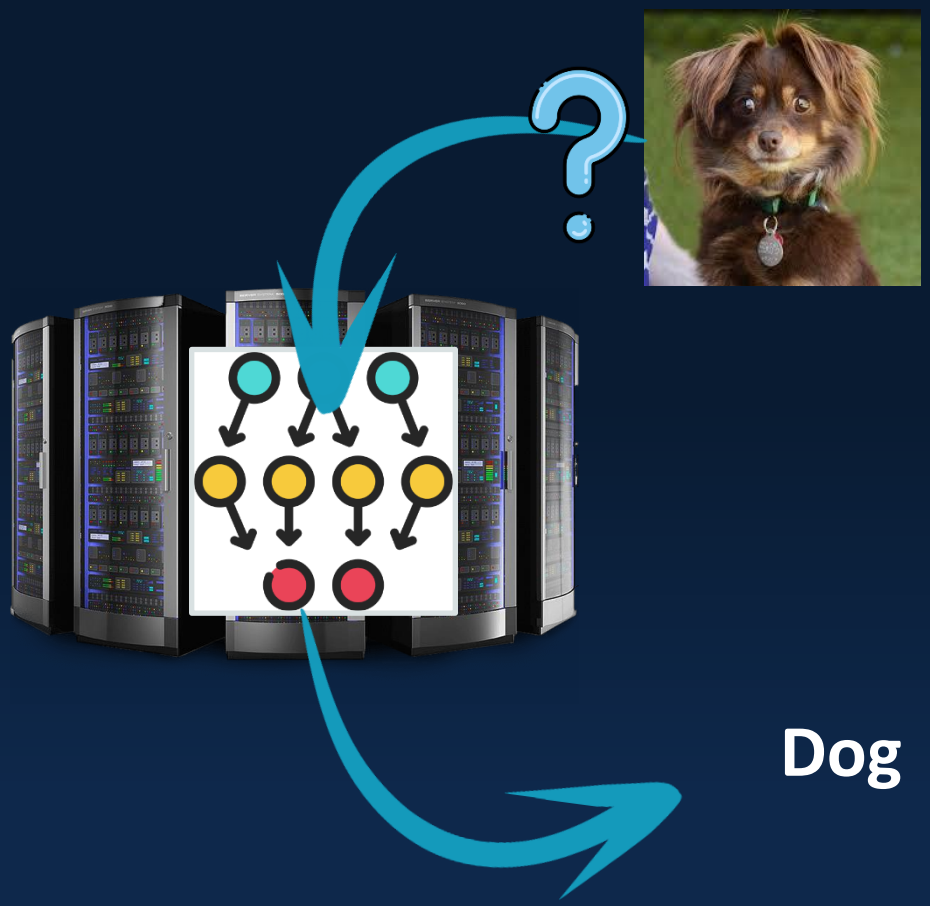
2000 Machine Learning



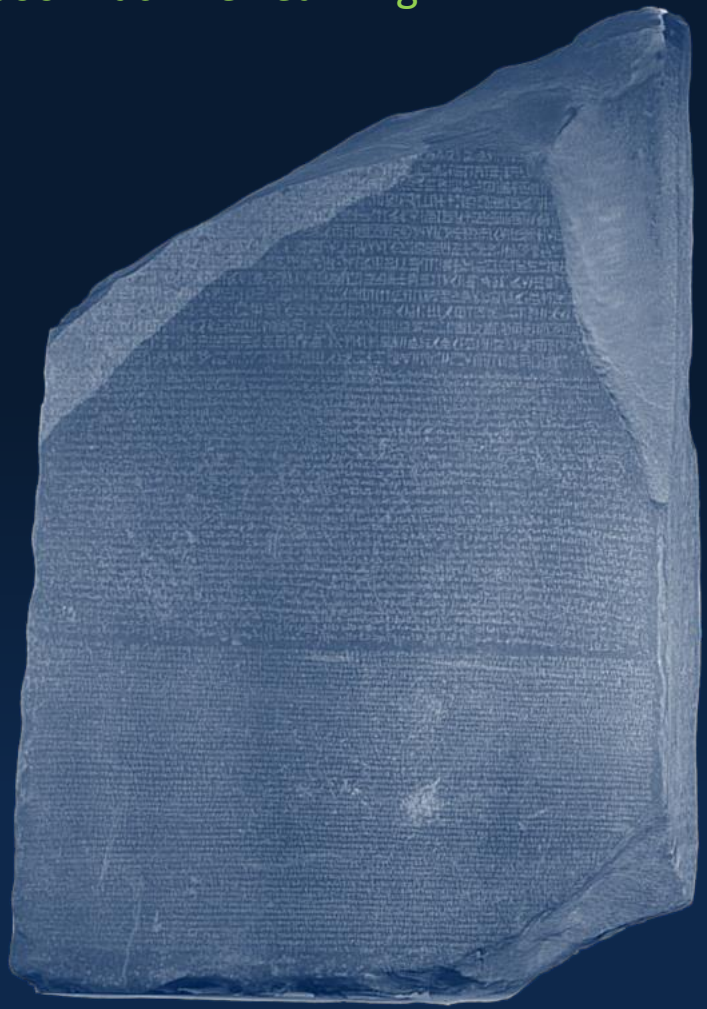


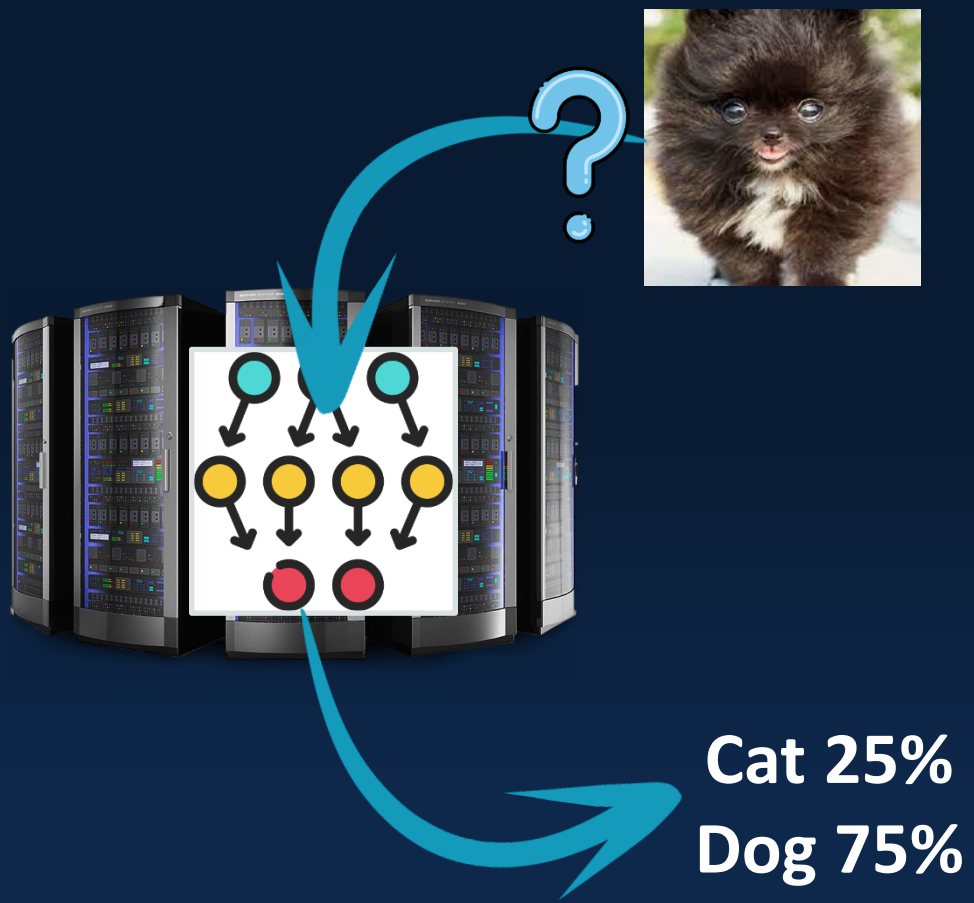
2000 Machine Learning





2000 Machine Learning

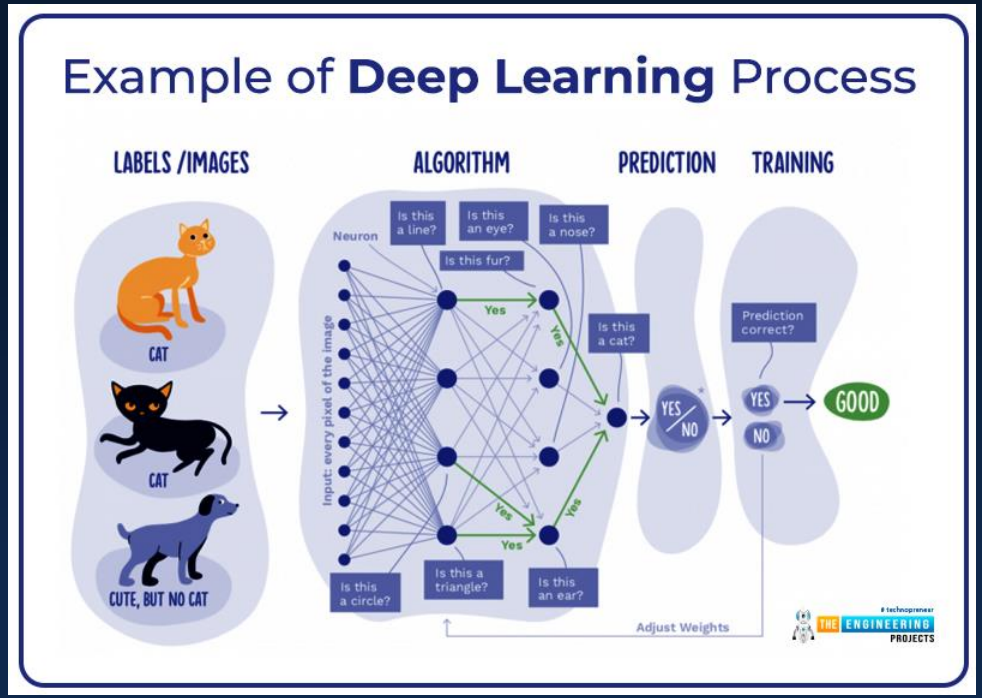




2005 Deep Learning

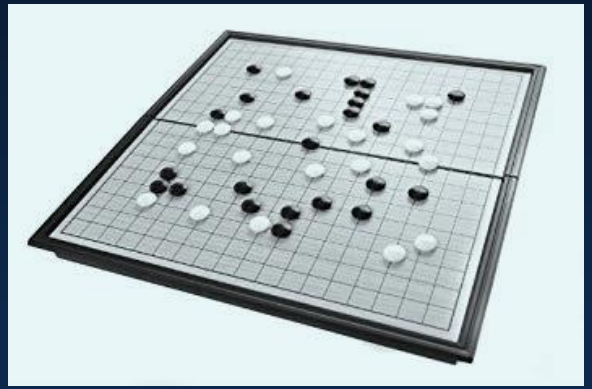
Geoffrey Everest Hinton (born 6 December 1947) is a British-Canadian computer scientist, cognitive scientist, and cognitive psychologist known for his work on artificial neural networks, which earned him the title "the Godfather of AI".

https://en.wikipedia.org/wiki/Geoffrey_Hinton



https://www.theengineeringprojects.com/2022/11/introduction-to-deep-learning.html

Deep Learning
Apprentissage Profond



MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing package e.g. R
- Databases SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

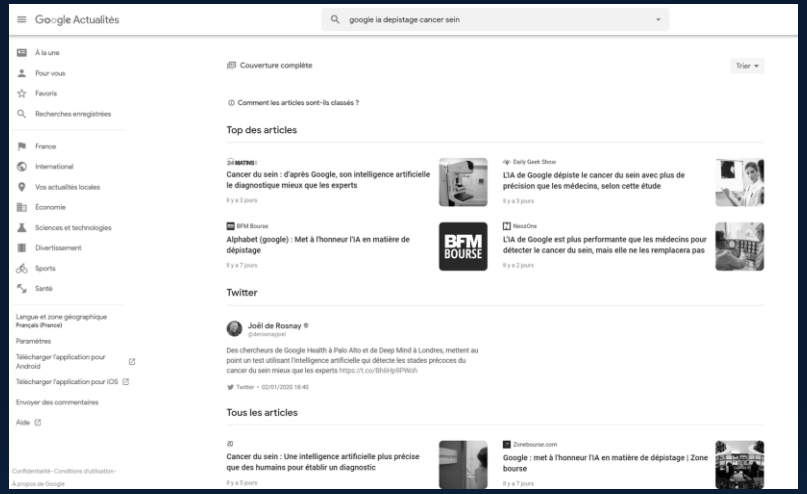
COMMUNICATION & VISUALIZATION

- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization: customer tracking and on-site analytics: predictive analytics and econometrics: data warehousing and big data systems: marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing DISTILLERY

2020



Article

International evaluation of an AI system for breast cancer screening

https://doi.org/10.1038/s41586-019-1799-6
Received: 27 July 2019
Accepted: 5 November 2019
Published online: 1 January 2020

Scott Mayer McKinney[1,14*], Marcin Sieniek[1,14], Varun Godbole[1,14], Jonathan Godwin[2,14], Natasha Antropova[2], Hutan Ashrafian[3,4], Trevor Back[2], Mary Chesus[2], Greg C. Corrado[1], Ara Darzi[3,4,5], Mozziyar Etemadi[6], Florencia Garcia-Vicente[6], Fiona J. Gilbert[7], Mark Halling-Brown[8], Demis Hassabis[2], Sunny Jansen[9], Alan Karthikesalingam[10], Christopher J. Kelly[10], Dominic King[10], Joseph R. Ledsam[2], David Melnick[6], Hormuz Mostofi[1], Lily Peng[1], Joshua Jay Reicher[11], Bernardino Romera-Paredes[2], Richard Sidebottom[12,13], Mustafa Suleyman[2], Daniel Tse[1*], Kenneth C. Young[8], Jeffrey De Fauw[2,15] & Shravya Shetty[1,15*]

Screening mammography aims to identify breast cancer at earlier stages of the disease, when treatment can be more successful[1]. Despite the existence of screening programmes worldwide, the interpretation of mammograms is affected by high rates of false positives and false negatives[2]. Here we present an artificial intelligence (AI) system that is capable of surpassing human experts in breast cancer prediction. To assess its performance in the clinical setting, we curated a large representative dataset from the UK and a large enriched dataset from the USA. We show an absolute reduction of 5.7% and 1.2% (USA and UK) in false positives and 9.4% and 2.7% in false negatives. We provide evidence of the ability of the system to generalize from the UK to the USA. In an independent study of six radiologists, the AI system outperformed all of the human readers: the area under the receiver operating characteristic curve (AUC-ROC) for the AI system was greater than the AUC-ROC for the average radiologist by an absolute margin of 11.5%. We ran a simulation in which the AI system participated in the double-reading process that is used in the UK, and found that the AI system maintained non-inferior performance and reduced the workload of the second reader by 88%. This robust assessment of the AI system paves the way for clinical trials to improve the accuracy and efficiency of breast cancer screening.

Breast cancer is the second leading cause of death from cancer in women[3], but early detection and treatment can considerably improve outcomes[1,4,5]. As a consequence, many developed nations have implemented large-scale mammography screening programmes. Major medical and governmental organizations recommend screening for all women starting between the ages of 40 and 50[6–8]. In the USA and UK combined, over 42 million exams are performed each year[9,10].

Despite the widespread adoption of mammography, interpretation of these images remains challenging. The accuracy achieved by experts in cancer detection varies widely, and the performance of even the best clinicians leaves room for improvement[11,12]. False positives can lead to patient anxiety[13], unnecessary follow-up and invasive diagnostic procedures. Cancers that are missed at screening may not be identified until they are more advanced and less amenable to treatment[14].

AI may be uniquely poised to help with this challenge. Studies have demonstrated the ability of AI to meet or exceed the performance of human experts on several tasks of medical-image analysis[15–19]. As a shortage of mammography professionals threatens the availability and adequacy of breast-screening services around the world[20–23], the scalability of AI could improve access to high-quality care for all.

Computer-aided detection (CAD) software for mammography was introduced in the 1990s, and several assistive tools have been approved for medical use[24]. Despite early promise[25,26], this generation of software failed to improve the performance of readers in real-world settings[12,27,28]. More recently, the field has seen a renaissance owing to the success of deep learning. A few studies have characterized systems for breast cancer prediction with stand-alone performance that approaches that of human experts[29,30]. However, the existing work has several limitations. Most studies are based on small, enriched datasets with limited follow-up, and few have compared performance to readers in actual clinical practice—instead relying on laboratory-based simulations of the reading environment. So far there has been little evidence of the ability of AI systems to translate between different screening populations and settings without additional training data[31]. Critically, the pervasive use of follow-up intervals that are no longer than 12 months[29,30,32,33]

[1]Google Health, Palo Alto, CA, USA. [2]DeepMind, London, UK. [3]Department of Surgery and Cancer, Imperial College London, London, UK. [4]Institute of Global Health Innovation, Imperial College London, London, UK. [5]Cancer Research UK Imperial Centre, Imperial College London, London, UK. [6]Northwestern Medicine, Chicago, IL, USA. [7]Department of Radiology, Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK. [8]Royal Surrey County Hospital, Guildford, UK. [9]Verily Life Sciences, South San Francisco, CA, USA. [10]Google Health, London, UK. [11]Stanford Health Care and Palo Alto Veterans Affairs, Palo Alto, CA, USA. [12]The Royal Marsden Hospital, London, UK. [13]Thirlestaine Breast Centre, Cheltenham, UK. [14]These authors contributed equally: Scott Mayer McKinney, Marcin T. Sieniek, Varun Godbole, Jonathan Godwin. [15]These authors jointly supervised this work: Jeffrey De Fauw, Shravya Shetty. *e-mail: scottmayer@google.com; tsed@google.com; sshetty@google.com

https://simulator.drdata.io/

Le Parisien

XV de France : «Ce ne sont pas les chiffres qui nous dirigent»... Comment Fabien Galthié utilise les datas

Le XV de France, qui affronte le Japon dimanche (14 heures) à Toulouse lors de son dernier match de la tournée d'automne, s'appuie dans sa préparation sur une multitude de données chiffrées, collectées, triées puis fournies par la société SAS France.

Fabien Galthié avant le match contre l'Afrique du Sud au Velodrome, à Marseille. Icon Sport/Johnny Fidelin

Les chiffres, fournis notamment par un puce GPS incrustée dans les maillots, remontent et s'inscrivent dans une base de données où figurent 1438 matchs (XV de France, U20 et Top 14 compris). Les ballons connectés et les protège-dents connectés élargissent également le champ des investigations. ***« Attention, ce ne sont pas les chiffres qui nous dirigent mais nous qui dirigeons les chiffres »,*** *lance Fabien Galthié. D'ailleurs, sur un match, je ne suis connecté à rien. Je regarde, j'écoute, je suis au cœur de l'environnement, il n'y a pas un seul chiffre qui m'impacte. Les décisions, on les prend avec notre intuition et notre savoir-faire. » La science, elle, est là pour répondre aux interrogations.*

https://www.leparisien.fr/sports/rugby/xv-de-france-ce-ne-sont-pas-les-chiffres-qui-nous-dirigent-comment-fabien-galthie-utilise-les-datas-16-11-2022-EOIYH3I7GVBYJMNJEM2JW4555A.php

OpenAI

OpenAI is an AI research and deployment company. Our mission is to ensure that artificial general intelligence benefits all of humanity.

OpenAI's mission is to ensure that artificial general intelligence (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—benefits all of humanity.

We will attempt to directly build safe and beneficial AGI, but will also consider our mission fulfilled if our work aids others to achieve this outcome.

OpenAI Charter

We're releasing a charter that describes the principles we use to execute on OpenAI's mission. This document reflects the strategy we've refined over the past two years, including feedback from many people internal and external to OpenAI. The timeline to AGI remains uncertain, but our charter will guide us in acting in the best interests of humanity throughout its development.

Newsroom

Announcements

ChatGPT: Optimizing Language Models for Dialogue
November 30, 2022 — Announcements, Research

DALL·E API Now Available in Public Beta
November 3, 2022 — Announcements, API

DALL·E Now Available Without Waitlist
September 28, 2022 — Announcements

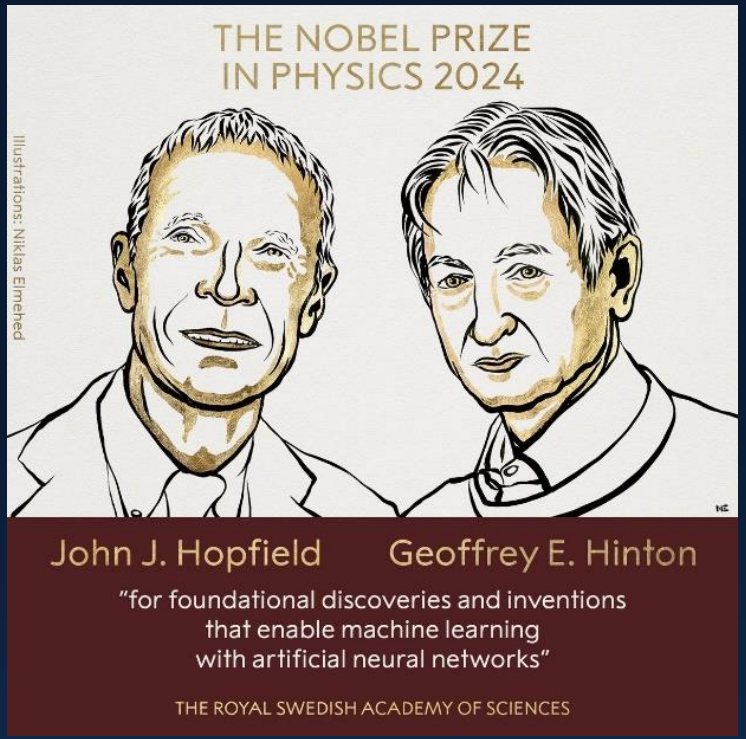
https://chat.openai.com/chat

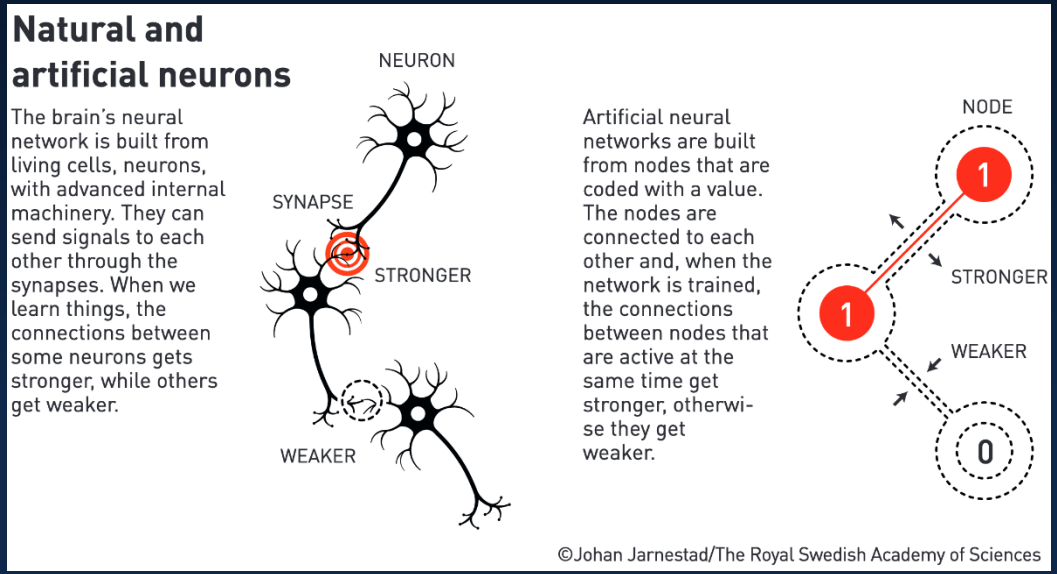
https://openai.com/dall-e-2/

https://openai.com

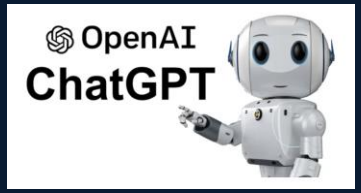
2022 Generative Artificial Intelligence

2022 Generative Artificial Intelligence





https://culturesciencesphysique.ens-lyon.fr/actualites/prix-nobel-de-physique-2024
https://www.clubic.com/actualite-539927-prix-nobel-de-physique-2024-les-laureats-recompenses-pour-leurs-travaux-sur-cette-techno-indispensable-a-l-ia.html



Agent conversationnel

You
Pourrais-tu rédiger une brève introduction pour un cours sur l'intelligence artificielle appliquée à l'ingénierie biomédicale ?

ChatGPT
Bien sûr ! Voici une proposition de texte pour introduire un cours sur l'intelligence artificielle appliquée à l'ingénierie biomédicale :

Introduction au cours : Intelligence Artificielle en Ingénierie Biomédicale

Bienvenue dans ce cours passionnant où nous explorerons les intersections innovantes entre l'intelligence artificielle (IA) et l'ingénierie biomédicale. Ce domaine en pleine expansion utilise les avancées de l'IA pour transformer les soins de santé, en améliorant les diagnostics, en optimisant les traitements et en révolutionnant la gestion des maladies.

Au cours de ce programme, nous aborderons les principes fondamentaux de l'IA, y compris l'apprentissage automatique, le traitement du langage naturel et la vision par ordinateur, et comment ces technologies peuvent être appliquées spécifiquement dans le contexte biomédical. Nous étudierons des études de cas réelles montrant l'impact de l'IA sur la création de dispositifs médicaux intelligents, le développement de techniques de diagnostic précoces et l'élaboration de solutions personnalisées pour le traitement des patients.

Ce cours est conçu pour fournir aux étudiants une compréhension profonde des défis techniques, éthiques et réglementaires associés à l'intégration de l'IA dans les applications biomédicales. En combinant théorie et pratique, nous préparerons les étudiants à devenir les pionniers de cette révolution technologique en santé.

Préparez-vous à plonger dans un monde où la technologie et la médecine convergent pour améliorer la qualité de vie des personnes à l'échelle mondiale. Ensemble, nous découvrirons comment l'intelligence artificielle peut être utilisée pour résoudre certains des problèmes les plus urgents en médecine aujourd'hui.

J'espère que cette introduction vous conviendra pour votre cours !

MIDJOURNEY
All about imagination

Midjourney Bot

/Imagine

Ultra realistic photo of a class of students taking a machine-learning workshop and listening to a human teacher

Consulter le journal

Le Monde

G. Dray

Actualités Économie Vidéos Débats Culture Le Goût du Monde Services

ÉCONOMIE • L'INVESTITURE DE DONALD TRUMP

Le second mandat de Donald Trump commence par un investissement de centaines de milliards de dollars pour l'intelligence artificielle

Le nouveau président américain a dévoilé, mardi 21 janvier, le projet Stargate, destiné à bâtir les centres de données géants de la future génération d'IA, élaboré par Oracle, OpenAI et SoftBank. Soit 100 milliards de dollars investis tout de suite, portés à 500 milliards de dollars, d'ici à 2029.

Par Arnaud Leparmentier (New York, correspondant)

Publié aujourd'hui à 07h46, modifié à 08h30 • Lecture 3 min.

Offrir l'article

https://www.lemonde.fr/economie/article/2025/01/22/le-mandat-trump-commence-par-une-pluie-de-centaines-de-milliards-dollars-destines-l-intelligence-artificielle_6509500_3234.html?lmd_medium=email&lmd_campaign=trf_newsletters_lmfr&lmd_creation=a_la_une&lmd_send_date=undefined&lmd_email_link=a-la-une-points-forts-lire-la-suite&M_BT=55310159973680

2025

Le Monde

SCIENCE & MÉDECINE – SUPPLÉMENT — VIVRE AVEC UN CŒUR OU UN REIN DE COCHON, C'EST POUR BIENTÔT ?

IA : le chinois DeepSeek ébranle le secteur

▸ La start-up chinoise a rendu public un modèle aux performances comparables aux leaders OpenAI ou Google, mais d'un coût beaucoup moins élevé

▸ Pour expliquer sa surprenante efficacité, DeepSeek dit avoir utilisé plusieurs techniques innovantes pour optimiser la quantité de calculs

▸ Cette annonce a provoqué un véritable séisme sur les marchés financiers, faisant éclater la bulle de l'intelligence artificielle qui porte le Nasdaq

▸ Lundi, le fabricant de semi-conducteurs Nvidia, qui valait 3 500 milliards de dollars, a vu sa valeur perdre 590 milliards en une journée

▸ Aux observateurs qui appellent à la prudence, les responsables de DeepSeek rappellent que leurs modèles sont « ouverts »

CINÉMA BOB DYLAN IMPRIME SA LÉGENDE À L'ÉCRAN

▸ « Un parfait inconnu », le biopic de James Mangold, sort mercredi en salle

▸ Timothée Chalamet incarne le chanteur, Monica Barbaro « est » Joan Baez

RDC Les forces du M23 et du Rwanda sont entrées dans Goma

ÉDITORIAL
DÉSIGNER CLAIREMENT LES PROTAGONISTES

Politique
François Bayrou se tourne vers sa droite pour éviter la censure

Agriculture
La Coordination rurale dans l'ombre de l'extrême droite

Commémoration
A Auschwitz, la mise en garde des derniers survivants

Mayotte
Après le cyclone Chido, une rentrée scolaire très perturbée

Crise climatique
La transition écologique, nouvelle cible des populistes

Reportage
A bas bruit, des Israéliens choisissent de s'exiler

Idées
Le temps de la « vassalisation heureuse » ?

LE BILAN DU MONDE

ÉCONOMIE • INTELLIGENCE ARTIFICIELLE

La start-up chinoise DeepSeek bouleverse le secteur de l'intelligence artificielle

En développant DeepSeek-R1, un modèle aussi performant que ceux des leaders américains comme OpenAI ou Google, mais avec moins de ressources et en open source, DeepSeek fait vaciller la Silicon Valley.

Par Alexandre Piquard

Publié hier à 20h01, modifié à 09h29 · Lecture 4 min.

Implementing the EU AI Act in Radiology: ESR's Recommendations

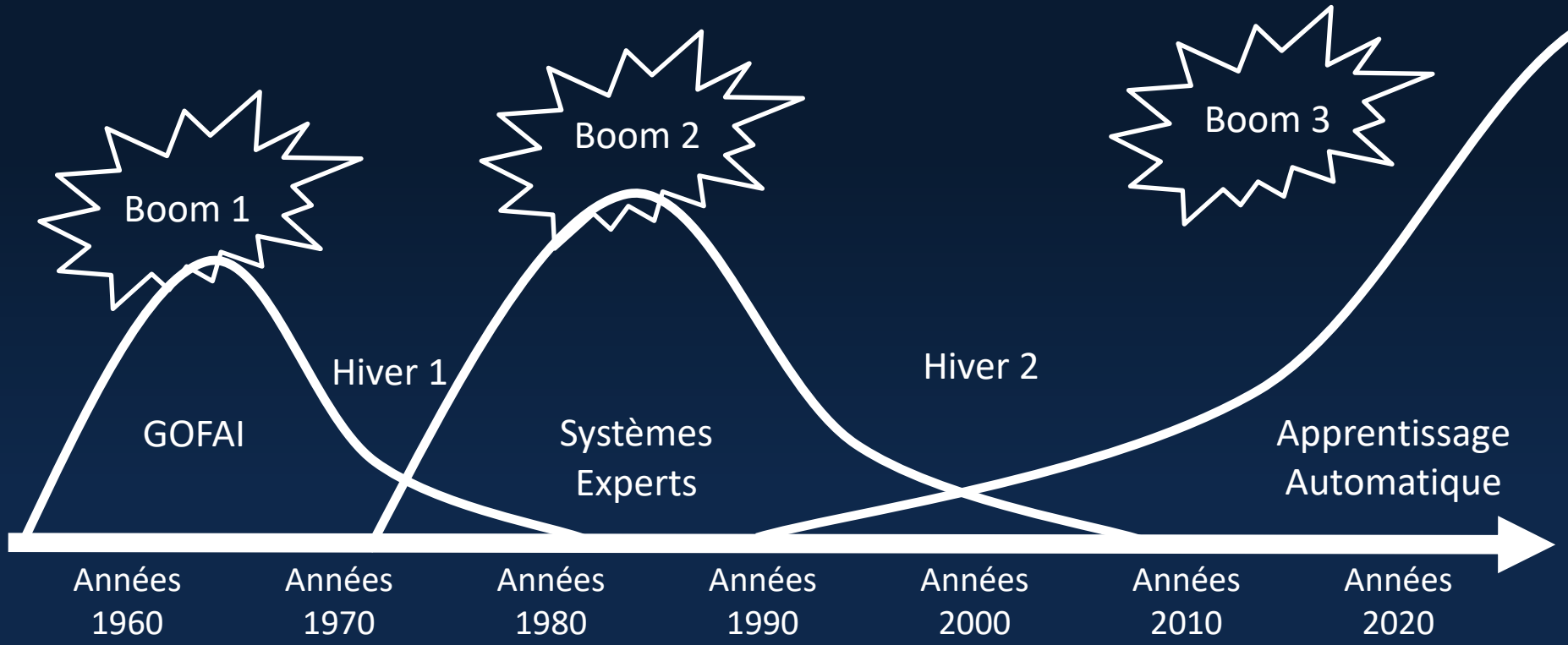
In EXEC Tue, 18 Mar 2025

AI in Lung Cancer Screening: Enhancing Accuracy and Reducing Workload

In Imaging Mon, 17 Mar 2025

https://healthmanagement.org/c/hospital/news/implementing-the-eu-ai-act-in-radiology-esrs-recommendations
https://healthmanagement.org/c/imaging/news/ai-in-lung-cancer-screening-enhancing-accuracy-and-reducing-workload

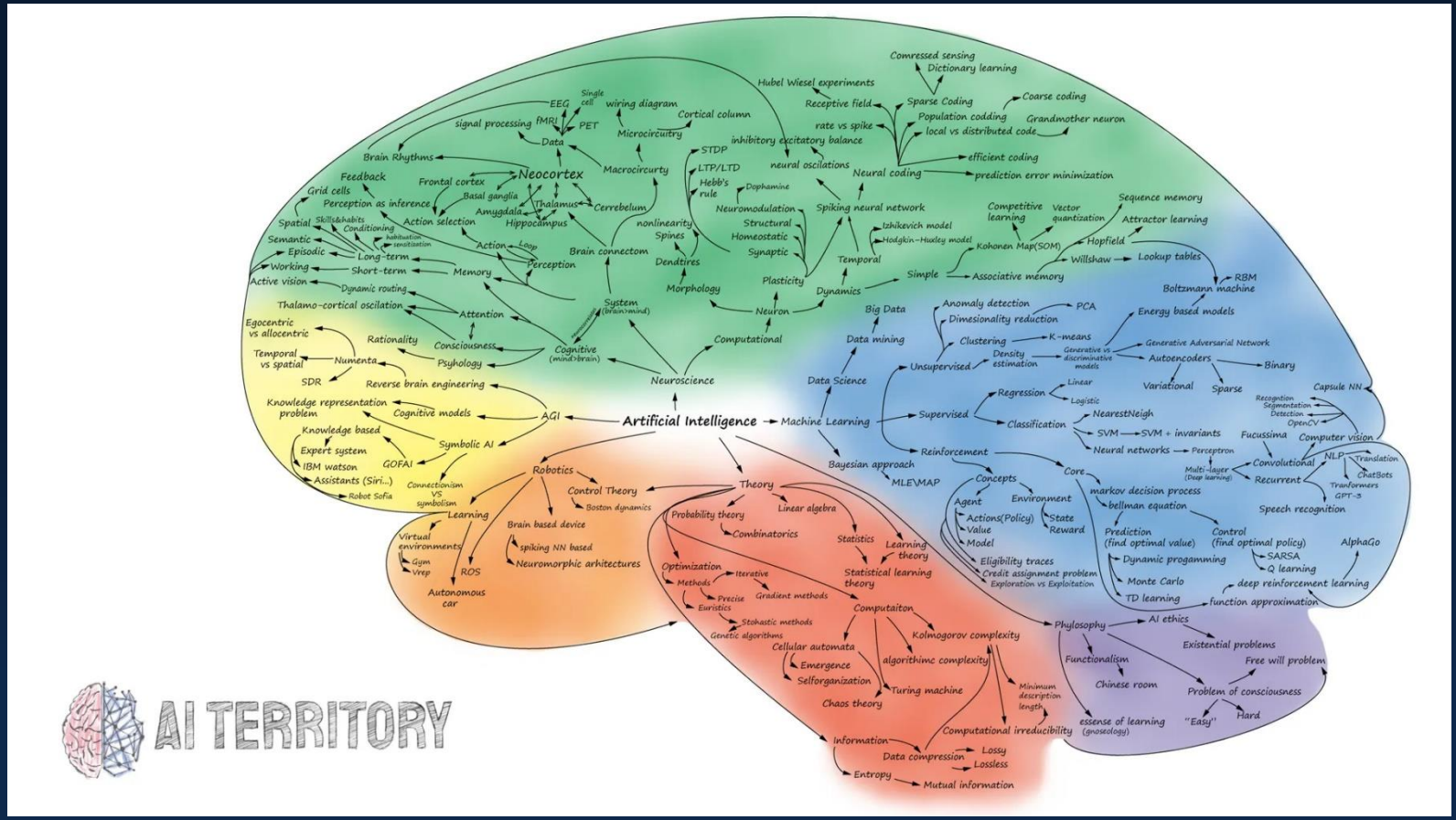
Historique de l’Intelligence Artificielle



https://www.technologystories.org/ai-evolution/#_ftnref2

Key definitions

The Map of Artificial Intelligence



https://medium.com/swlh/the-map-of-artificial-intelligence-2020-2c4f446f4e43

What is Artificial Intelligence (AI)?

Artificial intelligence refers to any tool used by a machine to “reproduce human-like behaviors, such as reasoning, planning, and creativity.”

What is machine learning (ML)?

Machine learning is a field of artificial intelligence that aims to give machines the ability to “learn” from data using mathematical models. More specifically, it is the process by which relevant information is extracted from a training data set.The goal of this phase is to obtain the parameters of a model that will achieve the best performance, particularly when performing the task assigned to the model. Once learning is complete, the model can then be deployed in production.

What is deep learning (DL)?

Deep learning is a subset of machine learning. Deep learning is a machine learning process that uses neural networks with multiple layers of hidden neurons. These algorithms have a large number of parameters, so they require a very large amount of data in order to be trained.

Qu'est-ce que l'intelligence artificielle? (Artificial Intelligence)

L’intelligence artificielle représente tout outil utilisé par une machine afin de « reproduire des comportements liés aux humains, tels que le raisonnement, la planification et la créativité ».

Qu'est-ce que l'apprentissage automatique ? (Machine Learning)

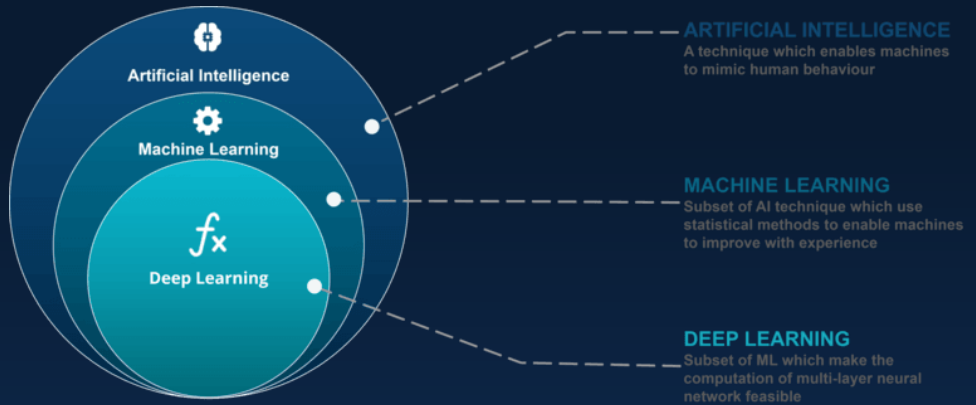
L’apprentissage automatique est un champ d’étude de l’intelligence artificielle qui vise à donner aux machines la capacité d’« apprendre » à partir de données, via des modèles mathématiques. Plus précisément, il s’agit du procédé par lequel les informations pertinentes sont tirées d’un ensemble de données d’entraînement.

Le but de cette phase est l’obtention des paramètres d’un modèle qui atteindront les meilleures performances, notamment lors de la réalisation de la tâche attribuée au modèle. Une fois l’apprentissage réalisé, le modèle pourra ensuite être déployé en production.

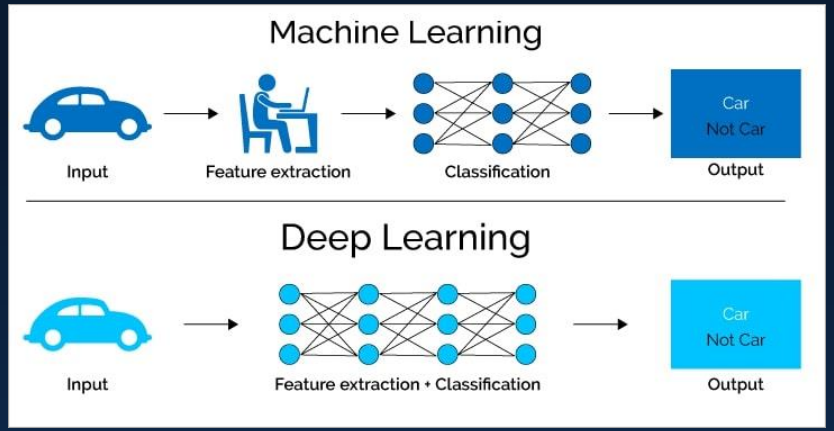
Qu'est-ce que l'apprentissage profond ? (Deep Learning)

L'apprentissage profond est un sous-ensemble de l'apprentissage automatique. L’apprentissage profond est un procédé d’apprentissage automatique utilisant des réseaux de neurones possédants plusieurs couches de neurones cachées. Ces algorithmes possédant de très nombreux paramètres, ils demandent un nombre très important de données afin d’être entraînés.

Qu'est-ce que l'apprentissage automatique ? (Machine Learning)
Qu'est-ce que l'apprentissage profond ? (Deep Learning)



https://datawider.com/how-deep-learning-is-different-from-machine-learning/



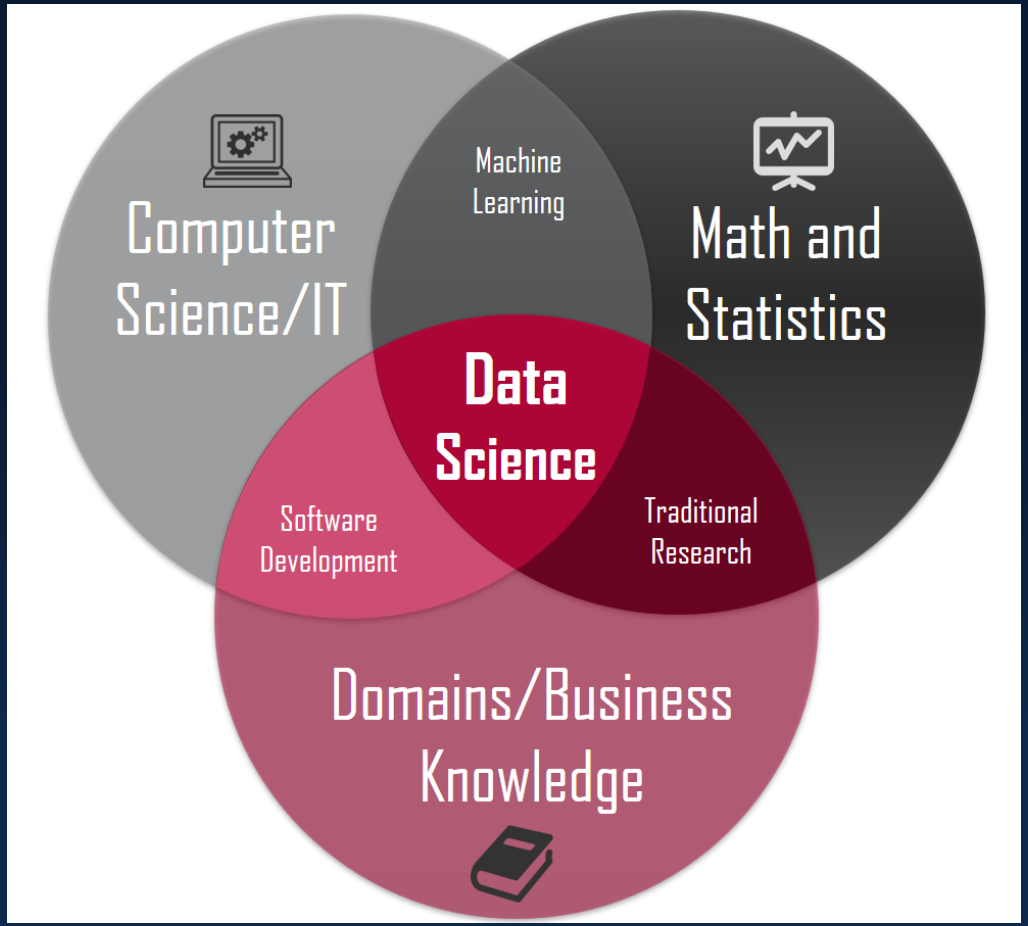
https://levity.ai/blog/difference-machine-learning-deep-learning

Qu'est-ce que la science des données ?

La science des données est l'étude de l'extraction automatisée de connaissance à partir de grands ensembles de données.

Plus précisément, la science des données est un domaine interdisciplinaire qui utilise des méthodes, des processus, des algorithmes et des systèmes scientifiques pour extraire des connaissances et des idées à partir de nombreuses données structurées ou non . Elle est souvent associée aux données massives et à l'analyse des données.

