DATA SCIENCE &
ARTIFICIAL INTELLIGENCE

https://dsai.units.it/

SCIENTIFIC &
DATA-INTENSIVE COMPUTING

https://sdic.units.it/

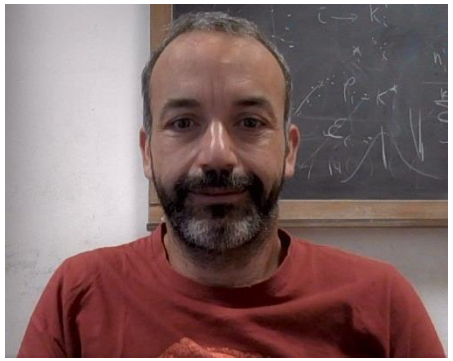
DATA SCIENCE &
SCIENTIFIC COMPUTING

https://dssc.units.it/


# Unsupervised Machine Learning

Introduction lecture

# About us…



Alex Rodriguez*

Ilaria Vascotto

*Office: C5_3.18 email: alejandro.rodriguezgarcia@units.it

# Today…

- Practical things:
  - Lecture schedule
  - Calendar
  - Lab. Activities
  - Material
  - Evaluation
- Unsupervised Machine Learning
  - What is?
  - What is for?
  - Why?

# Lessons: when and where

**CD2023 362SM UNSUPERVISED LEARNING**

**TEAMS code** : 6whfo58


**Monday 12-13** Aula 1B [Edificio D-Economia]

**Tuesday 9-11** Aula 1B [Edificio D-Economia]

**Friday 11-13** Aula 4C [Edificio H2, H2bis (e colleg. con H3)] (lab)

# About your background…

1. Artificial Intelligence and Data Analysis
2. Mathematics
3. Statistics
4. Computer Science
5. Physics/Chemistry
6. Others…

# Prerequisites

- **Theory**: basic notions of algebra and analysis (matrices, derivatives etc.)
- **Labs**: Python

# Content of the course

1. Basic notions about Unsupervised Machine Learning.
2. Dimensionality Reduction methods: General theory, classical methods and advance techniques.
3. Intrinsic dimension estimation methods
4. Density Estimation methods: Histograms, kernel density estimation and k-NN. Advanced methods.
5. Clustering: General Theory. Classification of clustering methods. Classical algorithms. Overview about recent developments and new methods. Clustering validation.

# Calendar

- We are not lecturing:
  - Oct. 16,17,20 (to be confirmed).
  - Nov. 3.
  - Dec. 8.
- Expected finish by Dec. 15

# Lab. activities

- **Coding the methods**
  - Understanding (efficacy ) >> Fast programs (efficiency)
  - The use of libraries is allowed for numerical manipulation and linear algebra operations, the methods should be coded from scratch.
- **Python programming on Jupyter notebooks**
  - In many lectures, we will apply the methods to toy models in order to understand the main features/problems.
  - In some labs, we will use more realistic models, so some data pre-treatment could be needed.

# Materials

- There’s no a single book that covers all the contents of this course.
- *Suggestion:* Make a common document with the notes. If you do it, let me know so I can participate and correct (if needed).
- I will provide in TEAMS the papers corresponding to the methods that we will treat during the course.
- I will record the lectures (please, If I forgot to start recording, do not hesitate to bring it to my attention) and the recording would be uploaded to TEAMS.
- Communications for the course would be given during the lectures or, if needed, post in TEAMS.
- I would provide you with the slides/notes ***at the end of each section***.

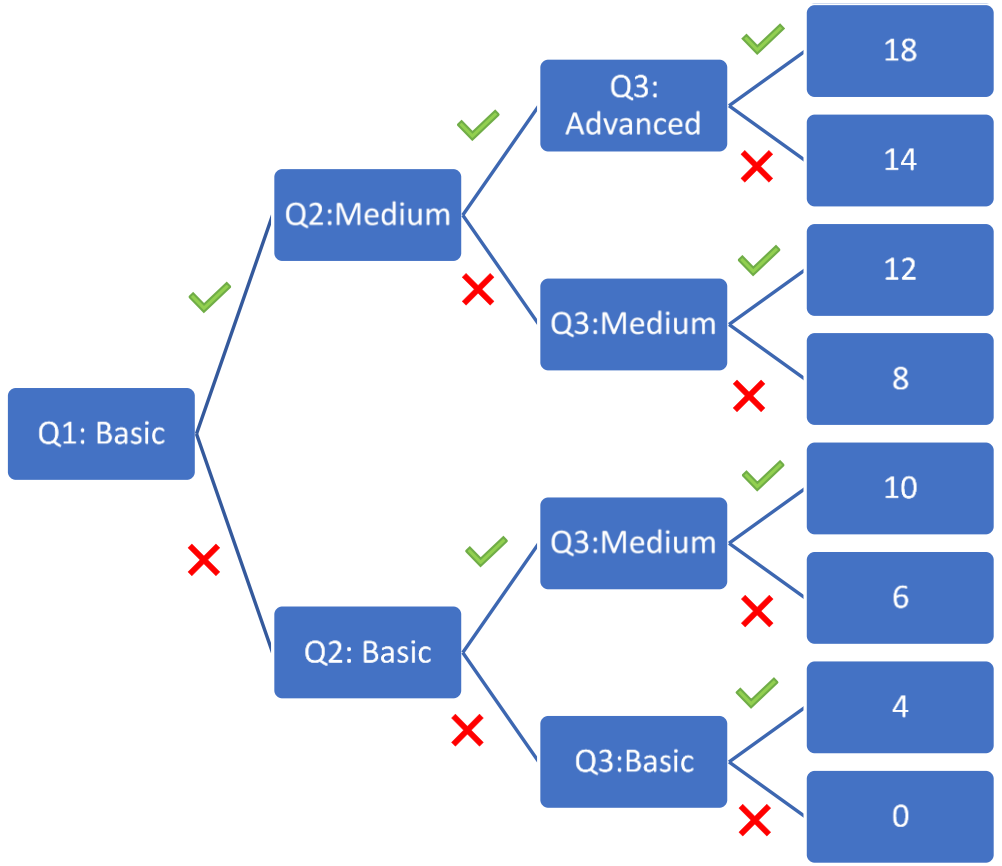
# Evaluation

- **The exam will consist in two parts:**
  1. *Presentation of a **project** where you explain a method from a research paper and compare it with the methods that we have seen during the exercises on the course.*
  2. *Three additional **questions** will be asked to assess the understanding of the topics of the course.*
- **Project from a research paper (10 minutes, max. 12 points):**
  - I will give a list of papers by the end of the course (You can suggest me a paper/method from outside the list).
  - To implement yourself the method is highly recommended, although not mandatory when the implementation is too difficult.
  - Explain the method and compare the performance and pros/cons with those from the exercises.
- **Questions (max. 18 points):** The idea is to assess that you have a solid understanding of the basis of the methods, so , if you have studied, you can take it easy ☺.

# Evaluation (II)

**First question:** max. 8 points
**Second question:** max. 6 points
**Third question:** max. 4 points

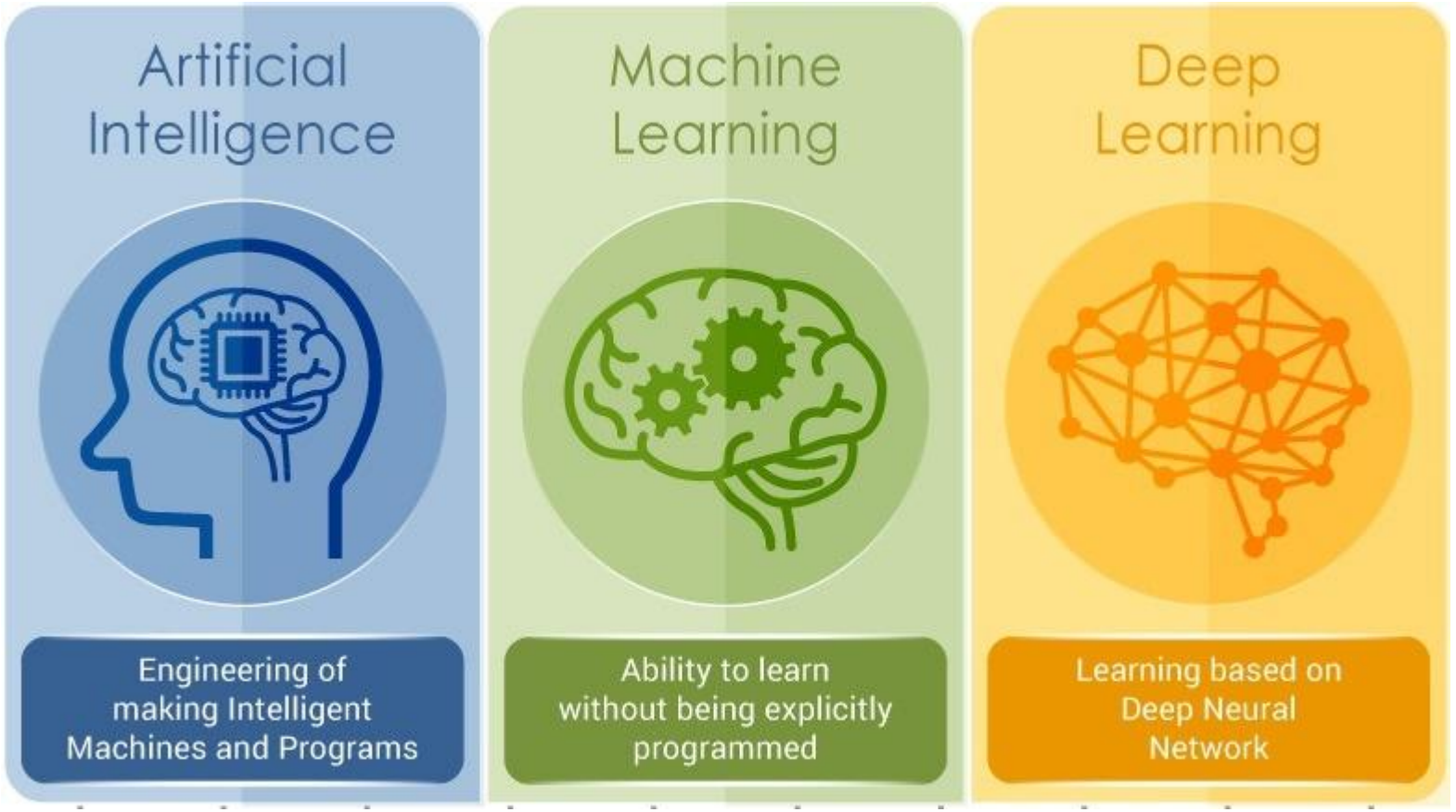
The first question would be chosen among the basic ones. If the student replies correctly the difficulty of the questions would increase, otherwise we will do questions of equivalent difficulty.