Elle utilise des techniques et des théories tirées de nombreux domaines dans le contexte des mathématiques, des statistiques, de l'informatique, de la théorie et des technologies de l'information, parmi lesquelles : l'apprentissage automatique, la compression de données et le calcul à haute performance.



https://fr.wikipedia.org/wiki/Science_des_donn%C3%A9es

https://towardsdatascience.com/introduction-to-statistics-e9d72d818745

Méthode d’apprentissage automatique : une illustration

Apprentissage Automatique Supervisé

Apprendre à l'ordinateur à discerner une peinture du mouvement:

- Impressionnisme,
- Surréalisme,
- Expressionisme abstrait.

Apprentissage Automatique Supervisé

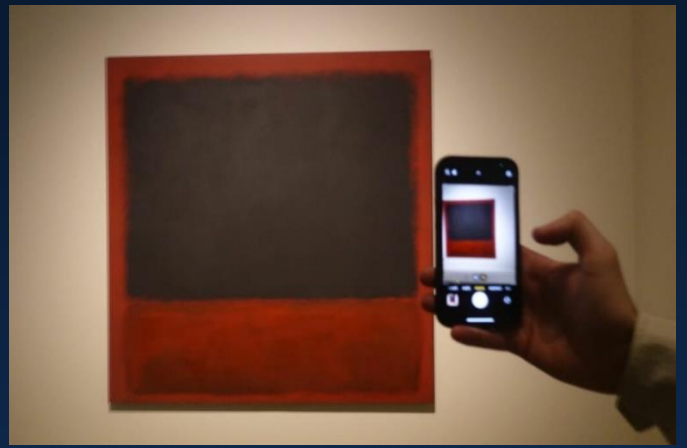
Apprendre à l'ordinateur à discerner une peinture du mouvement:

- Impressionnisme,
- Surréalisme,
- Expressionisme abstrait.

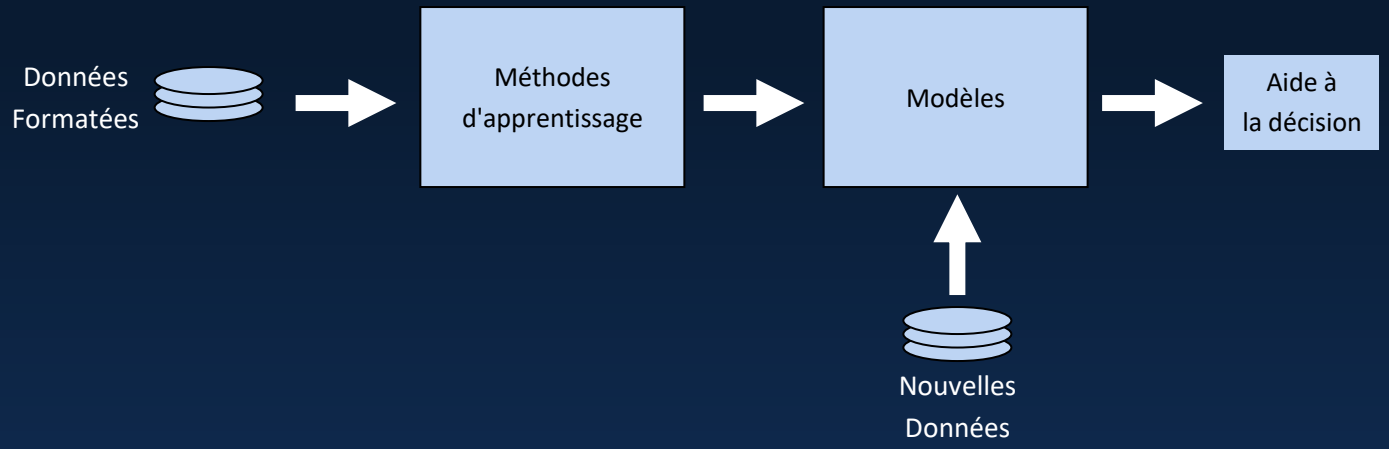
Apprentissage Automatique Supervisé

Apprendre à l'ordinateur à discerner une peinture du mouvement:

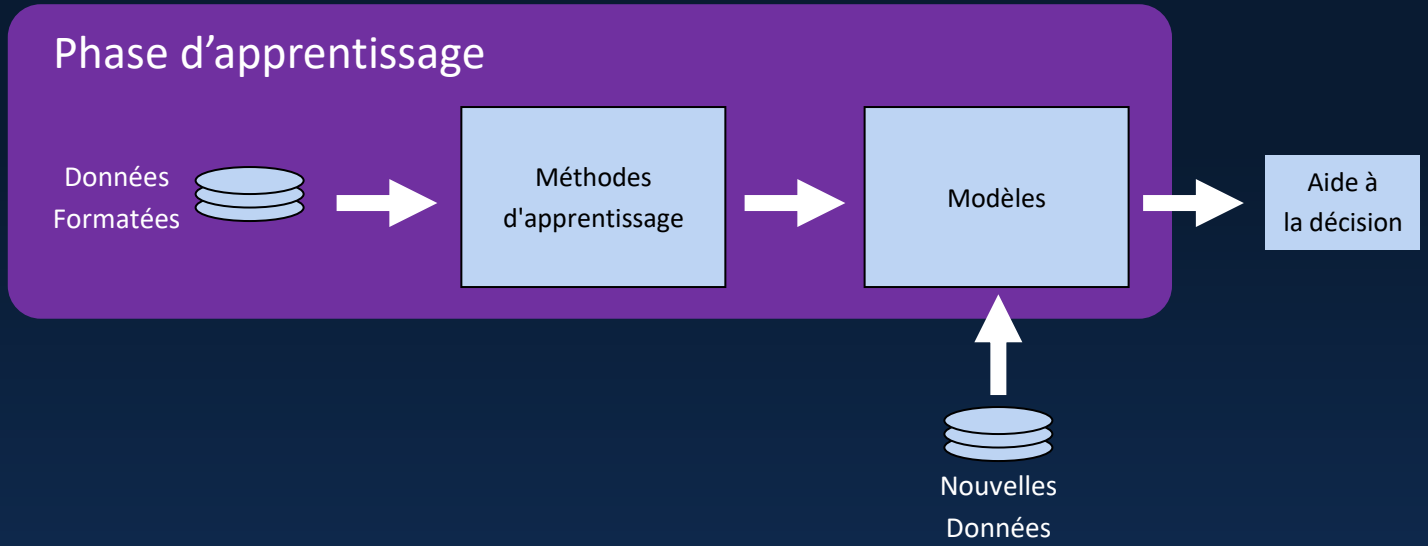
- Impressionnisme,
- Surréalisme,
- Expressionisme abstrait.



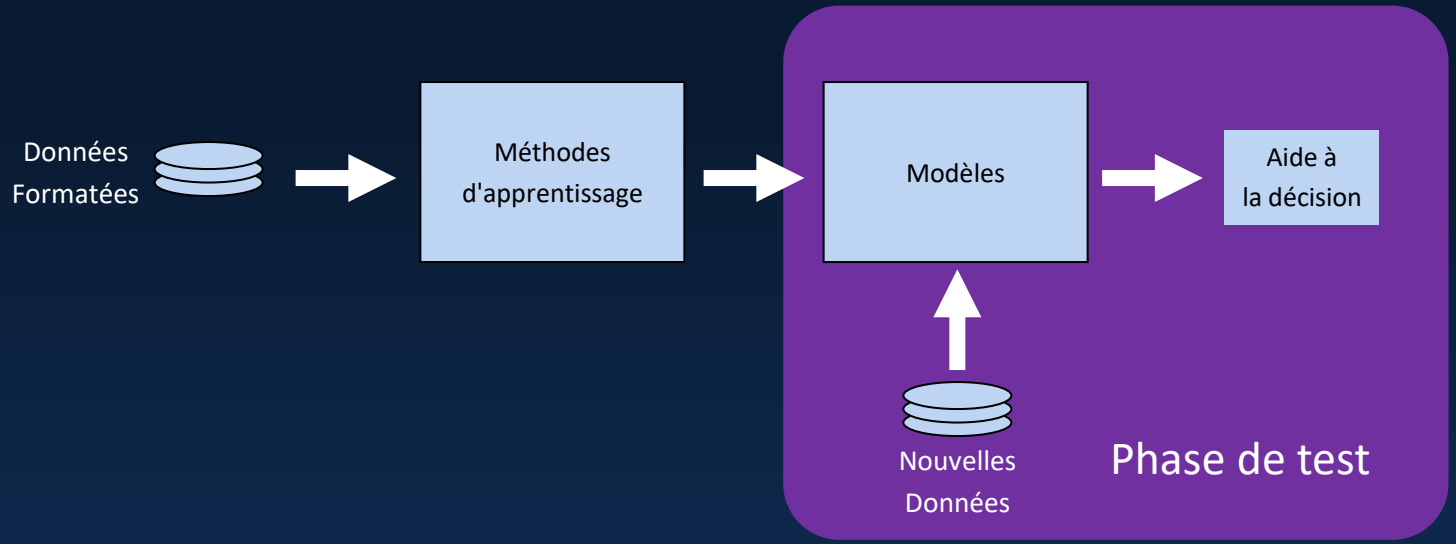
Processus d'induction à partir des données

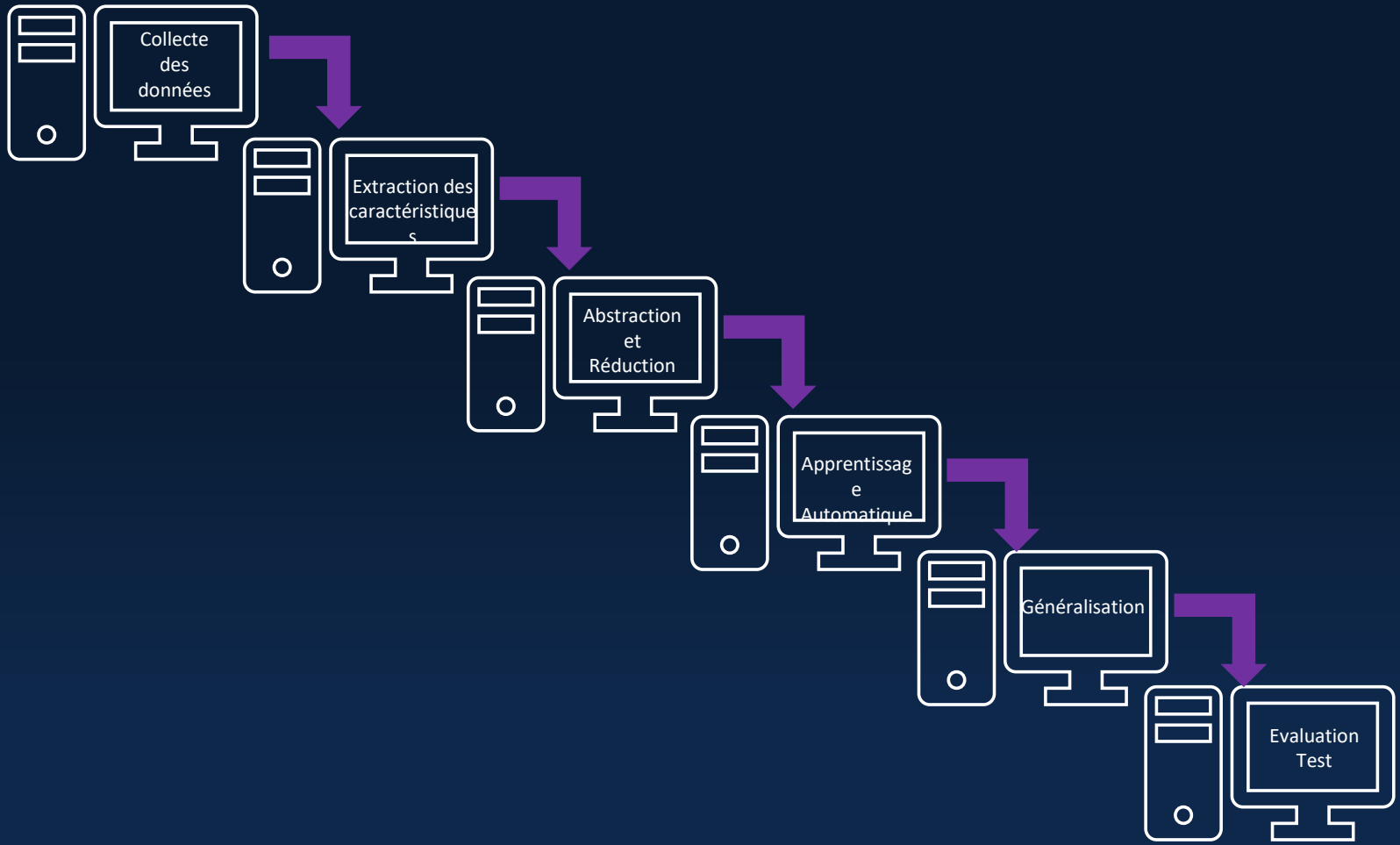


Processus d'induction à partir des données



Processus d'induction à partir des données




Collecte des données
Extraction des caractéristiques
Abstraction et Réduction
Apprentissage Automatique
Généralisation
Evaluation Test

Constitution de la base d’apprentissage

1. Impressionnisme (vers 1860-1890)

L'Impressionnisme est un mouvement qui met l’accent sur la lumière, la couleur et les sensations visuelles immédiates, souvent en plein air.
Artistes et œuvres associées :

Claude Monet – Impression, soleil levant (1872), Nymphéas (série, 1899-1926)
Pierre-Auguste Renoir – Bal du moulin de la Galette (1876), Déjeuner des canotiers (1881)
Edgar Degas – L'Étoile (1878), Les Repasseuses (1884)
Camille Pissarro – Boulevard Montmartre, effet de nuit (1897), Gelée blanche (1873)
Berthe Morisot – Le Berceau (1872), Jeune femme en toilette de bal (1879)
Alfred Sisley – Inondation à Port-Marly (1876), La Seine à Bougival (1876)
Gustave Caillebotte – Les Raboteurs de parquet (1875), Rue de Paris, temps de pluie (1877)
Mary Cassatt – La Toilette (1891), Le Thé (1880)
Frédéric Bazille – Atelier de Bazille (1870), La Réunion de famille (1867)
Eva Gonzalès – Nanny et Enfant (1877-78), Le Chignon (1877)

Constitution de la base d’apprentissage

2. Surréalisme (vers 1920-1960)

Le Surréalisme explore l’inconscient, les rêves et l’irrationnel, souvent avec une forte charge symbolique.
Artistes et œuvres associées :

Salvador Dalí – La Persistance de la mémoire (1931), Le Rêve (1931)
René Magritte – La Trahison des images (1929), L’Empire des lumières (1954)
Max Ernst – L’Éléphant Célèbes (1921), La Ville entière (1935-36)
Joan Miró – Le Carnaval d'Arlequin (1924-25), Bleu II (1961)
André Masson – Automatisme (1925), La Métamorphose (1929)
Yves Tanguy – Mama, Papa est blessé ! (1927), L’Extinction des espèces (1938)
Leonora Carrington – La Débutante (1937), The Lovers (1950)
Dorothea Tanning – Birthday (1942), Eine Kleine Nachtmusik (1943)
Victor Brauner – Le Surréaliste (1947), Le Grand Transparent (1947)
Frida Kahlo – Les Deux Fridas (1939), La Colonne brisée (1944)

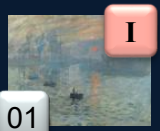
Constitution de la base d’apprentissage

3. Expressionnisme abstrait (vers 1940-1960)

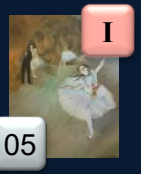
Ce mouvement, né aux États-Unis, privilégie l’expression spontanée des émotions à travers des formes abstraites et gestuelles.
Artistes et œuvres associées :

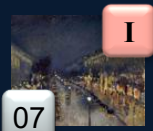
Jackson Pollock – Number 1A, 1948 (1948), Lavender Mist (1950)
Mark Rothko – Orange, Red, Yellow (1961), No. 61 (Rust and Blue) (1953)
Willem de Kooning – Woman I (1952), Excavation (1950)
Franz Kline – Mahoning (1956), Chief (1950)
Clyfford Still – PH-385 (1949), PH-1033 (1976)
Robert Motherwell – Elegy to the Spanish Republic No. 110 (1971), At Five in the Afternoon (1949)
Lee Krasner – The Seasons (1957), Gaea (1966)
Joan Mitchell – Hemlock (1956), City Landscape (1955)
Barnett Newman – Vir Heroicus Sublimis (1951), Onement VI (1953)
Ad Reinhardt – Abstract Painting (1966), Number 107 (1950)

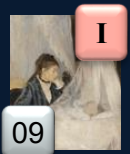
Constitution de la base d'apprentissage


I
01

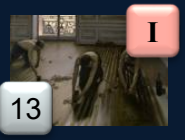

I
03

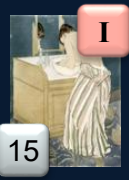

I
05


I
07


I
09


I
11


I
13


I
15

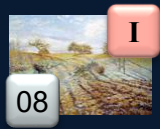

I
17

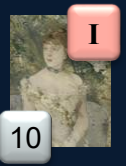

I
19

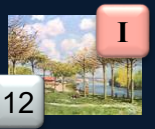

I
02


I
04

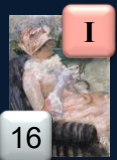

I
06


I
08

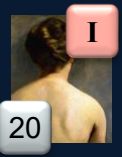

I
10

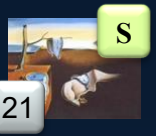

I
12


I
14

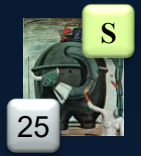

I
16

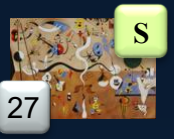

I
18

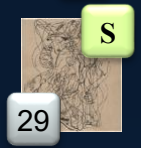

I
20

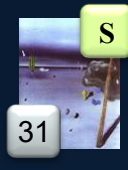

S
21


S
23
ceci n'est pas une pipe.

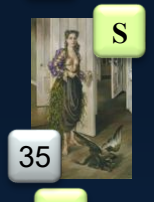

S
25

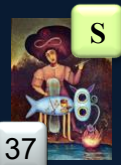

S
27


S
29

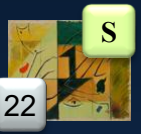

S
31

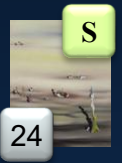

S
33

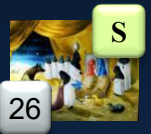

S
35


S
37

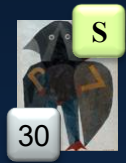

S
39

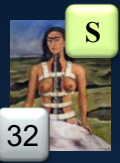

S
22

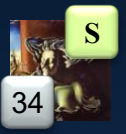

S
24

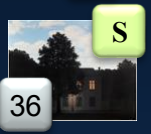

S
26

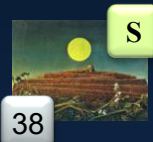

S
28

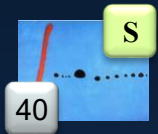

S
30


S
32


S
34


S
36


S
38


S
40

E
41

E
43

E
45

E
47

E
49

E
51

E
53

E
55

E
57

E
59

E
42

E
44

E
46

E
48

E
50

E
52

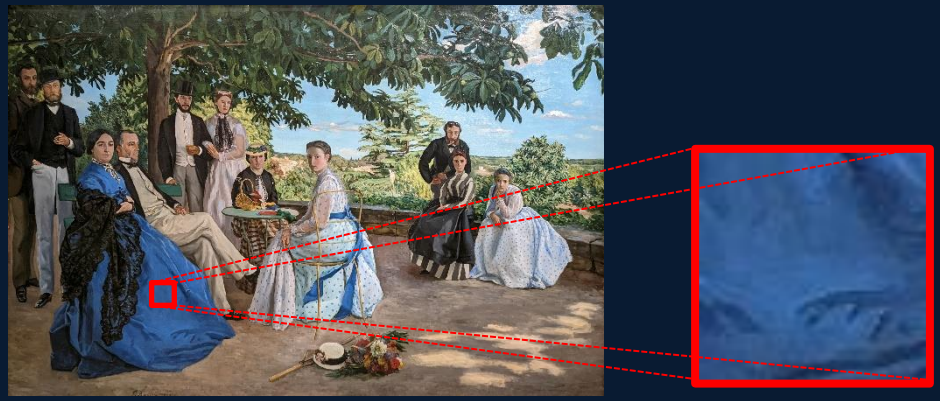
E
54

E
56

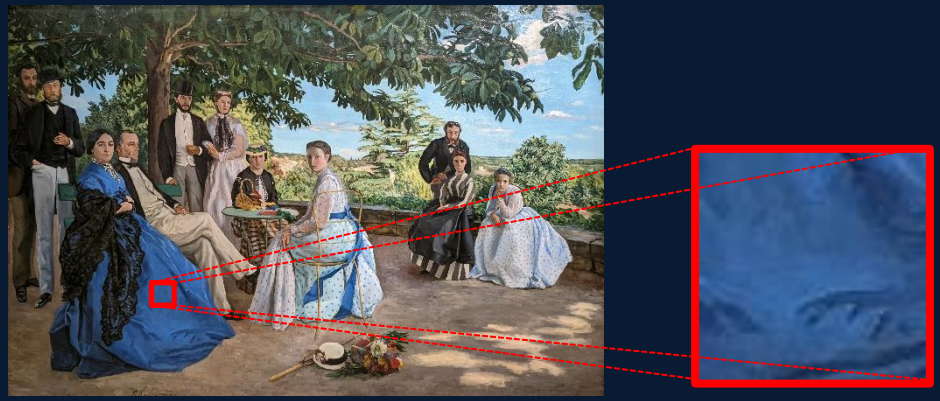
E
58

E
60

Numéro Descripteur	Descripteur	Type	Description	Interprétation	Exemple
1	Id	Métadonnée	Numéro d'Identification	Numéro d'identification unique	18
2	Style	Classe	Nom du Mouvement artistique du tableau {Impressionnisme, Surréalisme, Expressionnisme-Abstrait}	Classe à prédire	Impressionnisme
3	Annee	Métadonnée	Année d'achèvement du tableau		1867
4	Prenom_Nom_Titre	Métadonnée			Frederic_Bazille_La_Reunion_de_famille
5	No_Prenom_Nom_Titre_Annee	Métadonnée			18_Frederic_Bazille_ La_Reunion_de_famille_1867



Numéro Descripteur	Descripteur	Type	Description	Interprétation	Exemple
6	Moyenne_Rouge	Descripteur	Moyenne du canal rouge (0-1)	0 = Pas de rouge, 1 = Image très rouge	0.43
7	Moyenne_Vert	Descripteur	Moyenne du canal vert (0-1)	0 = Pas de vert, 1 = Image très verte	0.43
8	Moyenne_Bleu	Descripteur	Moyenne du canal bleu (0-1)	0 = Pas de bleu, 1 = Image très bleue	0.40
9	Saturation_Moyenne	Descripteur	Intensité moyenne des couleurs (0-1)	0 = Image en niveaux de gris, 1 = Couleurs très saturées	0.29
10	Variance_Saturation	Descripteur	Variation de la saturation (dispersion des couleurs)	Faible = Couleurs homogènes, élevée = Mélange de couleurs vives	0.03
11	Teinte_Moyenne	Descripteur	Couleur dominante dans l'image (0-360°)	0° = Rouge, 120° = Vert, 240° = Bleu	95.46
12	Contraste	Descripteur	Variation moyenne d'intensité entre pixels voisins	Faible = Image lisse, élevée = Image très contrastée	0.54
13	Entropie	Descripteur	Complexité de l'image (niveau de détail)	Faible = Image simple, élevée = Image très détaillée	7.76
14	Energie	Descripteur	Uniformité globale de l'image	Faible = texture très variée, élevée, texture très répétitive	252.03
15	GLCM_Contraste	Descripteur	Contraste basé sur la matrice de co-occurrence	Faible = Peu de variations de texture, élevée = Texture marquée	0.23
16	GLCM_Correlation	Descripteur	Corrélation entre pixels adjacents	Proche de 1 = Zones homogènes, proche de 0 = Texture désordonnée	0.97
17	GLCM_energie	Descripteur	Énergie (homogénéité) de la texture	Faible = Texture variée, élevée = Texture répétitive	0.11
18	GLCM_Homogeneite	Descripteur	Régularité des textures	Faible = Texture chaotique, élevée = Texture uniforme	0.91
19	Gabor_Moyenne	Descripteur	Réponse moyenne des filtres de Gabor (analyse de texture)	Faible = Texture faible, élevée = Texture marquée	701.31
20	Gabor_ecart-Type	Descripteur	Dispersion des réponses de Gabor	Faible = Texture homogène, élevée = Texture variée	1361.15
21	Nombre_Visages	Descripteur	Nombre de visages détectés	0 = Aucun visage, >0 = Nombre de visages trouvés	17.00
22	Densite_Contours	Descripteur	Nombre de contours détectés par Canny	Faible = Image floue, élevée = Image avec bords nets	0.10



Numéro Descripteur	Descripteur	Type	Description	Interprétation	Exemple
23	LBP_1	Descripteur	Local Binary Pattern 1	Mesure la texture locale	0.07
24	LBP_2	Descripteur	Local Binary Pattern 2	Mesure la texture locale	0.13
25	LBP_3	Descripteur	Local Binary Pattern 3	Mesure la texture locale	0.12
26	LBP_4	Descripteur	Local Binary Pattern 4	Mesure la texture locale	0.37
27	LBP_5	Descripteur	Local Binary Pattern 5	Mesure la texture locale	0.65
28	LBP_6	Descripteur	Local Binary Pattern 6	Mesure la texture locale	0.49
29	LBP_7	Descripteur	Local Binary Pattern 7	Mesure la texture locale	0.24
30	LBP_8	Descripteur	Local Binary Pattern 8	Mesure la texture locale	0.17
31	LBP_9	Descripteur	Local Binary Pattern 9	Mesure la texture locale	0.18
32	LBP_10	Descripteur	Local Binary Pattern 10	Mesure la texture locale	0.22
33	CouleurDominante_1_R	Descripteur	Composante rouge de la couleur dominante 1	0-1 : Plus proche de 1 = Plus de rouge	0.73
34	CouleurDominante_1_G	Descripteur	Composante verte de la couleur dominante 1	0-1 : Plus proche de 1 = Plus de vert	0.77
35	CouleurDominante_1_B	Descripteur	Composante bleue de la couleur dominante 1	0-1 : Plus proche de 1 = Plus de bleu	0.76
36	CouleurDominante_2_R	Descripteur	Composante rouge de la couleur dominante 2	0-1 : Plus proche de 1 = Plus de rouge	0.20
37	CouleurDominante_2_G	Descripteur	Composante verte de la couleur dominante 2	0-1 : Plus proche de 1 = Plus de vert	0.21
38	CouleurDominante_2_B	Descripteur	Composante bleue de la couleur dominante 2	0-1 : Plus proche de 1 = Plus de bleu	0.17
39	CouleurDominante_3_R	Descripteur	Composante rouge de la couleur dominante 3	0-1 : Plus proche de 1 = Plus de rouge	0.47
40	CouleurDominante_3_G	Descripteur	Composante verte de la couleur dominante 3	0-1 : Plus proche de 1 = Plus de vert	0.45
41	CouleurDominante_3_B	Descripteur	Composante bleue de la couleur dominante 3	0-1 : Plus proche de 1 = Plus de bleu	0.41

Numéro Descripteur	Descripteur	Type	Description	Interprétation	Exemple
42	HistR_1	Descripteur	Bin 1 de l'histogramme du canal rouge	0-1 : Distribution des pixels rouges dans l'image	0.08
43	HistG_1	Descripteur	Bin 1 de l'histogramme du canal vert	0-1 : Distribution des pixels verts dans l'image	0.38
44	HistB_1	Descripteur	Bin 1 de l'histogramme du canal bleu	0-1 : Distribution des pixels bleus dans l'image	0.31
45	HistGray_1	Descripteur	Bin 1 de l'histogramme des niveaux de gris	0-1 : Distribution des niveaux de gris	0.22
46	HistR_2	Descripteur	Bin 2 de l'histogramme du canal rouge	0-1 : Distribution des pixels rouges dans l'image	0.02
47	HistG_2	Descripteur	Bin 2 de l'histogramme du canal vert	0-1 : Distribution des pixels verts dans l'image	0.08
48	HistB_2	Descripteur	Bin 2 de l'histogramme du canal bleu	0-1 : Distribution des pixels bleus dans l'image	0.39
49	HistGray_2	Descripteur	Bin 2 de l'histogramme des niveaux de gris	0-1 : Distribution des niveaux de gris	0.31
50	HistR_3	Descripteur	Bin 3 de l'histogramme du canal rouge	0-1 : Distribution des pixels rouges dans l'image	0.20
51	HistG_3	Descripteur	Bin 3 de l'histogramme du canal vert	0-1 : Distribution des pixels verts dans l'image	0.03
52	HistB_3	Descripteur	Bin 3 de l'histogramme du canal bleu	0-1 : Distribution des pixels bleus dans l'image	0.12
53	HistGray_3	Descripteur	Bin 3 de l'histogramme des niveaux de gris	0-1 : Distribution des niveaux de gris	0.40
54	HistR_4	Descripteur	Bin 4 de l'histogramme du canal rouge	0-1 : Distribution des pixels rouges dans l'image	0.29
55	HistG_4	Descripteur	Bin 4 de l'histogramme du canal vert	0-1 : Distribution des pixels verts dans l'image	0.11
56	HistB_4	Descripteur	Bin 4 de l'histogramme du canal bleu	0-1 : Distribution des pixels bleus dans l'image	0.08
57	HistGray_4	Descripteur	Bin 4 de l'histogramme des niveaux de gris	0-1 : Distribution des niveaux de gris	0.08
58	HistR_5	Descripteur	Bin 5 de l'histogramme du canal rouge	0-1 : Distribution des pixels rouges dans l'image	0.40
59	HistG_5	Descripteur	Bin 5 de l'histogramme du canal vert	0-1 : Distribution des pixels verts dans l'image	0.30
60	HistB_5	Descripteur	Bin 5 de l'histogramme du canal bleu	0-1 : Distribution des pixels bleus dans l'image	0.21
61	HistGray_5	Descripteur	Bin 5 de l'histogramme des niveaux de gris	0-1 : Distribution des niveaux de gris	0.02

01	02	03	04	05
18	Impressionnisme	1867	Frederic_Bazille_La_Reunion_de_famille	18_Frederic_Bazille_La_Reunion_de_famille_1867

06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
0.43	0.43	0.40	0.29	0.03	95.46	0.54	7.76	252	0.23	0.97	0.11	0.91	701	1361	17.00	0.10	0.07	0.13	0.12	0.37	0.65	0.49	0.24	0.17	0.18	0.22	0.73

34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
0.77	0.76	0.20	0.21	0.17	0.47	0.45	0.41	0.08	0.38	0.31	0.22	0.02	0.08	0.39	0.31	0.20	0.03	0.12	0.40	0.29	0.11	0.08	0.08	0.40	0.30	0.21	0.02

01	02	03	04	05
18	Impressionnisme	1867	Frederic_Bazille_La_Reunion_de_famille	18_Frederic_Bazille_La_Reunion_de_famille_1867

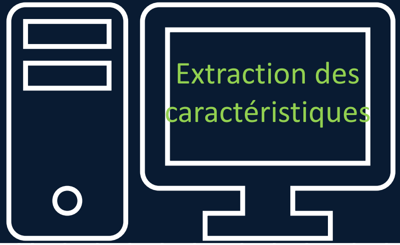
06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
0.43	0.43	0.40	0.29	0.03	95.46	0.54	7.76	252	0.23	0.97	0.11	0.91	701	1361	17.00	0.10	0.07	0.13	0.12	0.37	0.65	0.49	0.24	0.17	0.18	0.22	0.73

34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
0.77	0.76	0.20	0.21	0.17	0.47	0.45	0.41	0.08	0.38	0.31	0.22	0.02	0.08	0.39	0.31	0.20	0.03	0.12	0.40	0.29	0.11	0.08	0.08	0.40	0.30	0.21	0.02

02
Impressionnisme

06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
0.43	0.43	0.40	0.29	0.03	95.46	0.54	7.76	252	0.23	0.97	0.11	0.91	701	1361	17.00	0.10	0.07	0.13	0.12	0.37	0.65	0.49	0.24	0.17	0.18	0.22	0.73

34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
0.77	0.76	0.20	0.21	0.17	0.47	0.45	0.41	0.08	0.38	0.31	0.22	0.02	0.08	0.39	0.31	0.20	0.03	0.12	0.40	0.29	0.11	0.08	0.08	0.40	0.30	0.21	0.02

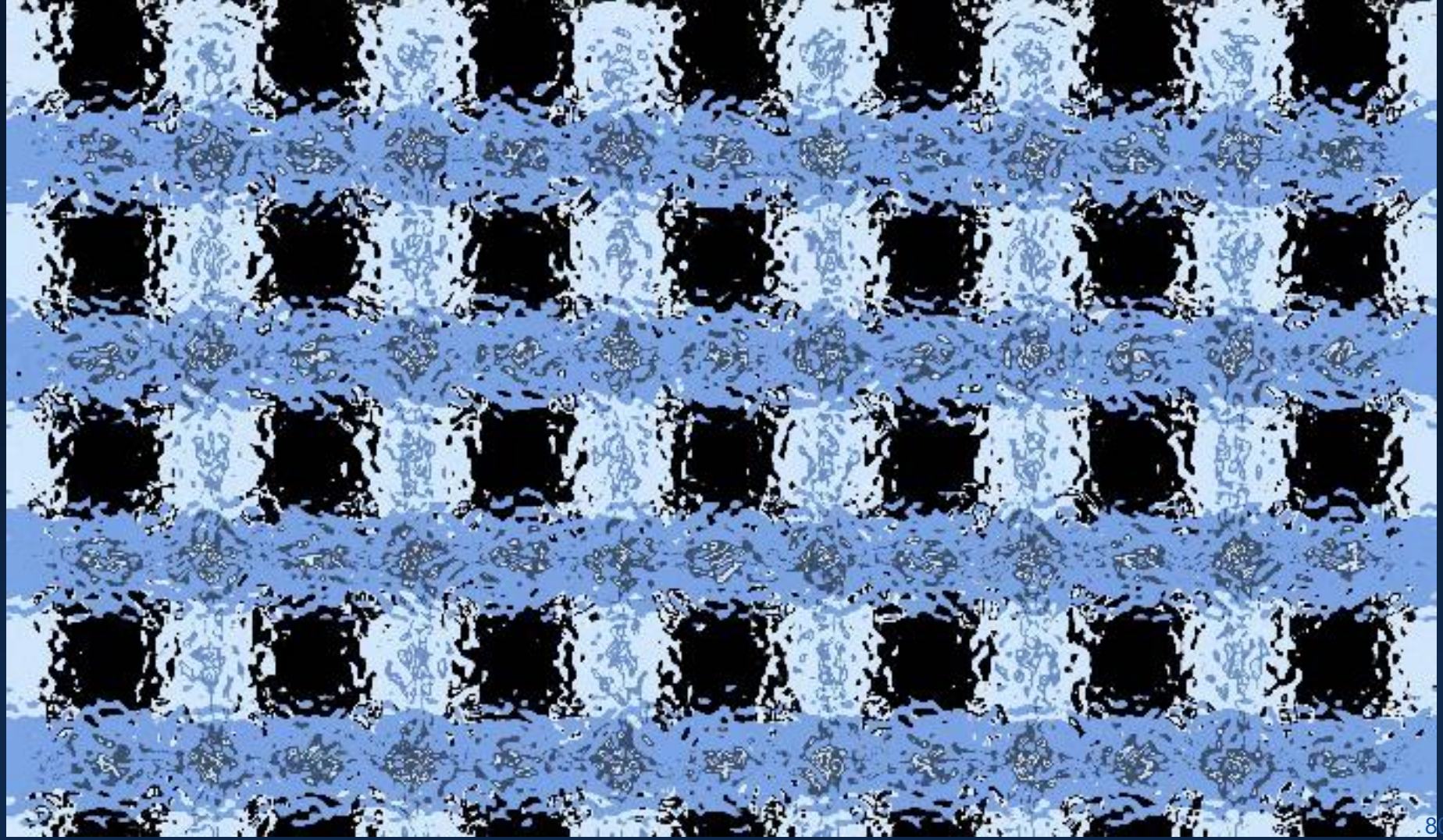


Ensemble de Tableaux d'Apprentissage :
Matrice de 60 lignes et 56 colonnes de descripteurs et 1 colonne de Classe