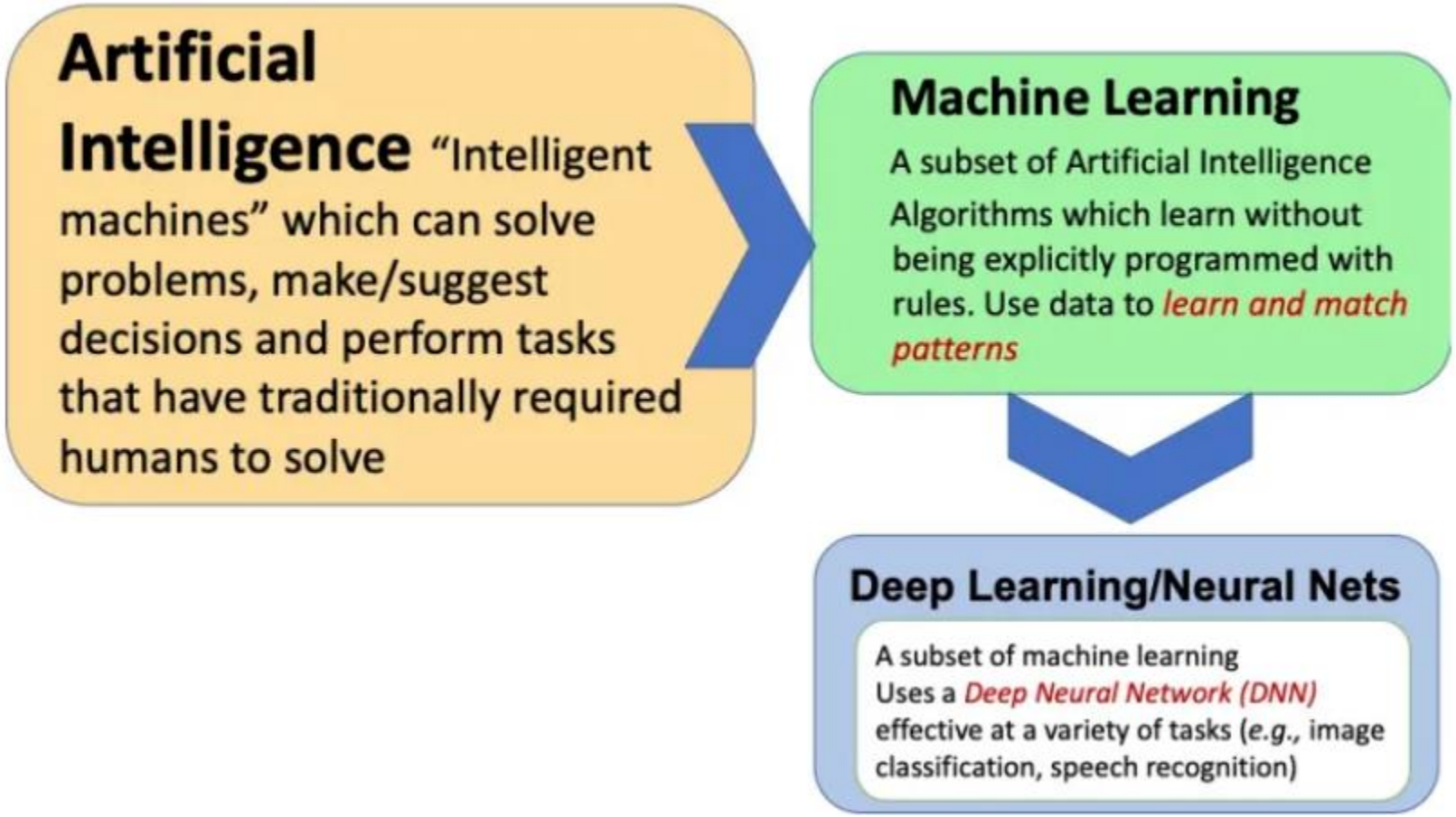


So let's start!!!

# What is machine learning?

Machine learning is part of artificial intelligence that deals with building methods that '**learn**', that is, methods that use data to improve performance on some set of tasks (e.g. image recognition). It is seen as a part of artificial intelligence.




Artificial Intelligence "Intelligent machines" which can solve problems, make/suggest decisions and perform tasks that have traditionally required humans to solve
Machine Learning
A subset of Artificial Intelligence
Algorithms which learn without being explicitly programmed with rules. Use data to learn and match patterns
Deep Learning/Neural Nets
A subset of machine learning
Uses a Deep Neural Network (DNN)
effective at a variety of tasks (e.g., image classification, speech recognition)

# What is machine learning?

- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn from data without being explicitly programmed.