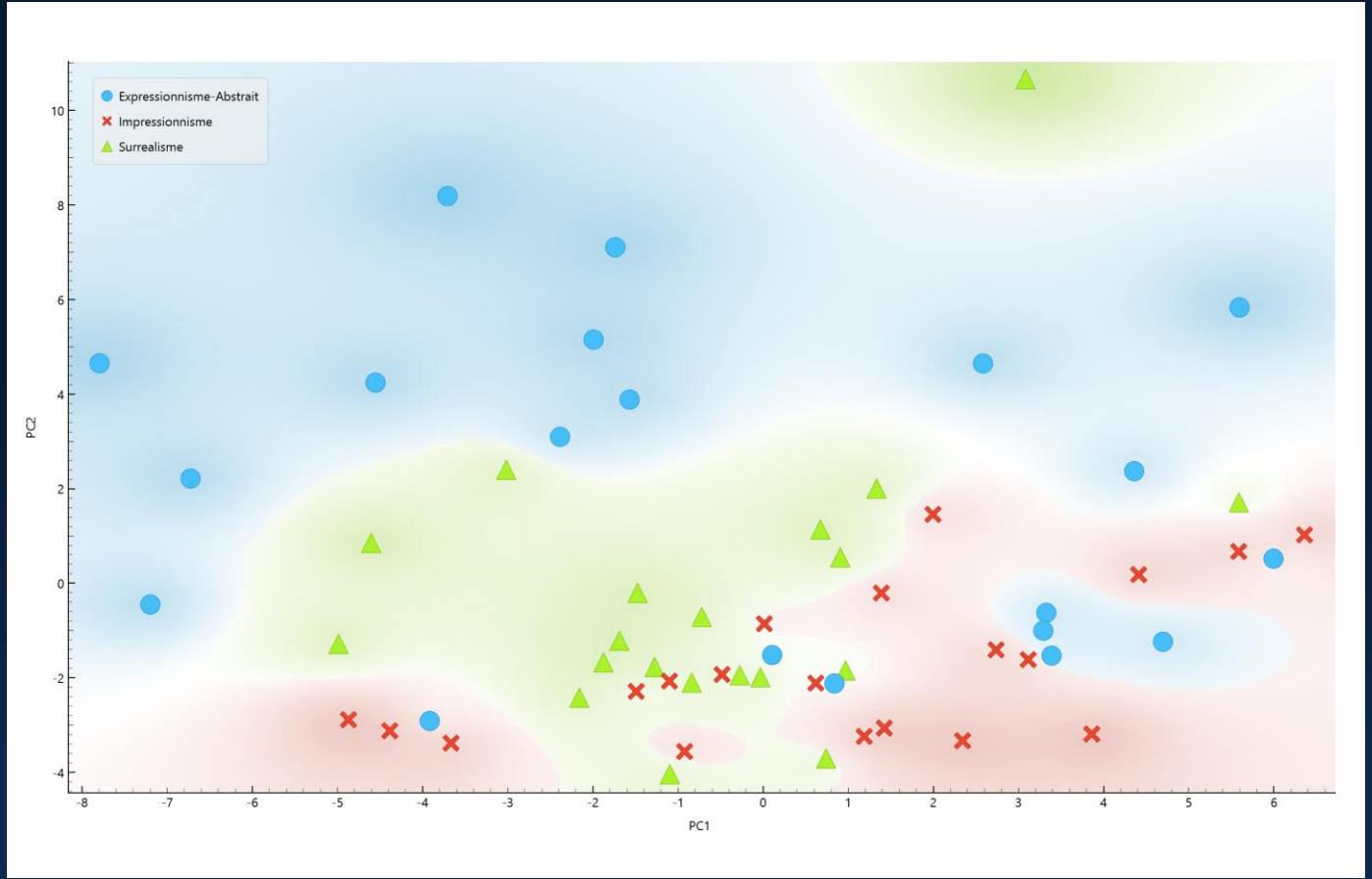
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	I	0.53	0.57	0.54	0.15	0.01	136.37	0.15	6.41	255.00	0.29	0.72	0.27	0.87	332	508	0.00	0.19	0.25	0.31	0.21	0.26	0.31	0.28	0.25	0.31	0.32	0.54
2	I	0.38	0.32	0.19	0.49	0.04	79.83	0.44	6.89	254.91	0.18	0.94	0.23	0.91	661	1219	0.00	0.13	0.04	0.10	0.12	0.39	0.71	0.48	0.21	0.11	0.10	0.14
3	I	0.40	0.39	0.36	0.24	0.02	127.38	0.52	7.57	255.00	0.29	0.95	0.11	0.88	609	1092	5.00	0.12	0.22	0.25	0.23	0.33	0.40	0.33	0.26	0.27	0.30	0.47
4	I	0.47	0.41	0.35	0.30	0.03	66.63	0.51	7.57	255.00	0.12	0.98	0.13	0.94	348	644	11.00	0.11	0.06	0.13	0.12	0.38	0.59	0.54	0.26	0.17	0.22	0.20
5	I	0.53	0.51	0.41	0.27	0.03	65.43	0.26	6.96	255.00	0.13	0.94	0.23	0.94	248	488	0.00	0.13	0.13	0.26	0.17	0.35	0.39	0.42	0.32	0.28	0.31	0.40
6	I	0.51	0.45	0.40	0.23	0.02	47.06	0.26	6.87	254.90	0.21	0.89	0.19	0.90	410	742	0.00	0.14	0.27	0.29	0.21	0.25	0.27	0.27	0.25	0.30	0.35	0.56
7	I	0.30	0.28	0.29	0.32	0.04	160.06	0.54	7.16	253.01	0.33	0.90	0.17	0.88	556	1028	0.00	0.12	0.12	0.22	0.17	0.37	0.49	0.48	0.28	0.26	0.27	0.31
8	I	0.68	0.66	0.59	0.25	0.05	114.68	0.33	7.65	253.42	1.47	0.76	0.05	0.67	697	1072	0.00	0.15	0.41	0.30	0.13	0.11	0.10	0.11	0.14	0.30	0.42	0.64
9	I	0.64	0.60	0.53	0.19	0.01	41.93	0.37	7.52	254.37	0.85	0.87	0.07	0.71	381	549	1.00	0.23	0.40	0.30	0.14	0.11	0.10	0.11	0.14	0.30	0.41	0.65
10	I	0.53	0.52	0.43	0.21	0.01	57.79	0.27	7.15	254.89	0.25	0.91	0.14	0.88	318	502	2.00	0.19	0.20	0.28	0.24	0.35	0.37	0.36	0.30	0.28	0.27	0.45
11	I	0.63	0.61	0.54	0.16	0.01	61.30	0.21	6.86	255.00	0.21	0.90	0.24	0.90	487	1054	0.00	0.11	0.14	0.28	0.18	0.35	0.45	0.41	0.29	0.28	0.25	0.38
12	I	0.68	0.67	0.57	0.29	0.03	110.13	0.29	7.55	254.97	0.14	0.97	0.15	0.93	652	1241	6.00	0.10	0.03	0.08	0.08	0.34	0.72	0.53	0.20	0.10	0.11	0.11
13	I	0.30	0.27	0.22	0.35	0.04	46.28	0.58	7.15	251.03	0.08	0.98	0.19	0.96	252	574	9.00	0.12	0.07	0.16	0.12	0.37	0.53	0.52	0.28	0.21	0.27	0.26
14	I	0.54	0.56	0.53	0.19	0.02	134.23	0.59	7.59	239.09	0.28	0.98	0.10	0.91	641	1516	0.00	0.11	0.16	0.23	0.18	0.33	0.45	0.38	0.26	0.25	0.38	0.40
15	I	0.62	0.61	0.53	0.23	0.01	88.98	0.36	7.57	254.86	0.43	0.93	0.09	0.85	712	1546	0.00	0.10	0.28	0.32	0.17	0.23	0.31	0.24	0.20	0.33	0.35	0.56
16	I	0.52	0.49	0.47	0.15	0.01	132.13	0.49	7.77	254.87	0.51	0.93	0.07	0.81	438	601	0.00	0.20	0.25	0.32	0.21	0.28	0.33	0.29	0.25	0.32	0.31	0.51
17	I	0.42	0.42	0.44	0.18	0.02	165.58	0.50	7.64	254.85	0.22	0.96	0.11	0.90	354	981	2.00	0.11	0.20	0.29	0.21	0.33	0.34	0.34	0.28	0.30	0.32	0.47
18	I	0.43	0.43	0.40	0.29	0.03	95.46	0.54	7.76	252.03	0.23	0.97	0.11	0.91	701	1361	17.00	0.10	0.07	0.13	0.12	0.37	0.65	0.49	0.24	0.17	0.18	0.22
19	I	0.34	0.33	0.26	0.27	0.01	62.08	0.48	7.12	254.97	0.23	0.93	0.17	0.89	262	464	1.00	0.13	0.31	0.32	0.17	0.20	0.21	0.20	0.20	0.32	0.39	0.60
20	I	0.49	0.41	0.29	0.42	0.04	50.24	0.62	7.77	245.46	0.27	0.97	0.09	0.87	221	361	0.00	0.16	0.35	0.31	0.15	0.13	0.11	0.13	0.16	0.31	0.42	0.65
21	S	0.54	0.47	0.42	0.33	0.04	146.25	0.56	7.21	255.00	0.17	0.98	0.16	0.93	447	1169	0.00	0.09	0.08	0.18	0.13	0.35	0.49	0.49	0.27	0.23	0.36	0.28
22	S	0.32	0.25	0.21	0.46	0.03	154.37	0.86	7.02	254.54	0.09	0.99	0.22	0.96	353	1224	4.00	0.05	0.02	0.08	0.04	0.30	0.58	0.61	0.23	0.18	0.32	0.13
23	S	0.80	0.78	0.74	0.14	0.04	59.20	0.38	4.38	247.34	0.83	0.93	0.55	0.93	863	2398	0.00	0.05	0.04	0.05	0.01	0.04	0.11	0.14	0.04	0.06	0.97	0.09
24	S	0.36	0.37	0.38	0.11	0.01	155.38	0.98	6.67	234.91	0.30	0.98	0.28	0.95	515	1840	0.00	0.06	0.13	0.25	0.11	0.29	0.29	0.43	0.28	0.30	0.45	0.44
25	S	0.37	0.42	0.37	0.23	0.02	119.25	0.60	7.75	250.08	0.45	0.94	0.08	0.83	747	1726	0.00	0.12	0.26	0.31	0.20	0.26	0.31	0.27	0.24	0.31	0.33	0.54
26	S	0.33	0.36	0.27	0.43	0.04	112.72	0.50	7.38	253.29	0.83	0.79	0.08	0.72	783	1096	0.00	0.27	0.13	0.23	0.24	0.45	0.52	0.41	0.25	0.24	0.17	0.29
27	S	0.57	0.45	0.33	0.48	0.06	56.27	0.39	7.32	254.81	0.23	0.95	0.15	0.93	608	1424	1.00	0.07	0.07	0.14	0.06	0.26	0.30	0.52	0.18	0.25	0.64	0.21
28	S	0.36	0.65	0.91	0.62	0.01	203.03	0.14	5.86	254.70	0.12	0.90	0.39	0.94	321	490	2.00	0.18	0.02	0.07	0.08	0.30	0.54	0.68	0.21	0.10	0.28	0.09
29	S	0.69	0.62	0.53	0.23	0.00	35.94	0.27	6.22	255.00	0.50	0.86	0.47	0.89	963	2082	5.00	0.11	0.31	0.31	0.17	0.20	0.26	0.21	0.20	0.31	0.36	0.59
30	S	0.54	0.47	0.21	0.64	0.06	60.98	0.39	7.18	252.40	0.27	0.94	0.15	0.89	415	1088	1.00	0.13	0.26	0.29	0.18	0.24	0.23	0.27	0.24	0.31	0.41	0.56
31	S	0.55	0.55	0.67	0.28	0.01	218.07	0.42	7.69	254.86	0.14	0.98	0.13	0.94	457	1046	1.00	0.13	0.02	0.07	0.06	0.31	0.65	0.61	0.21	0.12	0.17	0.10
32	S	0.60	0.58	0.52	0.15	0.01	55.20	0.34	7.02	254.82	0.08	0.98	0.20	0.97	219	532	0.00	0.09	0.03	0.08	0.05	0.22	0.19	0.51	0.16	0.17	0.76	0.11
33	S	0.48	0.42	0.35	0.34	0.05	66.95	0.30	6.90	251.79	0.14	0.94	0.24	0.95	395	1007	1.00	0.10	0.08	0.20	0.08	0.28	0.43	0.51	0.22	0.26	0.46	0.31
34	S	0.51	0.47	0.26	0.64	0.09	87.40	0.54	7.86	243.13	0.26	0.97	0.10	0.91	734	1512	0.00	0.10	0.08	0.18	0.11	0.31	0.66	0.47	0.22	0.20	0.22	0.25
35	S	0.44	0.43	0.34	0.26	0.01	64.44	0.40	7.25	255.00	0.11	0.97	0.17	0.95	363	932	4.00	0.08	0.07	0.16	0.09	0.31	0.50	0.56	0.24	0.23	0.35	0.26
36	S	0.30	0.24	0.16	0.44	0.03	69.34	0.40	6.39	255.00	0.10	0.92	0.39	0.95	338	847	0.00	0.09	0.06	0.18	0.09	0.30	0.47	0.54	0.24	0.23	0.42	0.26
37	S	0.38	0.31	0.29	0.49	0.04	175.35	0.66	7.56	251.44	0.49	0.92	0.10	0.85	671	1442	0.00	0.13	0.20	0.30	0.19	0.31	0.44	0.33	0.24	0.30	0.30	0.44
38	S	0.42	0.41	0.40	0.22	0.05	122.15	0.66	6.89	252.12	0.25	0.98	0.19	0.92	605	1581	0.00	0.09	0.10	0.20	0.14	0.36	0.53	0.47	0.29	0.23	0.25	0.30
39	S	0.52	0.49	0.42	0.32	0.08	86.90	0.45	7.65	250.72	0.35	0.94	0.10	0.88	542	1096	2.00	0.10	0.31	0.28	0.16	0.21	0.27	0.23	0.19	0.29	0.41	0.58
40	S	0.47	0.43	0.40	0.39	0.05	117.44	0.50	7.79	249.69	0.97	0.84	0.05	0.70	796	1092	0.00	0.25	0.19	0.25	0.25	0.39	0.46	0.37	0.26	0.25	0.22	0.39
41	E	0.59	0.54	0.48	0.19	0.02	89.32	0.40	7.52	255.00	1.32	0.79	0.07	0.72	1473	1967	0.00	0.17	0.29	0.28	0.19	0.28	0.36	0.29	0.21	0.28	0.33	0.53
42	E	0.53	0.45	0.42	0.26	0.02	41.60	0.45	7.75	251.54	0.49	0.92	0.07	0.81	1447	2005	3.00	0.16	0.07	0.13	0.16	0.44	0.64	0.49	0.21	0.13	0.09	0.16
43	E	0.80	0.37	0.21	0.74	0.05	45.49	0.24	5.13	255.00	0.08	0.94	0.73	0.98	278	1413	1.00	0.05	0.01	0.03	0.01	0.10	0.09	0.30	0.06	0.08	0.94	0.06
44	E	0.20	0.17	0.33	0.60	0.02	274.63	0.45	6.10	254.76	0.05	0.95	0.32	0.97	190	445	0.00	0.13	0.01	0.09	0.03	0.26	0.35	0.69	0.24	0.19	0.47	0.14
45	E	0.58	0.56	0.48	0.25	0.04	75.87	0.28	7.31	255.00	0.37	0.88	0.12	0.85	608	1063	1.00	0.15	0.24	0.30	0.18	0.27	0.41	0.31	0.23	0.30	0.31	0.50
46	E	0.64	0.59	0.41	0.43	0.06	82.83	0.47	7.81	244.20	3.27	0.66	0.04	0.62	1402	2304	0.00	0.17	0.34	0.33	0.16	0.19	0.22	0.20	0.19	0.33	0.35	0.60
47	E	0.46	0.46	0.45	0.06	0.00	121.73	0.83	6.50	253.91	0.31	0.98	0.24	0.92	432	1160	0.00	0.08	0.17	0.25	0.10	0.21	0.19	0.30	0.20	0.30	0.61	0.48
48	E	0.43	0.42	0.40	0.22	0.04	62.89	0.91	6.21	254.57	0.28	0.99	0.26	0.92	356	1081	0.00	0.07	0.17	0.30	0.12	0.24	0.22	0.31	0.24	0.33	0.48	0.51
49	E	0.56	0.25	0.26	0.55	0.04	189.49	0.24	6.25	254.99	0.10	0.90	0.37	0.95	218	381	0.00	0.18	0.09	0.20	0.14	0.37	0.44	0.53	0.32	0.22	0.27	0.31
50	E	0.83	0.64	0.53	0.37	0.07	30.10	0.28	5.68	255.00	0.11	0.97	0.45	0.97	432	1082	0.00	0.07	0.07	0.13	0.05	0.15	0.19	0.32	0.13	0.17	0.85	0.22
51	E	0.31	0.30	0.29	0.08	0.01	82.72	1.29	5.78	164.27	0.27	0.99	0.49	0.97	393	1650	0.00	0.03	0.16	0.29	0.09	0.23	0.20	0.31	0.25	0.32	0.52	0.51
52	E	0.33	0.33	0.33	0.25	0.01	168.54	1.04	4.25	254.99	0.13	0.99	0.44	0.96	336	1473	0.00	0.04	0.02	0.04	0.02	0.06	0.08	0.16	0.05	0.06	0.98	0.06
53	E	0.70	0.58	0.51	0.34	0.03	91.80	0.34	7.41	255.00	0.62	0.89	0.11	0.83	1021	1924	0.00	0.12	0.27	0.28	0.16	0.25	0.35	0.28	0.20	0.30	0.37	0.54
54	E	0.52	0.44	0.44	0.25	0.02	187.75	0.48	7.21	255.00	0.36	0.94	0.15	0.87	705	1309	0.00	0.10	0.21	0.28	0.17	0.28	0.37	0.32	0.25	0.30	0.36	0.49
55	E	0.65	0.67	0.64	0.15	0.04	91.70	0.43	7.34	254.51	0.77	0.93	0.12	0.82	1040	1959	5.00	0.10	0.23	0.28	0.16	0.26	0.35	0.33	0.24	0.30	0.39	0.49
56	E	0.65	0.57	0.53	0.27	0.05	120.16	0.39	7.71	254.97	0.86	0.87	0.07	0.78	1099	1752	0.00	0.13	0.29	0.31	0.17	0.24	0.31	0.26	0.19	0.31	0.36	0.56
57	E	0.73	0.20	0.18	0.75	0.00	27.75	0.12	4.15	255.00	0.03	0.92	0.85	0.99	132	755	0.00	0.02	0.00	0.04	0.00	0.10	0.02	0.28	0.03	0.11	0.94	0.07
58	E	0.06	0.10	0.39	0.87	0.01	232.13	0.84	3.98	254.96	0.11	0.93	0.71	0.95	89	577	0.00	0.03	0.16	0.28	0.13	0.25	0.20	0.31	0.27	0.31	0.51	0.50
59	E	0.11	0.10	0.11	0.09	0.01	151.37	0.18	3.87	249.46	0.04	0.78	0.81	0.98	49	143	0.00	0.03	0.10	0.16	0.03	0.16	0.07	0.29	0.17	0.27	0.78	0.38
60	E	0.87	0.87	0.84	0.04	0.00	62.16	0.04	5.13	255.00	0.07	0.86	0.44	0.97	238	305	0.00	0.14	0.07	0.25	0.08	0.25	0.31	0.48	0.27	0.29	0.46	0.41

	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
1	0.64	0.65	0.59	0.39	0.45	0.46	0.51	0.56	0.54	0.00	0.07	0.77	0.16	0.00	0.00	0.02	0.76	0.22	0.00	0.00	0.02	0.88	0.09	0.00	0.00	0.02	0.81	0.17	0.00
2	0.41	0.35	0.25	0.75	0.63	0.26	0.27	0.22	0.15	0.01	0.65	0.23	0.10	0.01	0.02	0.74	0.17	0.06	0.00	0.25	0.70	0.05	0.00	0.00	0.01	0.75	0.18	0.06	0.00
3	0.48	0.46	0.40	0.75	0.71	0.64	0.19	0.21	0.22	0.10	0.41	0.30	0.17	0.03	0.05	0.45	0.36	0.12	0.02	0.04	0.56	0.31	0.09	0.01	0.05	0.45	0.35	0.13	0.02
4	0.77	0.69	0.62	0.22	0.20	0.18	0.55	0.44	0.36	0.09	0.29	0.29	0.28	0.04	0.03	0.45	0.35	0.16	0.01	0.05	0.55	0.29	0.11	0.00	0.05	0.40	0.36	0.18	0.02
5	0.69	0.72	0.69	0.53	0.52	0.42	0.44	0.38	0.25	0.00	0.05	0.75	0.20	0.00	0.00	0.15	0.64	0.21	0.00	0.01	0.46	0.36	0.17	0.00	0.00	0.13	0.67	0.20	0.00
6	0.64	0.58	0.53	0.36	0.31	0.28	0.53	0.44	0.37	0.00	0.19	0.59	0.22	0.00	0.00	0.32	0.62	0.06	0.00	0.00	0.46	0.51	0.02	0.00	0.00	0.28	0.65	0.08	0.00
7	0.17	0.17	0.23	0.61	0.55	0.45	0.39	0.35	0.30	0.14	0.55	0.25	0.05	0.01	0.12	0.63	0.22	0.02	0.00	0.03	0.81	0.15	0.01	0.00	0.11	0.63	0.23	0.02	0.00
8	0.83	0.85	0.86	0.39	0.29	0.20	0.66	0.62	0.49	0.01	0.07	0.25	0.54	0.14	0.03	0.09	0.27	0.44	0.17	0.06	0.19	0.28	0.25	0.22	0.02	0.09	0.28	0.45	0.16
9	0.63	0.58	0.51	0.23	0.20	0.17	0.82	0.78	0.70	0.03	0.15	0.16	0.57	0.10	0.04	0.14	0.24	0.54	0.04	0.06	0.14	0.40	0.39	0.00	0.04	0.14	0.23	0.55	0.04
10	0.70	0.68	0.58	0.35	0.36	0.27	0.53	0.52	0.42	0.00	0.18	0.49	0.33	0.00	0.00	0.17	0.55	0.28	0.00	0.00	0.38	0.55	0.07	0.00	0.00	0.19	0.54	0.27	0.00
11	0.40	0.38	0.32	0.71	0.71	0.64	0.60	0.58	0.49	0.00	0.06	0.33	0.60	0.01	0.00	0.07	0.37	0.56	0.00	0.00	0.15	0.54	0.31	0.00	0.00	0.07	0.38	0.55	0.00
12	0.42	0.40	0.25	0.84	0.84	0.86	0.67	0.64	0.43	0.00	0.07	0.30	0.46	0.17	0.00	0.08	0.28	0.50	0.12	0.01	0.35	0.20	0.19	0.25	0.00	0.09	0.32	0.46	0.13
13	0.16	0.14	0.08	0.57	0.52	0.46	0.33	0.31	0.24	0.16	0.55	0.27	0.01	0.01	0.18	0.56	0.24	0.01	0.00	0.33	0.50	0.15	0.01	0.00	0.18	0.56	0.24	0.02	0.01
14	0.85	0.87	0.81	0.12	0.14	0.16	0.52	0.53	0.51	0.19	0.15	0.17	0.28	0.21	0.16	0.18	0.16	0.27	0.23	0.13	0.22	0.17	0.34	0.14	0.16	0.18	0.16	0.28	0.22
15	0.91	0.86	0.78	0.33	0.38	0.37	0.70	0.66	0.51	0.00	0.26	0.20	0.32	0.21	0.00	0.17	0.31	0.41	0.11	0.01	0.18	0.52	0.25	0.04	0.00	0.21	0.28	0.39	0.11
16	0.79	0.74	0.72	0.19	0.20	0.20	0.54	0.50	0.47	0.09	0.25	0.21	0.37	0.07	0.09	0.25	0.31	0.33	0.03	0.08	0.28	0.33	0.28	0.03	0.08	0.25	0.29	0.34	0.03
17	0.74	0.75	0.81	0.23	0.23	0.23	0.50	0.50	0.52	0.09	0.37	0.38	0.14	0.02	0.06	0.40	0.37	0.15	0.02	0.08	0.38	0.34	0.15	0.05	0.07	0.39	0.37	0.15	0.02
18	0.73	0.77	0.76	0.20	0.21	0.17	0.47	0.45	0.41	0.08	0.38	0.31	0.22	0.02	0.08	0.39	0.31	0.20	0.03	0.12	0.40	0.29	0.11	0.08	0.08	0.40	0.30	0.21	0.02
19	0.64	0.58	0.47	0.20	0.20	0.15	0.36	0.34	0.26	0.04	0.63	0.22	0.11	0.00	0.03	0.68	0.25	0.05	0.00	0.11	0.70	0.18	0.01	0.00	0.04	0.68	0.23	0.05	0.00
20	0.86	0.72	0.51	0.39	0.35	0.25	0.17	0.14	0.08	0.09	0.40	0.15	0.21	0.15	0.13	0.42	0.18	0.23	0.05	0.23	0.44	0.25	0.07	0.00	0.12	0.42	0.16	0.25	0.05
21	0.69	0.47	0.28	0.79	0.82	0.83	0.28	0.21	0.21	0.00	0.40	0.21	0.22	0.17	0.00	0.50	0.17	0.18	0.15	0.00	0.63	0.06	0.18	0.12	0.00	0.46	0.23	0.17	0.14
22	0.77	0.79	0.67	0.19	0.11	0.10	0.49	0.37	0.25	0.13	0.57	0.17	0.09	0.04	0.46	0.32	0.10	0.08	0.03	0.55	0.30	0.08	0.07	0.01	0.37	0.40	0.12	0.08	0.03
23	0.94	0.92	0.87	0.11	0.13	0.12	0.59	0.56	0.49	0.10	0.05	0.03	0.05	0.77	0.08	0.07	0.03	0.05	0.77	0.08	0.08	0.03	0.54	0.26	0.09	0.07	0.03	0.05	0.77
24	0.09	0.08	0.09	0.81	0.84	0.86	0.40	0.36	0.34	0.49	0.12	0.03	0.24	0.11	0.50	0.12	0.01	0.26	0.11	0.50	0.12	0.01	0.26	0.11	0.50	0.12	0.01	0.26	0.10
25	0.14	0.17	0.15	0.67	0.75	0.67	0.39	0.46	0.39	0.17	0.39	0.23	0.20	0.01	0.14	0.33	0.25	0.26	0.02	0.17	0.38	0.26	0.19	0.00	0.15	0.35	0.26	0.23	0.01
26	0.19	0.21	0.17	0.37	0.42	0.33	0.71	0.70	0.41	0.12	0.55	0.24	0.08	0.02	0.06	0.53	0.33	0.07	0.02	0.14	0.65	0.20	0.01	0.00	0.07	0.57	0.29	0.07	0.01
27	0.68	0.62	0.52	0.59	0.36	0.17	0.17	0.14	0.11	0.05	0.11	0.33	0.51	0.00	0.08	0.29	0.42	0.20	0.00	0.17	0.38	0.35	0.10	0.00	0.06	0.20	0.50	0.24	0.00
28	0.31	0.61	0.91	0.54	0.22	0.18	0.38	0.70	0.96	0.01	0.65	0.31	0.01	0.01	0.01	0.03	0.19	0.77	0.00	0.01	0.02	0.00	0.05	0.91	0.01	0.01	0.66	0.31	0.00
29	0.76	0.67	0.57	0.26	0.24	0.22	0.94	0.91	0.88	0.00	0.13	0.03	0.81	0.03	0.00	0.13	0.11	0.73	0.03	0.01	0.14	0.76	0.08	0.02	0.00	0.13	0.07	0.77	0.03
30	0.68	0.63	0.39	0.65	0.47	0.11	0.16	0.23	0.13	0.13	0.10	0.22	0.54	0.00	0.06	0.23	0.47	0.24	0.00	0.36	0.44	0.20	0.00	0.00	0.06	0.25	0.50	0.19	0.00
31	0.70	0.77	0.91	0.29	0.21	0.25	0.50	0.48	0.63	0.00	0.24	0.33	0.41	0.01	0.03	0.26	0.23	0.39	0.09	0.02	0.21	0.15	0.30	0.32	0.01	0.25	0.24	0.42	0.08
32	0.74	0.72	0.64	0.23	0.21	0.20	0.49	0.46	0.40	0.01	0.15	0.25	0.59	0.00	0.02	0.16	0.27	0.55	0.00	0.01	0.22	0.37	0.40	0.00	0.01	0.16	0.26	0.56	0.00
33	0.60	0.57	0.48	0.44	0.30	0.21	0.38	0.44	0.43	0.02	0.15	0.74	0.08	0.01	0.02	0.36	0.58	0.04	0.01	0.05	0.46	0.47	0.02	0.00	0.02	0.29	0.66	0.03	0.01
34	0.58	0.47	0.13	0.79	0.82	0.71	0.17	0.18	0.12	0.12	0.18	0.32	0.28	0.09	0.10	0.28	0.36	0.17	0.09	0.48	0.24	0.11	0.11	0.05	0.10	0.28	0.38	0.17	0.08
35	0.46	0.43	0.32	0.68	0.65	0.57	0.26	0.27	0.20	0.00	0.40	0.37	0.22	0.00	0.00	0.44	0.38	0.18	0.00	0.00	0.64	0.28	0.07	0.00	0.00	0.46	0.36	0.18	0.00
36	0.38	0.32	0.20	0.24	0.18	0.13	0.64	0.53	0.33	0.00	0.78	0.19	0.03	0.00	0.10	0.79	0.11	0.01	0.00	0.24	0.74	0.02	0.00	0.00	0.01	0.88	0.10	0.01	0.00
37	0.73	0.69	0.60	0.21	0.13	0.16	0.44	0.36	0.31	0.09	0.50	0.24	0.14	0.04	0.28	0.40	0.20	0.11	0.02	0.16	0.58	0.17	0.06	0.02	0.16	0.50	0.21	0.12	0.01
38	0.13	0.16	0.18	0.32	0.26	0.23	0.76	0.77	0.75	0.13	0.42	0.11	0.34	0.00	0.10	0.53	0.03	0.34	0.00	0.12	0.51	0.03	0.34	0.00	0.08	0.54	0.04	0.34	0.00
39	0.53	0.50	0.43	0.80	0.80	0.74	0.30	0.25	0.18	0.07	0.17	0.43	0.30	0.03	0.04	0.32	0.36	0.26	0.02	0.15	0.28	0.33	0.23	0.00	0.04	0.30	0.39	0.26	0.02
40	0.21	0.19	0.17	0.78	0.72	0.68	0.49	0.44	0.42	0.08	0.30	0.32	0.24	0.06	0.09	0.31	0.41	0.16	0.03	0.12	0.39	0.28	0.20	0.02	0.08	0.31	0.40	0.18	0.03
41	0.81	0.75	0.65	0.26	0.25	0.25	0.57	0.52	0.45	0.00	0.25	0.22	0.48	0.06	0.00	0.28	0.26	0.45	0.01	0.00	0.31	0.44	0.24	0.01	0.00	0.27	0.26	0.46	0.02
42	0.25	0.19	0.16	0.53	0.45	0.41	0.78	0.69	0.66	0.04	0.22	0.37	0.32	0.05	0.07	0.29	0.42	0.21	0.01	0.09	0.33	0.40	0.16	0.01	0.06	0.27	0.41	0.24	0.02
43	0.86	0.35	0.11	0.67	0.33	0.29	0.99	0.99	0.98	0.00	0.00	0.08	0.64	0.28	0.00	0.82	0.14	0.00	0.04	0.51	0.40	0.05	0.00	0.04	0.00	0.04	0.91	0.00	0.04
44	0.21	0.06	0.12	0.22	0.33	0.68	0.19	0.19	0.34	0.01	0.99	0.00	0.00	0.00	0.40	0.60	0.01	0.00	0.00	0.20	0.48	0.14	0.18	0.00	0.28	0.71	0.01	0.00	0.00
45	0.59	0.55	0.44	0.32	0.36	0.33	0.77	0.72	0.65	0.01	0.16	0.37	0.43	0.04	0.00	0.14	0.51	0.35	0.01	0.01	0.27	0.53	0.19	0.00	0.00	0.13	0.54	0.33	0.01
46	0.61	0.56	0.35	0.19	0.15	0.11	0.87	0.83	0.59	0.09	0.12	0.17	0.39	0.22	0.11	0.13	0.19	0.42	0.15	0.16	0.26	0.39	0.17	0.02	0.10	0.14	0.20	0.45	0.11
47	0.10	0.10	0.11	0.89	0.88	0.87	0.54	0.54	0.53	0.43	0.10	0.03	0.14	0.30	0.42	0.11	0.03	0.17	0.27	0.37	0.15	0.03	0.21	0.23	0.42	0.11	0.03	0.16	0.28
48	0.12	0.11	0.09	0.92	0.91	0.88	0.48	0.45	0.47	0.36	0.24	0.02	0.04	0.34	0.41	0.19	0.02	0.05	0.33	0.48	0.12	0.02	0.17	0.21	0.41	0.19	0.02	0.05	0.33
49	0.63	0.16	0.21	0.63	0.33	0.32	0.33	0.27	0.25	0.01	0.13	0.44	0.43	0.00	0.10	0.85	0.05	0.01	0.00	0.01	0.95	0.03	0.01	0.00	0.00	0.65	0.34	0.01	0.00
50	0.87	0.82	0.72	0.75	0.31	0.20	0.81	0.61	0.48	0.00	0.00	0.01	0.76	0.22	0.00	0.34	0.04	0.62	0.00	0.00	0.36	0.04	0.61	0.00	0.00	0.03	0.34	0.63	0.00
51	0.94	0.93	0.91	0.05	0.05	0.05	0.87	0.82	0.75	0.68	0.02	0.00	0.04	0.26	0.68	0.01	0.01	0.14	0.16	0.68	0.02	0.01	0.15	0.14	0.68	0.01	0.01	0.13	0.17
52	0.47	0.47	0.44	0.10	0.11	0.14	0.93	0.87	0.79	0.68	0.03	0.01	0.03	0.24	0.63	0.08	0.01	0.11	0.16	0.01	0.71	0.01	0.27	0.00	0.65	0.06	0.01	0.10	0.18
53	0.70	0.45	0.45	0.88	0.78	0.70	0.33	0.34	0.24	0.03	0.09	0.19	0.39	0.31	0.00	0.23	0.32	0.43	0.02	0.00	0.28	0.35	0.37	0.00	0.00	0.17	0.34	0.47	0.03
54	0.20	0.15	0.20	0.71	0.68	0.62	0.53	0.33	0.37	0.01	0.27	0.26	0.45	0.00	0.10	0.38	0.15	0.38	0.00	0.00	0.43	0.32	0.25	0.00	0.04	0.36	0.22	0.38	0.00
55	0.88	0.88	0.85	0.20	0.25	0.22	0.56	0.60	0.56	0.07	0.18	0.11	0.28	0.35	0.02	0.20	0.13	0.33	0.32	0.03	0.22	0.11	0.45	0.18	0.03	0.21	0.12	0.32	0.32
56	0.64	0.48	0.41	0.84	0.79	0.75	0.31	0.25	0.25	0.02	0.14	0.22	0.45	0.16	0.01	0.26	0.26	0.40	0.07	0.03	0.30	0.25	0.39	0.03	0.01	0.22	0.29	0.41	0.08
57	0.75	0.21	0.18	0.71	0.18	0.17	0.81	0.65	0.57	0.00	0.00	0.01	0.99	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.93	0.06	0.01	0.00
58	0.04	0.07	0.37	0.66	0.80	0.74	0.05	0.10	0.41	0.97	0.00	0.00	0.02	0.00	0.95	0.02	0.00	0.02	0.00	0.00	0.30	0.68	0.02	0.00	0.89	0.09	0.00	0.02	0.00
59	0.12	0.12	0.14	0.11	0.10	0.10	0.05	0.05	0.05	0.92	0.08	0.00	0.00	0.00	0.92	0.08	0.00	0.00	0.00	0.72	0.28	0.00	0.00	0.00	0.91	0.09	0.00	0.00	0.00
60	0.90	0.90	0.87	0.82	0.82	0.78	0.86	0.86	0.83	0.00	0.00	0.00	0.55	0.45	0.00	0.00	0.00	0.54	0.46	0.00	0.00	0.00	0.82	0.18	0.00	0.00	0.00	0.57	0.43

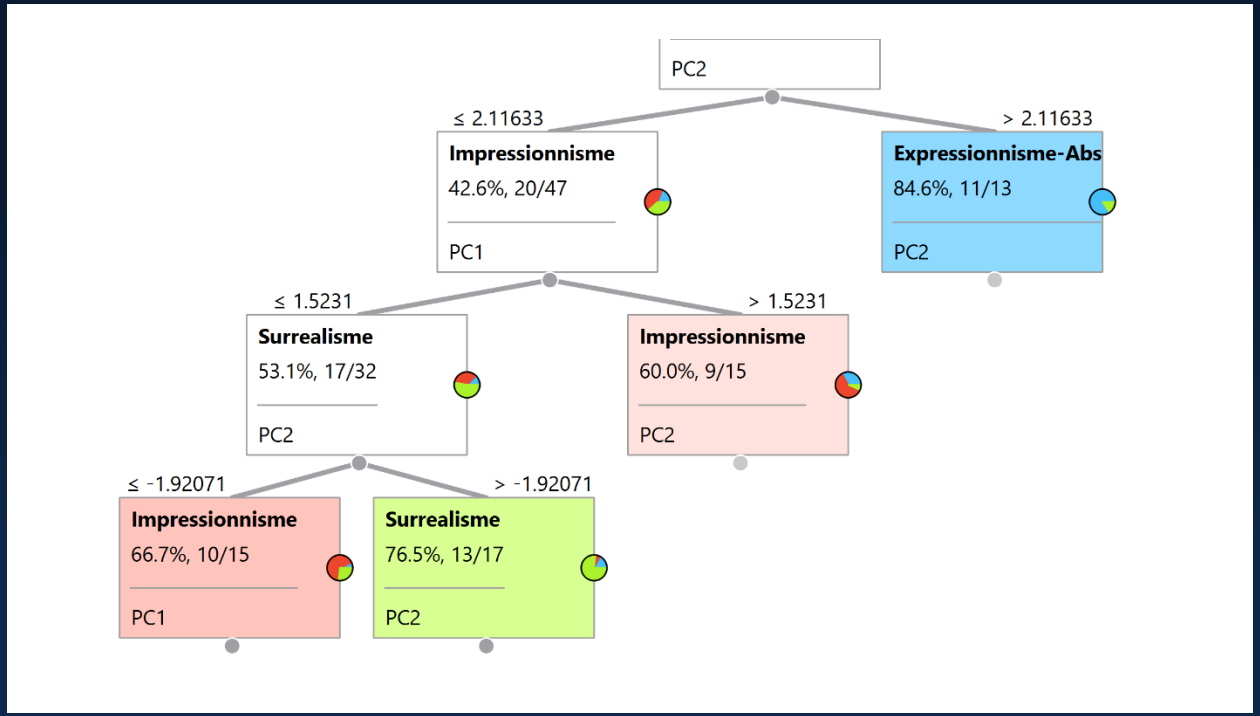
https://www.demilked.com/anamorphosis-anamorphic-cylinder-art/



Id	Style	PC1	PC2
01	Impressionnisme	3.86	-3.19
02	Impressionnisme	-4.39	-3.12
03	Impressionnisme	-0.92	-3.56
04	Impressionnisme	-1.10	-2.07
05	Impressionnisme	1.19	-3.24
06	Impressionnisme	1.42	-3.06
07	Impressionnisme	-3.67	-3.38
08	Impressionnisme	6.36	1.02
09	Impressionnisme	5.59	0.67
10	Impressionnisme	2.34	-3.33
11	Impressionnisme	3.11	-1.61
12	Impressionnisme	1.39	-0.21
13	Impressionnisme	-4.87	-2.88
14	Impressionnisme	2.00	1.46
15	Impressionnisme	4.41	0.18
16	Impressionnisme	2.74	-1.41
17	Impressionnisme	0.62	-2.11
18	Impressionnisme	-0.48	-1.93
19	Impressionnisme	-1.49	-2.29
20	Impressionnisme	0.01	-0.86
21	Surrealisme	-1.69	-1.22
22	Surrealisme	-4.60	0.85
23	Surrealisme	3.08	10.66
24	Surrealisme	-3.02	2.40
25	Surrealisme	-0.27	-1.94
26	Surrealisme	-1.10	-4.04
27	Surrealisme	-0.72	-0.71
28	Surrealisme	0.67	1.14
29	Surrealisme	5.59	1.71
30	Surrealisme	-0.03	-1.99
31	Surrealisme	0.91	0.55
32	Surrealisme	1.33	2.00
33	Surrealisme	-0.84	-2.10
34	Surrealisme	-1.88	-1.67
35	Surrealisme	-2.16	-2.42
36	Surrealisme	-4.99	-1.28
37	Surrealisme	-1.27	-1.77
38	Surrealisme	-1.47	-0.20
39	Surrealisme	0.97	-1.85
40	Surrealisme	0.74	-3.70
41	Expressionnisme-Abstrait	4.70	-1.24
42	Expressionnisme-Abstrait	0.84	-2.12
43	Expressionnisme-Abstrait	-1.99	5.15
44	Expressionnisme-Abstrait	-7.20	-0.45
45	Expressionnisme-Abstrait	3.39	-1.53
46	Expressionnisme-Abstrait	6.00	0.52
47	Expressionnisme-Abstrait	-1.57	3.88
48	Expressionnisme-Abstrait	-2.38	3.10
49	Expressionnisme-Abstrait	-3.92	-2.91
50	Expressionnisme-Abstrait	2.58	4.65
51	Expressionnisme-Abstrait	-1.74	7.10
52	Expressionnisme-Abstrait	-3.71	8.19
53	Expressionnisme-Abstrait	3.33	-0.62
54	Expressionnisme-Abstrait	0.11	-1.52
55	Expressionnisme-Abstrait	4.36	2.37
56	Expressionnisme-Abstrait	3.29	-1.00
57	Expressionnisme-Abstrait	-4.55	4.24
58	Expressionnisme-Abstrait	-6.73	2.21
59	Expressionnisme-Abstrait	-7.79	4.65
60	Expressionnisme-Abstrait	5.60	5.83



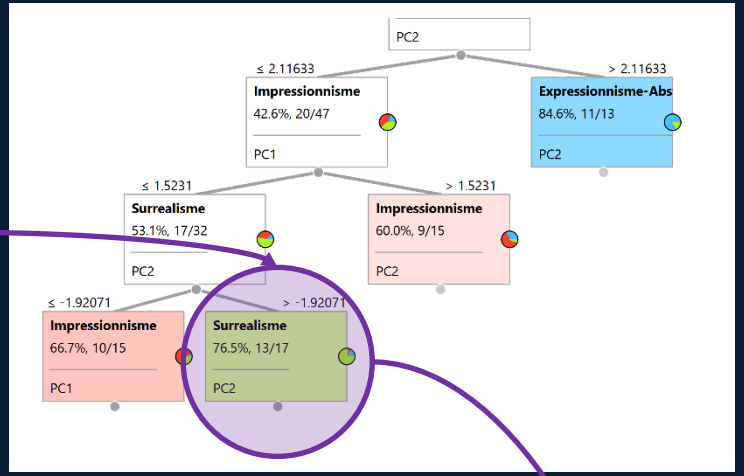
Arbre de décision



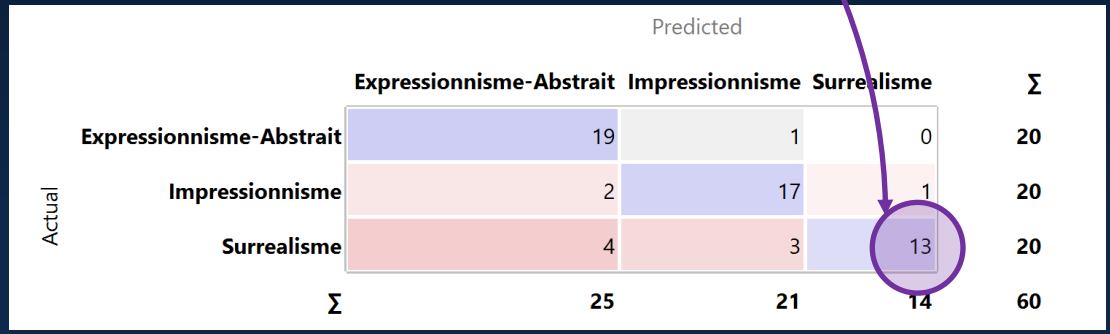
Id	Style	PC1	PC2
34	Surrealisme	-1.88	-1.67

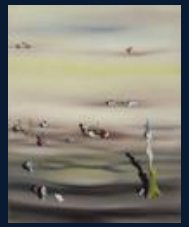
Id	Style	PC1	PC2
01	Impressionnisme	3.86	-3.19
02	Impressionnisme	-4.39	-3.12
03	Impressionnisme	-0.92	-3.56
04	Impressionnisme	-1.10	-2.07
05	Impressionnisme	1.19	-3.24
06	Impressionnisme	1.42	-3.06
07	Impressionnisme	-3.67	-3.38
08	Impressionnisme	6.36	1.02
09	Impressionnisme	5.59	0.67
10	Impressionnisme	2.34	-3.33
11	Impressionnisme	3.11	-1.61
12	Impressionnisme	1.39	-0.21
13	Impressionnisme	-4.87	-2.88
14	Impressionnisme	2.00	1.46
15	Impressionnisme	4.41	0.18
16	Impressionnisme	2.74	-1.41
17	Impressionnisme	0.62	-2.11
18	Impressionnisme	-0.48	-1.93
19	Impressionnisme	-1.49	-2.29
20	Impressionnisme	0.01	-0.86
21	Surrealisme	-1.69	-1.22
22	Surrealisme	-4.60	0.85
23	Surrealisme	3.08	10.66
24	Surrealisme	-3.02	2.40
25	Surrealisme	-0.27	-1.94
26	Surrealisme	-1.10	-4.04
27	Surrealisme	-0.72	-0.71
28	Surrealisme	0.67	1.14
29	Surrealisme	5.59	1.71
30	Surrealisme	-0.03	-1.99
31	Surrealisme	0.91	0.55
32	Surrealisme	1.33	2.00
34	Surrealisme	-1.88	-1.67
36	Surrealisme	-4.99	-1.28
37	Surrealisme	-1.27	-1.77
38	Surrealisme	-1.47	-0.20
39	Surrealisme	0.97	-1.85
40	Surrealisme	0.74	-3.70
41	Expressionnisme-Abstrait	4.70	-1.24
42	Expressionnisme-Abstrait	0.84	-2.12
43	Expressionnisme-Abstrait	-1.99	5.15
44	Expressionnisme-Abstrait	-7.20	-0.45
45	Expressionnisme-Abstrait	3.39	-1.53
46	Expressionnisme-Abstrait	6.00	0.52
47	Expressionnisme-Abstrait	-1.57	3.88
48	Expressionnisme-Abstrait	-2.38	3.10
49	Expressionnisme-Abstrait	-3.92	-2.91
50	Expressionnisme-Abstrait	2.58	4.65
51	Expressionnisme-Abstrait	-1.74	7.10
52	Expressionnisme-Abstrait	-3.71	8.19
53	Expressionnisme-Abstrait	3.33	-0.62
54	Expressionnisme-Abstrait	0.11	-1.52
55	Expressionnisme-Abstrait	4.36	2.37
56	Expressionnisme-Abstrait	3.29	-1.00
57	Expressionnisme-Abstrait	-4.55	4.24
58	Expressionnisme-Abstrait	-6.73	2.21
59	Expressionnisme-Abstrait	-7.79	4.65
60	Expressionnisme-Abstrait	5.60	5.83

Arbre de décision



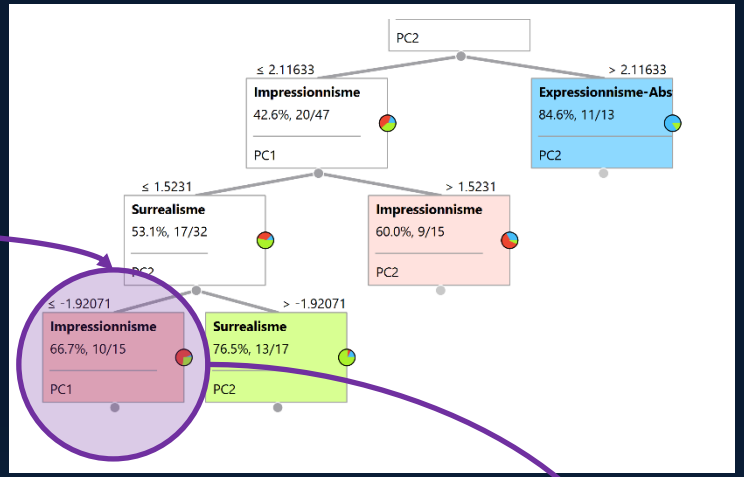
Matrice de confusion





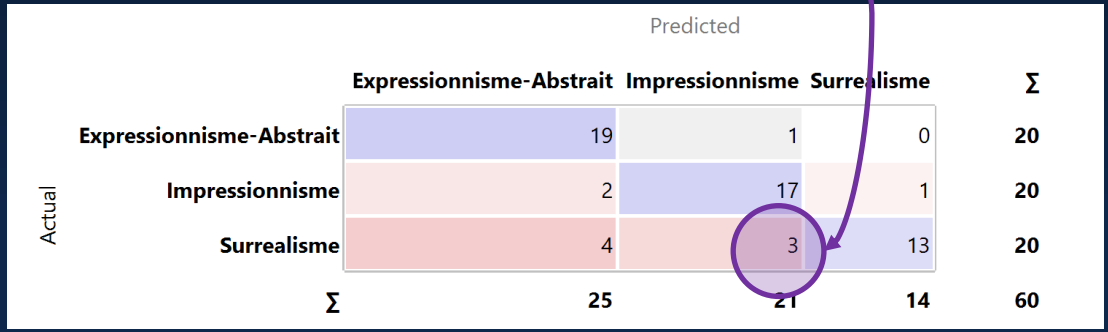
Id	Style	PC1	PC2
33	Surrealisme	-0.84	-2.10

Arbre de décision



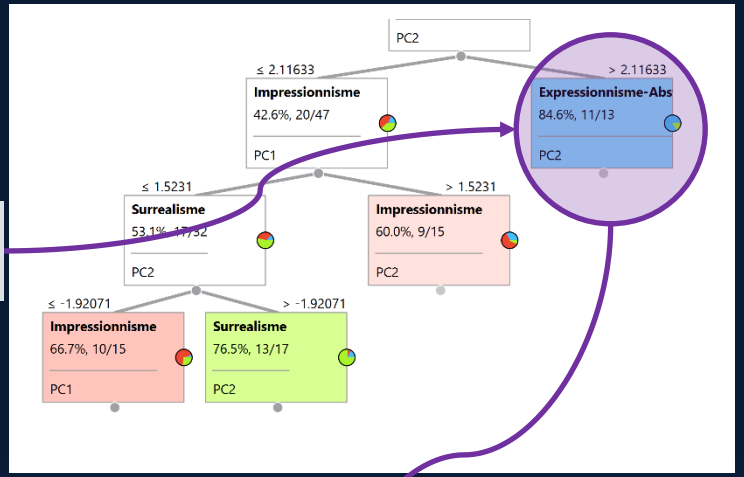
Id	Style	PC1	PC2
01	Impressionnisme	3.86	-3.19
02	Impressionnisme	-4.39	-3.12
03	Impressionnisme	-0.92	-3.56
04	Impressionnisme	-1.10	-2.07
05	Impressionnisme	1.19	-3.24
06	Impressionnisme	1.42	-3.06
07	Impressionnisme	-3.67	-3.38
08	Impressionnisme	6.36	1.02
09	Impressionnisme	5.59	0.67
10	Impressionnisme	2.34	-3.33
11	Impressionnisme	3.11	-1.61
12	Impressionnisme	1.39	-0.21
13	Impressionnisme	-4.87	-2.88
14	Impressionnisme	2.00	1.46
15	Impressionnisme	4.41	0.18
16	Impressionnisme	2.74	-1.41
17	Impressionnisme	0.62	-2.11
18	Impressionnisme	-0.48	-1.93
19	Impressionnisme	-1.49	-2.29
20	Impressionnisme	0.01	-0.86
21	Surrealisme	-1.69	-1.22
22	Surrealisme	-4.60	0.85
23	Surrealisme	3.08	10.66
24	Surrealisme	-3.02	2.40
25	Surrealisme	-0.27	-1.94
26	Surrealisme	-1.10	-4.04
27	Surrealisme	-0.72	-0.71
28	Surrealisme	0.67	1.14
29	Surrealisme	5.59	1.71
30	Surrealisme	-0.03	-1.99
31	Surrealisme	0.91	0.55
33	Surrealisme	-0.84	-2.10
34	Surrealisme	-1.88	-1.67
35	Surrealisme	-2.16	-2.42
36	Surrealisme	-4.99	-1.28
37	Surrealisme	-1.27	-1.77
38	Surrealisme	-1.47	-0.20
39	Surrealisme	0.97	-1.85
40	Surrealisme	0.74	-3.70
41	Expressionnisme-Abstrait	4.70	-1.24
42	Expressionnisme-Abstrait	0.84	-2.12
43	Expressionnisme-Abstrait	-1.99	5.15
44	Expressionnisme-Abstrait	-7.20	-0.45
45	Expressionnisme-Abstrait	3.39	-1.53
46	Expressionnisme-Abstrait	6.00	0.52
47	Expressionnisme-Abstrait	-1.57	3.88
48	Expressionnisme-Abstrait	-2.38	3.10
49	Expressionnisme-Abstrait	-3.92	-2.91
50	Expressionnisme-Abstrait	2.58	4.65
51	Expressionnisme-Abstrait	-1.74	7.10
52	Expressionnisme-Abstrait	-3.71	8.19
53	Expressionnisme-Abstrait	3.33	-0.62
54	Expressionnisme-Abstrait	0.11	-1.52
55	Expressionnisme-Abstrait	4.36	2.37
56	Expressionnisme-Abstrait	3.29	-1.00
57	Expressionnisme-Abstrait	-4.55	4.24
58	Expressionnisme-Abstrait	-6.73	2.21
59	Expressionnisme-Abstrait	-7.79	4.65
60	Expressionnisme-Abstrait	5.60	5.83