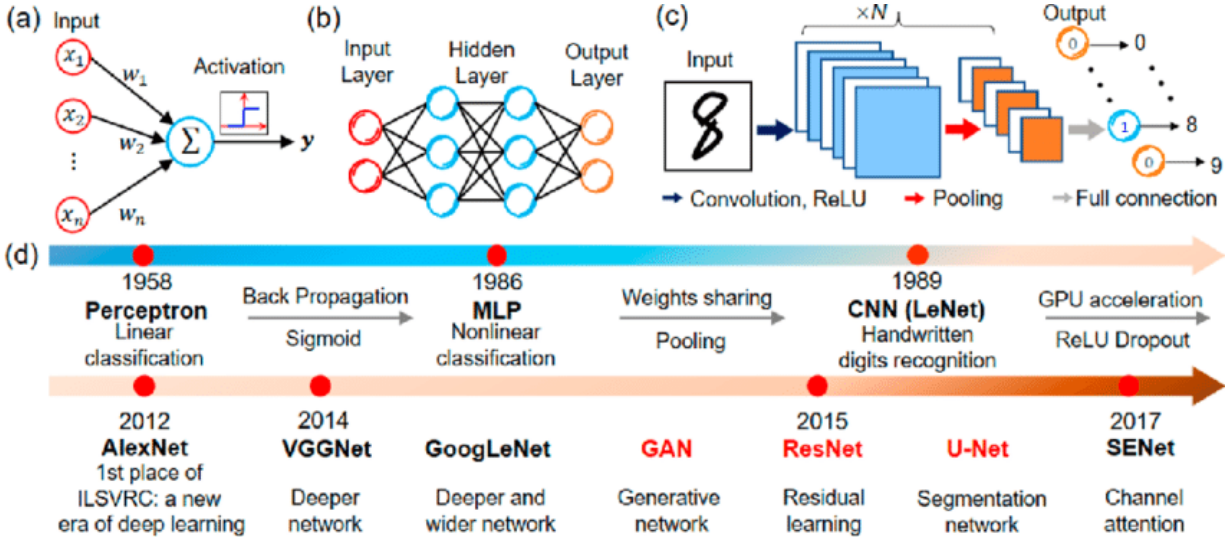
- Main difference with traditional Computer Science: learn a program that deals with data (Machine Learning) vs have a coded program that run the data (Computer Science). *Optimization not logic*.

# What made possible the ML success?

- Data abundance


- Algorithmic development


- Better computers



# What's Unsupervised Machine Learning

- **Unsupervised learning** is a type of algorithm that learns patterns from untagged data.
- Other levels of supervision:
  - Supervised Machine Learning: Data tagged by an expert. $p(y|x)$
  - Reinforcement Learning: Data untagged, but rules for performance evaluation are provided.
  - Semi-supervised Learning: Small portion of the data is tagged.

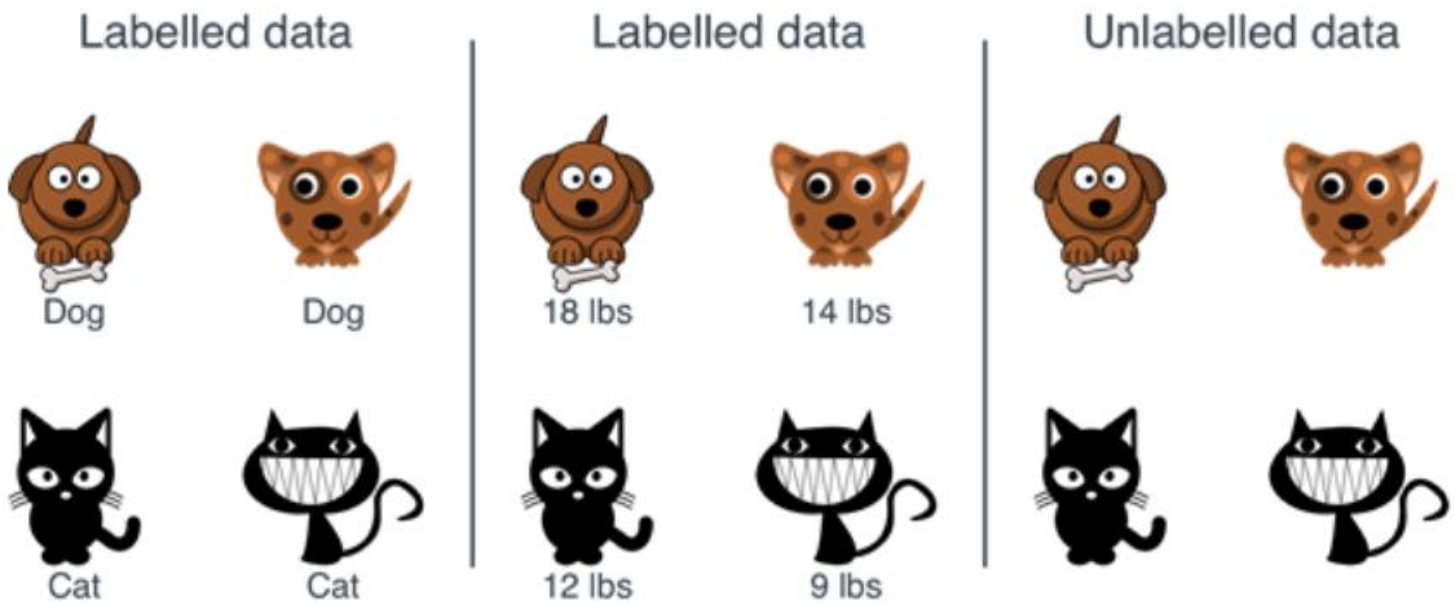
# Learning modes in ML

**SUPERVISED LEARNING**

**UNSUPERVISED LEARNING**

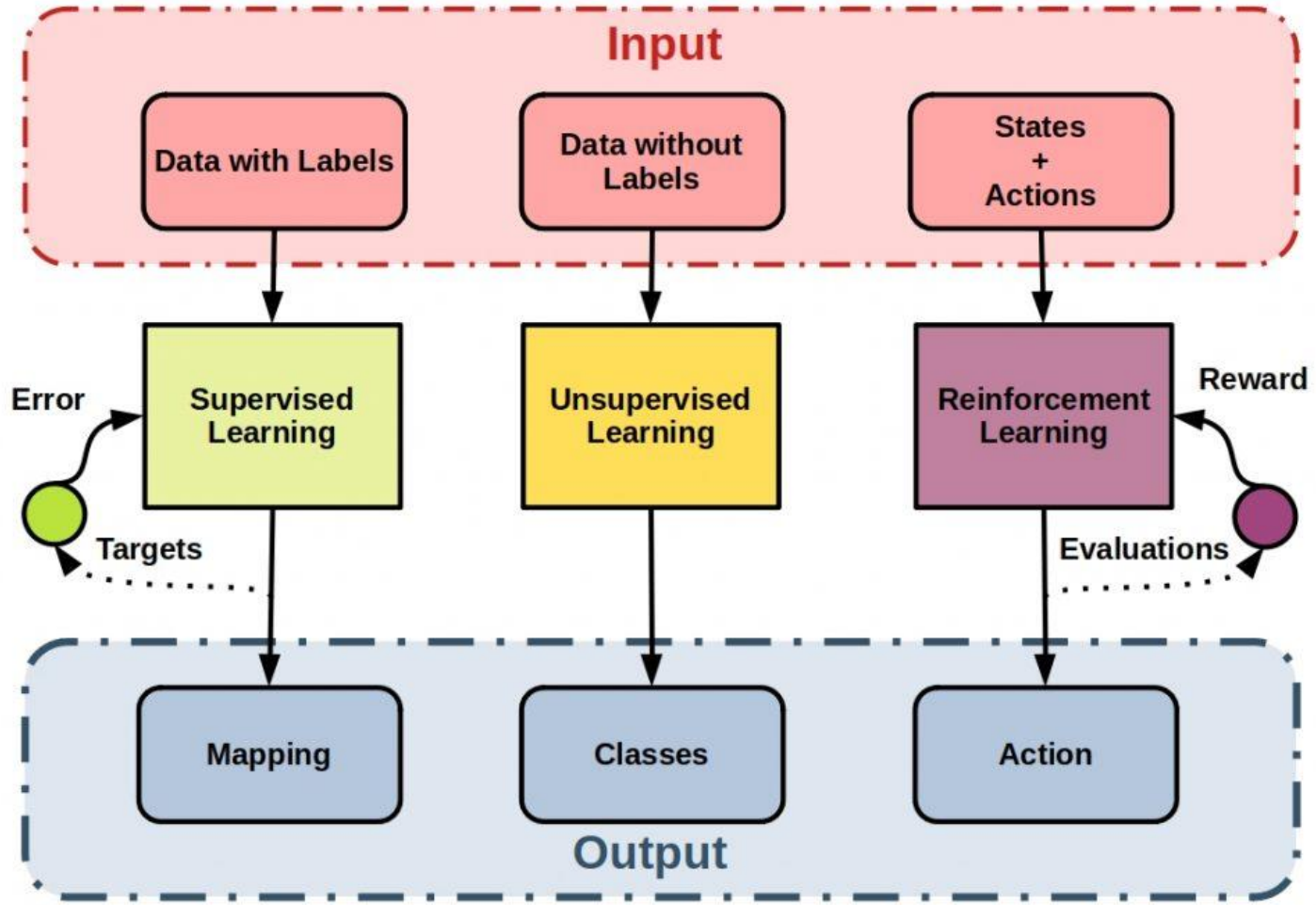
**REINFORCEMENT LEARNING**

# Learning modes in ML




The  labels are the provided by the teacher


ALGO
Dog
Cat

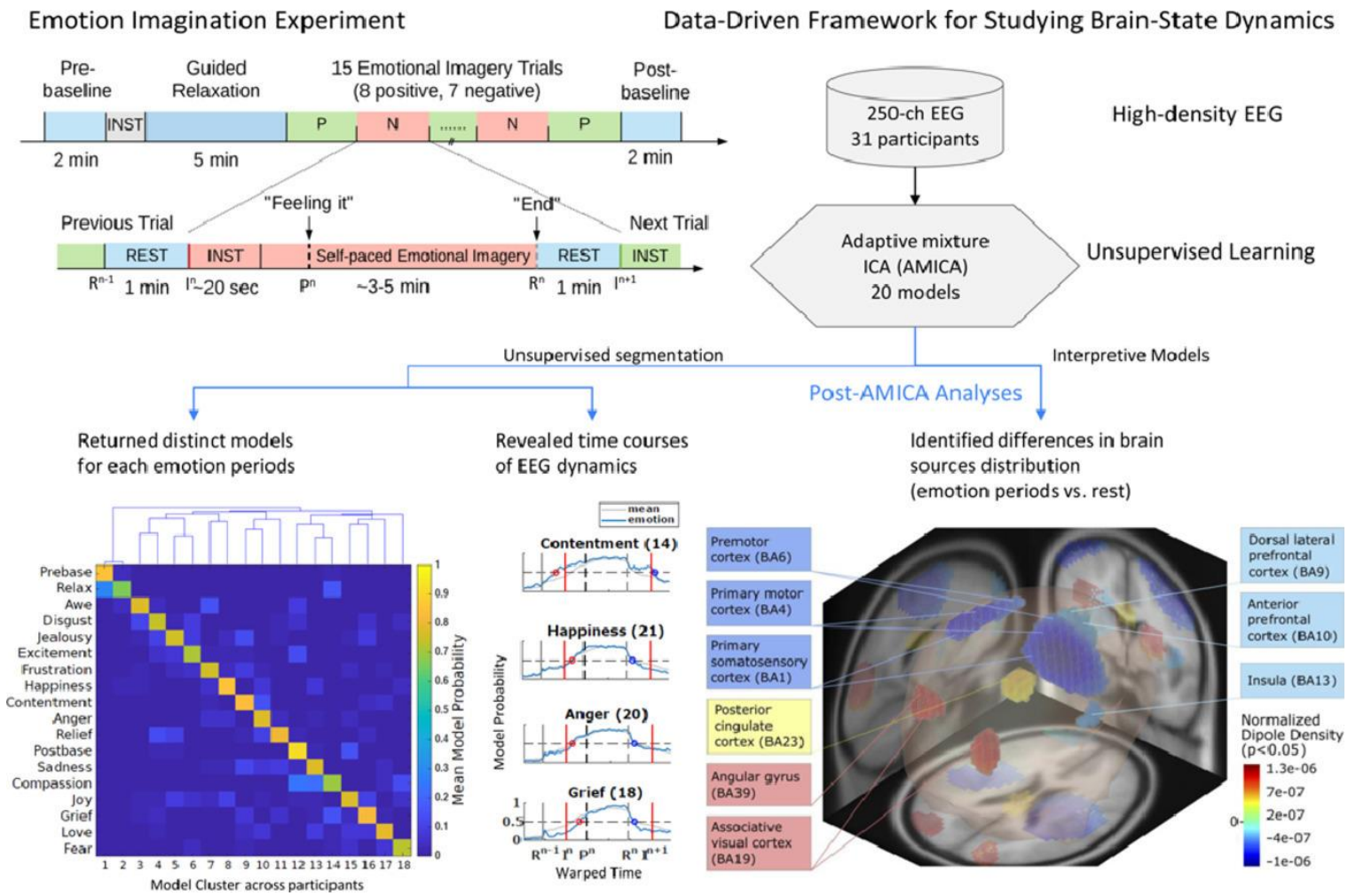

Input
Data with Labels
Data without Labels
States + Actions
Error
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Reward
Targets
Evaluations
Mapping
Classes
Action
Output

# Which patterns?

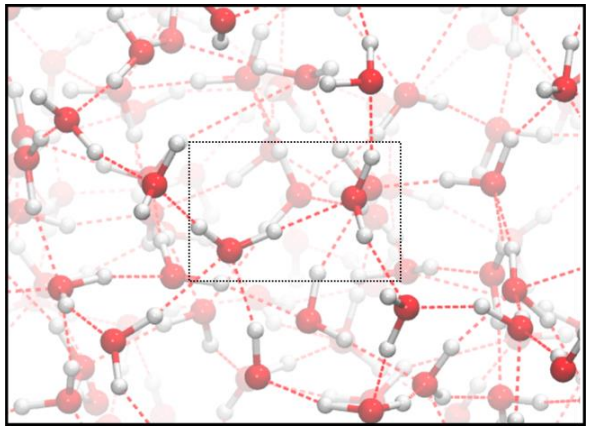
- In supervised learning we learn a conditional distribution $p(y|x)$ while in Unsupervised Learning we learn directly properties of the underlying probability distribution: $p(x)$.
- Different properties (patterns) that we learn from $p(x)$ correspond to different Unsupervised Learning task (we will see that on the next lecture).