Matrice de confusion



Id	Style	PC1	PC2
47	Expressionnisme -Abstrait	-1.57	3.88

Arbre de décision



Id	Style	PC1	PC2
01	Impressionnisme	3.86	-3.19
02	Impressionnisme	-4.39	-3.12
03	Impressionnisme	-0.92	-3.56
04	Impressionnisme	-1.10	-2.07
05	Impressionnisme	1.19	-3.24
06	Impressionnisme	1.42	-3.06
07	Impressionnisme	-3.67	-3.38
08	Impressionnisme	6.36	1.02
09	Impressionnisme	5.59	0.67
10	Impressionnisme	2.34	-3.33
11	Impressionnisme	3.11	-1.61
12	Impressionnisme	1.39	-0.21
13	Impressionnisme	-4.87	-2.88
14	Impressionnisme	2.00	1.46
15	Impressionnisme	4.41	0.18
16	Impressionnisme	2.74	-1.41
17	Impressionnisme	0.62	-2.11
18	Impressionnisme	-0.48	-1.93
19	Impressionnisme	-1.49	-2.29
20	Impressionnisme	0.01	-0.86
21	Surrealisme	-1.69	-1.22
22	Surrealisme	-4.60	0.85
23	Surrealisme	3.08	10.66
24	Surrealisme	-3.02	2.40
25	Surrealisme	-0.27	-1.94
26	Surrealisme	-1.10	-4.04
27	Surrealisme	-0.72	-0.71
28	Surrealisme	0.67	1.14
29	Surrealisme	5.59	1.71
30	Surrealisme	-0.03	-1.99
31	Surrealisme	0.91	0.55
32	Surrealisme	1.33	2.00
33	Surrealisme	-0.84	-2.10
34	Surrealisme	-1.88	-1.67
35	Surrealisme	-2.16	-2.42
36	Surrealisme	-4.99	-1.28
37	Surrealisme	-1.27	-1.77
38	Surrealisme	-1.47	-0.20
39	Surrealisme	0.97	-1.85
40	Surrealisme	0.74	-3.70
41	Expressionnisme-Abstrait	4.70	-1.24
42	Expressionnisme-Abstrait	0.84	-2.12
43	Expressionnisme-Abstrait	-1.99	5.15
44	Expressionnisme-Abstrait	-7.20	-0.45
45	Expressionnisme-Abstrait	3.39	-1.53
46	Expressionnisme-Abstrait	6.00	0.52
47	Expressionnisme-Abstrait	-1.57	3.88
48	Expressionnisme-Abstrait	-2.36	3.10
49	Expressionnisme-Abstrait	-3.92	-2.91
50	Expressionnisme-Abstrait	2.58	4.65
51	Expressionnisme-Abstrait	-1.74	7.10
52	Expressionnisme-Abstrait	-3.71	8.19
53	Expressionnisme-Abstrait	3.33	-0.62
54	Expressionnisme-Abstrait	0.11	-1.52
55	Expressionnisme-Abstrait	4.36	2.37
56	Expressionnisme-Abstrait	3.29	-1.00
57	Expressionnisme-Abstrait	-4.55	4.24
58	Expressionnisme-Abstrait	-6.73	2.21
59	Expressionnisme-Abstrait	-7.79	4.65
60	Expressionnisme-Abstrait	5.60	5.83

Matrice de confusion

Actual \ Predicted	Expressionnisme-Abstrait	Impressionnisme	Surrealisme	Σ
Expressionnisme-Abstrait	19	1	0	20
Impressionnisme	2	17	1	20
Surrealisme	4	3	13	20
Σ	25	21	14	60

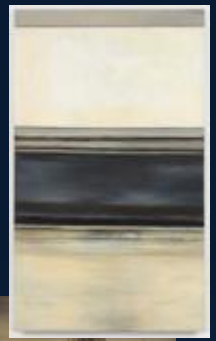
Constitution de la base de Test

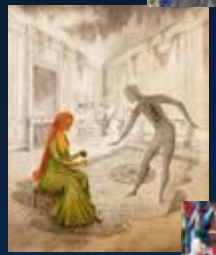
1. Impressionnisme
    - Paul Cézanne – La Montagne Sainte-Victoire (1904-1906)
    - Paul Gauguin – Vision après le sermon (La Lutte de Jacob avec l'Ange) (1888)
    - Georges Seurat – Un dimanche après-midi à l'Île de la Grande Jatte (1884-1886)
    - Armand Guillaumin – Soleil couchant à Ivry (1873)
    - Gustave Moreau – La Licorne (1885)

2. Surréalisme
    - Paul Delvaux – Le Voyage Légendaire (1974)
    - Remedios Varo – Les Feuilles mortes (1963)
    - Kay Sage – Tomorrow is Never (1955)
    - Wilfredo Lam – La Jungle (1943)
    - Toyen – Le Paravent (1966)

3. Expressionnisme abstrait
    - Hedda Sterne – Number 6 (1951)
    - Norman Lewis – Evening Rendezvous (1962)
    - Perle Fine – Summer I (1968)
    - Grace Hartigan – The King is Dead (1950)
    - Michael Goldberg – Park Avenue Façade (1957)

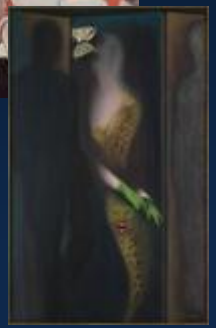
Constitution de la base de Test

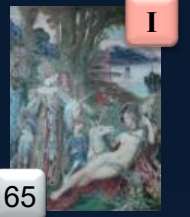


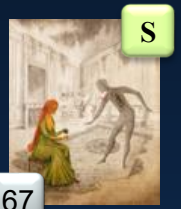


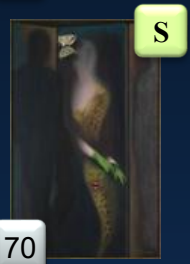


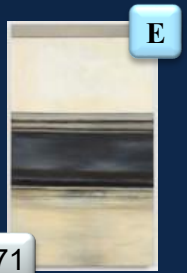


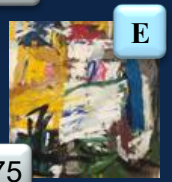
Constitution de la base de Test







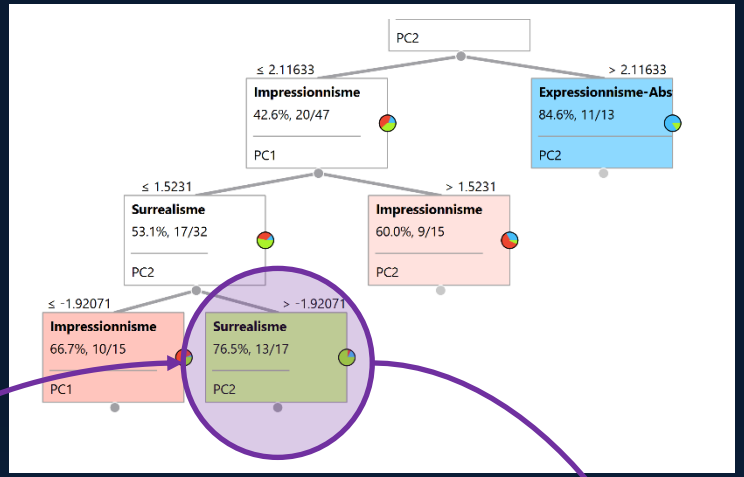




Apprentissage Automatique

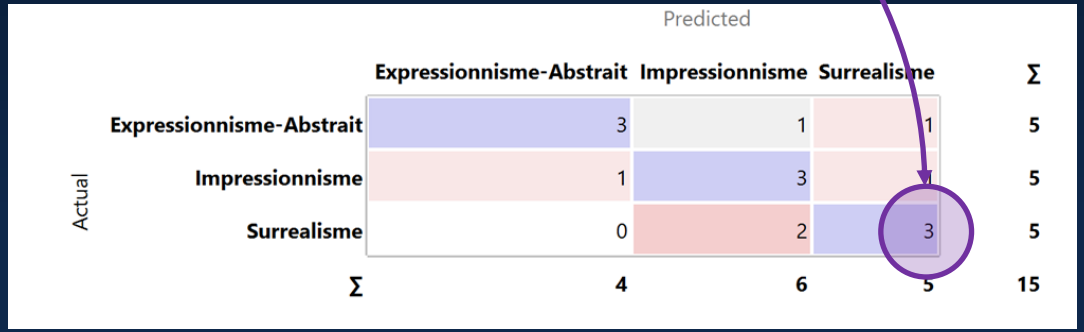
Phase de Test

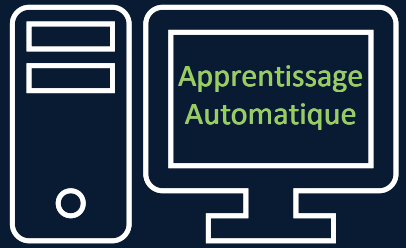
Arbre de décision



Id	Style	PC1	PC2
61	Impressionnisme	1.34	-1.47
62	Impressionnisme	-1.82	-1.61
63	Impressionnisme	-0.64	-3.34
64	Impressionnisme	0.87	-2.39
65	Impressionnisme	0.13	-3.13
66	Surrealisme	-2.74	-1.23
67	Surrealisme	5.43	0.94
68	Surrealisme	-2.43	-2.71
69	Surrealisme	-0.76	-2.05
70	Surrealisme	-6.80	2.00
71	Expressionnisme-Abstrait	1.75	3.66
72	Expressionnisme-Abstrait	-2.86	-3.03
73	Expressionnisme-Abstrait	1.24	-1.78
74	Expressionnisme-Abstrait	0.09	-0.21
75	Expressionnisme-Abstrait	2.18	0.34

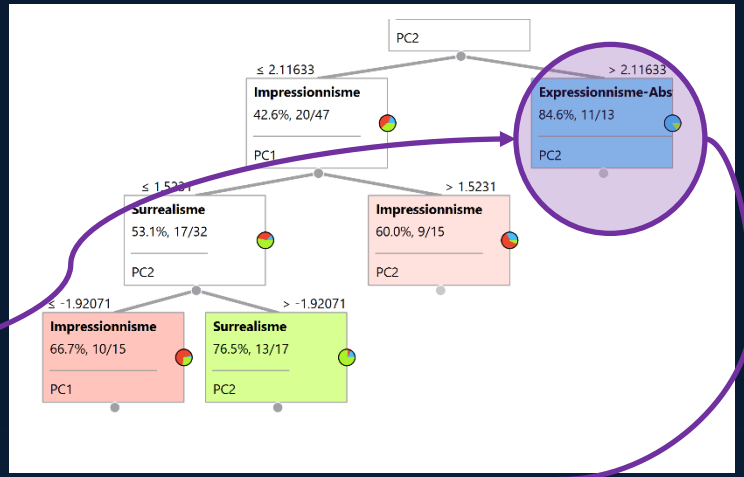
Matrice de confusion





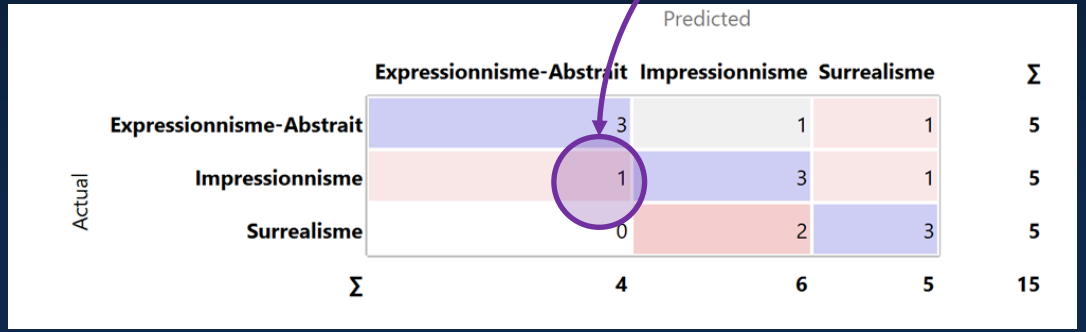
Phase de Test

Arbre de décision



Id	Style	PC1	PC2
61	Impressionnisme	1.34	-1.47
62	Impressionnisme	-1.82	-1.61
63	Impressionnisme	-0.64	-3.34
64	Impressionnisme	0.87	-2.39
65	Impressionnisme	0.13	-3.13
66	Surrealisme	-2.74	-1.23
67	Surrealisme	5.43	0.94
68	Surrealisme	-2.43	-2.71
69	Surrealisme	-0.76	-2.05
70	Surrealisme	-6.80	2.00
71	Expressionnisme-Abstrait	1.75	3.66
72	Expressionnisme-Abstrait	-2.86	-3.03
73	Expressionnisme-Abstrait	1.24	-1.78
74	Expressionnisme-Abstrait	0.09	-0.21
75	Expressionnisme-Abstrait	2.18	0.34

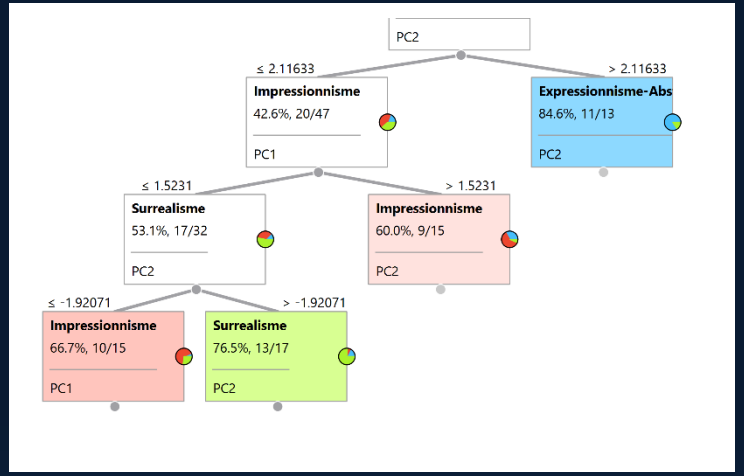
Matrice de confusion



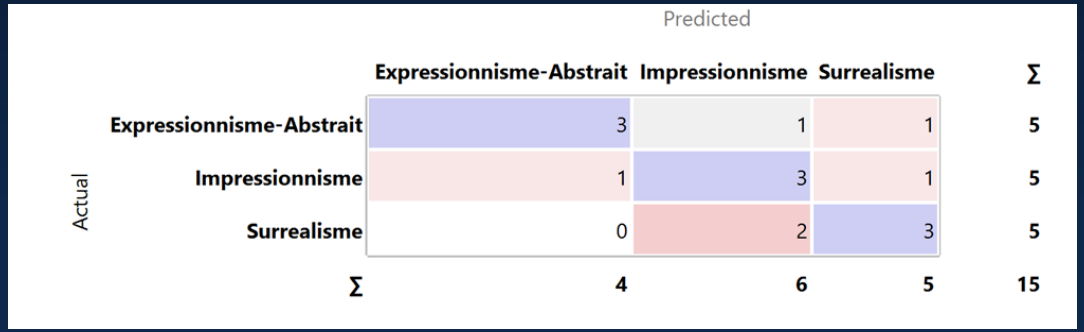
Phase de Test

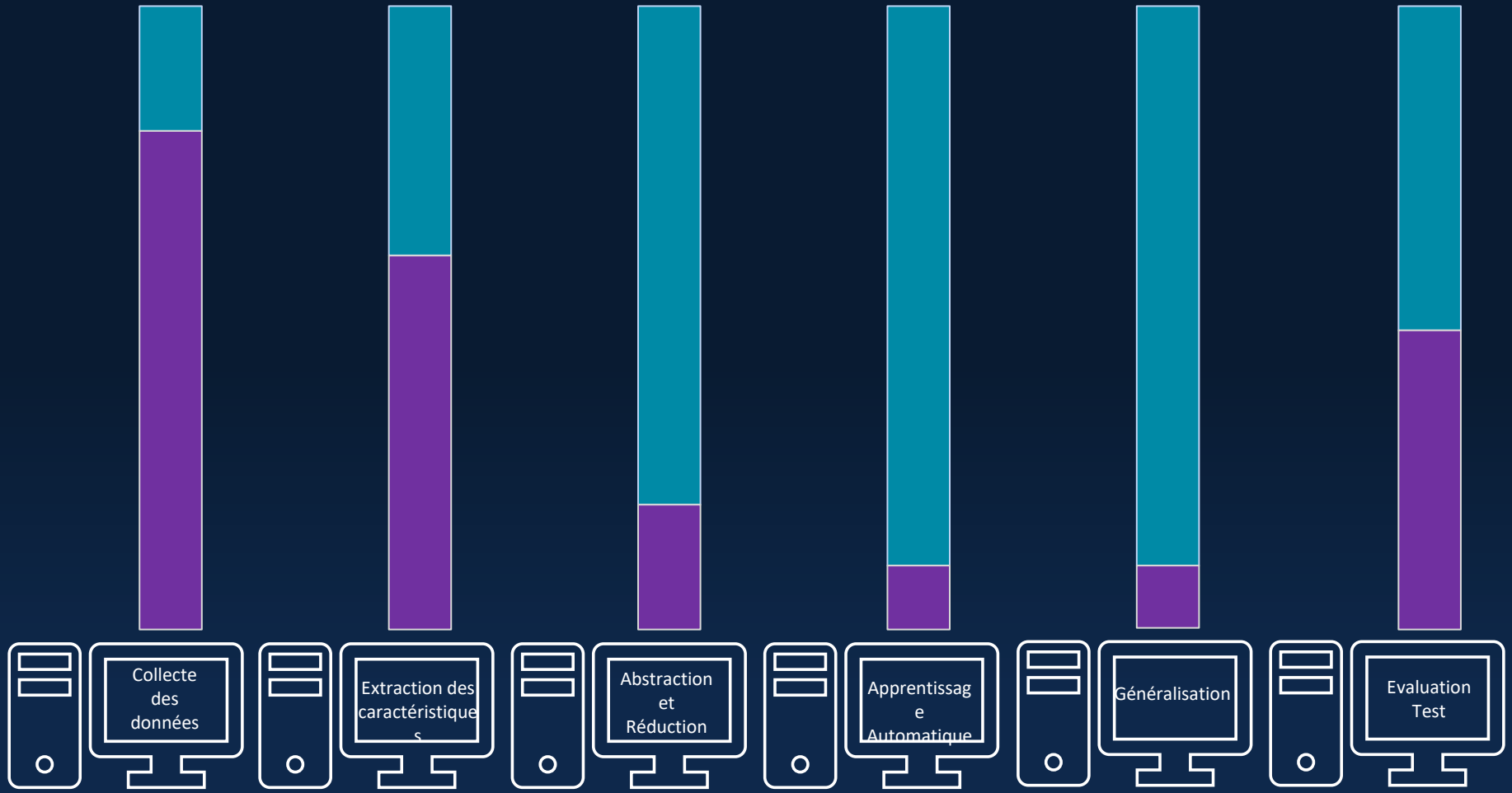
Id	Style	PC1	PC2
61	Impressionnisme	1.34	-1.47
62	Impressionnisme	-1.82	-1.61
63	Impressionnisme	-0.64	-3.34
64	Impressionnisme	0.87	-2.39
65	Impressionnisme	0.13	-3.13
66	Surrealisme	-2.74	-1.23
67	Surrealisme	5.43	0.94
68	Surrealisme	-2.43	-2.71
69	Surrealisme	-0.76	-2.05
70	Surrealisme	-6.80	2.00
71	Expressionnisme-Abstrait	1.75	3.66
72	Expressionnisme-Abstrait	-2.86	-3.03
73	Expressionnisme-Abstrait	1.24	-1.78
74	Expressionnisme-Abstrait	0.09	-0.21
75	Expressionnisme-Abstrait	2.18	0.34

Arbre de décision



Matrice de confusion

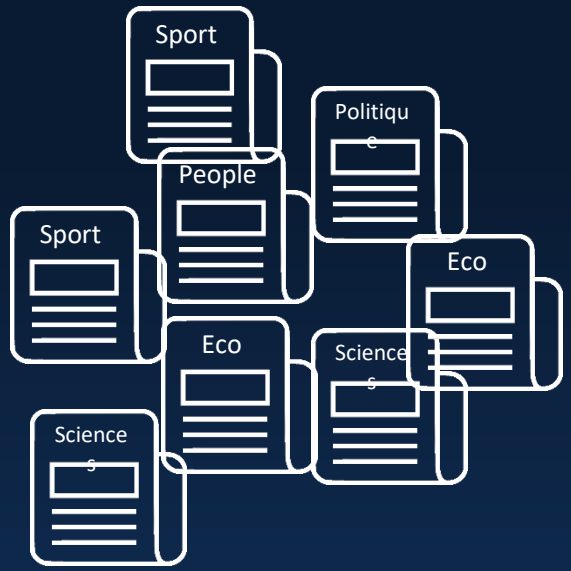



Collecte des données
Extraction des caractéristiques
Abstraction et Réduction
Apprentissage Automatique
Généralisation
Evaluation Test
Experts Métier
Experts Apprentissage Automatique – Science des données

Exemples

Classification de textes - Recommandation

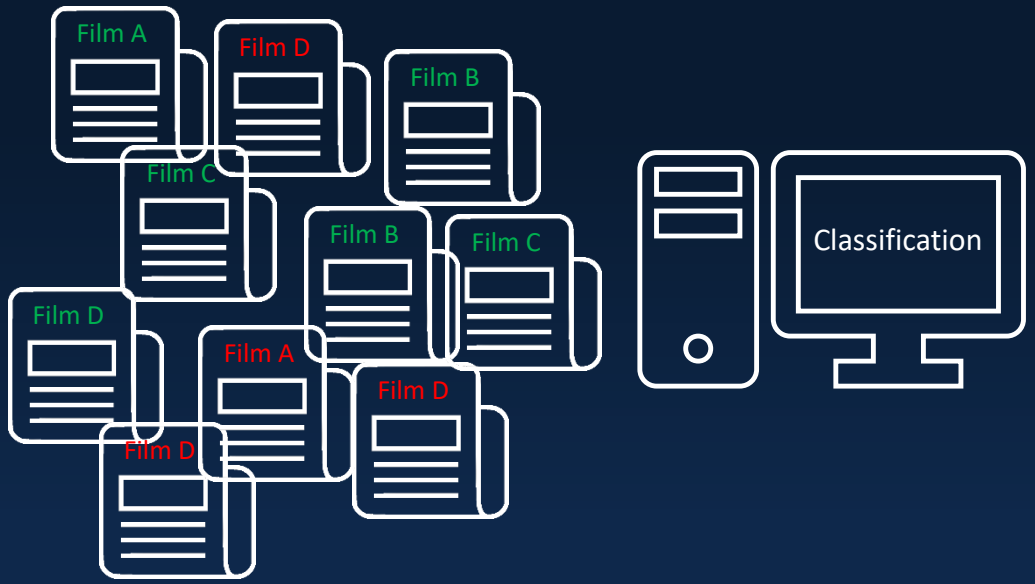
Articles de presse



Startup

Classification de textes - Recommandation

Opinions



Clustering sémantique de chercheurs

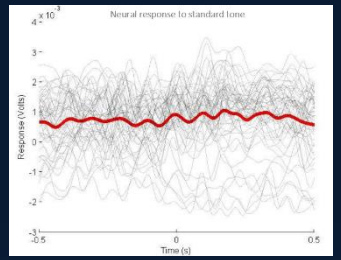
Organismes de recherche

Classification d’échantillons sanguins

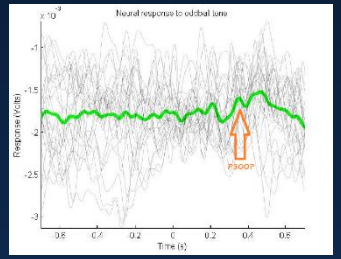
Grand groupe international

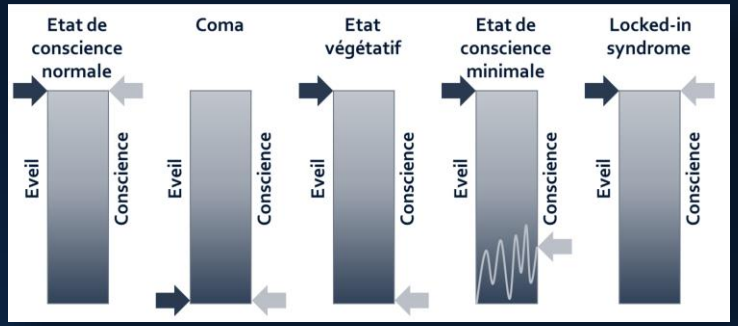
Classification de signaux cérébraux / états de conscience

Fréquent



Rare

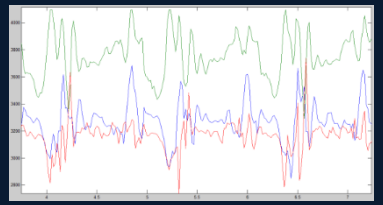




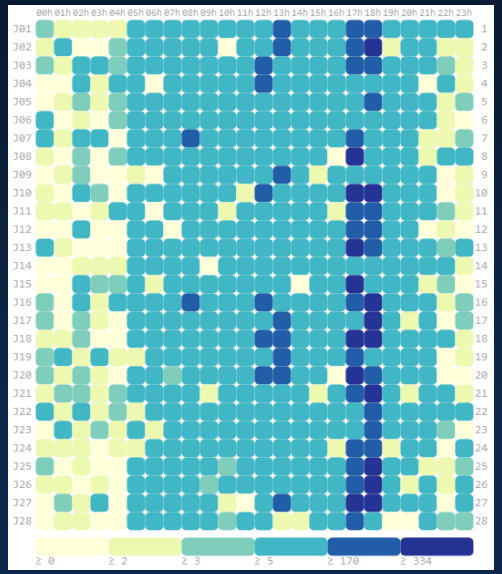
CHU

Suivi de l'activité physique

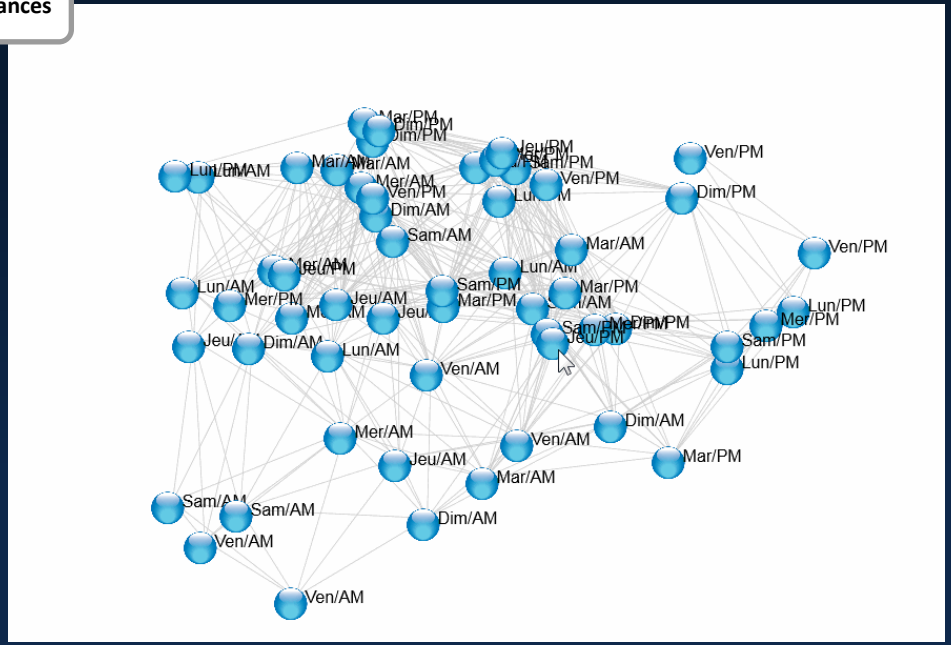
Données



Informations



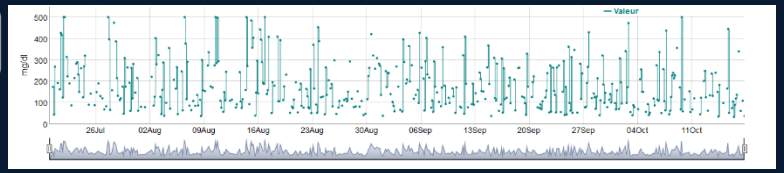
Connaissances



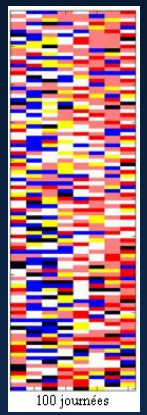
CHU

Suivi d'un patient diabétique

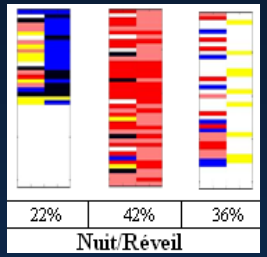
Données



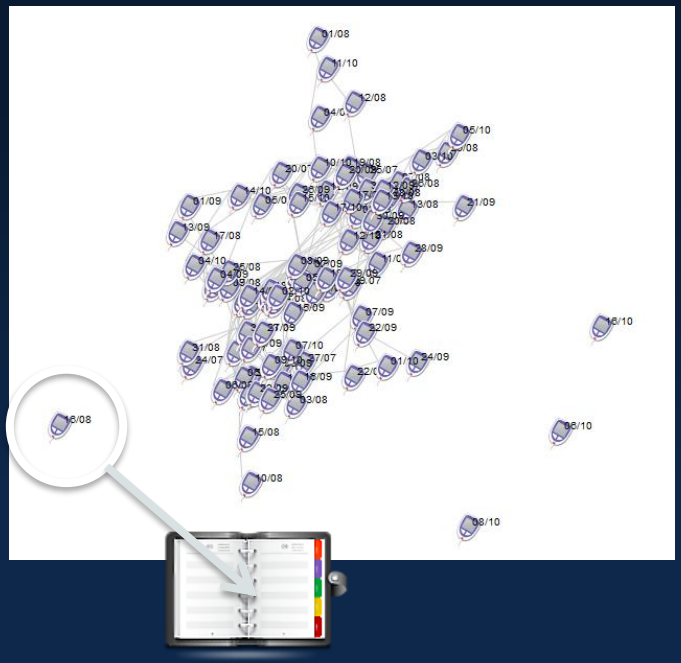
Informations



Clustering



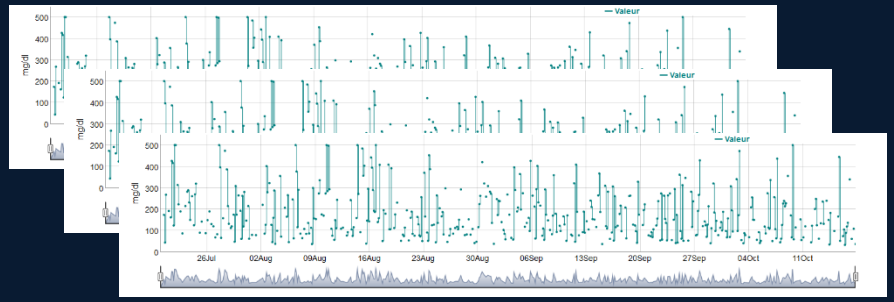
Connaissances



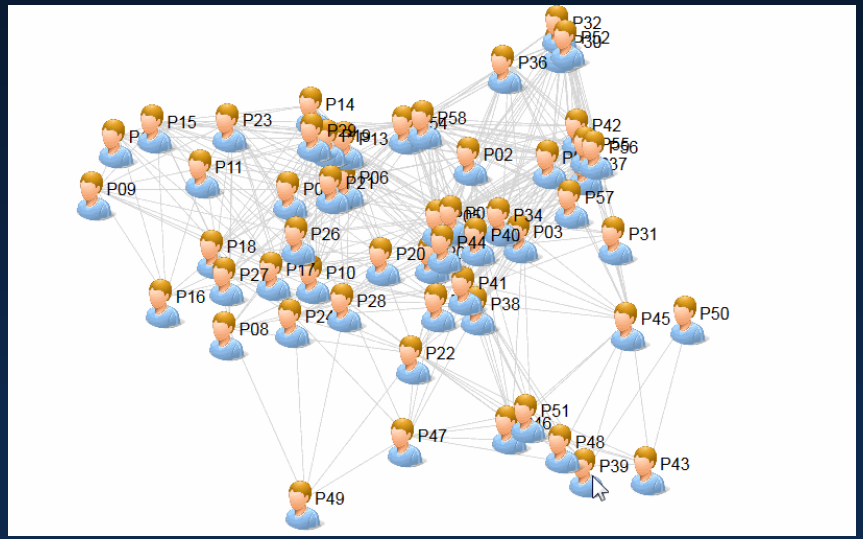
CHU

Suivi de patients diabétiques

Données



Informations



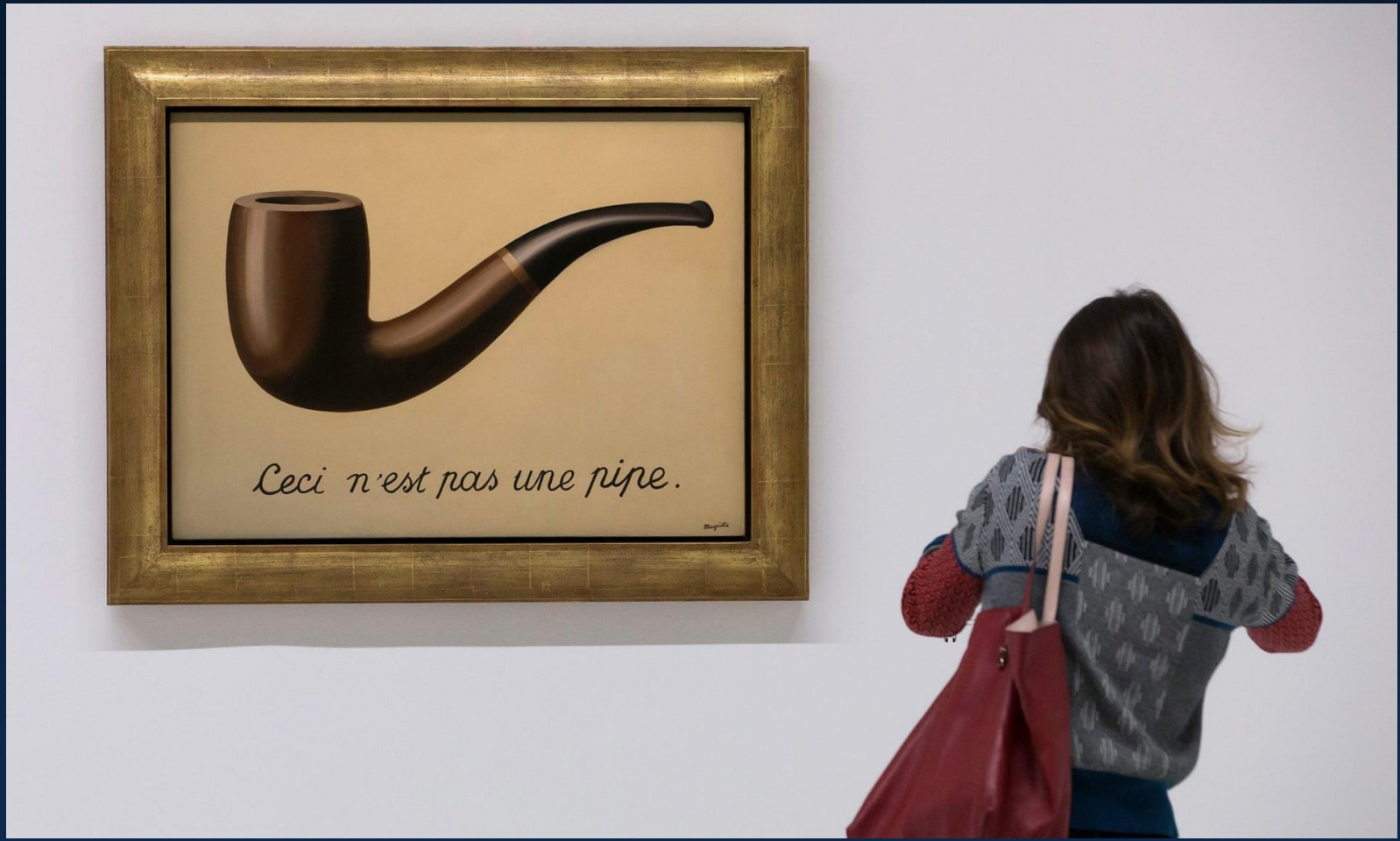
Connaissances

CHU

En guise de conclusion


"L'intelligence ce n'est pas ce que l'on sait mais ce qu'on l'on fait quand on ne sait pas."
JEAN PIAGET

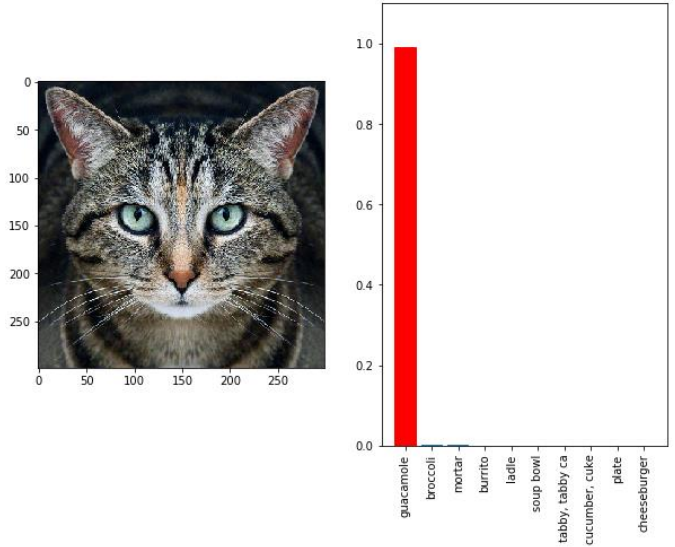
INTELLIGENCE
IS THE ABILITY TO
ADAPT TO CHANGE.
~
STEPHEN HAWKING



La Trahison des images
René Magritte 1929

Pour aller plus loin

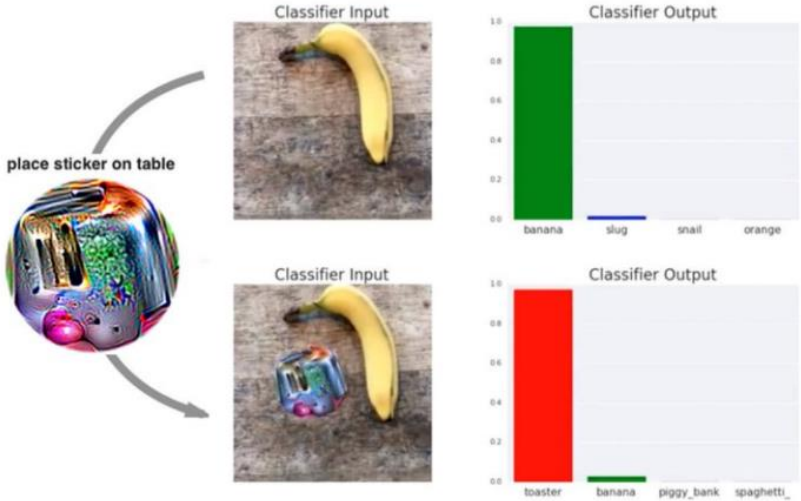
Les réseaux neuronaux se laissent facilement berner



Computer Vision system fooled into thinking it's a photo of guacamole. (Photo credit MIT CSAIL)

Adversarial Patch

A paper (Adversarial Patch) recently published by Google has lifted the bar. The paper shows how it is possible to show the system any image and it will classify it as a toaster.



https://medium.com/deep-learning-cafe/neural-networks-easily-fooled-e19bf575b527

Les réseaux neuronaux se laissent facilement berner

Stop

Yield

Speed Limit

(a) Normal

(b) Attack

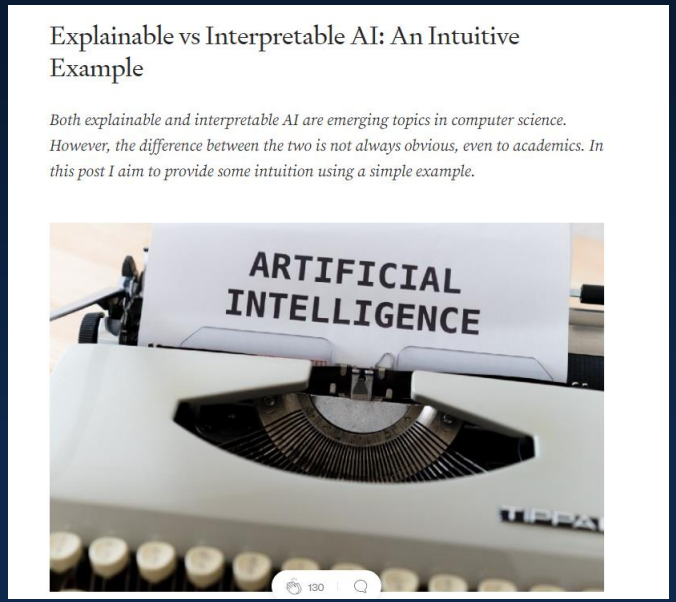
Automotive ML systems have been tricked into mistaking stop signs for speed limit signs

https://portswigger.net/daily-swig/trojannet-a-simple-yet-effective-attack-on-machine-learning-models

Explainable AI (XAI) - Interpretable AI - Explainable Machine Learning (XML),

https://en.wikipedia.org/wiki/Explainable_artificial_intelligence

https://medium.com/swlh/explainable-vs-interpretable-ai-an-intuitive-example-6baf8fc6d402



Explainable vs Interpretable AI: An Intuitive Example

Both explainable and interpretable AI are emerging topics in computer science. However, the difference between the two is not always obvious, even to academics. In this post I aim to provide some intuition using a simple example.

Les 5 débats de l'Intelligence artificielle
https://www.opinionact.com/ressources/ia-les-5-debats-de-lintelligence-artificielle

Débat philosophico-scientifique : qu'est-ce que l'intelligence ?

Une question pour le moins passionnelle est celle de savoir si l'on pourra un jour imiter, voire dépasser l'intelligence humaine. C'est le fameux point de singularité défendu par le controversé Ray Kurz
dépassés par les rob
intelligences.

Le directeur de l'ingé
2050. Les scientifiqu
critiquent la robustes
D'autres argumenten
l'intelligence humain
non verbale mais aus
échouer les IA au cél

Le débat éthique : faut-il s'inquiéter du transhumanisme ?

La conviction des transhumanistes est qu'il faut passer d'une logique de réparation du corps humain à une logique d'amélioration de manière à éradiquer définitivement la
maladie, voire le vieillissem
sont émus publiquement d
recherche n'était pas encac

En France, des personnalite
risque de dérive éthique et
Le futurologue Joël de Ross
Pour lui, l'homme au cervea
l'homme a fait pour son pro
son quotidien.

Le débat économique : création ou destruction d'emplois ?

Robot avocat, journaliste, enquêteur, médecin, chauffeur... Toute tâche qui nécessite de traiter une grande masse d'informations à synthétiser, recouper pour identifier, calculer, déduire ou encore décider est inévitablement menacée de robotisation. Les dernières
études estiment que 10% des emplois
tâches) de substitution par les intellig
vu de manière plus ou moins optimist

Pour Daniel Cohen ou encore Jéremy
d'emplois car le substitut de valeur se
services, rognant ainsi les emplois int
rappelle que, comme le furent l'invent
artificielle est une « *general purpose t*
l'économie. Il estime donc que ce n'es
pourra juger des innovations secondai

Le débat juridique : faut-il une personnalité juridique à l'IA ?

Les partisans de la création d'une personnalité juridique font le parallèle avec la personnalité morale des entreprises, qui n'exclut pas celle des personnes physiques.
Pour d'autres, faire payer les robots, c'est
vraie question semble être aujourd'hui
intelligences artificielles.

Laurence Devillers édicte 11 commande
traçabilité des algorithmes (« *tu pourra*
le demande ») et leur contrôle (« *tu se*
as appris »).

Le débat sociétal : que devient le travail dans un monde de robots ?

La fragmentation du travail intellectuel en tâches, déjà entamée par l'ubérisation via le modèle des plateformes, va s'accélérer avec la robotisation. Les experts s'accordent à dire que la logique de poste de travail où des salariés en CDI sont rémunérés en fonction du nombre d'heures accomplies n'a plus de sens dans ce monde post-révolutions numériques. Les métiers vont devoir tous évoluer pour apprendre à travailler avec l'IA, d'où un enjeu crucial de formation et d'accompagnement des salariés.

Jean-Gabriel Ganascia évoque une évolution du travail au sens initial « labor » vers la notion chère à Hannah Arendt de travail « œuvre », où l'homme va exercer toute sa puissance créatrice et émotionnelle. Identifier, reconnaître les différentes formes de contribution sociétale (accompagner un proche malade, être membre d'une association, prendre des nouvelles des enfants d'un client...) et les valoriser seront des sujets majeurs, malheureusement trop vite évacués par la solution politique d'un revenu universel.

L'Éthique de l'IA : Questions & Dilemmes

https://octopeek.com/fr/lethique-de-lia-questions-dilemmes/

Le Biais Systémique dans les Modèles d'IA

L'un des plus gran
celui des différent
souvent des consi
décisions sont bia
raisonnement et

En pratique, un al
(entrée, sortie) afi
sortie souhaités. I
(c'est-à-dire de dé
Les biais découler
données utilisées
conception des al

Pour illustrer ce p
comportement bi
"asian girls » renv
renvoie principale
Un autre exemple
performances pré
lorsqu'il s'agit d'id
Si cette anomalie
dévastatrices dan
telles que la vidéc
l'homme justifian

Le Dilemme du Tramway dans les Véhicules Autonomes

Alors que de pl
qui contrôlent
au dilemme du

L'analogue d'ur
mortelle ne pe
des occupants

Quelle est donc
décision doit-e
plus élevé) plu

Naturellement,
l'utilitarisme es
l'utilité, l'utilité
accident inévit
de noter que m
de la définition
l'utilité totale, l

À l'heure actue
facteurs écono
peut-il espérer
propre voiture
décision est dé

La Créativité Artificielle

Il existe une catégorie d'algorith
de nouvelles données selon une
n'a pas fallu longtemps pour que
ouvrant ainsi la voie à l'émergen

Dans le monde des arts, une IA g
artiste afin d'apprendre à réalise
artistique dudit artiste. En 2016,
surnommé « the Next Rembrand
l'apprentissage, plusieurs siècles

Une fois l'effet de surprise estom
désigné comme l'auteur d'une te
modèle ? Ou est-ce l'artiste origi
ne se limite pas au monde des ar
exemple, en 2019, l'Office Europ
désignée comme l'inventeur. Ce
de la propriété intellectuelle, et s
pouvons-nous actualiser notre d
concernées : l'IA, les personnes

En outre, l'art de l'IA est-il une m
limites de la créativité humaine
comprendre la nature de l'implic
préserver la valeur de l'art dans

Accusé, levez-vous! L'AI dans le Tribunal

Un autre cas intéressant d'application de l'IA se trouve dans l
algorithmes pour évaluer les cas, remplaçant essentiellemen
technologie de l'intelligence artificielle dans des domaines a
une valeur tangible certes, principalement sous la forme d'u
de questions éthiques qui méritent d'être explorées.

Le système judiciaire est lent, et cela peut être attribué à plu
processus juridiques, le volume élevé de dossiers juridiques
quelques-unes. Cela crée une opportunité pour l'utilisation d
d'automatisation des tâches répétitives et chronophages pou
premier juge artificiel; un personnage féminin numérique dot
faciales pour gérer les tâches répétitives de base dans un no
toutefois pas la seule cible de ces solutions.
Les algorithmes et les outils d'analyse de données sont égale
accélérer la recherche juridique et miner des informations pr
n'ont pas les moyens d'engager un avocat, DoNotPay est là p
autoproclamé est un conseiller juridique doté d'une intelligen
relevant des cours de petites créances.

L'un des arguments les plus importants utilisés pour défendr
capacité de cette technologie à être "neutre » dans son juge
neutre, étant donné qu'elle souffre de biais intégrés et de cor
essentiellement nos propres préjugés, comme nous l'avons e
Un autre problème imposant est le manque de transparence
souvent considérés comme une boîte noire. Même si ces rés
comprendre le pourquoi de leurs prédictions pose problème,
permettre de faire aveuglément confiance à une machine lor
un jugement sans comprendre quels facteurs ont contribué à
Enfin, l'entraînement de ces modèles s'appuie sur des protoc
ensembles de données massifs, ce qui soulève également de

Désinformation et Fake News

« It is only when they go wrong that machines remind you how powerful they are. »
Clive James

La désinformation a toujours été utilisée comme une arme politique pour manipuler les peuples. Aujourd'hui, elle est encore amplifiée par l'essor des modèles spécialisés d'apprentissage profond qui peuvent produire un nouveau texte, trafiquer des images, ou même falsifier des discours. Ces algorithmes ont atteint un niveau d'avancement où un humain parviendrait difficilement à distinguer leurs créations de signaux réels.
Le modèle GPT-3 d'OpenAI, qui signifie Generative Pre-trained Transformer, est un énorme réseau de neurones qui a été entraîné sur un corpus massif de textes afin de capturer les régularités statistiques du langage naturel. Le modèle pouvait ensuite être facilement utilisé pour générer des paragraphes de texte après avoir reçu quelques mots initiaux.
Il n'a pas fallu longtemps pour que GPT-3 soit exploité pour produire toutes sortes de fake news, les unes aussi plausibles que les autres, ce qui a déclenché de nombreuses sonnettes d'alarme concernant cette horrible utilisation de la technologie et ses éventuelles répercussions.
En ce qui concerne les images et les vidéos, un développement récent de la recherche sur l'apprentissage profond a permis de falsifier des vidéos en remplaçant le visage d'une personne source dans une vidéo par le visage d'une personne cible dans une image.
Cette technologie, appelée DeepFake, nous permet essentiellement de mettre de nouveaux mots dans la bouche d'une personnalité publique, de jouer la star dans notre film préféré ou d'accomplir toute autre application que vous pouvez imaginer. Et si ce n'était pas suffisant, elle peut également s'étendre aux signaux audio, ce qui permet de tromper facilement des personnes au téléphone.
En 2019, un criminel a utilisé un deepFake pour imiter la voix du PDG d'une entreprise énergétique allemande afin d'exiger un virement bancaire frauduleux de 220 000 €, dans ce qui serait le premier cybercrime de l'histoire basé sur l'IA. Bien que techniquement intrigants, les DeepFakes introduisent une nouvelle série de problèmes sociaux, car ils discréditent complètement l'ouïe et la vue, les deux sens les plus fiables sur lesquels les humains s'appuient pour prendre des décisions.
Ces modèles d'apprentissage profond peuvent être utilisés contre les femmes en intégrant synthétiquement leur visage dans du contenu pornographique, piéger quelqu'un en le plaçant sur la scène d'un crime qu'il n'a pas commis ou diffuser une fausse déclaration de guerre du président d'une nation puissante.

La prolifération de la désinformation sur l'internet est devenue un fléau alarmant qui exige une attention immédiate. Le processus fastidieux de création de faux contenus multimédias, autrefois réservé aux experts, est désormais accessible au public et sans aucun pré-requis. Les plateformes de réseaux sociaux sont parmi les parties les plus concernées, car elles constituent le principal moyen de communication incontrôlé qui permet aux fake news de devenir virales. C'est pourquoi Google, Facebook, Twitter et d'autres géants de la technologie mènent une guerre contre la désinformation et la propagande dangereuse, mais leurs efforts ont jusqu'à présent eu des effets limités.

SCIENCES • INTELLIGENCE ARTIFICIELLE

IA : « Des humains sont essentiels pour entraîner les systèmes »

Derrière les programmes d'intelligence artificielle, il y a tout un travail de collecte, de vérification et d'annotation des données effectué par des « petites mains ». Une activité essentielle en grande partie externalisée dans les pays en développement, notamment à Madagascar, expliquent deux jeunes sociologues, Maxime Cornet et Clément Le Ludec, dans un entretien au « Monde ».

Propos recueillis par David Larousserie

Publié le 22 juin 2024 à 17h00, modifié le 24 juin 2024 à 16h29 · Lecture 5 min. · Read in English

: vidéo

A Madagascar, les petites mains bien réelles de l'intelligence artificielle

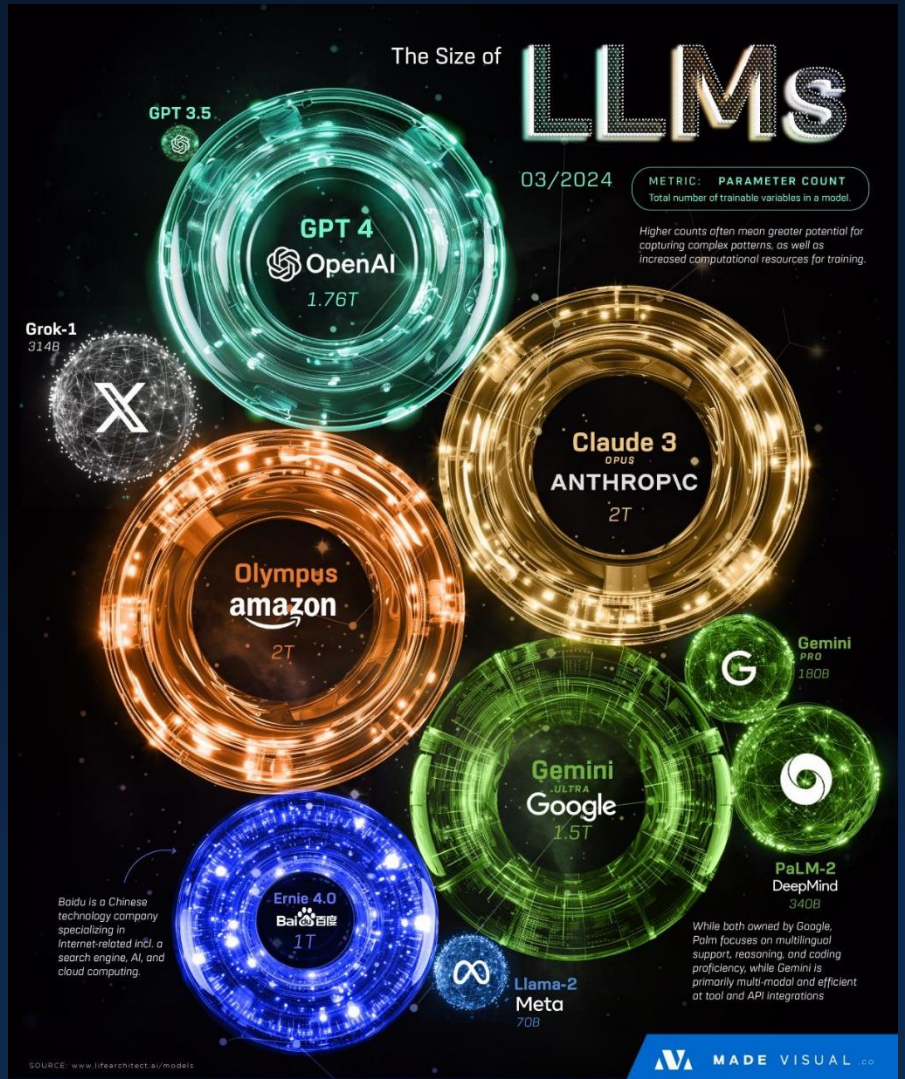
Publié le 29/04/2024 14:21

Temps de lecture : 10min - vidéo : 6min

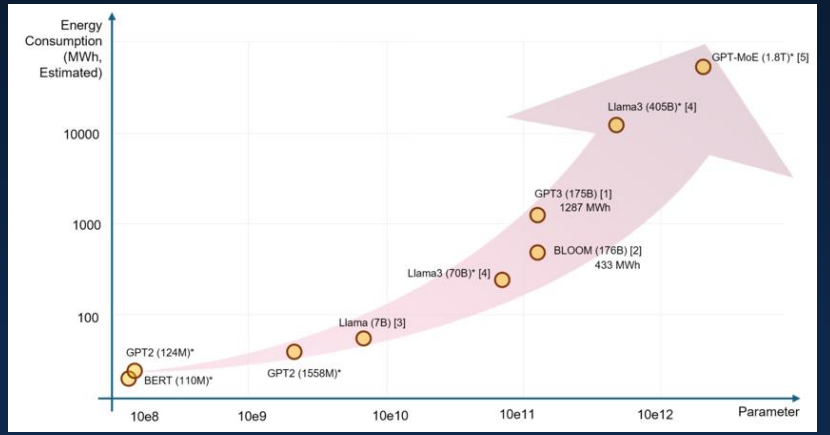
L'Oeil du 20 heures

France Télévisions

Un billion : mille millions ($1000 \times 10^6 = 10^9$)
Un trillion : mille billions ($1000 \times 10^9 = 10^{12}$)



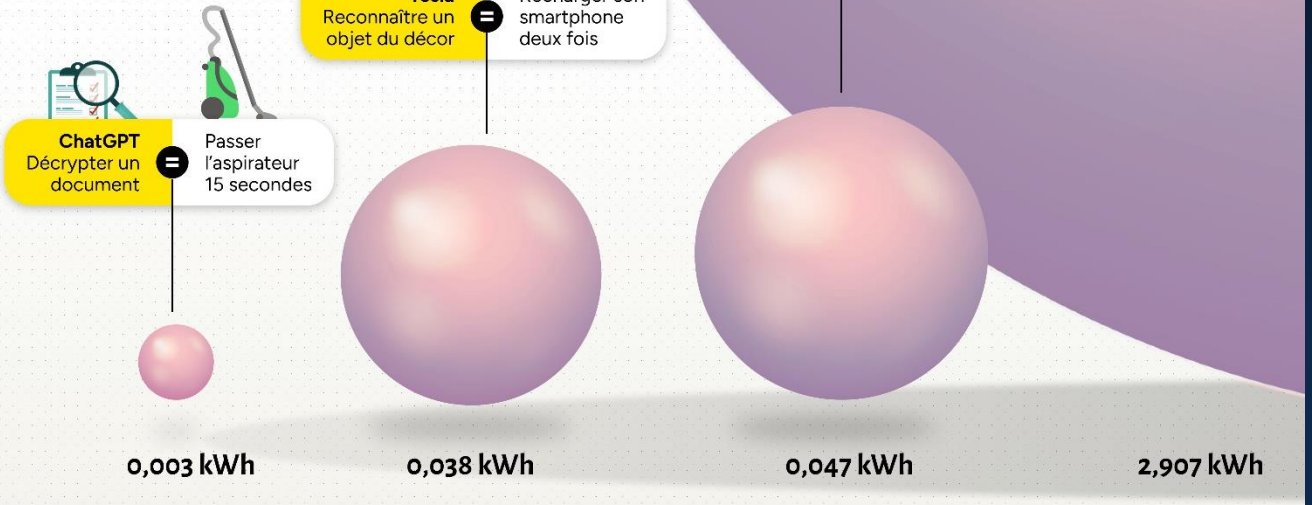
https://www.reddit.com/r/Infographics/comments/1c81n2o/the_size_of_llms_apr_2024/?rdt=60519



https://www.researchgate.net/figure/Reported-energy-consumption-of-training-different-LLM-models-with-respect-to-model_fig5_384115745

Le coût de l'usage massif de l'intelligence artificielle

Que pèsent **1000 requêtes à une IA générative** (parmi des milliards effectuées chaque jour) comparées à nos consommations d'énergie habituelles ?



Sources : Hugging Face, Carnegie Mellon University, Futura sciences

basta! basta.media

https://basta.media/ia-generative-a-le-potentiel-de-detruire-la-planete-mais-pas-comme-vous-le

IA, datacenters et cryptomonnaies font exploser la consommation mondiale d'électricité

Dans quelles proportions **la consommation d'électricité exclusivement liée à l'IA, aux centres de données et aux cryptos** augmente-t-elle dans le monde ?

2022 460 TWh

En 2022, cette consommation était **équivalente à un pays comme la France**

2026 (scénario bas)

+35% (620 TWh)

Soit **l'équivalent de 1,3 France en plus** dans le monde

2026 (scénario haut)

+130% (1 050 TWh)

Soit **l'équivalent de 2,3 France en plus** dans le monde

Source : International Energy Agency, janvier 2024

basta! basta.media

https://basta.media/ia-generative-a-le-potentiel-de-detruire-la-planete-mais-pas-comme-vous-le

IMT Mines Alès
École Mines-Télécom

SCIENCE DES DONNÉES - NIVEAU 1

MACHINE LEARNING

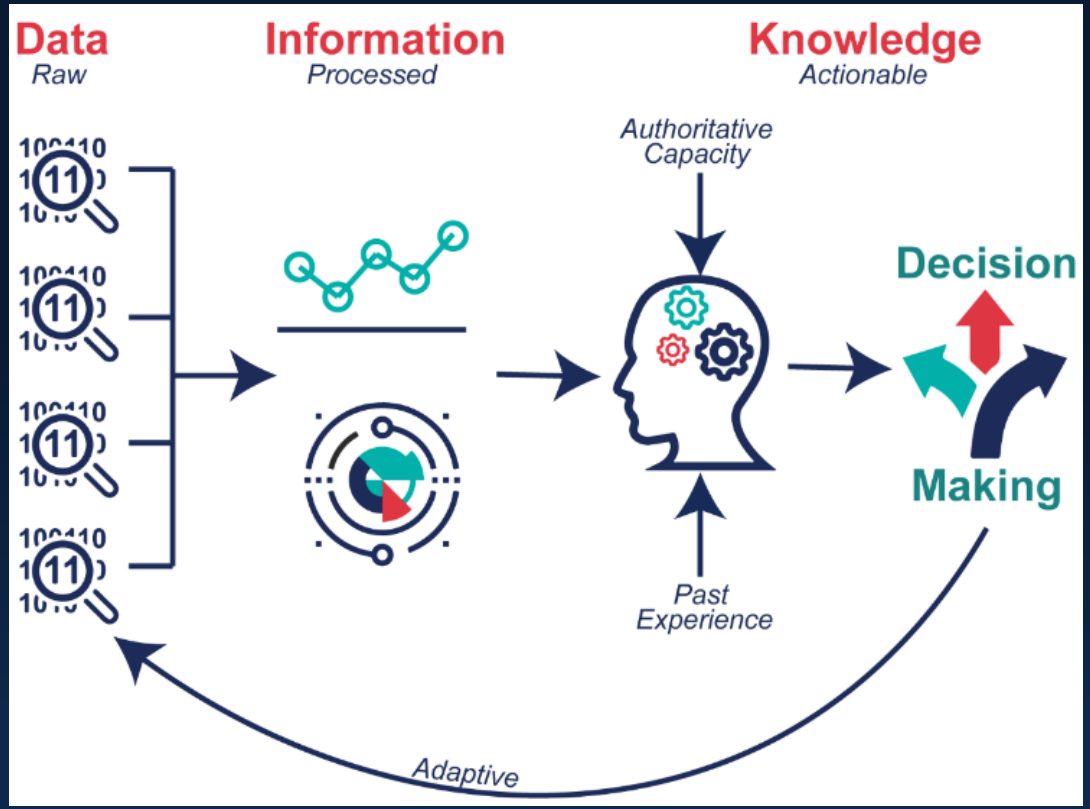
Gérard Dray

afaq ISO 9001 Qualité AFNOR CERTIFICATION

EuroMov Digital Health in Motion

Machine Learning Process

Data, Information, Knowledge, Decision



https://internetofwater.org/valuing-data/what-are-data-information-and-knowledge

Data

Data Frame (Observations x Features) of dimension (n x p)

		Features / Variables / Attributes / Characteristics / ...										
		Age	Income	Gender		Marital Status			...	Feature j	...	Feature p
		X^1	X^2	M	F	Single	Married	Widowed Divorced	...	X^j	...	X^p
				X^3	X^4	X^5	X^6	X^7				
Observations	X_1	x_1^1	x_1^2	x_1^3	x_1^4	x_1^5	x_1^6	x_1^7	...	x_1^j	...	x_1^p
	X_2	x_2^1	x_2^2	x_2^3	x_2^4	x_2^5	x_2^6	x_2^7	...	x_2^j	...	x_2^p
	...	...	...	...	...	...	...	...	...	...	...	...
	X_i	x_i^1	x_i^2	x_i^3	x_i^4	x_i^5	x_i^6	x_i^7	...	x_i^j	...	x_i^p
	...	...	...	...	...	...	...	...	...	...	...	...
	X_n	x_n^1	x_n^2	x_n^3	x_n^4	x_n^5	x_n^6	x_n^7	...	x_n^j	...	x_n^p

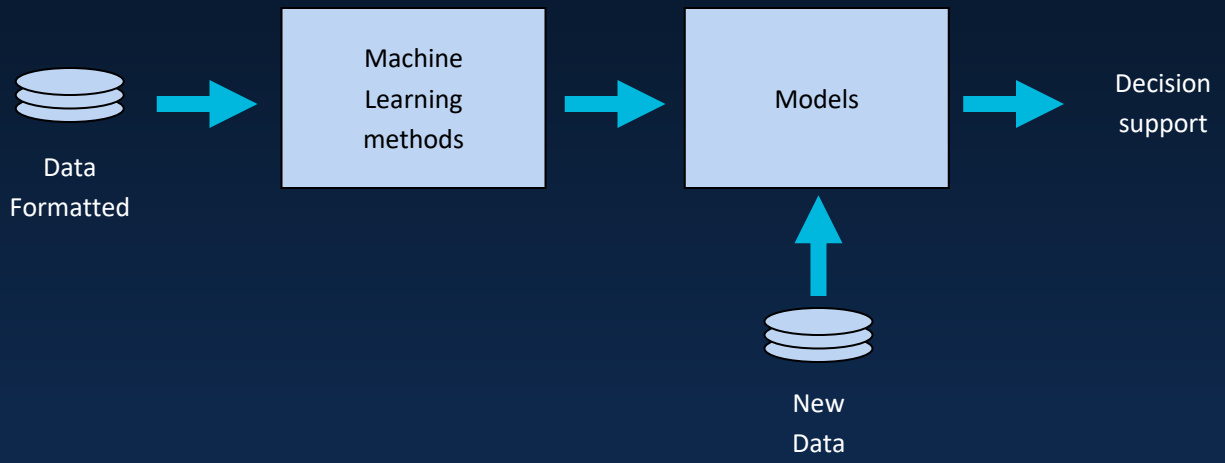
The features Age and Income are quantitative

The features Gender and Marital status are qualitative “One-Hot Encoding”

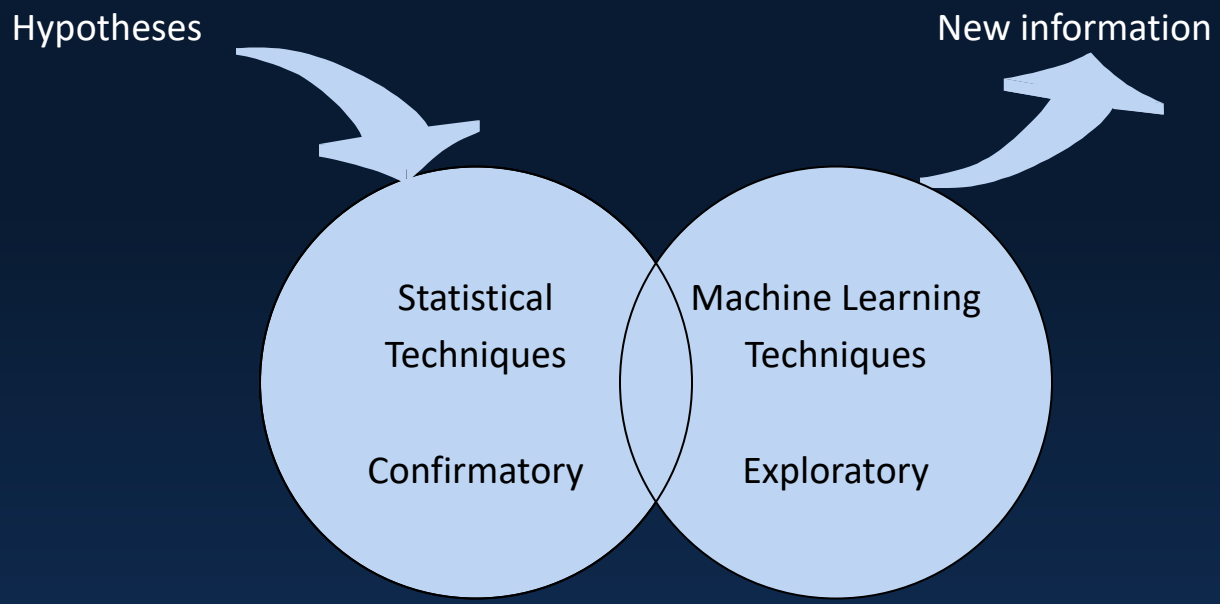
Modality of the features Marital status: (Single, Married, Widowed or Divorced)

The features : X^3, X^4, X^5, X^6, and X^7 are Boolean (1 = true, 0 = false)

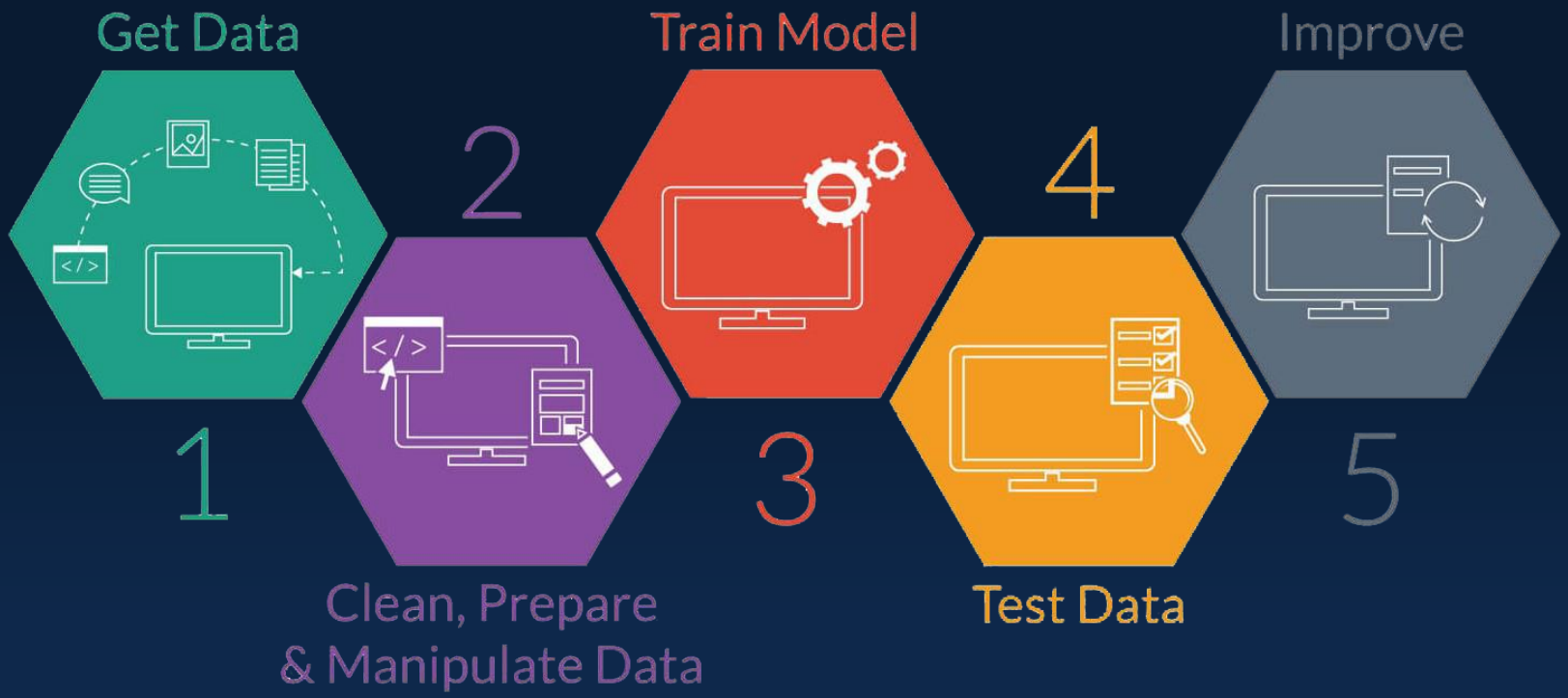
Data-driven induction process



Statistics vs. Machine Learning

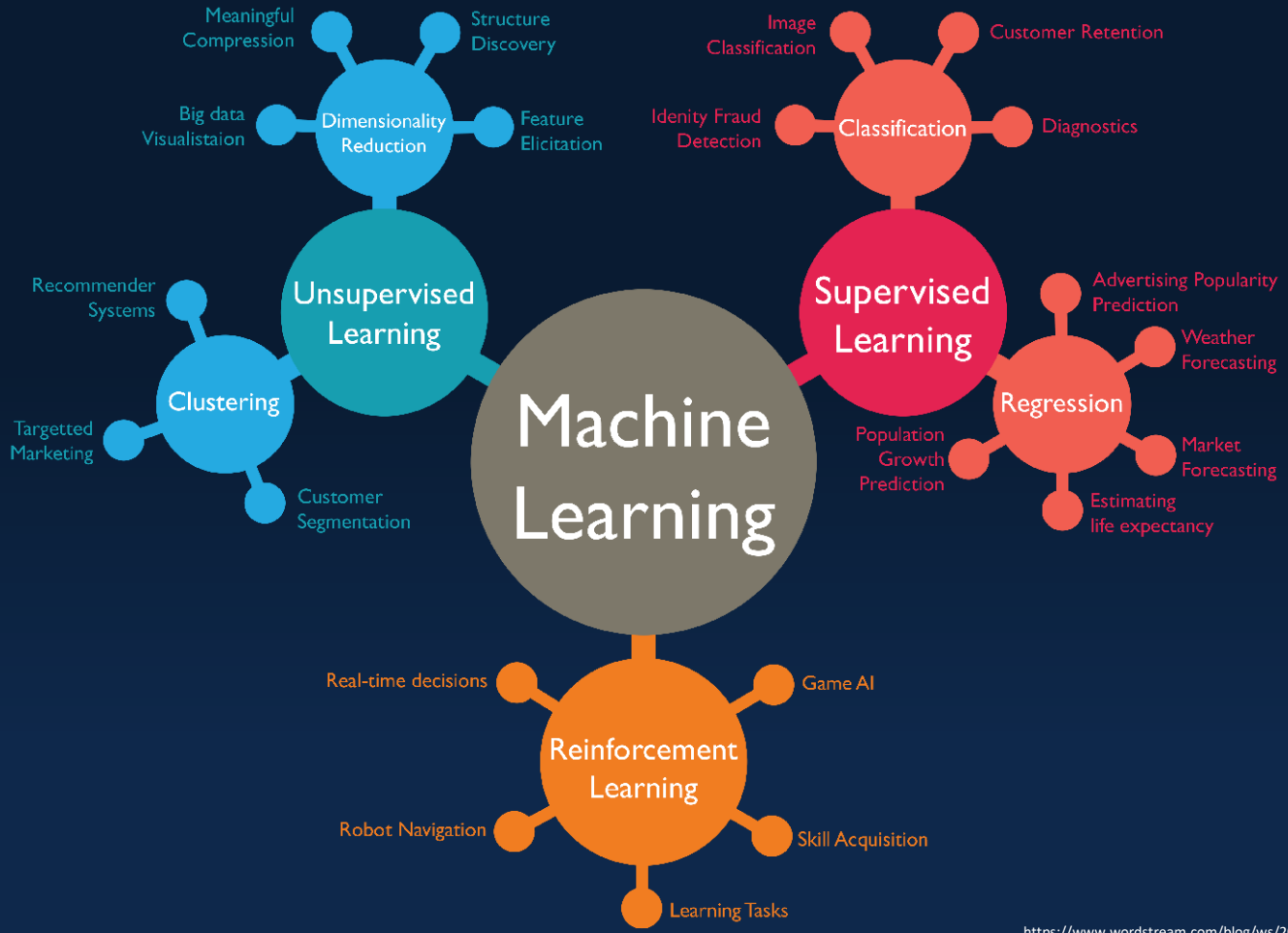


Key stages of machine learning



https://towardsdatascience.com/wtf-is-machine-learning-a-quick-guide-39457e49c65b

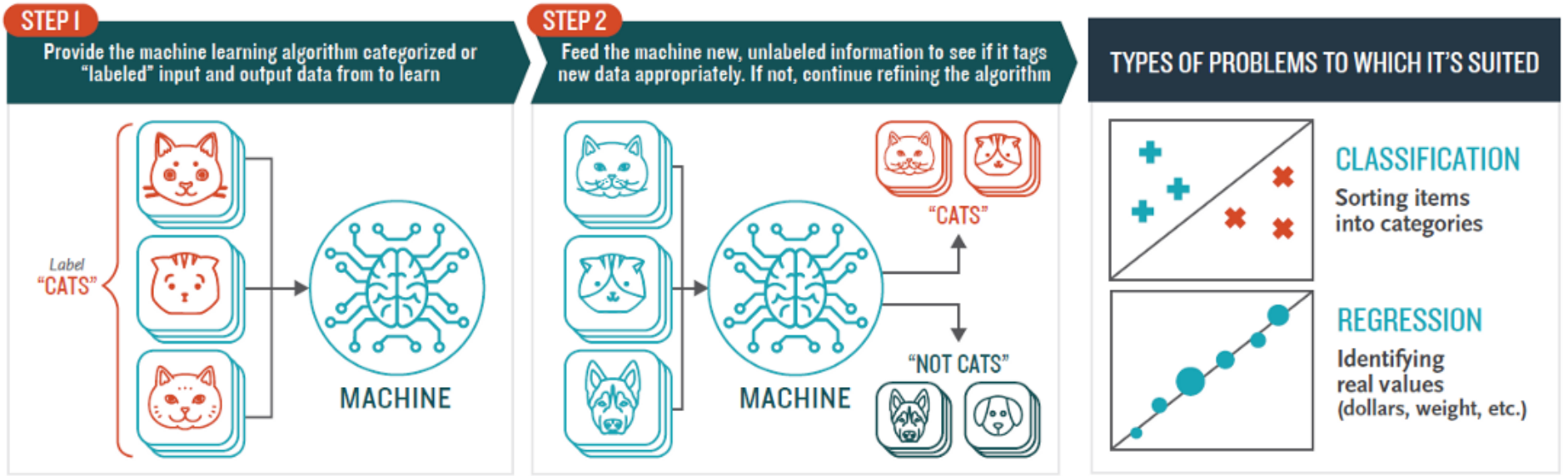
Machine Learning approaches



https://www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications

Supervised learning

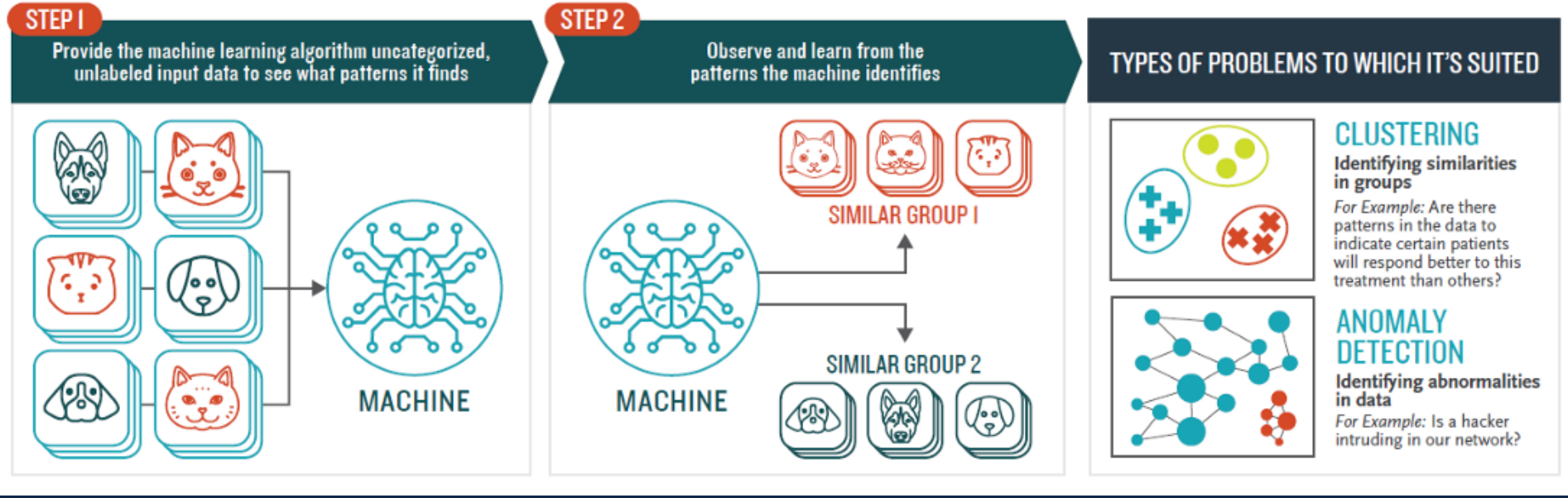
In supervised learning, we provide the data with the desired outcome (i.e., labeled data). For example, if we want our system to learn to detect cats, we will collect thousands of images, draw a bounding box around the cat, and provide the data set to the machine so that it can learn on its own.



https://towardsdatascience.com/wtf-is-machine-learning-a-quick-guide-39457e49c65b

Unsupervised learning

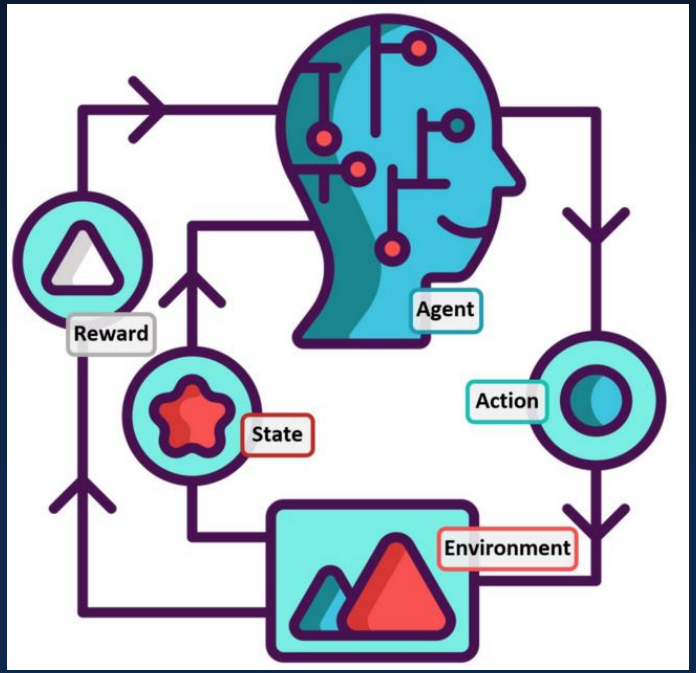
In unsupervised learning, we simply provide data and let the machine find patterns in the dataset. For example, we can provide three different shapes (circles, triangles, and squares) and let the machine group them together. Such a technique is called clustering.



https://towardsdatascience.com/wtf-is-machine-learning-a-quick-guide-39457e49c65b

Reinforcement learning

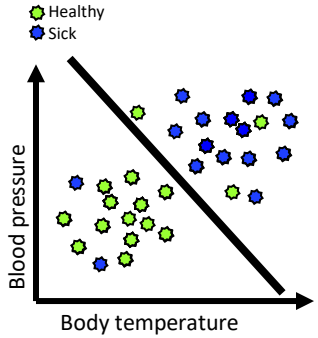
Reinforcement learning uses learning algorithms that learn from repeated experiences through trial and error. It thus replicates the "natural" mechanism of knowledge acquisition. Robots, chatbots, autonomous cars—its applications are numerous in artificial intelligence.



https://towardsdatascience.com/reinforcement-learning-fda8ff535bb6

Supervised Machine Learning: Classification

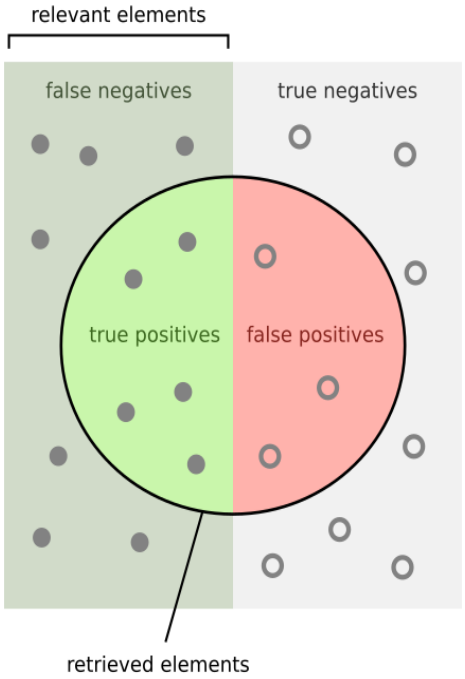
Machine learning: Classification



Confusion Matrix

		Predicted Classes: Positive	Predicted Classes: Negative
Real Classes	Positive	True Positives (TP)	False Negatives (FN)
Real Classes	Negative	False Positives (FP)	True Negatives (TN)

		Predicted Classes: Sick	Predicted Classes: Healthy
Real Classes	Sick	6	2
Real Classes	Healthy	3	5



How many retrieved items are relevant?

Precision = $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$

How many relevant items are retrieved?

Recall = $\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$

Classical algorithms:

- K-nearest neighbors
- Naive Bayes
- Decision tree
- Logistic regression
- Support vector machines
- Neural networks
- …

$$\text{Precision} = \frac{6}{6 + 3} = 6/9 = 0.67$$

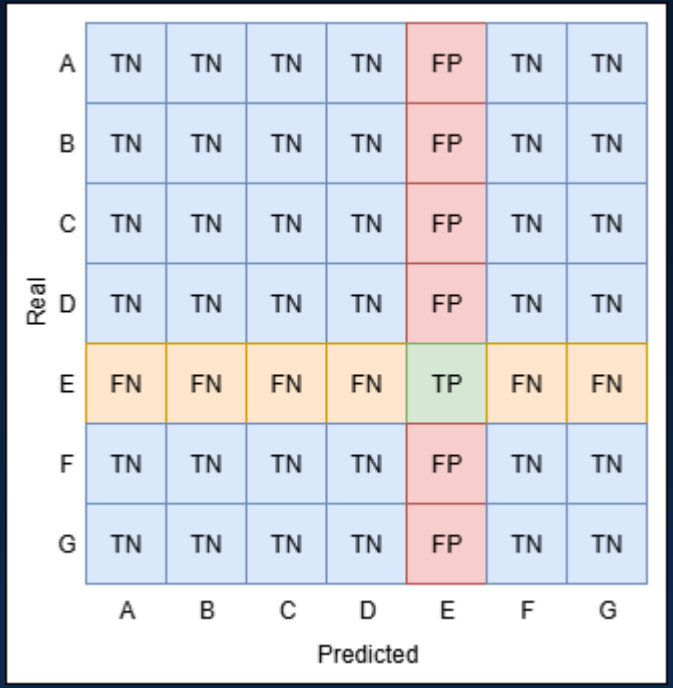
$$\text{Recall} = \frac{6}{6 + 2} = 6/8 = 0.75$$

F1-Measure / F1 Score

$$\frac{2 \times Precision \times Recall}{Precision + Recall} = 0.71$$

https://en.wikipedia.org/wiki/Precision_and_recall

Confusion Matrix



		Predicted Classes: Positive	Predicted Classes: Negative		
Real Classes	Positive	True Positives (TP)	False Negatives (FN)	Sensitivity / TP rate / Recall $\frac{TP}{TP+FN}$	
	Negative	False Positives (FP)	True Negatives (TN)	Specificity $\frac{TN}{TN+FP}$	FP Rate $\frac{FP}{TN+FP}$
		Precision $\frac{TP}{TP+FP}$	TN Rate $\frac{TN}{TN+FN}$		

Accuracy $\frac{TP+TN}{TP+TN+FP+FN}$

F1-Measure / F1 Score $\frac{2 \times Precision \times Recall}{Precision + Recall}$

https://en.wikipedia.org/wiki/Confusion_matrix

Bias-Variance Tradeoff

When we talk about prediction using machine learning models, it's important to understand prediction errors (i.e., bias and variance). The goal of any machine learning model is to find a model that minimizes the prediction errors on unseen data. There's a tradeoff of a model's ability to minimize prediction errors between bias and variance. Understanding these concepts would help us address the issues of overfitting and underfitting.

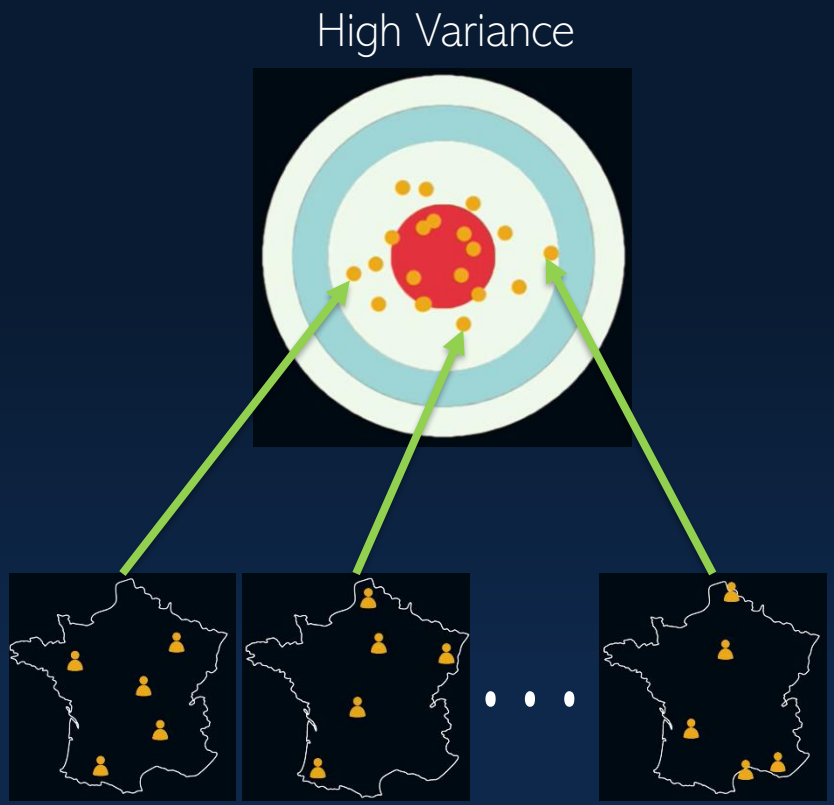
How many kilometers do you travel each day?