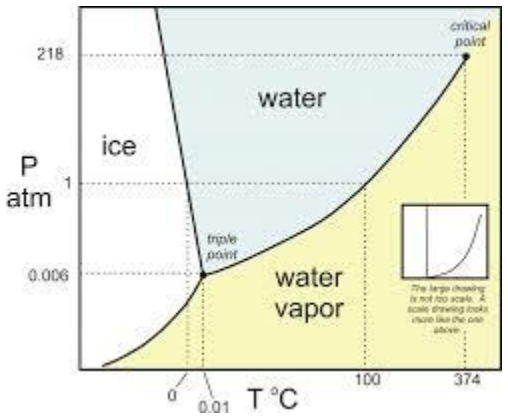
# Applications of Unsupervised Learning

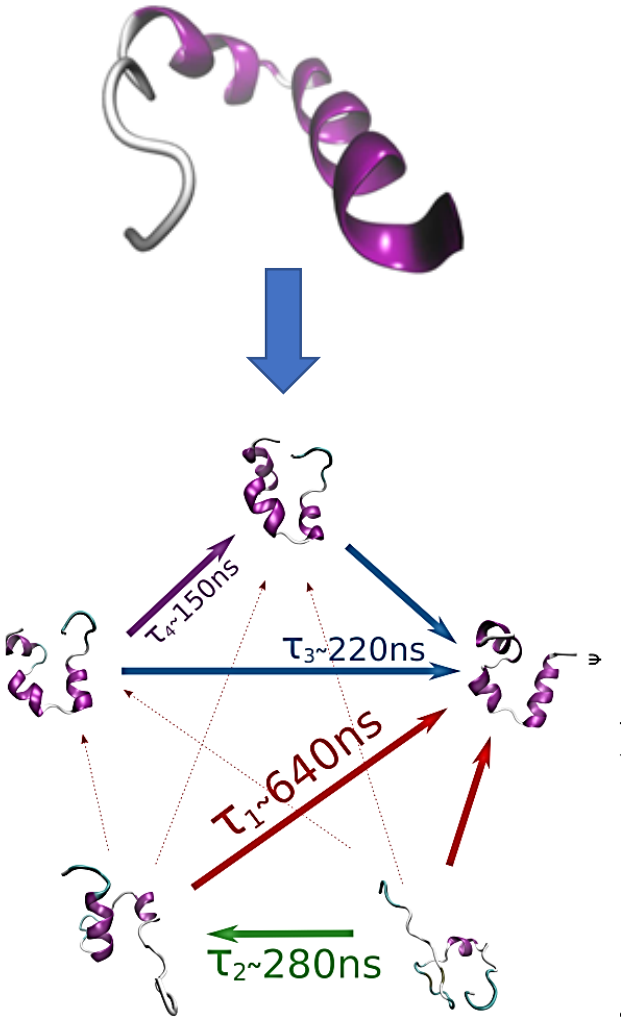
# Signal processing: Example from Neurosciences.



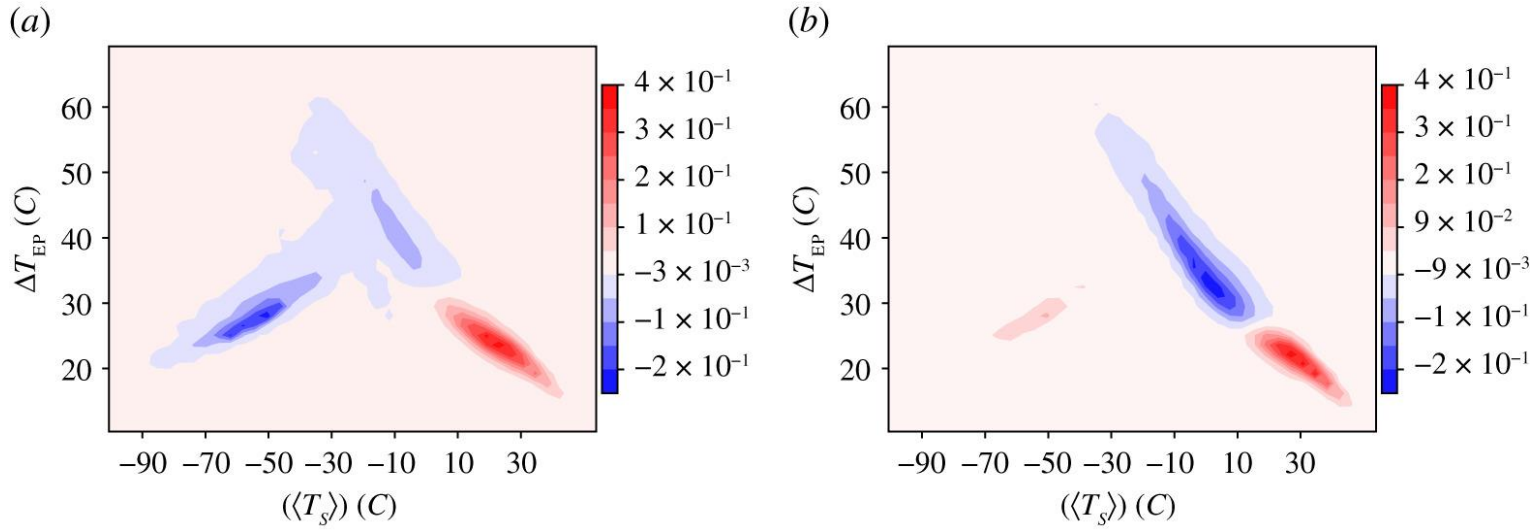
# Simulations insight (Physics/chemistry examples)