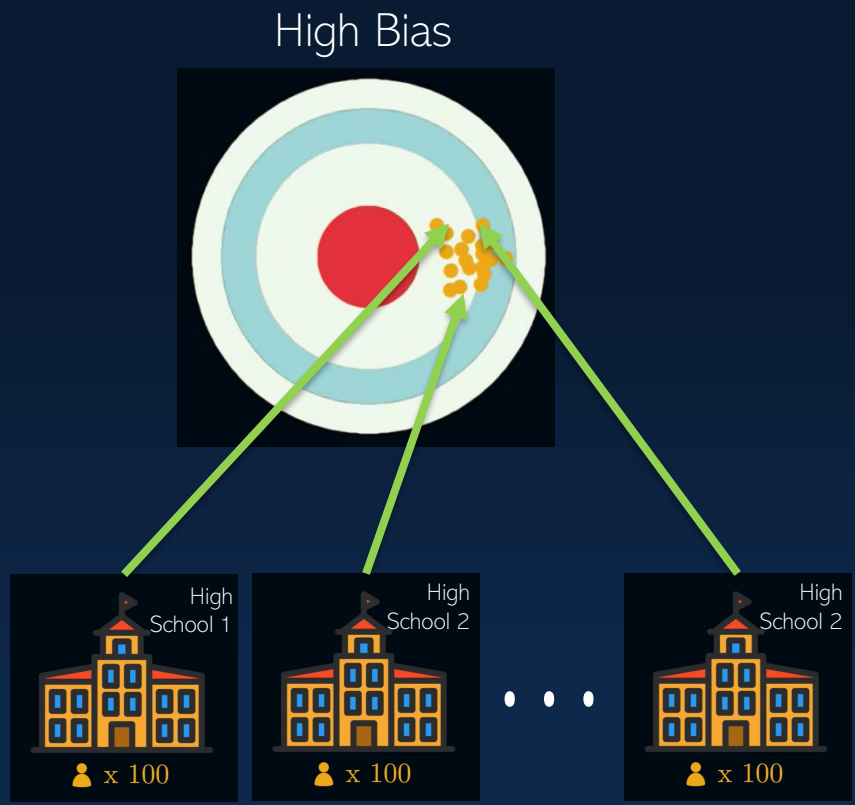
Estimation 1 : 22 km

Estimation 2 : 7.1 km

https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://youtu.be/bhJgOrnvqp0

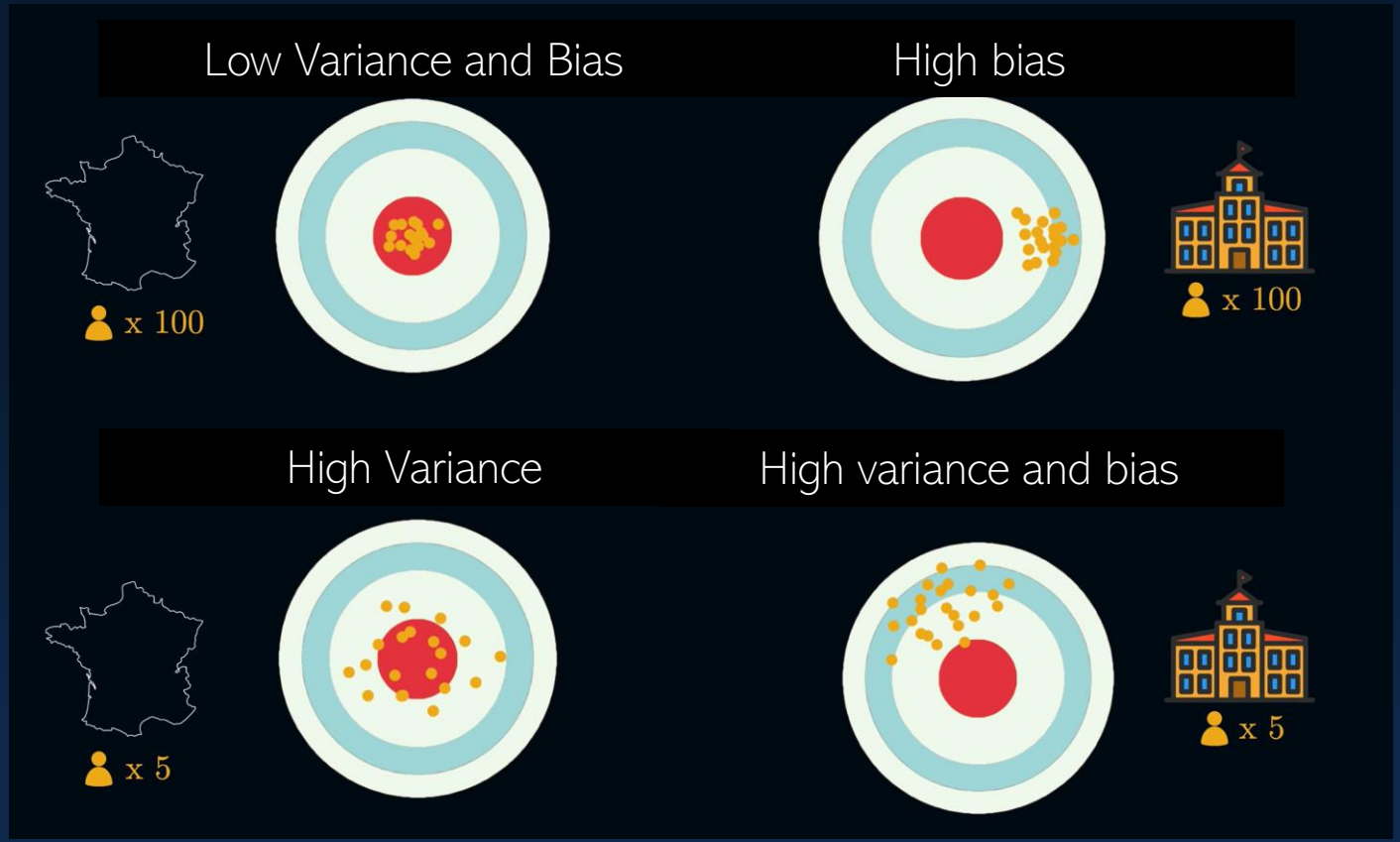
Bias-Variance Tradeoff





https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://youtu.be/bhJgOrnvqp0

Bias-Variance Tradeoff



https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://youtu.be/bhJgOrnvqp0

Bias-Variance Tradeoff

What is bias?

Bias is the difference between our model's average prediction and the correct value we are trying to predict. A model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on both the training and test data.

What is variance?

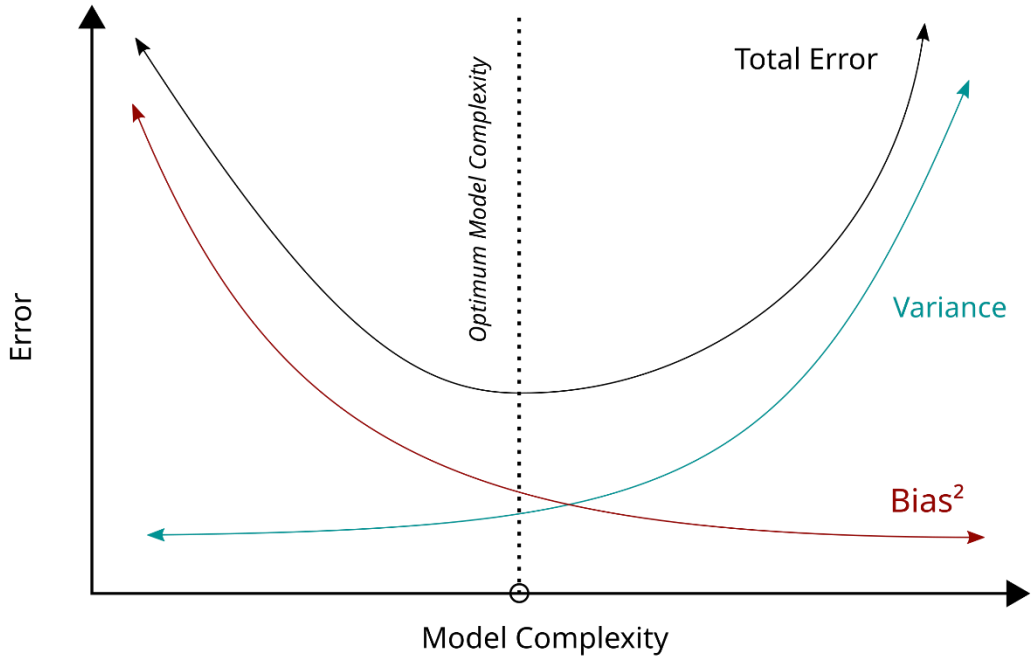
Variance is the variability of the model's prediction, which tells us how spread out our data is. Models with high variance pay close attention to the training data and do not generalize to data they have not seen before. As a result, these models perform very well on the training data but have high error rates on the test data.

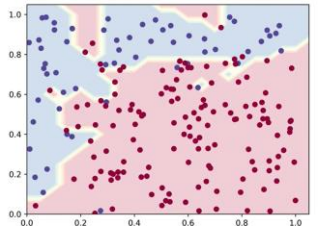
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://youtu.be/bhJgOrnvqp0

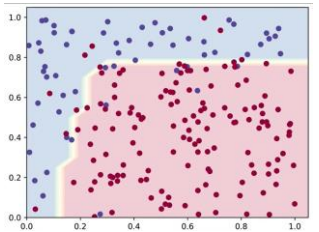
Bias-Variance Tradeoff

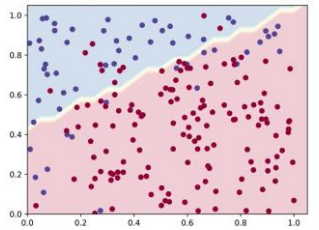
$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2$$

$$Err(x) = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$







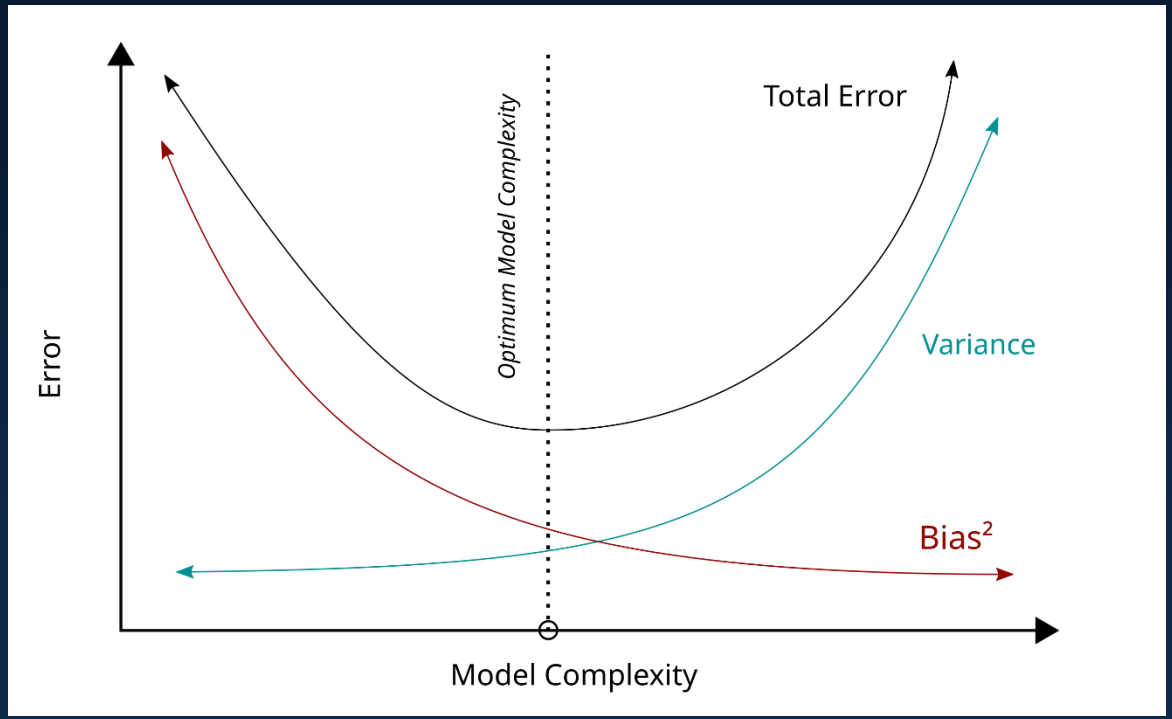


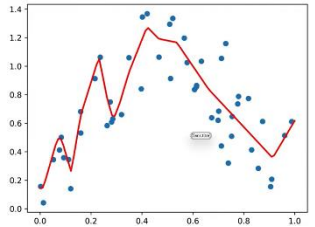
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229

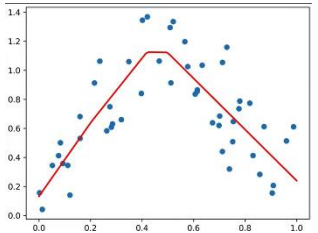
Bias-Variance Tradeoff

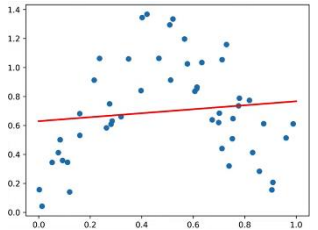
$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] + \sigma_e^2$$

$$Err(x) = \text{Bias}^2 + \text{Variance} + \text{Irreducible Error}$$



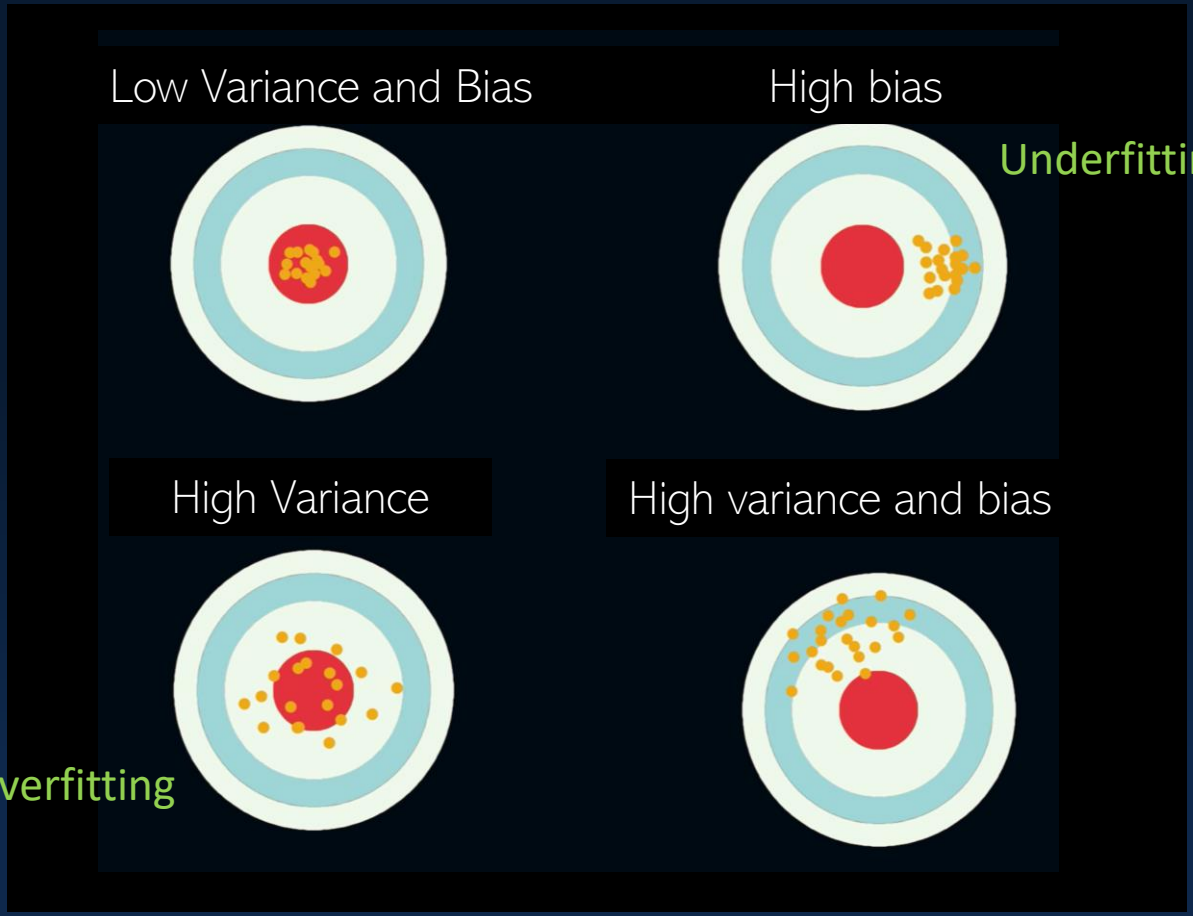






https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229

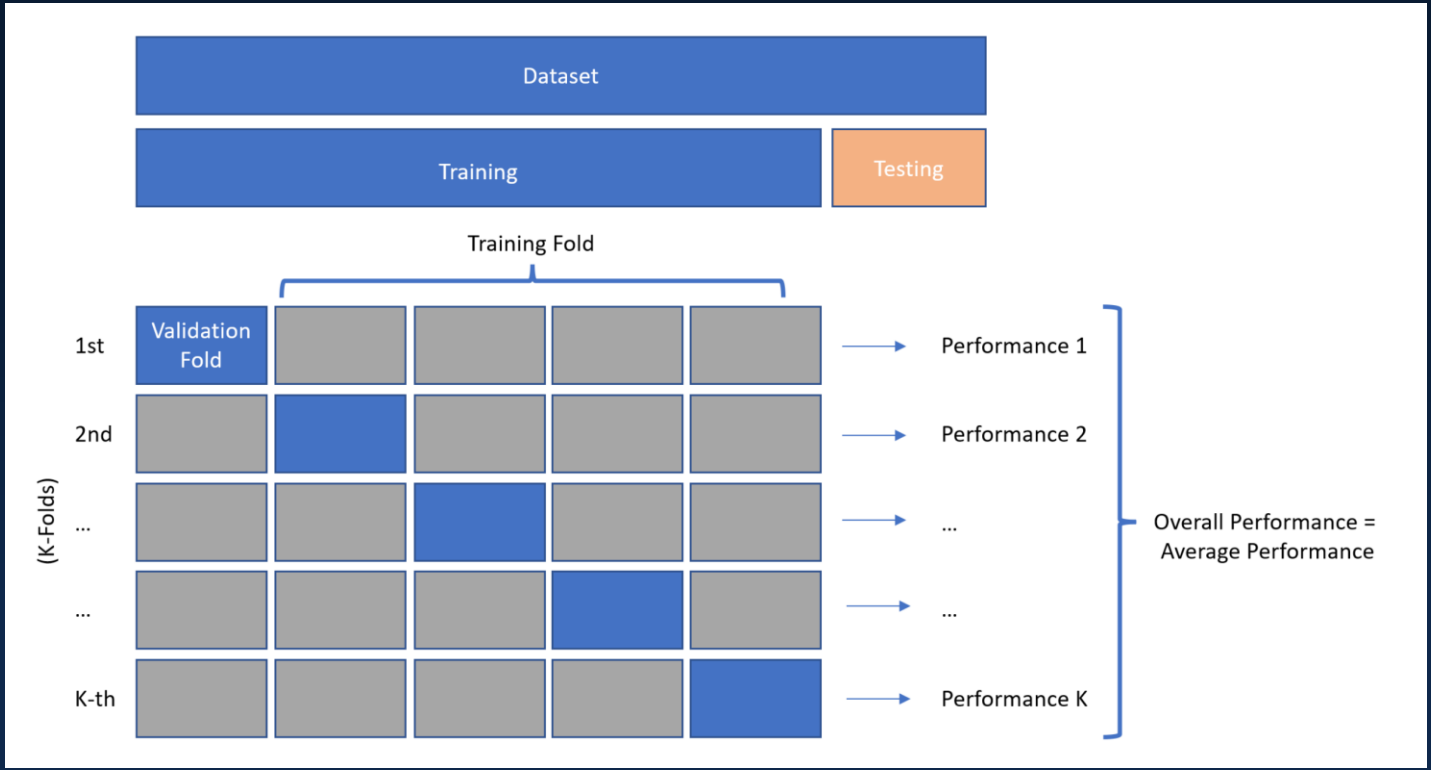
Bias-Variance Tradeoff



https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
https://youtu.be/bhJgOrnvqp0

Bias-Variance Tradeoff

K-Fold Cross-Validation is a popular tool used in training a machine learning model. The idea of this method is simple, we create a multiple set of training data from the observed data, then train the model and evaluate the model based on the validation data, which is similar to evaluating the model on the unseen data.



https://towardsdatascience.com/machine-learning-bias-variance-tradeoff-and-regularization-94846f945131/

Apprentissage automatique : Classification

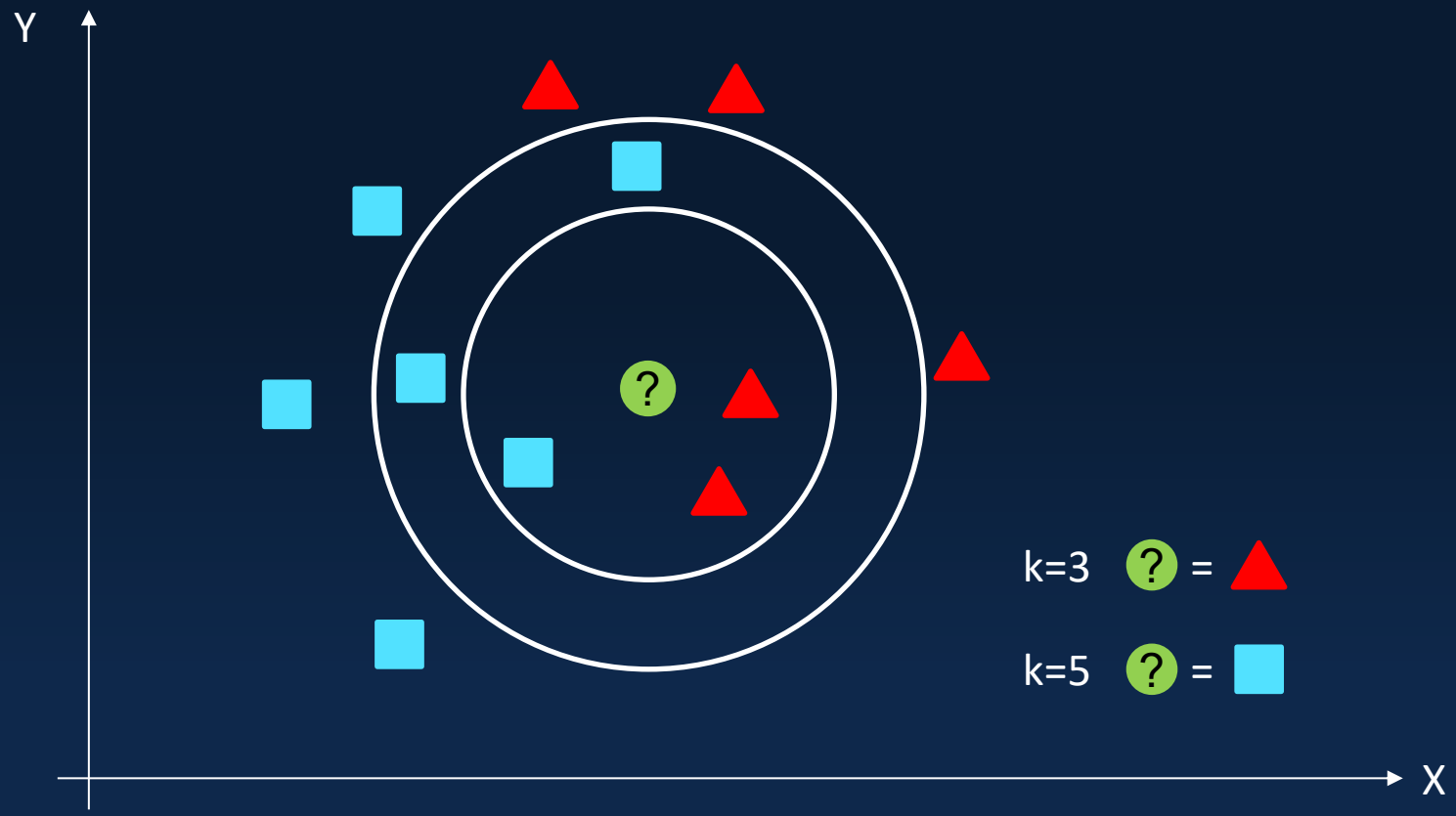
Am I going to play tennis?

Day	Outlook	Temperature °C	Temperature	Humidity %	Humidity	Windy	Play
1	Sunny	29	Hot	85	High	False	No
2	Sunny	27	Hot	90	High	True	No
3	Overcast	28	Hot	86	High	False	Yes
4	Rainy	21	Mild	96	High	False	Yes
5	Rainy	20	Cool	80	Normal	False	Yes
6	Rainy	18	Cool	70	Normal	True	No
7	Overcast	18	Cool	65	Normal	True	Yes
8	Sunny	22	Mild	95	High	False	No
9	Sunny	21	Cool	70	Normal	False	Yes
10	Rainy	24	Mild	80	Normal	False	Yes
11	Sunny	24	Mild	70	Normal	True	Yes
12	Overcast	22	Mild	90	High	True	Yes
13	Overcast	27	Hot	75	Normal	False	Yes
14	Rainy	22	Mild	91	High	True	No
15	Sunny	19	Cool	75	Normal	True	?
16	Overcast	22	Mild	89	Normal	False	?

Supervised Machine Learning: Classification

Method: k-Nearest Neighbors

Machine learning: Classification
k Nearest Neighbors (kNN)



Machine learning: Classification
k Nearest Neighbors (kNN)

Day	Outlook	Temperature	Humidity	Windy	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature Hot	Temperature Mild	Temperature Cool	Humidity High	Humidity Normal	Windy True	Windy False	Play
1	Sunny	Hot	High	False	1	0	0	1	0	0	1	0	0	1	No
2	Sunny	Hot	High	True	1	0	0	1	0	0	1	0	1	0	No
3	Overcast	Hot	High	False	0	1	0	1	0	0	1	0	0	1	Yes
4	Rainy	Mild	High	False	0	0	1	0	1	0	1	0	0	1	Yes
5	Rainy	Cool	Normal	False	0	0	1	0	0	1	0	1	0	1	Yes
6	Rainy	Cool	Normal	True	0	0	1	0	0	1	0	1	1	0	No
7	Overcast	Cool	Normal	True	0	1	0	0	0	1	0	1	1	0	Yes
8	Sunny	Mild	High	False	1	0	0	0	1	0	1	0	0	1	No
9	Sunny	Cool	Normal	False	1	0	0	0	0	1	0	1	0	1	Yes
10	Rainy	Mild	Normal	False	0	0	1	0	1	0	0	1	0	1	Yes
11	Sunny	Mild	Normal	True	1	0	0	0	1	0	0	1	1	0	Yes
12	Overcast	Mild	High	True	0	1	0	0	1	0	1	0	1	0	Yes
13	Overcast	Hot	Normal	False	0	1	0	1	0	0	0	1	0	1	Yes
14	Rainy	Mild	High	True	0	0	1	0	1	0	1	0	1	0	No
15	Sunny	Cool	Normal	True	1	0	0	0	0	1	0	1	1	0	?
16	Overcast	Mild	Normal	False	0	1	0	0	1	0	0	1	0	1	?

Coding of nominal features (1)
“One-Hot Encoding”

Machine learning: Classification
k Nearest Neighbors (kNN)

Day	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature Hot	Temperature Mild	Temperature Cool	Humidity High	Humidity Normal	Windy True	Windy False	Play	Dist. Hamming Day 15
1	1	0	0	1	0	0	1	0	0	1	No	6
2	1	0	0	1	0	0	1	0	1	0	No	4
3	0	1	0	1	0	0	1	0	0	1	Yes	8
4	0	0	1	0	1	0	1	0	0	1	Yes	8
5	0	0	1	0	0	1	0	1	0	1	Yes	4
6	0	0	1	0	0	1	0	1	1	0	No	2
7	0	1	0	0	0	1	0	1	1	0	Yes	2
8	1	0	0	0	1	0	1	0	0	1	No	6
9	1	0	0	0	0	1	0	1	0	1	Yes	2
10	0	0	1	0	1	0	0	1	0	1	Yes	6
11	1	0	0	0	1	0	0	1	1	0	Yes	2
12	0	1	0	0	1	0	1	0	1	0	Yes	6
13	0	1	0	1	0	0	0	1	0	1	Yes	6
14	0	0	1	0	1	0	1	0	1	0	No	6

Hamming distance

$$D_{Hamming}(X,Y) = \sum_{i=1}^{k} |x_i - y_i| = |1-1| + |0-0| + |0-0| + |1-0| + |0-0| + |0-1| + |1-0| + |0-1| + |0-1| + |1-0| = 6$$

Day	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature Hot	Temperature Mild	Temperature Cool	Humidity High	Humidity Normal	Windy True	Windy False
1	1	0	0	1	0	0	1	0	0	1
15	1	0	0	0	0	1	0	1	1	0

Machine learning: Classification
k Nearest Neighbors (kNN)

Day	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature Hot	Temperature Mild	Temperature Cool	Humidity High	Humidity Normal	Windy True	Windy False	Play	Dist. Hamming Day 15
1	1	0	0	1	0	0	1	0	0	1	No	6
2	1	0	0	1	0	0	1	0	1	0	No	4
3	0	1	0	1	0	0	1	0	0	1	Yes	8
4	0	0	1	0	1	0	1	0	0	1	Yes	8
5	0	0	1	0	0	1	0	1	0	1	Yes	4
6	0	0	1	0	0	1	0	1	1	0	No	2
7	0	1	0	0	0	1	0	1	1	0	Yes	2
8	1	0	0	0	1	0	1	0	0	1	No	6
9	1	0	0	0	0	1	0	1	0	1	Yes	2
10	0	0	1	0	1	0	0	1	0	1	Yes	6
11	1	0	0	0	1	0	0	1	1	0	Yes	2
12	0	1	0	0	1	0	1	0	1	0	Yes	6
13	0	1	0	1	0	0	0	1	0	1	Yes	6
14	0	0	1	0	1	0	1	0	1	0	No	6

Hamming distance

$$D_{Hamming}(X,Y) = \sum_{i=1}^{k} |x_i - y_i| = |1-1| + |0-0| + |0-0| + |0-0| + |1-0| + |0-1| + |0-0| + |1-1| + |1-1| + |0-0| = 2$$

Day	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature Hot	Temperature Mild	Temperature Cool	Humidity High	Humidity Normal	Windy True	Windy False
11	1	0	0	0	1	0	0	1	1	0
15	1	0	0	0	0	1	0	1	1	0

Machine learning: Classification
k Nearest Neighbors (kNN)

Day 15

Day	Play	Dist. Hamming Day 15
6	No	2
7	Yes	2
9	Yes	2
11	Yes	2
2	No	4
5	Yes	4
1	No	6
8	No	6
10	Yes	6
12	Yes	6
13	Yes	6
14	No	6
3	Yes	8
4	Yes	8
15	?	0

k = 1 ?
k = 3 ?
k = 5 ?

Machine learning: Classification
k Nearest Neighbors (kNN)

Day 16

Day	Play	Dist. Hamming Day 16
10	Yes	2
13	Yes	2
7	Yes	4
9	Yes	4
11	Yes	4
5	Yes	4
8	No	4
12	Yes	4
3	Yes	4
4	Yes	4
6	No	6
1	No	6
14	No	6
2	No	8
16	?	0

k = 1 ?
k = 3 ?
k = 5 ?

Machine learning: Classification
k Nearest Neighbors (kNN)

Day	Outlook	Windy	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature °C	Humidity %	Windy True	Windy False	Play
1	Sunny	False	1	0	0	29	85	0	1	No
2	Sunny	True	1	0	0	27	90	1	0	No
3	Overcast	False	0	1	0	28	86	0	1	Yes
4	Rainy	False	0	0	1	21	96	0	1	Yes
5	Rainy	False	0	0	1	20	80	0	1	Yes
6	Rainy	True	0	0	1	18	70	1	0	No
7	Overcast	True	0	1	0	18	65	1	0	Yes
8	Sunny	False	1	0	0	22	95	0	1	No
9	Sunny	False	1	0	0	21	70	0	1	Yes
10	Rainy	False	0	0	1	24	80	0	1	Yes
11	Sunny	True	1	0	0	24	70	1	0	Yes
12	Overcast	True	0	1	0	22	90	1	0	Yes
13	Overcast	False	0	1	0	27	75	0	1	Yes
14	Rainy	True	0	0	1	22	91	1	0	No
15	Sunny	True	1	0	0	19	75	1	0	?
16	Overcast	False	0	1	0	22	89	0	1	?

Coding of nominal features (2)

Machine learning: Classification
k Nearest Neighbors (kNN)

Day	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature	Humidity	Windy True	Windy False	Play	Dist. Euclidienne Day 15
1	1	0	0	29	85	0	1	No	14.21
2	1	0	0	27	90	1	0	No	17.00
3	0	1	0	28	86	0	1	Yes	14.35
4	0	0	1	21	96	0	1	Yes	21.19
5	0	0	1	20	80	0	1	Yes	5.48
6	0	0	1	18	70	1	0	No	5.29
7	0	1	0	18	65	1	0	Yes	10.15
8	1	0	0	22	95	0	1	No	20.27
9	1	0	0	21	70	0	1	Yes	5.57
10	0	0	1	24	80	0	1	Yes	7.35
11	1	0	0	24	70	1	0	Yes	7.07
12	0	1	0	22	90	1	0	Yes	15.36
13	0	1	0	27	75	0	1	Yes	8.25
14	0	0	1	22	91	1	0	No	16.34

Euclidean distance

$$D_{euclidienne}(X,Y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} = \sqrt{(1-1)^2 + (0-0)^2 + (0-0)^2 + (29-19)^2 + (85-75)^2 + (0-1)^2 + (1-0)^2} = 14.21$$

Day	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature	Humidity	Windy True	Windy False
1	1	0	0	29	85	0	1
15	1	0	0	19	75	1	0

Machine learning: Classification
k Nearest Neighbors (kNN)

Day	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature	Humidity	Windy True	Windy False	Play	Dist. Euclidienne Day 15
1	1	0	0	29	85	0	1	No	14.21
2	1	0	0	27	90	1	0	No	17.00
3	0	1	0	28	86	0	1	Yes	14.35
4	0	0	1	21	96	0	1	Yes	21.19
5	0	0	1	20	80	0	1	Yes	5.48
6	0	0	1	18	70	1	0	No	5.29
7	0	1	0	18	65	1	0	Yes	10.15
8	1	0	0	22	95	0	1	No	20.27
9	1	0	0	21	70	0	1	Yes	5.57
10	0	0	1	24	80	0	1	Yes	7.35
11	1	0	0	24	70	1	0	Yes	7.07
12	0	1	0	22	90	1	0	Yes	15.36
13	0	1	0	27	75	0	1	Yes	8.25
14	0	0	1	22	91	1	0	No	16.34

Euclidean distance

$$D_{euclidienne}(X,Y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} = \sqrt{(1-1)^2 + (0-0)^2 + (0-0)^2 + (24-19)^2 + (70-75)^2 + (1-1)^2 + (0-0)^2} = 7.07$$

Day	Outlook Sunny	Outlook Overcast	Outlook Rainy	Temperature	Humidity	Windy True	Windy False
11	1	0	0	24	70	1	0
15	1	0	0	19	75	1	0

Machine learning: Classification
k Nearest Neighbors (kNN)

Day 15

Day	Play	Dist. Euclidienne Day 15
6	No	5.29
5	Yes	5.48
9	Yes	5.57
11	Yes	7.07
10	Yes	7.35
13	Yes	8.25
7	Yes	10.15
1	No	14.21
3	Yes	14.35
12	Yes	15.36
14	No	16.34
2	No	17.00
8	No	20.27
4	Yes	21.19
15	?	0.00

k = 1 ?
k = 3 ?
k = 5 ?

Machine learning: Classification
k Nearest Neighbors (kNN)

Day 16

Day	Play	Dist. Euclidienne Day 16
12	Yes	1.73
14	No	2.83
2	No	5.48
8	No	6.16
3	Yes	6.71
4	Yes	7.21
1	No	8.19
5	Yes	9.33
10	Yes	9.33
13	Yes	14.87
9	Yes	19.08
11	Yes	19.21
6	No	19.52
7	Yes	24.37
16	?	0.00

k = 1 ?
k = 3 ?
k = 5 ?

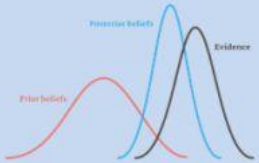
Supervised Machine Learning: Classification

Method: Naives Bayes

Machine learning: Classification
Naives Bayes Method

Naive Bayes

Supervised Learning Algorithm for Classification



Bayes Theorem :

$$P(B_i|A) = \frac{P(B_i) \times P(A|B_i)}{P(A)}$$

$$= \frac{P(B_i) \times P(A|B_i)}{\sum_{j=1}^{k} P(B_j) \times P(A|B_j)}$$

Thomas Bayes 1702-1762

Probabilistic model

$$P(C_k|x) = \frac{P(C_k) \times P(x|C_k)}{P(x)} = \propto p(C_k, x_1, x_2, \ldots, x_n) \propto P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

$$\hat{y} = P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

On note :

$x = \left[x^1, x^2, \ldots, x^i, \ldots, x^p\right]$ un vecteur de descripteurs

$C = \{c_1, \cdots, c_q\}$ l'ensemble des classes possibles

S : ensemble fini de couples de la forme (x, c_k)

n : nombre d'observations de S

n_k : nombre d'éléments de S appartenant à la classe c_k

La règle de classification de Bayes recommande de classer le vecteur x dans la classe c_k pour laquelle $P(c_k/x)$ est maximal.

Ce qui revient à maximiser : $\frac{p(x/c_k)p(c_k)}{p(x)}$

soit encore $p(x/c_k)p(c_k)$, $p(x)$ ne dépendant pas de c_k

Pour pouvoir appliquer la règle de Bayes il faut donc pouvoir estimer : $p(x/c_k)$ et $p(c_k)$

On estime $p(c_k)$ par : $\hat{p}(c_k) = \frac{n_k}{n}$

et $p(x/c_k)$ par : $\hat{p}(x/c_k) = \prod_{i=1}^{p} p(x^i/c_k)$

Ce qui revient à considérer les attributs comme indépendants les uns des autres.

La règle de classification de Bayes devient alors : classer le vecteur x dans la classe c_k qui maximise :

$$\prod_{i=1}^{p} p(x^i/c_k)\, p(c_k)$$

Exemple

$$S = S_1 \cup S_2$$
$$S_1 = \{01100, 11001, 10110, 10101, 10010\}$$
$$S_2 = \{01010, 11111, 11010, 11101, 10101\}$$
$$C = \{c_1, c_2\}$$

On demande de classer x=00111

$$\hat{p}(c_1) = \frac{n_1}{n} = \frac{5}{10} = \frac{1}{2}$$
$$\hat{p}(c_2) = \frac{n_2}{n} = \frac{5}{10} = \frac{1}{2}$$

On suppose les attributs indépendants

	c_1	c_2		c_1	c_2
$\hat{p}(x_1 = 0/c_j)$	1/5	1/5	$\hat{p}(x_1 = 1/c_j)$	4/5	4/5
$\hat{p}(x_2 = 0/c_j)$	3/5	1/5	$\hat{p}(x_2 = 1/c_j)$	2/5	4/5
$\hat{p}(x_3 = 0/c_j)$	2/5	2/5	$\hat{p}(x_3 = 1/c_j)$	3/5	3/5
$\hat{p}(x_4 = 0/c_j)$	3/5	2/5	$\hat{p}(x_4 = 1/c_j)$	2/5	3/5
$\hat{p}(x_5 = 0/c_j)$	3/5	2/5	$\hat{p}(x_5 = 1/c_j)$	2/5	3/5

$S = S_1 \cup S_2$
$S_1 = \{01100,11001,10110,10101,10010\}$
$S_2 = \{01010,11111,11010,11101,10101\}$
$C = \{c_1, c_2\}$

$$\prod_{i=1}^{p} p(x^i/c_k)\, p(c_k)$$

$$\hat{p}(00111/c_1) = \frac{1}{2} \times \frac{1}{5} \times \frac{3}{5} \times \frac{3}{5} \times \frac{2}{5} \times \frac{2}{5} = \frac{1}{2} \times \frac{36}{5^5} = 5.75\ 10^{-3}$$
$$\hat{p}(00111/c_2) = \frac{1}{2} \times \frac{1}{5} \times \frac{1}{5} \times \frac{3}{5} \times \frac{3}{5} \times \frac{3}{5} = \frac{1}{2} \times \frac{27}{5^5} = 4.3\ 10^{-3}$$

On classera donc x=00111 en classe c_1

Machine learning: Classification
Method: Naives Bayes

Day	Outlook	Temperature	Humidity	Windy	Play	
1	Sunny	Hot	High	False	No	
2	Sunny	Hot	High	True	No	
3	Overcast	Hot	High	False	Yes	
4	Rainy	Mild	High	False	Yes	
5	Rainy	Cool	Normal	False	Yes	
6	Rainy	Cool	Normal	True	No	
7	Overcast	Cool	Normal	True	Yes	
8	Sunny	Mild	High	False	No	
9	Sunny	Cool	Normal	False	Yes	
10	Rainy	Mild	Normal	False	Yes	
11	Sunny	Mild	Normal	True	Yes	
12	Overcast	Mild	High	True	Yes	
13	Overcast	Hot	Normal	False	Yes	
14	Rainy	Mild	High	True	No	
15	Sunny	Cool	Normal	True	?	
0.36	0.60	0.20	0.20	0.60	No	0.014400
0.64	0.22	0.33	0.67	0.33	Yes	0.016461
16	Overcast	Mild	Normal	False	?	
0.40	0.00	0.40	0.20	0.40	No	0.000000
0.60	0.44	0.44	0.67	0.67	Yes	0.087791

p(Play = no)	=	5	/	14	= 0.36
p(Play = yes)	=	9	/	14	= 0.64
p(Outlook = Overcast / Play = No)	=	0	/	5	= 0.00
p(Outlook = Rainy / Play = No)	=	2	/	5	= 0.40
p(Outlook = Sunny / Play = No)	=	3	/	5	= 0.60
p(Outlook = Overcast / Play = Yes)	=	4	/	9	= 0.44
p(Outlook = Rainy / Play = Yes)	=	3	/	9	= 0.33
p(Outlook = Sunny / Play = Yes)	=	2	/	9	= 0.22
p(Temperature = Cool / Play = No)	=	1	/	5	= 0.20
p(Temperature = Hot / Play = No)	=	2	/	5	= 0.40
p(Overcast = Mild / Play = No)	=	2	/	5	= 0.40
p(Temperature = Cool / Play = Yes)	=	3	/	9	= 0.33
p(Temperature = Hot / Play = Yes)	=	2	/	9	= 0.22
p(Temperature = Mild / Play = Yes)	=	4	/	9	= 0.44
p(Humidity = High / Play = No)	=	4	/	5	= 0.80
p(Humidity = Normal / Play = No)	=	1	/	5	= 0.20
p(Humidity = High/ Play = Yes)	=	3	/	9	= 0.33
p(Humidity = Normal / Play = Yes)	=	6	/	9	= 0.67
p(Windy = False / Play = No)	=	2	/	5	= 0.40
p(Windy = True / Play = No)	=	3	/	5	= 0.60
p(Windy = False/ Play = Yes)	=	6	/	9	= 0.67
p(Windy = True / Play = Yes)	=	3	/	9	= 0.33

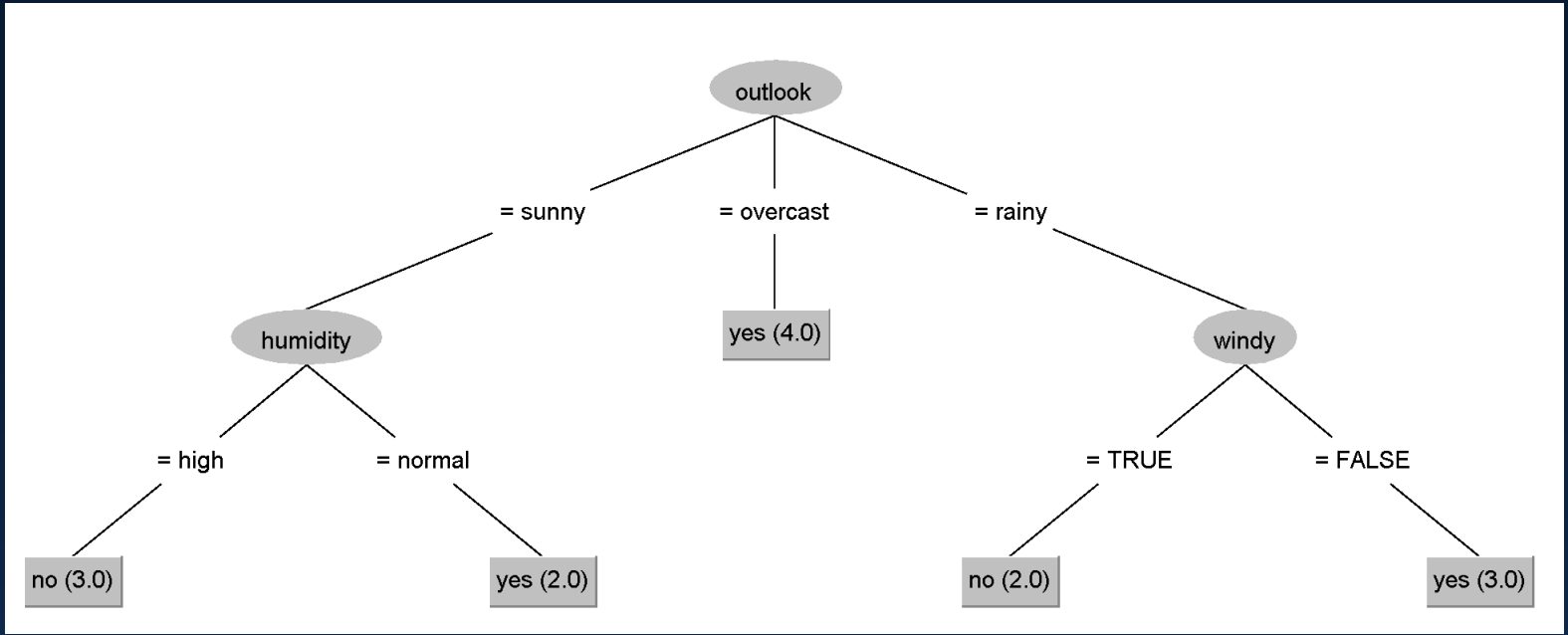
Supervised Machine Learning: Classification

Method: Decision Tree

Machine learning: Classification
Method: Decision Tree

Day	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
6	Rainy	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rainy	Mild	High	True	No