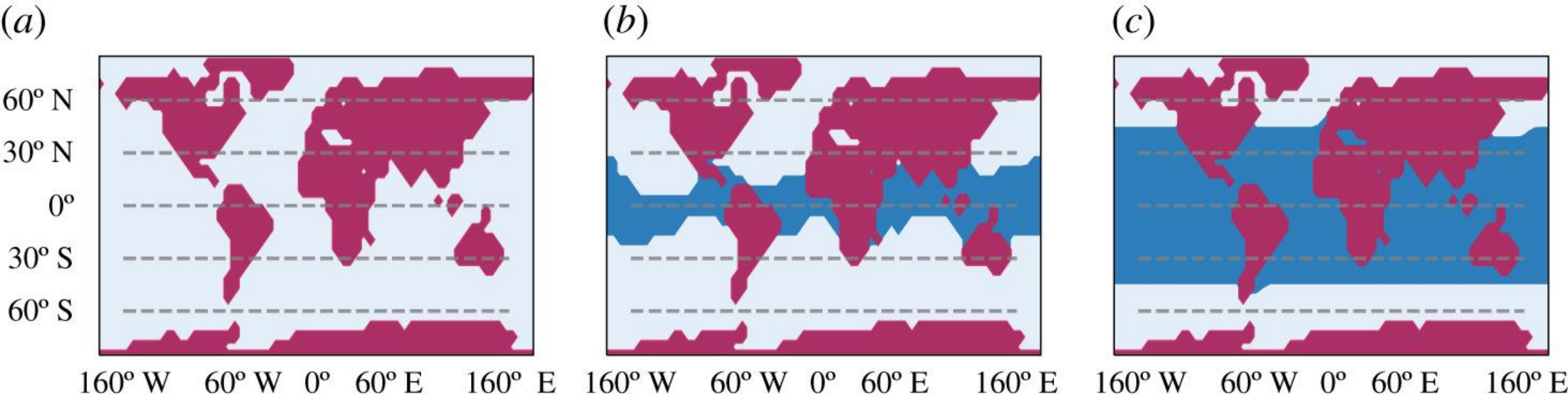




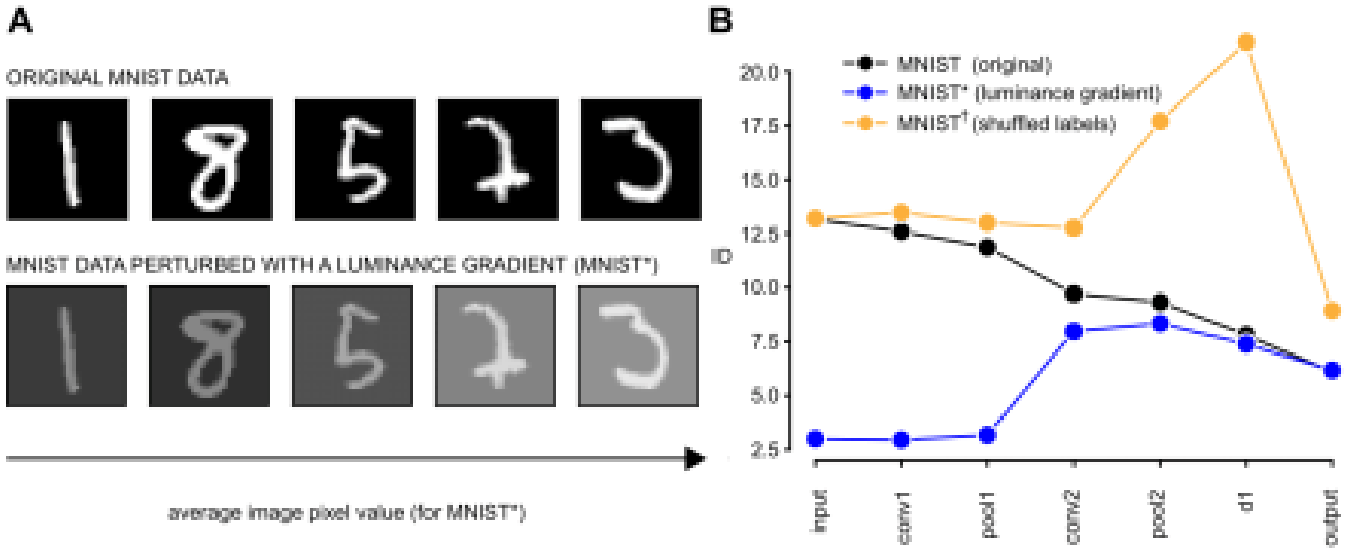



# Meteorological patterns



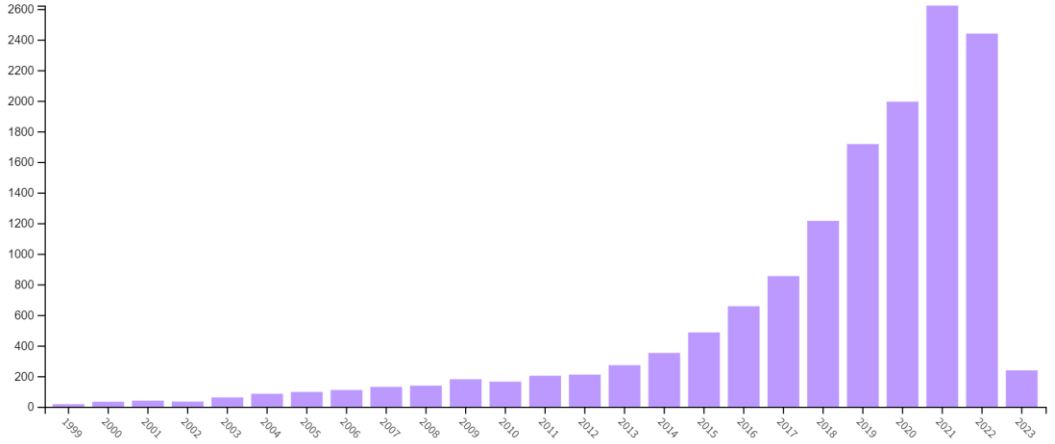



# Understanding Machine Learning methods



# Why shall you study Unsupervised Machine Learning?

- A lot of application fields…
- Furthermore, it forms part of many Machine Learning/Data Analytics pipelines.
- A lot of room for improvement, very active research field.




Citations according with Web Of Knowledge