Machine learning: Classification
Method: Decision Tree

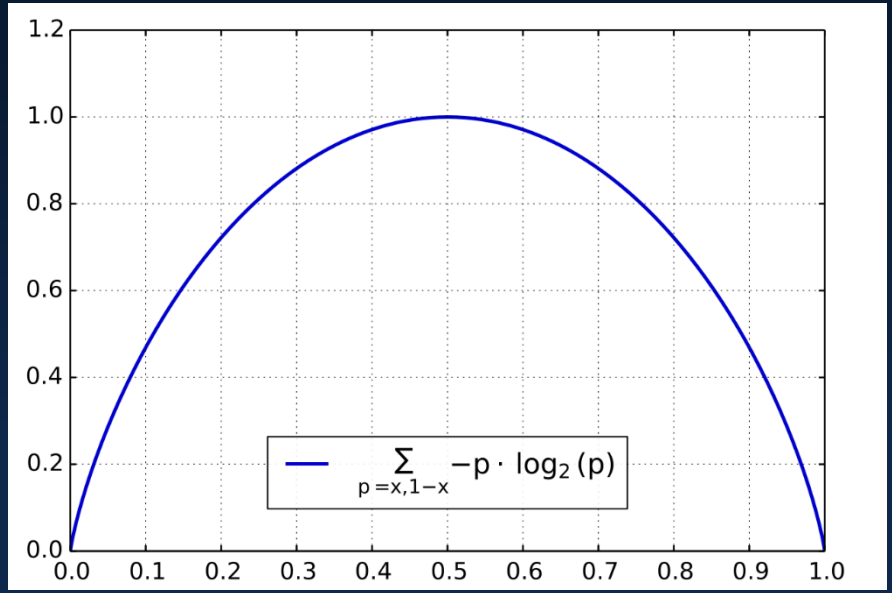


Machine learning: Classification
Method: Decision Tree

Entropy (noun)

1. Name given by Clausius to the state function denoted by S, which characterizes the state of “disorder” of a system.
2. In communication theory, a number that measures the uncertainty of the nature of a given message based on the preceding one. (Entropy is zero when there is no uncertainty.)

$$H(X) = -\sum_{i=1}^{n} P(x_i) log_b(P(x_i))$$



https://www.larousse.fr/dictionnaires/francais/entropie/30093

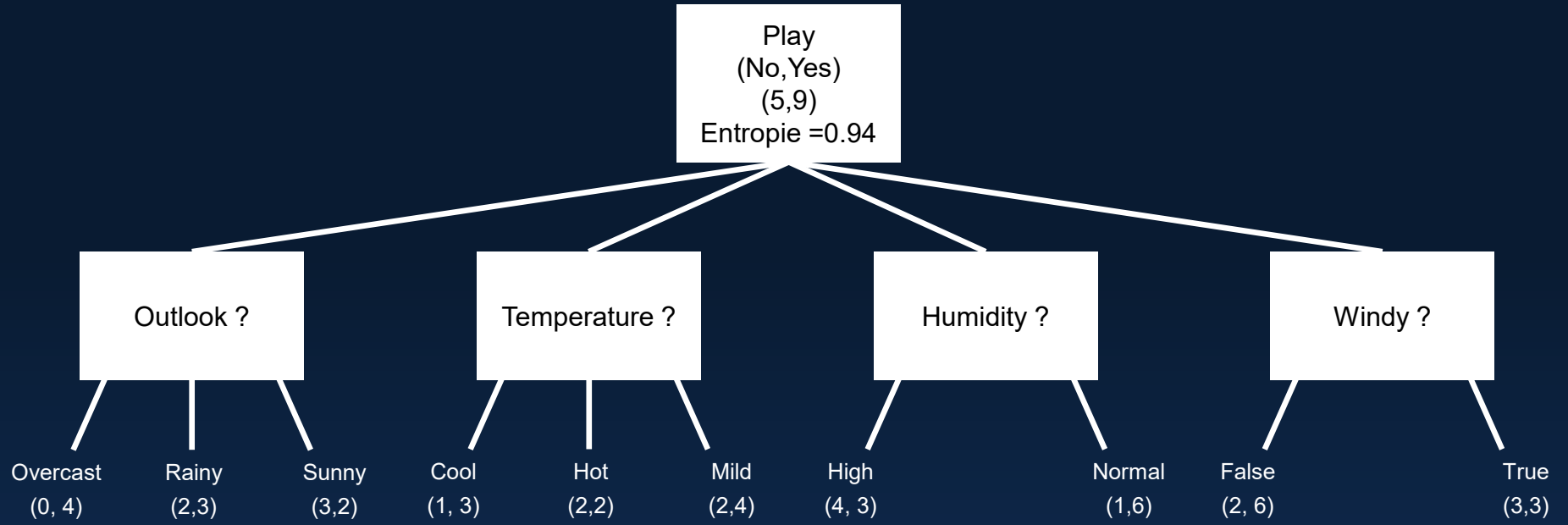
Machine learning: Classification
Method: Decision Tree

Day	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
6	Rainy	Cool	Normal	True	No
8	Sunny	Mild	High	False	No
14	Rainy	Mild	High	True	No
3	Overcast	Hot	High	False	Yes
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
7	Overcast	Cool	Normal	True	Yes
9	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes

$$\mathrm{H}(Play) = -P(Play = No)log_2\big(P(Play = No)\big) - P(Play = Yes)log_2\big(P(Play = Yes)\big)$$

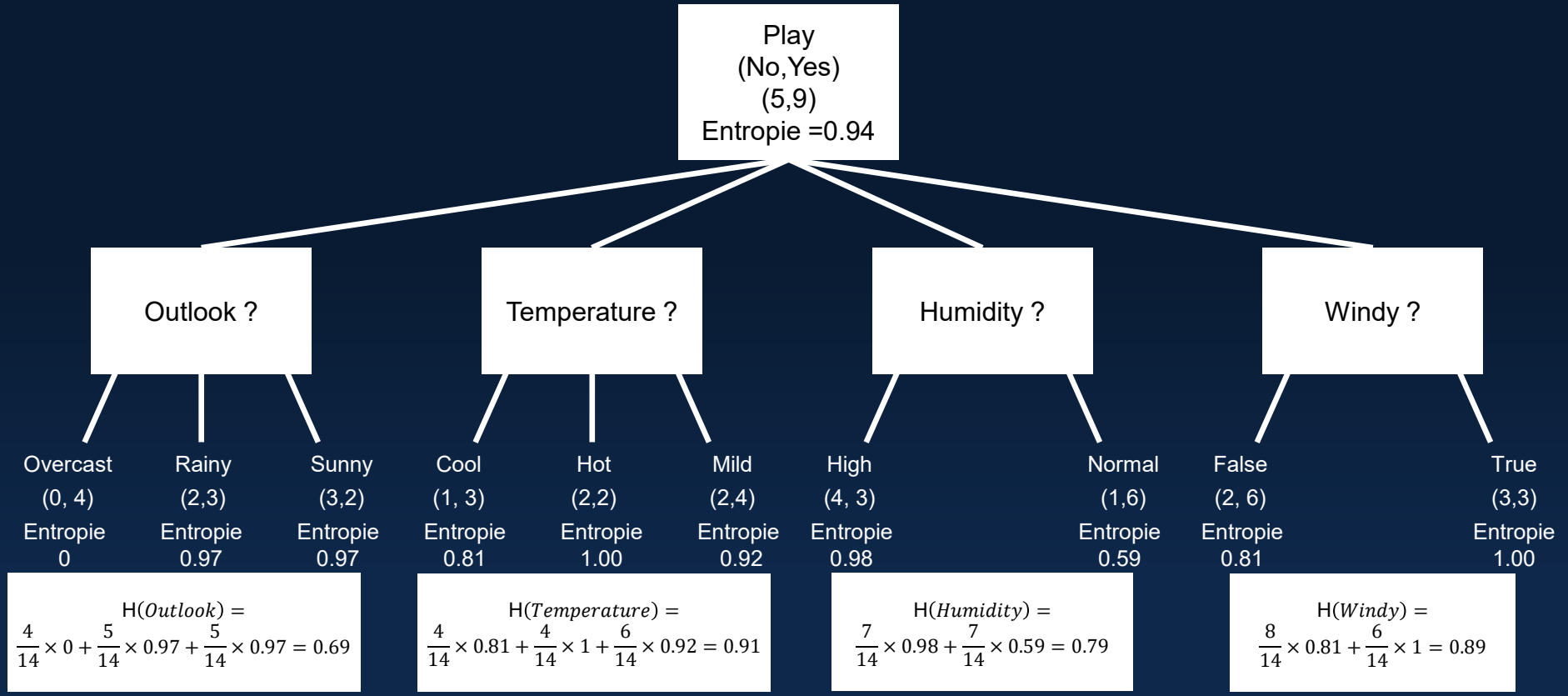
$$\mathrm{H}(Play) = -\frac{5}{14}log_2\left(\frac{5}{14}\right) - \frac{9}{14}log_2\left(\frac{9}{14}\right) = 0.94$$

Machine learning: Classification
Method: Decision Tree



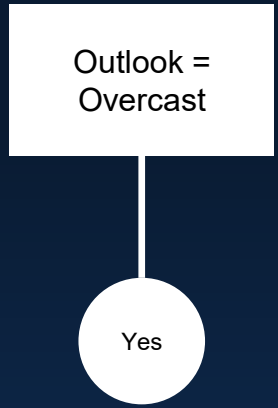
Which test will remove the most uncertainty?
Which test will reduce entropy the most?

Machine learning: Classification
Method: Decision Tree



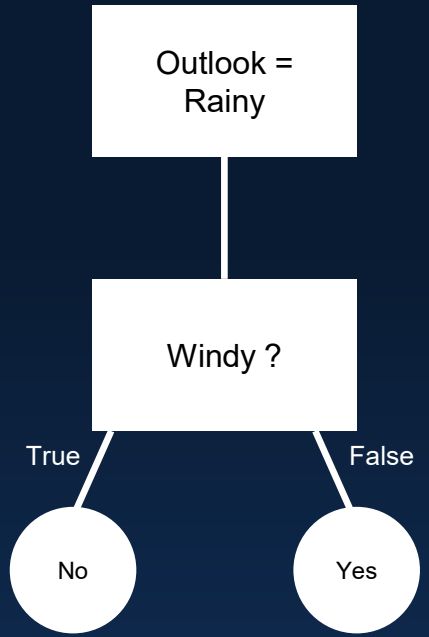
Machine learning: Classification
Method: Decision Tree

Day	Outlook	Temperature	Humidity	Windy	Play
3	Overcast	Hot	High	False	Yes
7	Overcast	Cool	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes



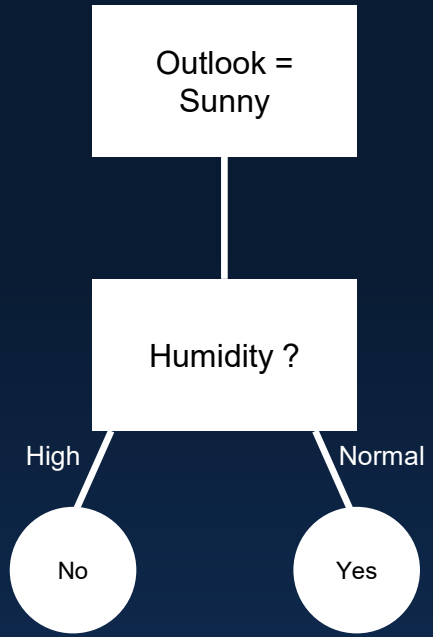
Machine learning: Classification
Method: Decision Tree

Day	Outlook	Temperature	Humidity	Windy	Play
6	Rainy	Cool	Normal	True	No
14	Rainy	Mild	High	True	No
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes



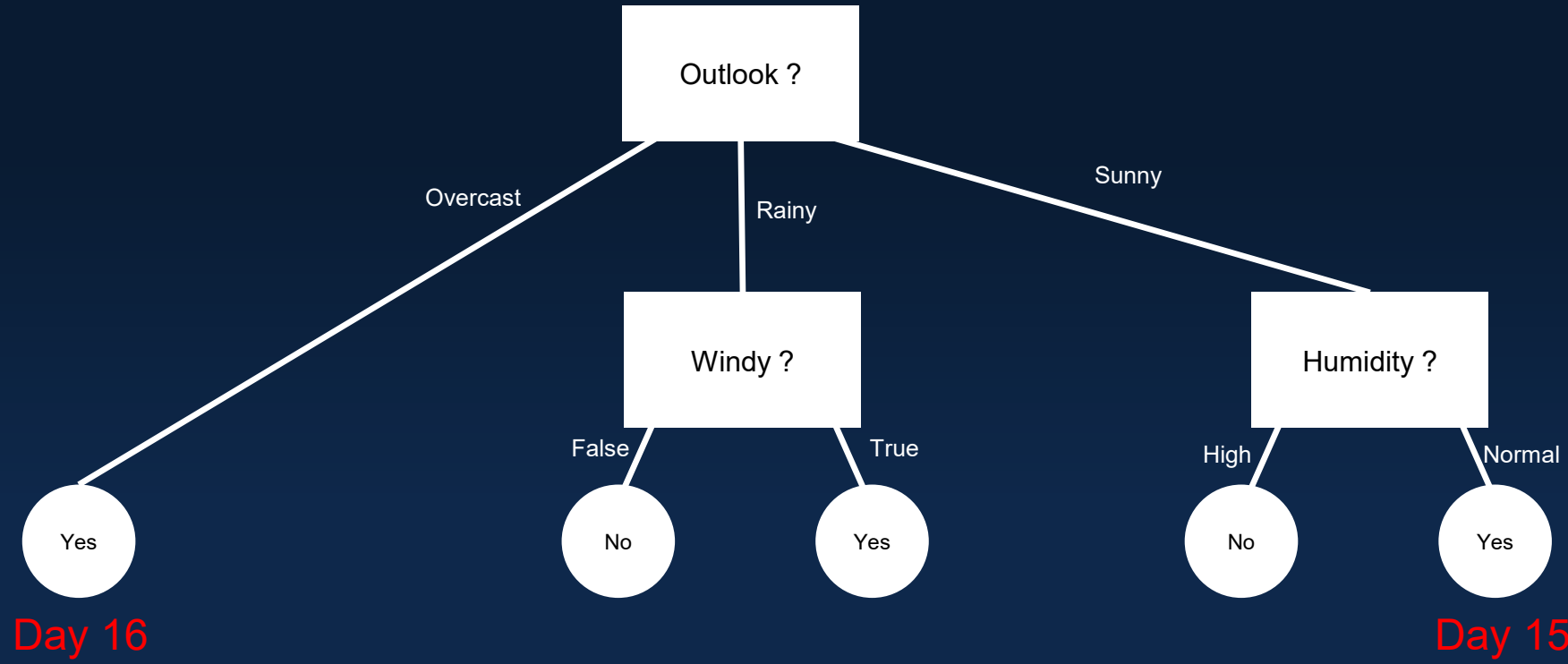
Machine learning: Classification
Method: Decision Tree

Day	Outlook	Temperature	Humidity	Windy	Play
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes



Machine learning: Classification
Method: Decision Tree

Day	Outlook	Temperature	Humidity	Windy	Play
15	Sunny	Cool	Normal	True	?
16	Overcast	Mild	Normal	False	?



Classification mise en pratique :  « Orange Data Mining » étape par étape

Classification mise en pratique : « Orange Data Mining » étape par étape

Étape 1 : Poser le problème

Étape 2 : Rechercher les données

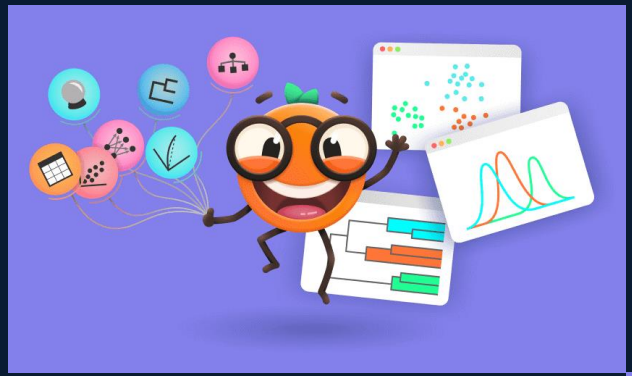
Étape 3 : Nettoyer les données

Étape 4 : Transformer les données

Étape 5 : Sélectionner les données pertinentes

Étape 6 : Rechercher les modèles

Étape 7 : Évaluer et valider les résultats



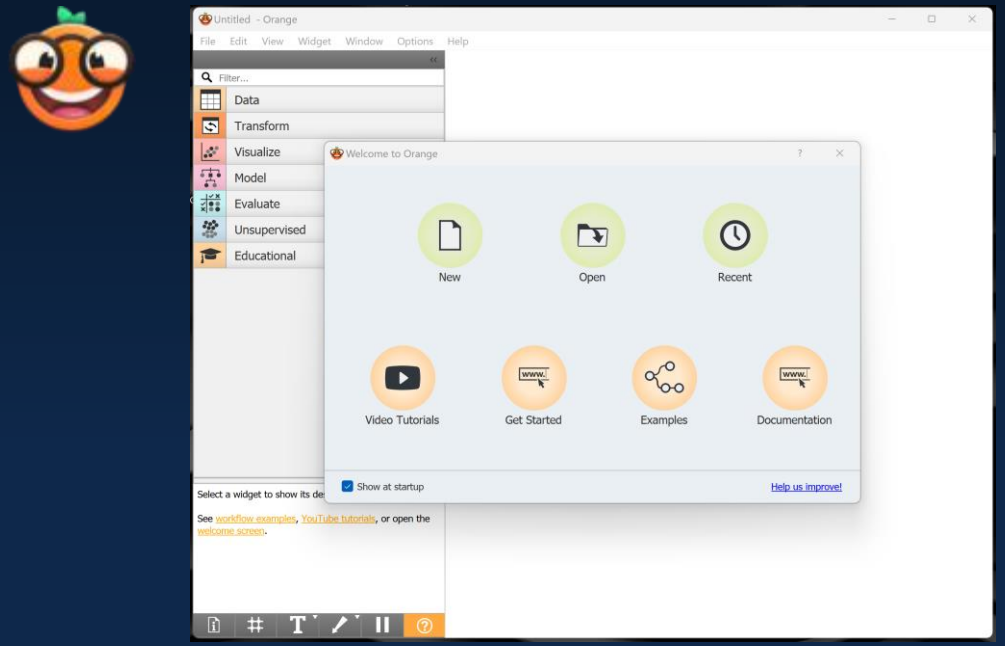
https://orangedatamining.com/

Data Mining Fruitful and Fun

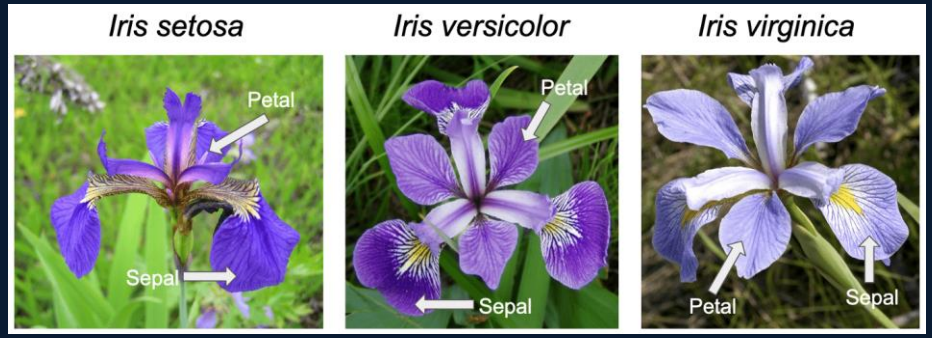
Open source machine learning and data visualization.

Download Orange 3.36.2

https://orangedatamining.com/download/



Étape 1: Poser le problème



Un exemple : Classification supervisée de fleurs d'iris

Connaissant :

- la longueur des sépales d'une fleur d'iris
- la largeur des sépales d'une fleur d'iris
- la longueur des pétales d'une fleur d'iris
- la largeur des pétales d'une fleur d'iris

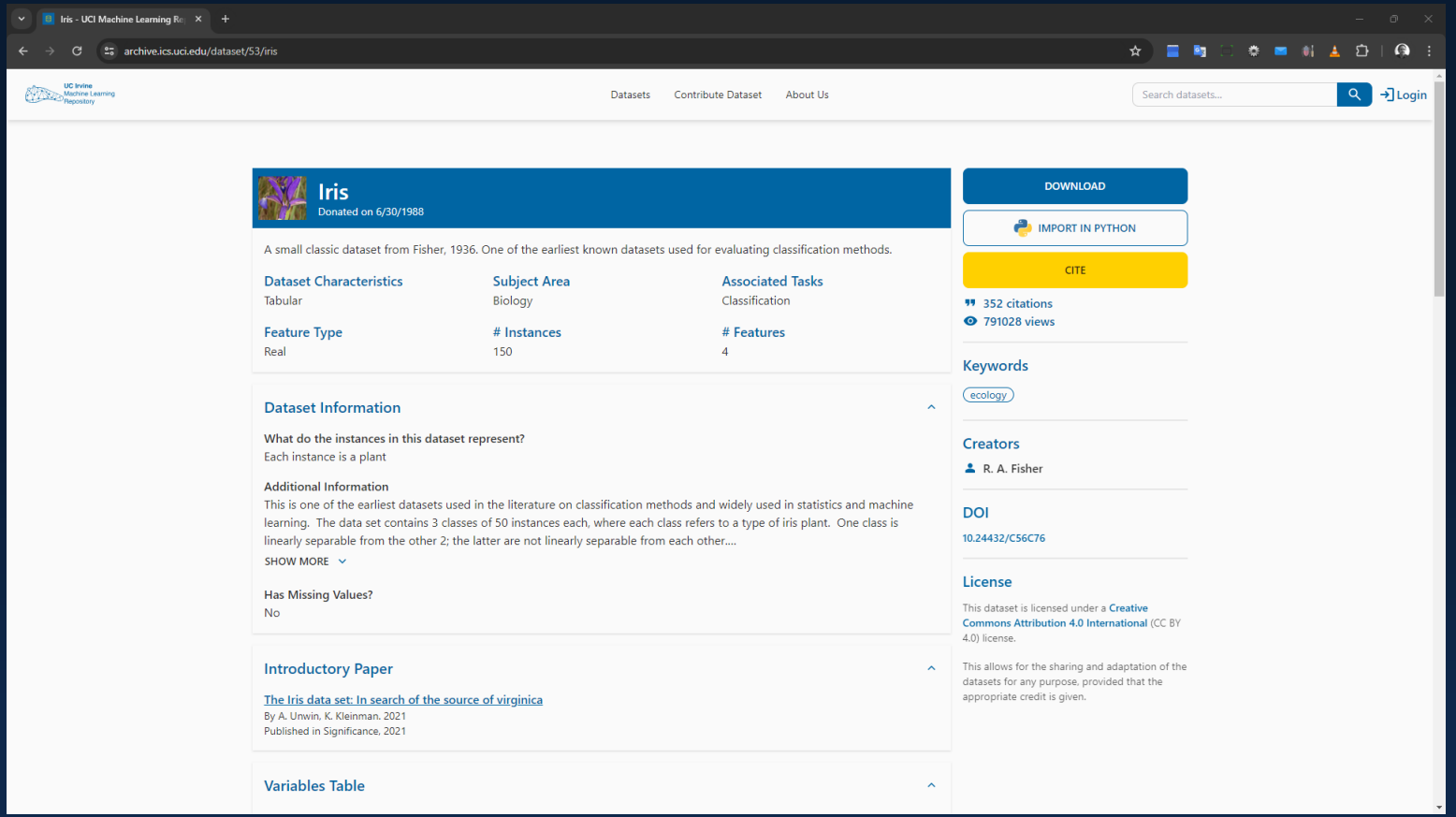
Est-il possible de classer les fleurs dans les catégories suivantes ?

- Iris Setosa - classe 1
- Iris Versicolour - classe 2
- Iris Virginica - classe 3

L'objectif est de fournir un modèle de classification supervisée qui accepte en entrée les 4 caractéristiques des sépales et des pétales des fleurs d'iris et qui fournit en sortie le numéro de la classe de la fleur (1, 2 ou 3).

Les résultats de la classification seront évalués sur une base de données de test représentant 25% des données collectées sous la forme d'une matrice de confusion.

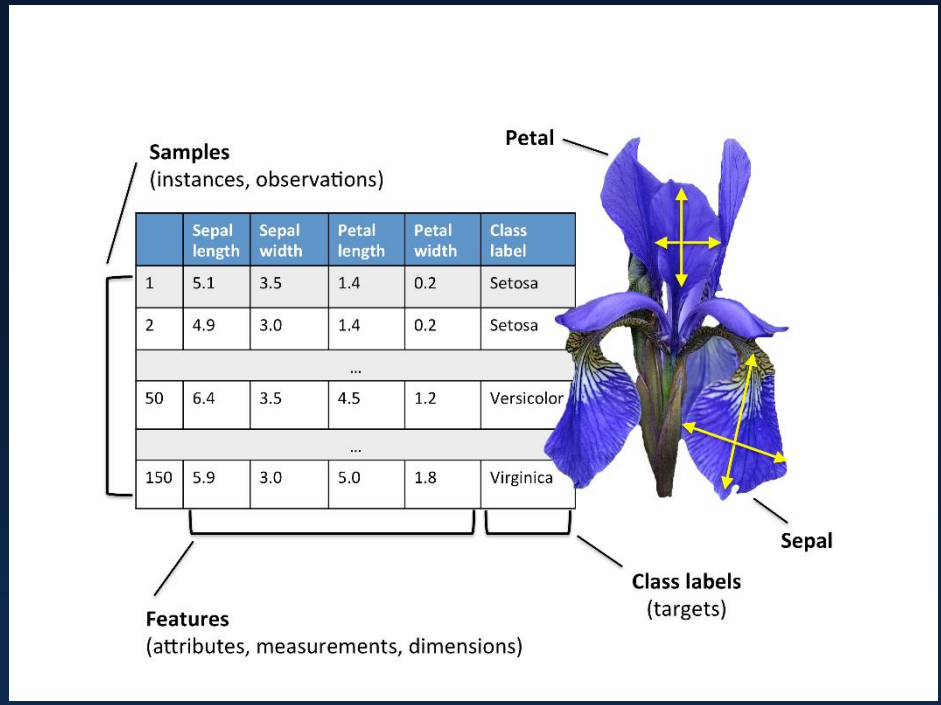
Étape 2 : Rechercher les données



Étape 2 : Rechercher les données

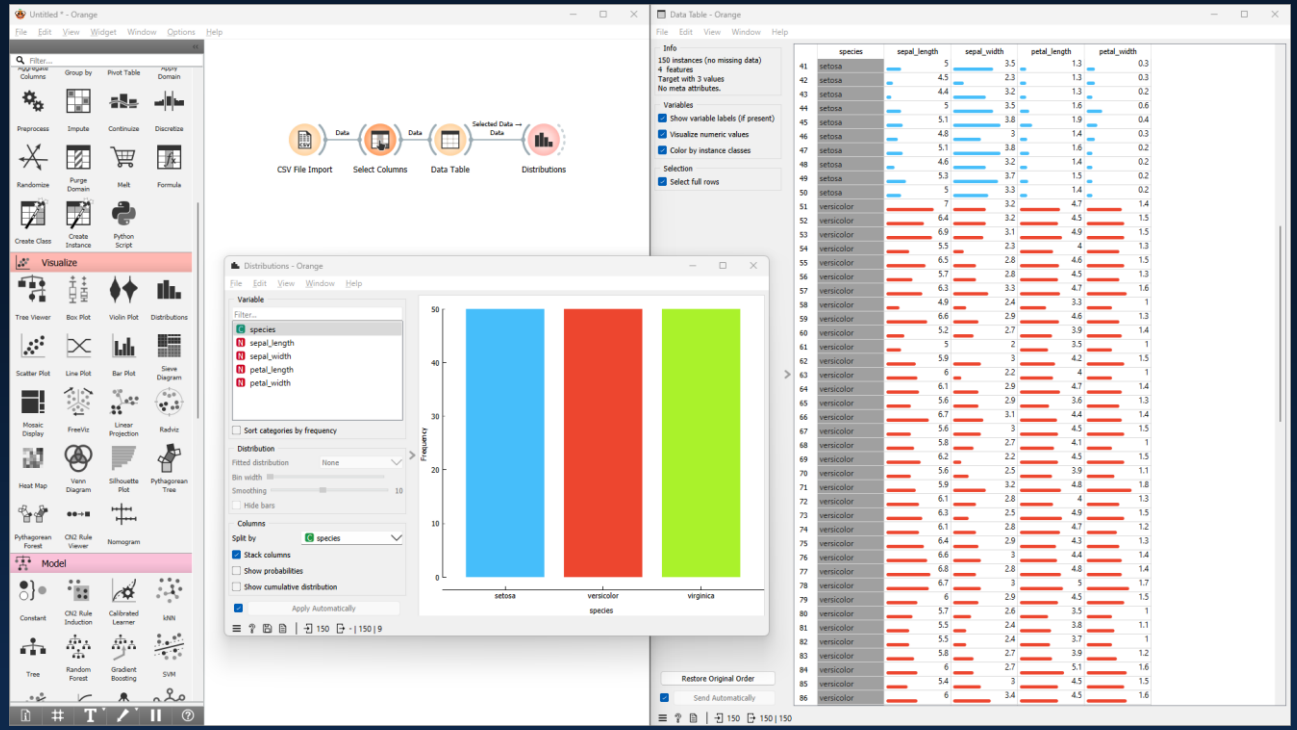
Un exemple : Classification de fleurs d'iris

Title: Iris Plants Database Updated Sept 21 2000 by C.Blake
File : iris.dat
Sources:
(a) Creator: R.A. Fisher
(b) Donor: Michael Marshall
(c) Date: July, 1988
Number of Instances: 150 (50 in each of three classes)
Number of Attributes: 4 numeric attributes and the class
Attribute Information:
1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
-- Iris Setosa
-- Iris Versicolour
-- Iris Virginica
Missing Attribute Values: None

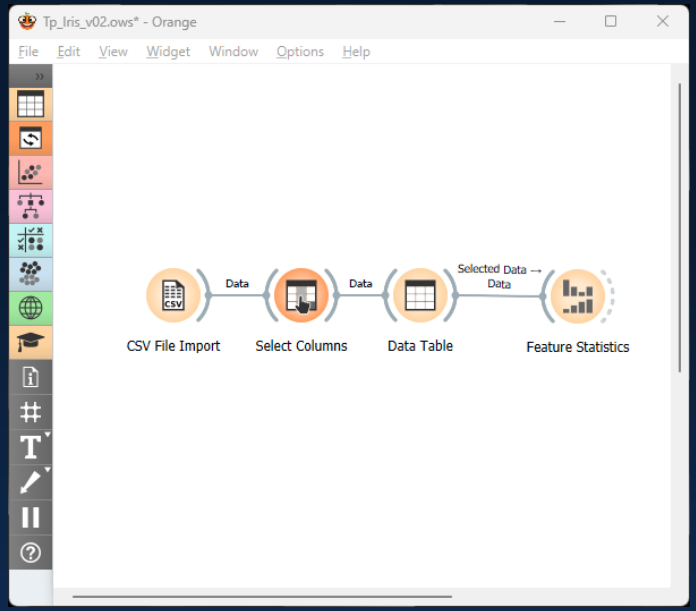


	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

Étape 2 : Rechercher les données



Étape 2 : Rechercher les données



Étape 3 : Nettoyer les données

Vérifier l’origine des données

Traiter les valeurs aberrantes

Traiter les valeurs manquantes

Traiter les valeurs nulles

Estimer la qualité des données

Étape 4 : Transformer les données

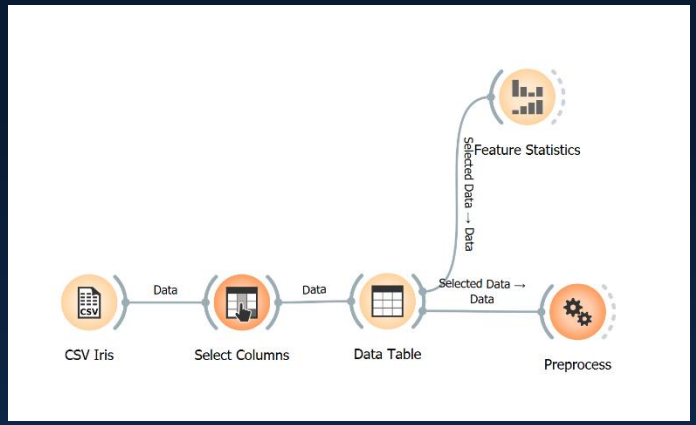
Cette étape peut consister à :

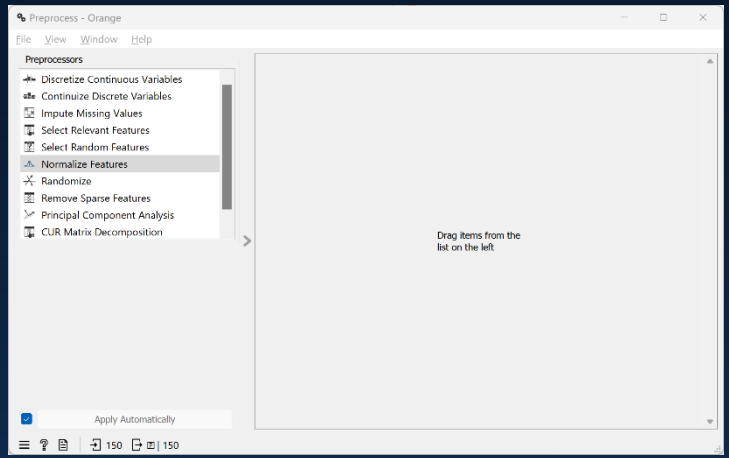
- Normaliser les données
- Coder les informations qualitatives
- Coder en ratios (pourcentages)
- Transformer les dates en durées
- Transcoder les données,

  exemple : code postal en coordonnées géographiques
- Exprimer des fréquences
- Exprimer des tendances
- Réaliser des combinaisons de variables

...

Étape 4 : Transformer les données





Étape 5 : Sélectionner les données pertinentes

		Caractères										
		Age	Revenu	Sexe		Situation Matrimoniale			…	Caractère j	…	Caractère p
		X^1	X^2	M	F	Marié	Célibataire	Veuf Divorcé	…	X^j	…	X^p
				X^3	X^4	X^5	X^6	X^7				
Individus	X_1	x_1^1	x_1^2	x_1^3	x_1^4	x_1^5	x_1^6	x_1^7	…	x_1^j	…	x_1^p
	X_2	x_2^1	x_2^2	x_2^3	x_2^4	x_2^5	x_2^6	x_2^7	…	x_2^j	…	x_2^p
	…	…	…	…	…	…	…	…	…	…	…	…
	X_i	x_i^1	x_i^2	x_i^3	x_i^4	x_i^5	x_i^6	x_i^7	…	x_i^j	…	x_i^p
	…	…	…	…	…	…	…	…	…	…	…	…
	X_n	x_n^1	x_n^2	x_n^3	x_n^4	x_n^5	x_n^6	x_n^7	…	x_n^j	…	x_n^p

Zone optimale	Temps de calcul long
Multiplication des apprentissages pour s'assurer de la stabilité	Trop peu d'exemples par rapport à la taille du problème

- Nombre d'individus +

- Nombre de caractères +

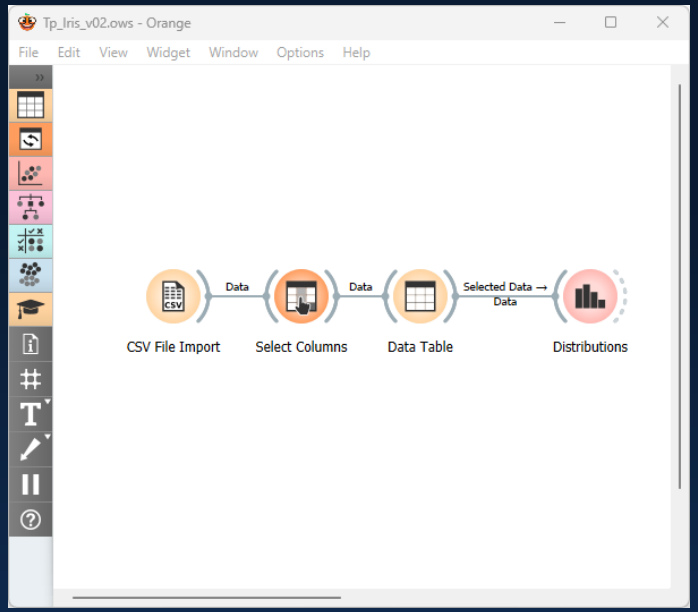
Réduire les dimensions

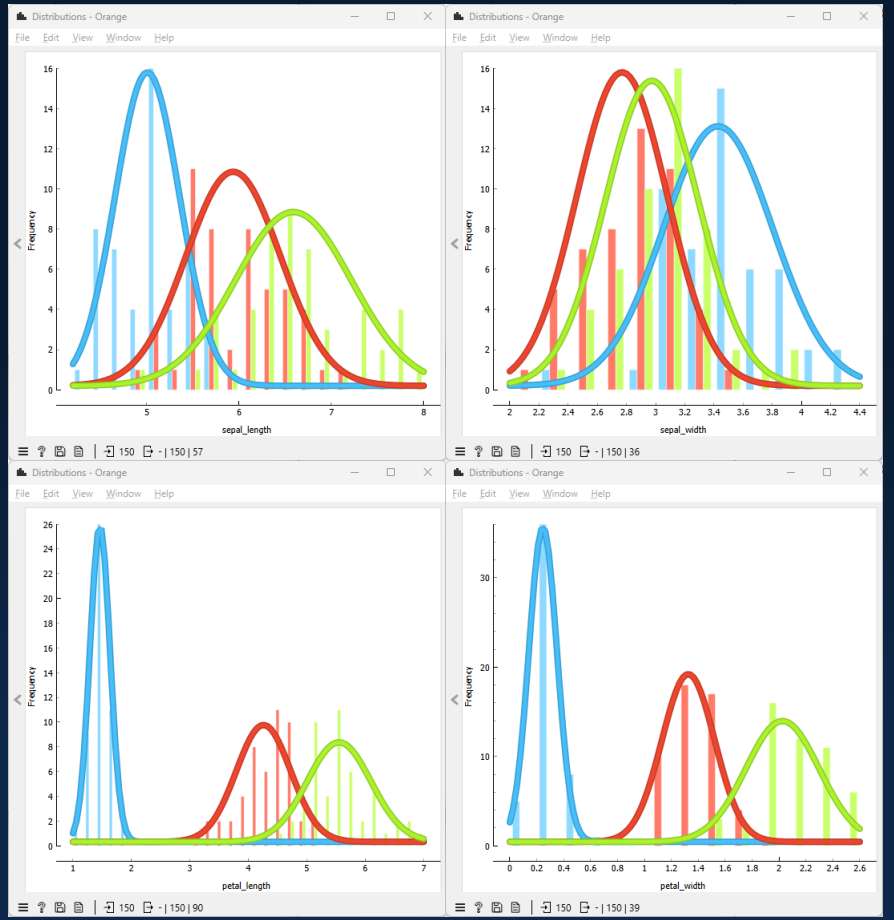
- Expertise humaine
- Analyses graphiques
- Information Mutuelle
- Analyses de corrélation
- Analyse en composantes principales
- …

Étape 5 : Sélectionner les données pertinentes

Réduire les dimensions

- Expertise humaine

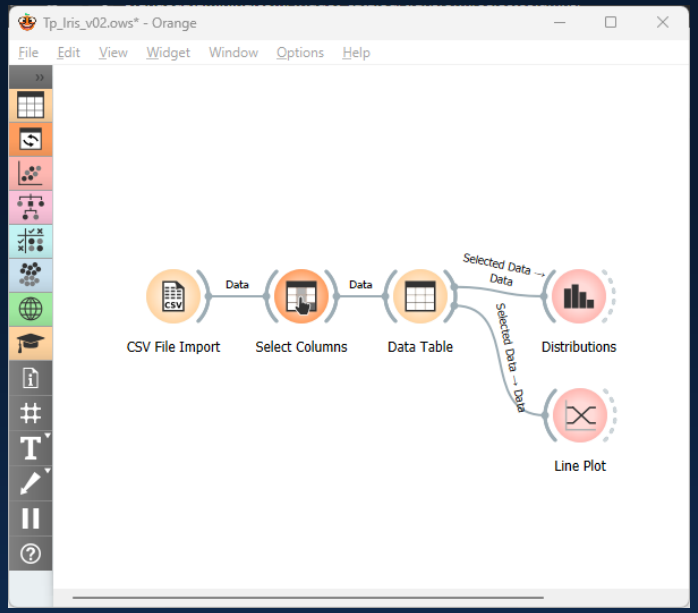


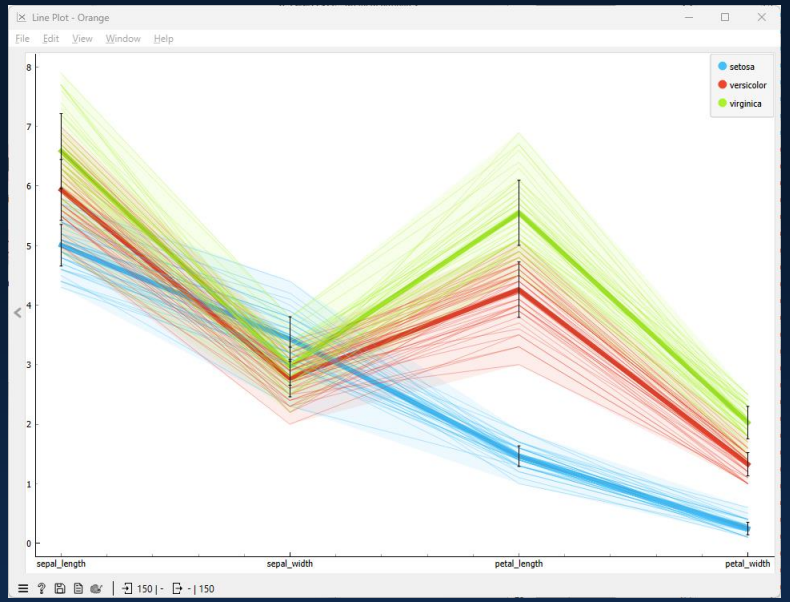


Étape 5 : Sélectionner les données pertinentes

Réduire les dimensions

- Expertise humaine

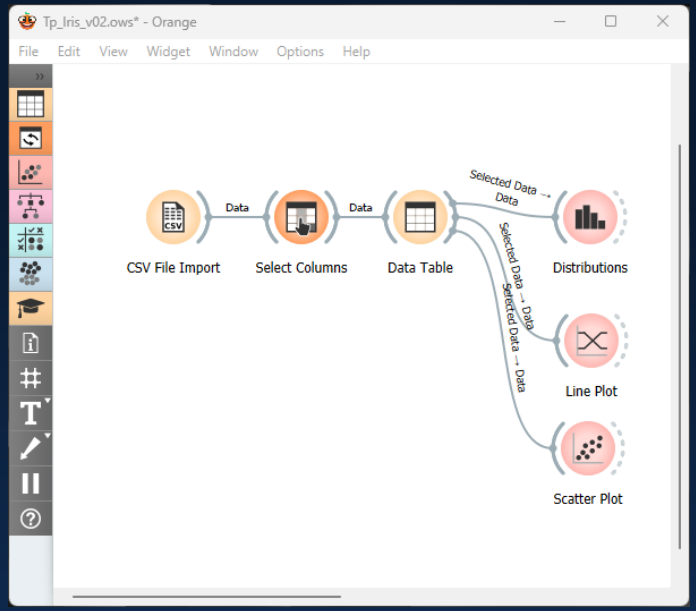


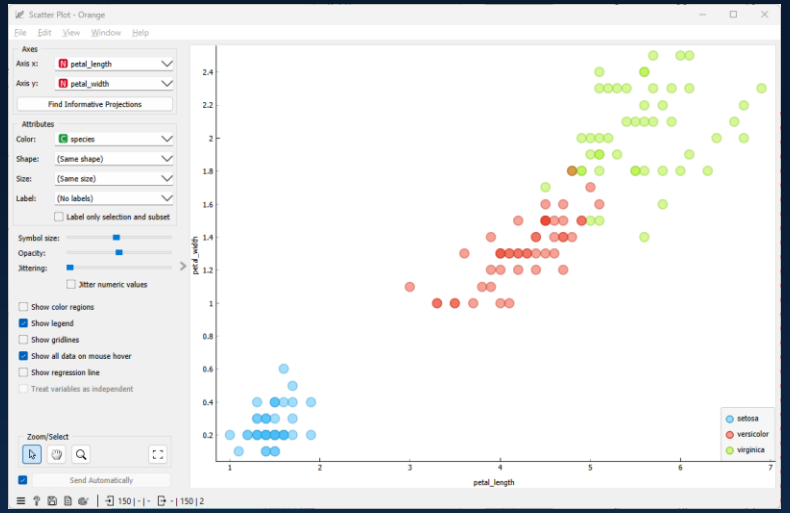


Étape 5 : Sélectionner les données pertinentes

Réduire les dimensions

- Expertise humaine

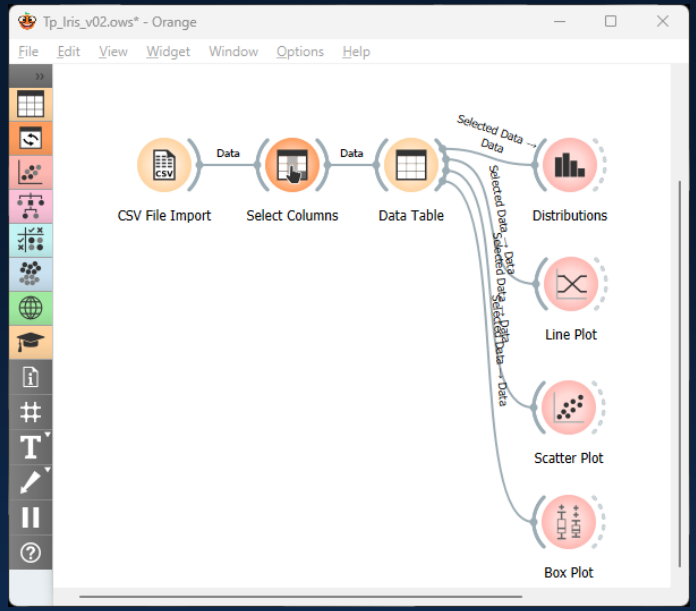


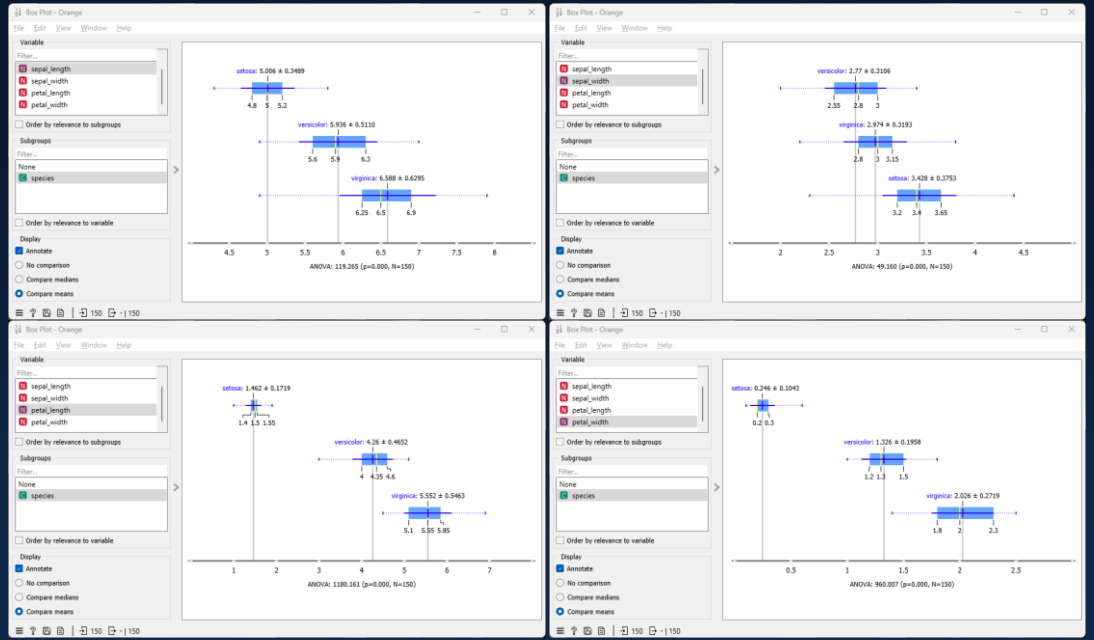


Étape 5 : Sélectionner les données pertinentes

Réduire les dimensions

- Expertise humaine

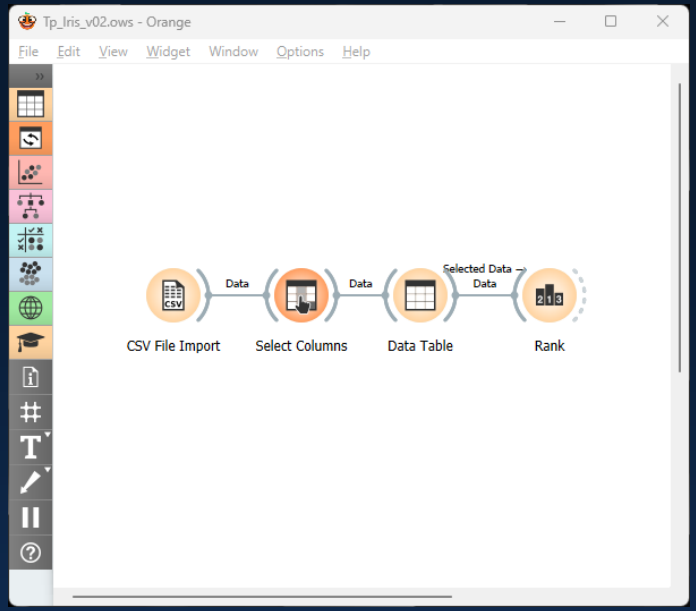


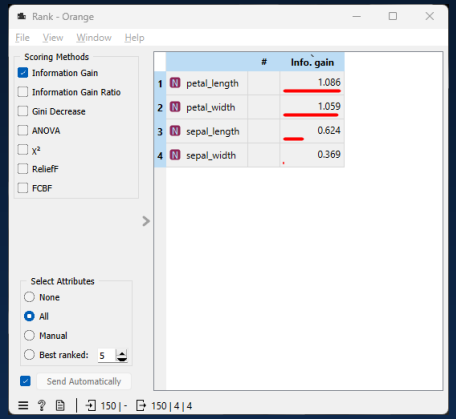


Étape 5 : Sélectionner les données pertinentes

Réduire les dimensions

- Information Mutuelle

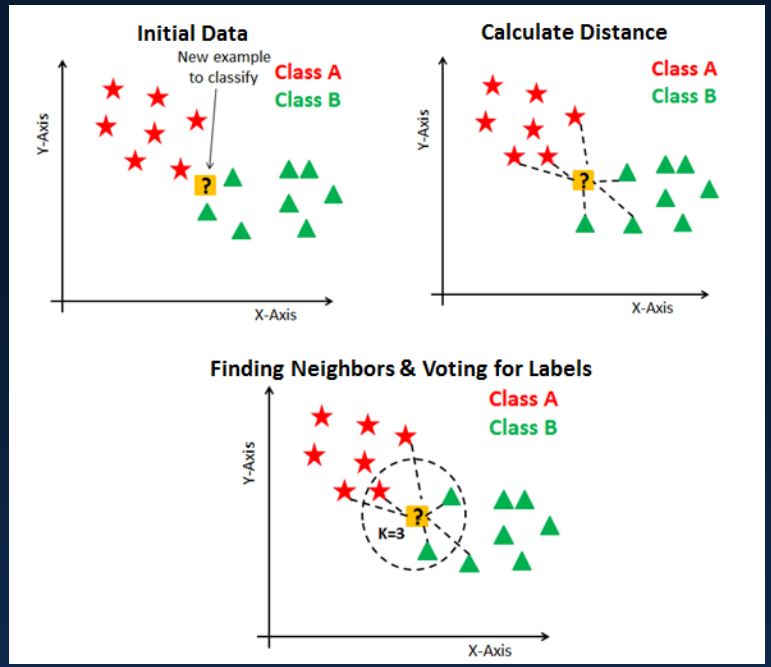




Étape 6 : Rechercher les modèles

Choix d'une méthode ou d'une technique :

- K plus proches voisins
- Naive Bayes
- Arbres de décision
- Réseaux de neurones
- Régression
- Règles
- Analyses factorielles
- Logique Floue
- Algorithmes génétiques
- Le raisonnement à base de cas
- Les réseaux Bayésiens
- ...



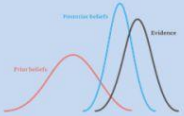
Étape 6 : Rechercher les modèles

Choix d'une méthode ou d'une technique :

- K plus proches voisins
- Naive Bayes
- Arbres de décision
- Réseaux de neurones
- Régression
- Règles
- Analyses factorielles
- Logique Floue
- Algorithmes génétiques
- Le raisonnement à base de cas
- Les réseaux Bayésiens
- ...

Naive Bayes

Supervised Learning Algorithm for Classification



Bayes Theorem :

$$P(B_i|A) = \frac{P(B_i) \times P(A|B_i)}{P(A)}$$

$$= \frac{P(B_i) \times P(A|B_i)}{\sum_{j=1}^{k} P(B_j) \times P(A|B_j)}$$

Thomas Bayes 1702-1762

Probabilistic model

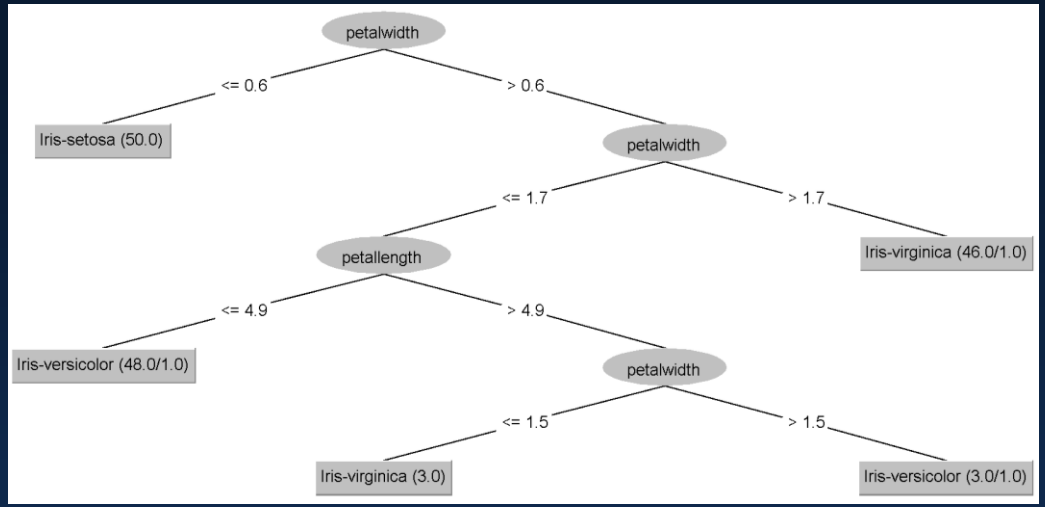
$$P(C_k|x) = \frac{P(C_k) \times P(x|C_k)}{P(x)} = \propto p(C_k, x_1, x_2, \ldots, x_n) \propto P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

$$\hat{y} = P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

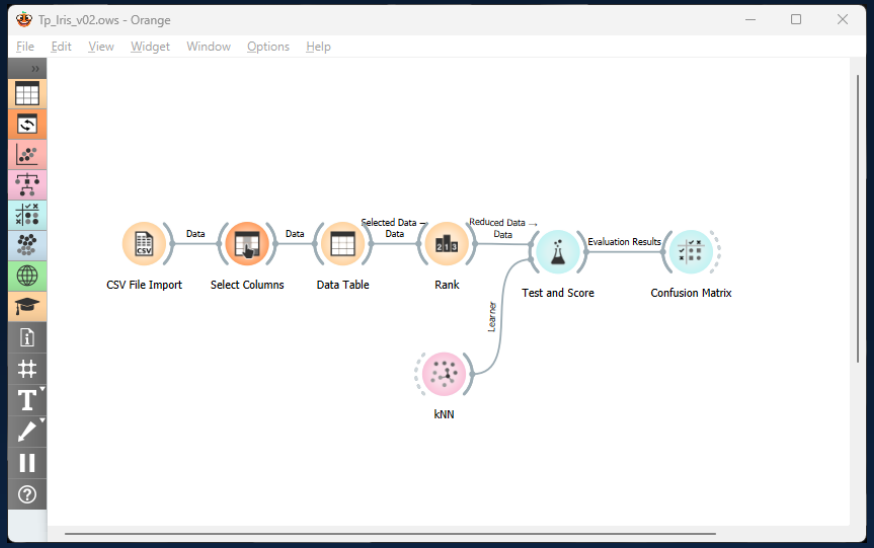
Étape 6 : Rechercher les modèles

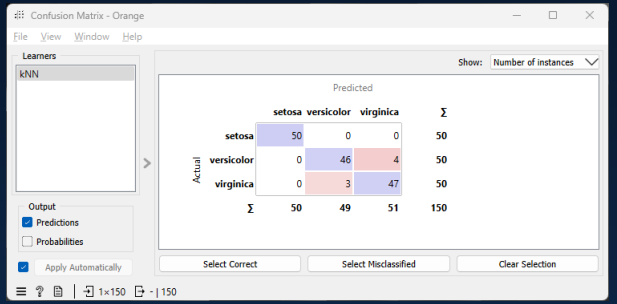
Choix d'une méthode ou d'une technique :

- K plus proches voisins
- Naive Bayes
- Arbres de décision
- Réseaux de neurones
- Régression
- Règles
- Analyses factorielles
- Logique Floue
- Algorithmes génétiques
- Le raisonnement à base de cas
- Les réseaux Bayésiens
- ...

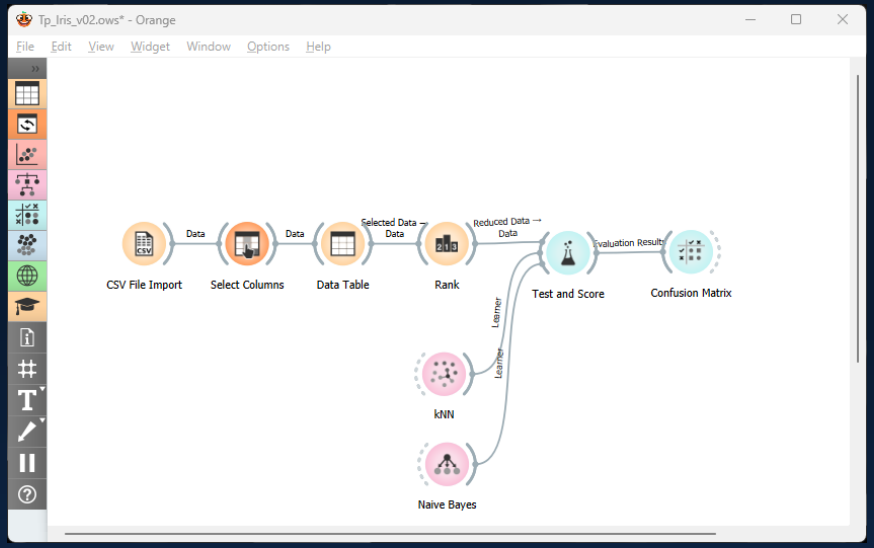


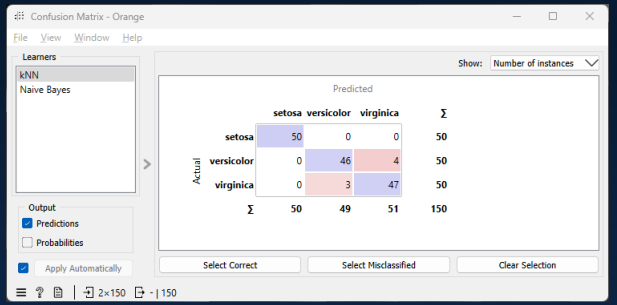
Étape 7 : Évaluer et valider les résultats



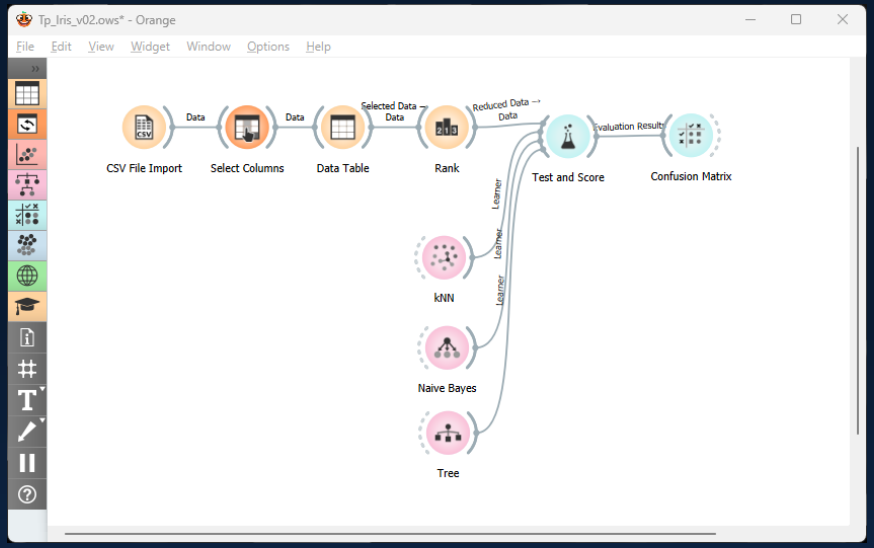


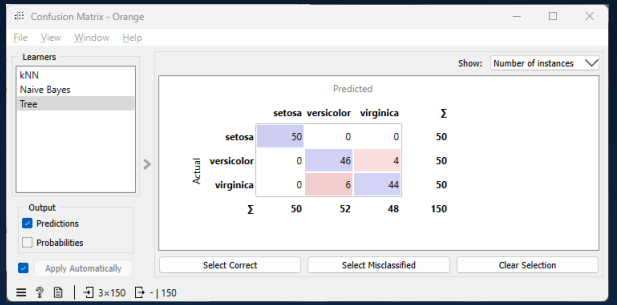
Étape 7 : Évaluer et valider les résultats



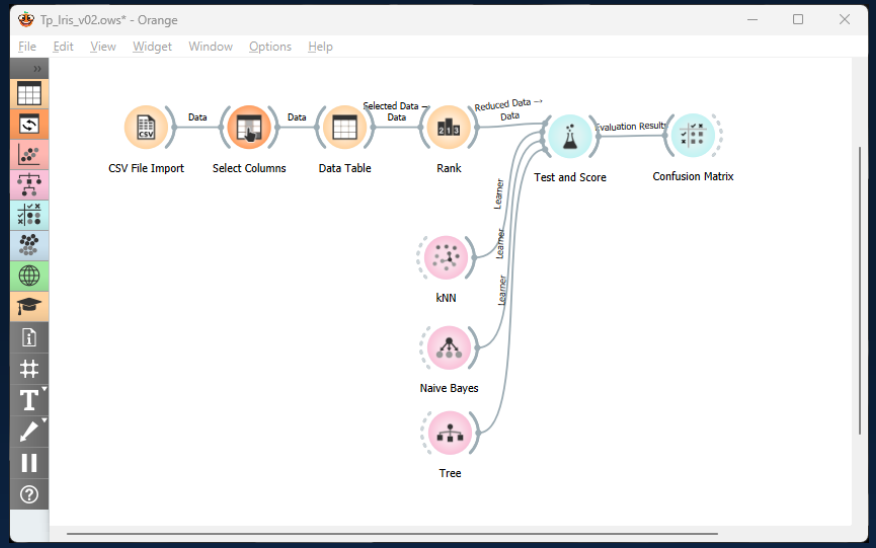


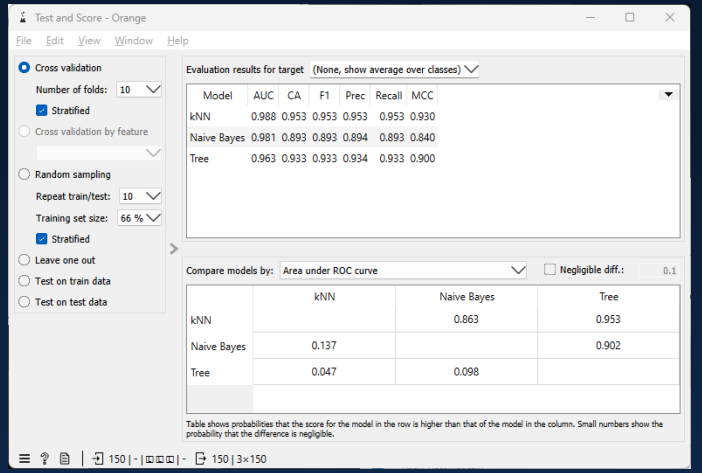
Étape 7 : Évaluer et valider les résultats





Étape 7 : Évaluer et valider les résultats





Machine Learning:

Feature Engineering

Machine learning: Feature Engineering

Feature engineering is a preprocessing step in supervised machine learning and statistical modeling which transforms raw data into a more effective set of inputs. Each input comprises several attributes, known as features. By providing models with relevant information, feature engineering significantly enhances their predictive accuracy and decision-making capability.

		Features / Variables / Attributes / Characteristics / ...										
		Age	Income	Gender		Marital Status			...	Feature j	...	Feature p
		X^1	X^2	M	F	Single	Married	Widowed Divorced	...	X^j	...	X^p
				X^3	X^4	X^5	X^6	X^7				
Observations	X_1	x_1^1	x_1^2	x_1^3	x_1^4	x_1^5	x_1^6	x_1^7	...	x_1^j	...	x_1^p
	X_2	x_2^1	x_2^2	x_2^3	x_2^4	x_2^5	x_2^6	x_2^7	...	x_2^j	...	x_2^p
	...	...	...	...	...	...	...	...	...	...	...	...
	X_i	x_i^1	x_i^2	x_i^3	x_i^4	x_i^5	x_i^6	x_i^7	...	x_i^j	...	x_i^p
	...	...	...	...	...	...	...	...	...	...	...	...
	X_n	x_n^1	x_n^2	x_n^3	x_n^4	x_n^5	x_n^6	x_n^7	...	x_n^j	...	x_n^p

https://en.wikipedia.org/wiki/Feature_engineering

Machine learning: Feature Engineering

Feature engineering techniques:

- Feature imputation
- Feature encoding
- Feature scaling
- Feature transformation
- Feature selection
- …

Machine learning: Feature Engineering
Feature Imputation

Missing data can skew your analysis, reduce model accuracy, and lead to biased results. Imputation fills these gaps, allowing you to use the complete dataset for analysis and modeling.

Imputation Techniques

Technique	Description	Pros	Cons	Use Case
Mean/Median/Mode Imputation	Replace missing values with mean, median, or mode	Simple to implement	May distort original data distribution	Useful when data is normally distributed
KNN Imputation	Use nearest neighbors to fill missing values	More accurate imputation	Computationally expensive	When dataset has relationships between features
Forward/Backward Fill	Use next/previous value to fill missing values	Maintains trend in data	Can propagate errors	Time series data
Interpolation	Estimate missing values based on other values	Handles continuous data well	May not be accurate for all types of data	Continuous data with trends

medium.com/@sanjayskumar4010/simple-guide-for-data-imputation-encoding-scaling-and-feature-transformation-techniques-cbb42dae1700

Machine learning: Feature Engineering
Feature Encoding

Machine learning models require numerical input. Encoding transforms categorical data into a numerical format, enabling models to process and learn from it.

Encoding Techniques

Technique	Description	Pros	Cons	Use Case
One-Hot Encoding	Converts categories to binary columns	Preserves all information	Can create high-dimensional data	Categorical data with no ordinal relationship
Label Encoding	Assigns unique number to each category	Simple and efficient	Implies ordinal relationship	Ordinal categorical data
Ordinal Encoding	Assigns numbers co Close ng order	Preserves ordinal relationship	Assumes order is meaningful	Ordered categorical data
Frequency Encoding	Uses frequency of categories	Handles high cardinality well	Loses category information	High cardinality categorical data

medium.com/@sanjayskumar4010/simple-guide-for-data-imputation-encoding-scaling-and-feature-transformation-techniques-cbb42dae1700

Machine learning: Feature Engineering
Feature Scaling

Feature scaling ensures that all features contribute equally to the model. It can significantly improve the performance of algorithms that rely on distance measurements or gradient-based optimization.

Scaling Techniques

Technique	Description	Pros	Cons	Use Case
Standardization	Transforms data to have mean of 0 and std dev of 1	Handles outliers well	Assumes normal distribution	Data with outliers
Min-Max Scaling	Scales data to a fixed range (0 to 1)	Preserves relationships	Sensitive to outliers	Data without outliers
Robust Scaling	Uses median and IQR	Not sensitive to outliers	May not scale all features equally	Data with many outliers

medium.com/@sanjayskumar4010/simple-guide-for-data-imputation-encoding-scaling-and-feature-transformation-techniques-cbb42dae1700

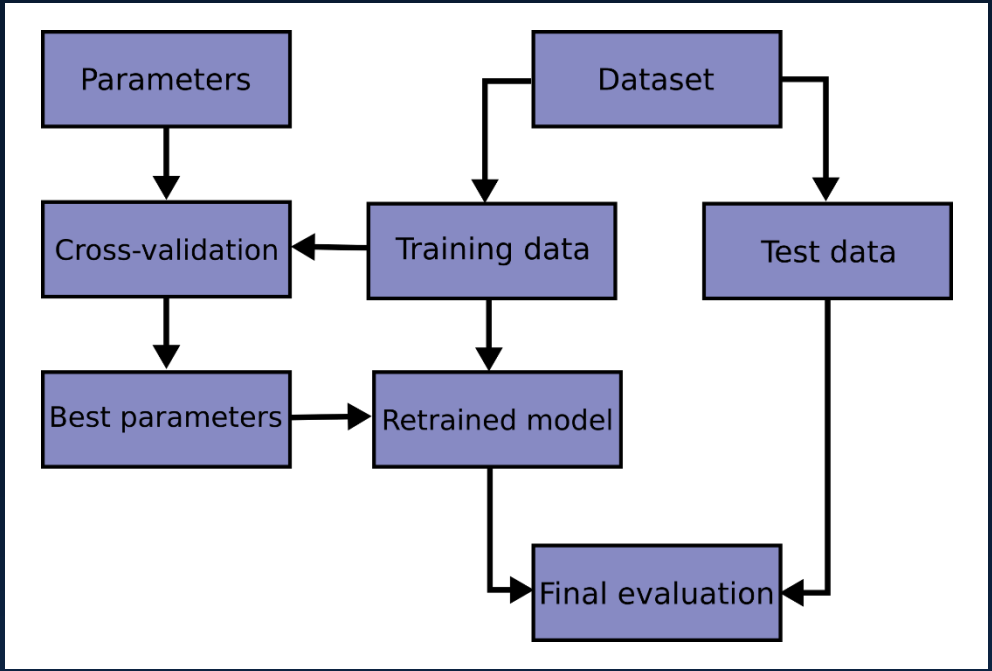
Machine learning: Feature Engineering
Feature Transformation

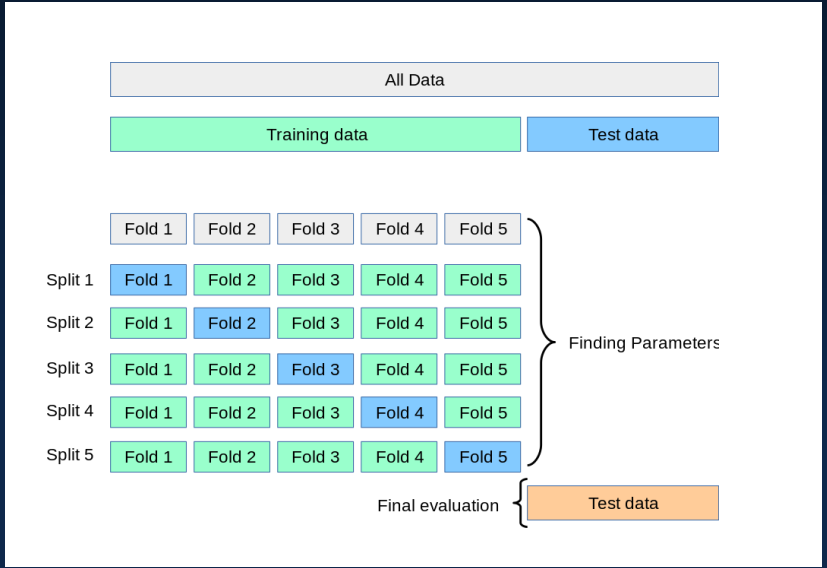
Feature transformation modifies the data into a more suitable format for modeling, helping to improve model performance and interpretability.

Feature Transformation Techniques

Technique	Description	Pros	Cons	Use Case
Log Transformation	Reduces skewness in data	Handles skewed data well	Not applicable for negative values	Skewed data
Box-Cox Transformation	Transforms data to be more normal	Flexible and powerful	Requires positive data	Data requiring normalization
Polynomial Features	Adds polynomial terms	Captures non-linear relationships	Can create very large feature space	Non-linear relationships in data

medium.com/@sanjayskumar4010/simple-guide-for-data-imputation-encoding-scaling-and-feature-transformation-techniques-cbb42dae1700





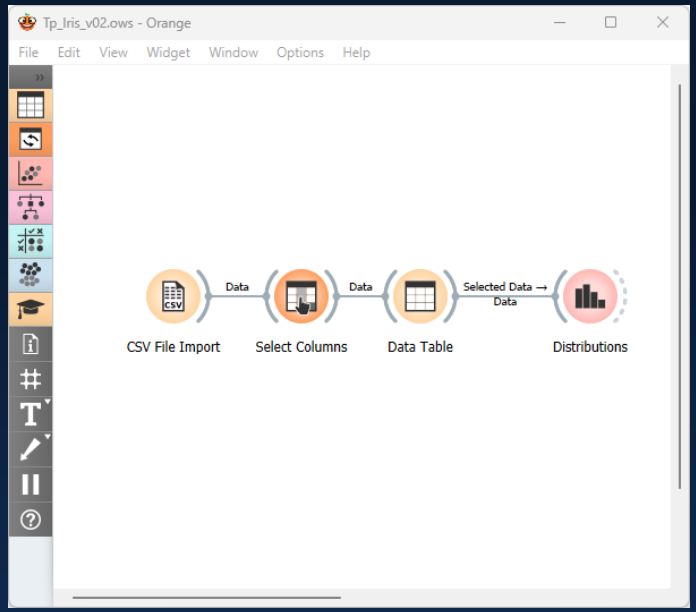
https://scikit-learn.org/stable/modules/cross_validation.html

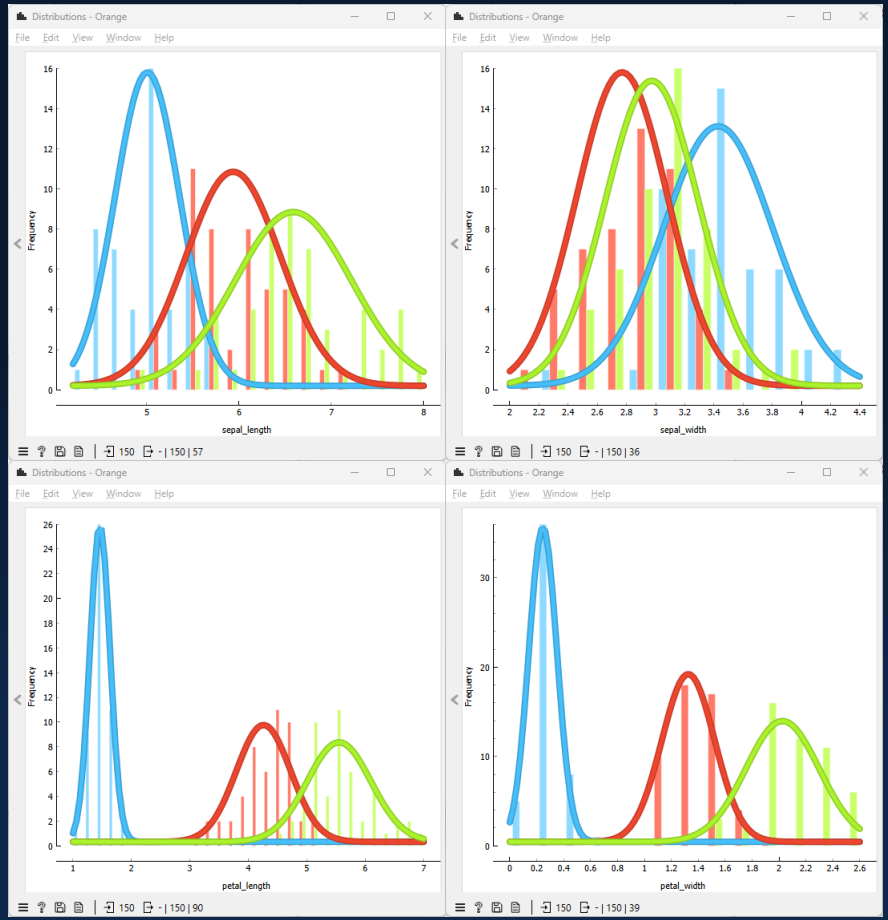
Machine learning: Feature Engineering
Feature selection

Feature selection

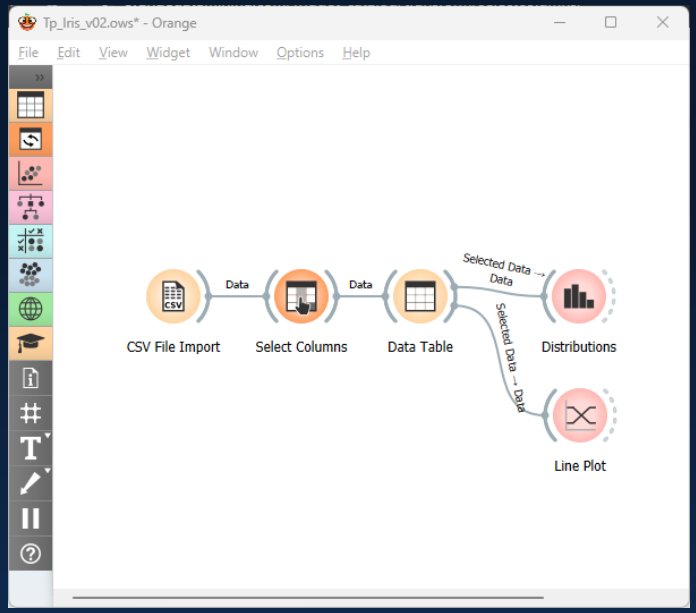
- Human expertise
- Graphical analyses
- Mutual information
- Correlation analyses
- Principal component analysis ...

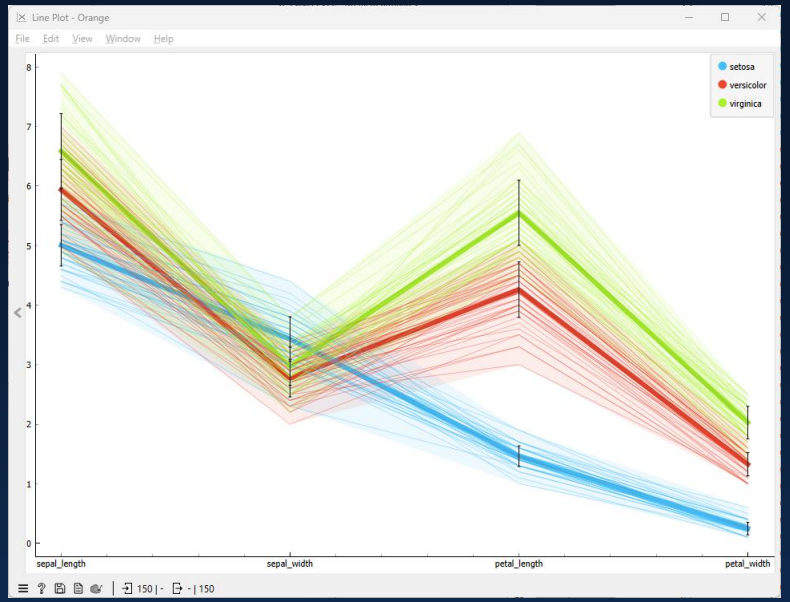
Machine learning: Feature Engineering
Feature selection



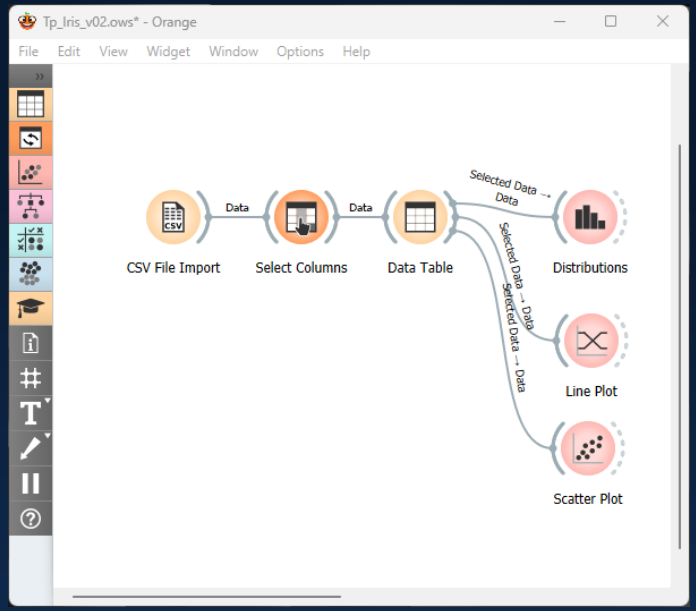


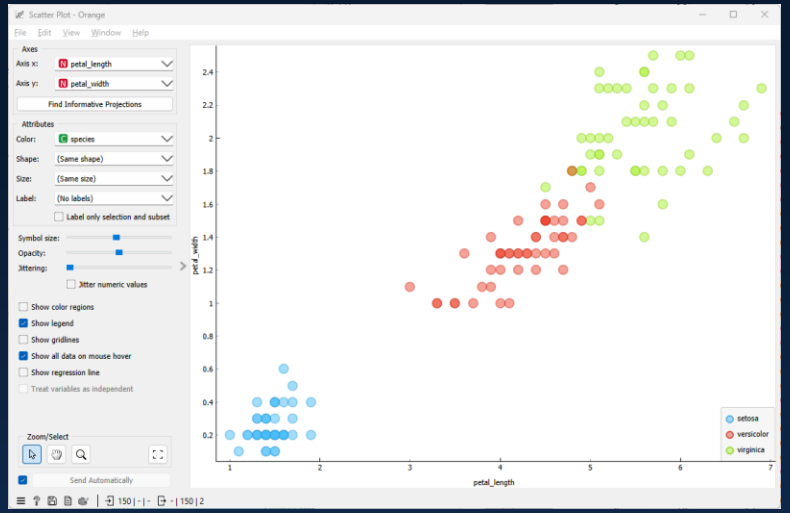
Machine learning: Feature Engineering
Feature selection



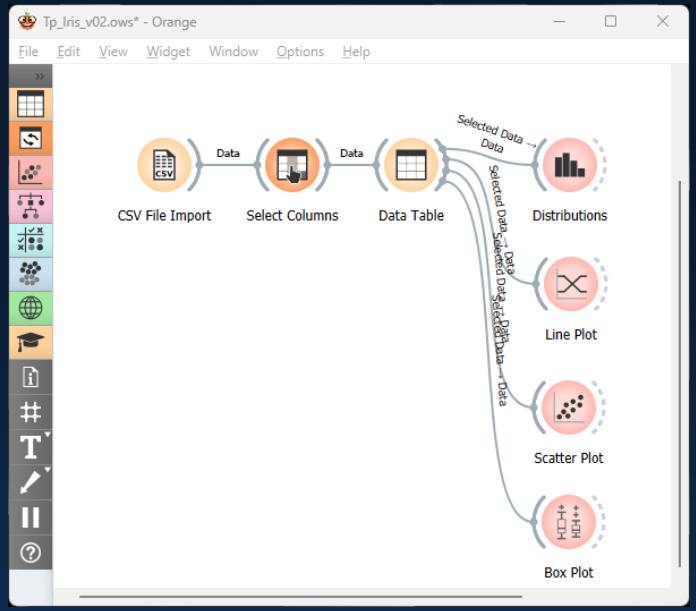


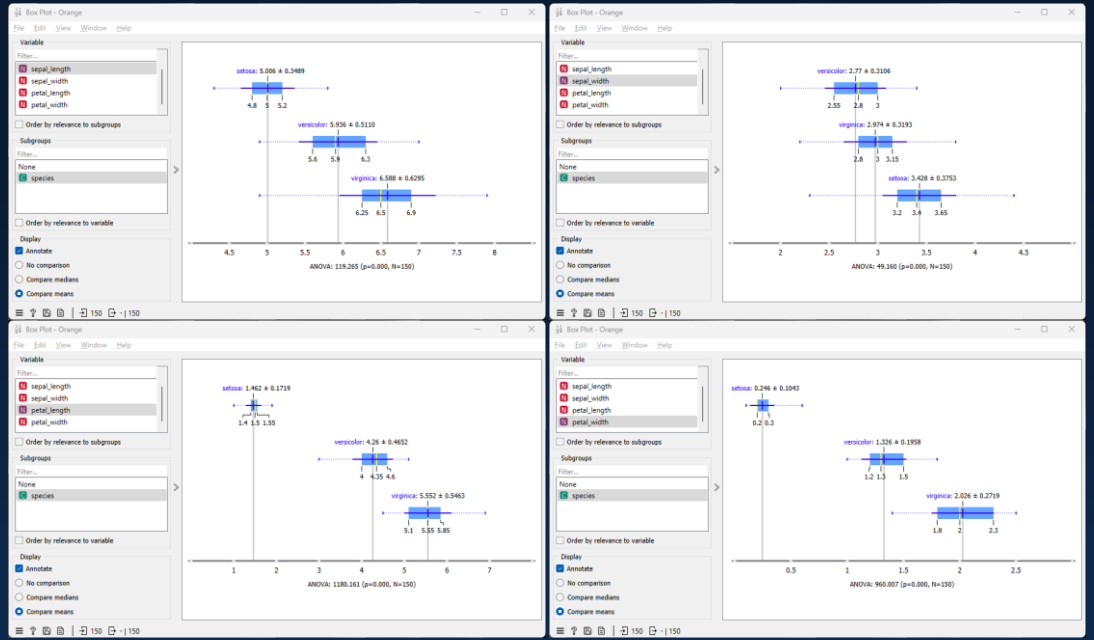
Machine learning: Feature Engineering
Feature selection





Machine learning: Feature Engineering
Feature selection



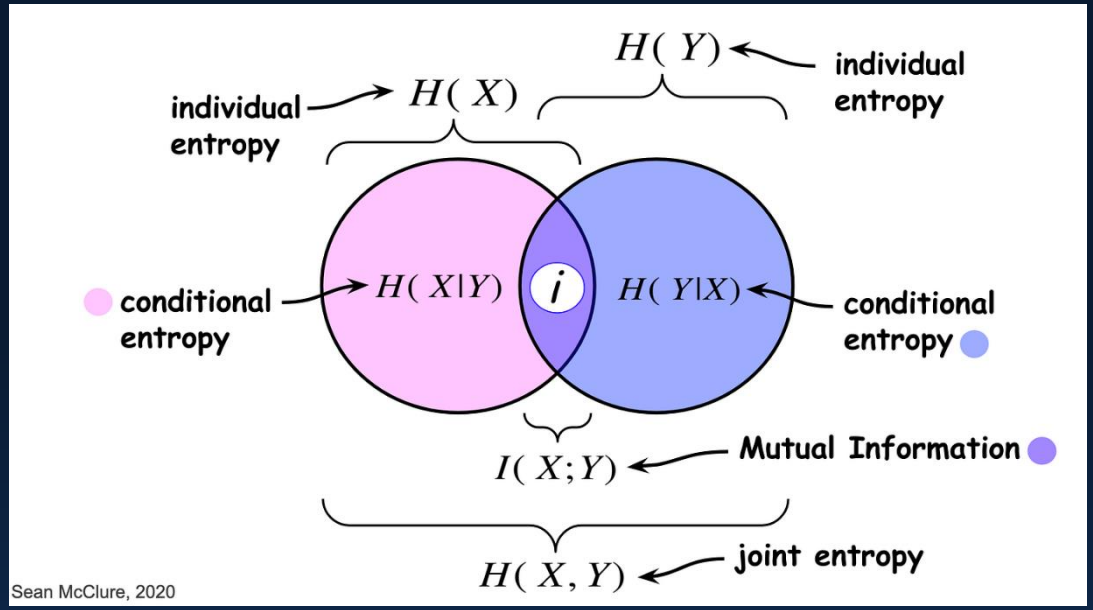


Machine learning: Feature Engineering
Feature selection

Y is a feature

X is the class

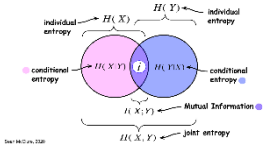
$$\begin{aligned} \mathrm{I}(X;Y) &\equiv \mathrm{H}(X) - \mathrm{H}(X \mid Y) \\ &\equiv \mathrm{H}(Y) - \mathrm{H}(Y \mid X) \end{aligned}$$



$$H[x] = E[h(x)] = -\sum_x p(x)h(x) = -\sum_x p(x)\log_2 p(x)$$

$$H[x|y] = -\sum_x \sum_y p(x,y)\log_2 p(x|y) = -\sum_x \sum_y p(x,y)\log_2 \frac{p(x,y)}{p(y)}$$

Machine learning: Feature Engineering
Feature selection



$$H[x] = E[h(x)] = -\sum_x p(x)h(x) = -\sum_x p(x)\log_2 p(x)$$

$$H[x|y] = -\sum_x \sum_y p(x,y)\log_2 p(x|y) = -\sum_x \sum_y p(x,y)\log_2 \frac{p(x,y)}{p(y)}$$

$$\mathrm{I}(X;Y) \equiv \mathrm{H}(X) - \mathrm{H}(X \mid Y)$$

		X = Play			X = Play				X = Play								
		No	Yes		No	Yes	Marginal p(x)		No	Yes							
Y = Outlook	Overcast	0	4	Overcast	0.00	0.29	0.29	Overcast	0.00	0.00							1
	Rainy	2	3	Rainy	0.14	0.21	0.36	Rainy	0.19	0.16							
	Sunny	3	2	Sunny	0.21	0.14	0.36	Sunny	0.16	0.19	H(Play)=	0.94	H(Play|Outlook)=	0.69	I(Play,Outlook)=	0.25	
				Marginal p(y)	0.36	0.64	1.00										

		X = Play			X = Play				X = Play								
		No	Yes		No	Yes	Marginal p(x)		No	Yes							
Y = Temperature	Cool	1	3	Cool	0.07	0.21	0.29	Cool	0.14	0.09							4
	Hot	2	2	Hot	0.14	0.14	0.29	Hot	0.14	0.14							
	Mild	2	4	Mild	0.14	0.29	0.43	Mild	0.23	0.17	H(X=Play)=	0.94	H(Play|Temperature)=	0.91	I(Play,Temperature)=	0.03	
				Marginal p(y)	0.36	0.64	1.00										

		X = Play			X = Play				X = Play								
		No	Yes		No	Yes	Marginal p(x)		No	Yes							
Y = Humidity	High	4	3	High	0.29	0.21	0.50	High	0.23	0.26							2
	Normal	1	6	Normal	0.07	0.43	0.50	Normal	0.20	0.10	H(X=Play)=	0.94	H(Play|Humidity)=	0.79	I(Play,Humidity)=	0.15	
				Marginal p(y)	0.36	0.64	1.00										

		X = Play			X = Play				X = Play								
		No	Yes		No	Yes	Marginal p(x)		No	Yes							
Y = Windy	False	2	6	False	0.14	0.43	0.57	False	0.29	0.18							3
	True	3	3	True	0.21	0.21	0.43	True	0.21	0.21	H(X=Play)=	0.94	H(Play|Windy)=	0.89	I(Play,Windy)=	0.05	
				Marginal p(y)	0.36	0.64	1.00										

Machine learning: Feature Engineering
Feature selection