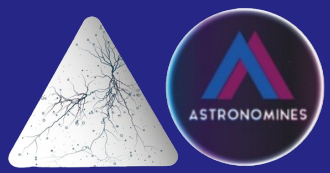


UNLOCKING AI
Essentiel Concepts for ML Mystery

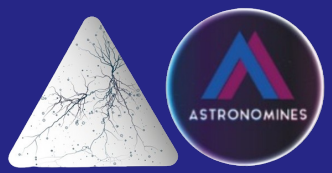
MAJDOUB Adnane

EAIC Co-Training Head

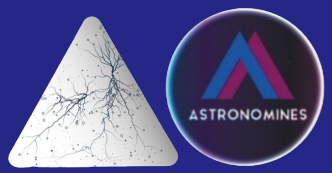
Second-year AI Engineering Student

ENSIAS AI Club x Astronomies

December 13, 2023



Summary

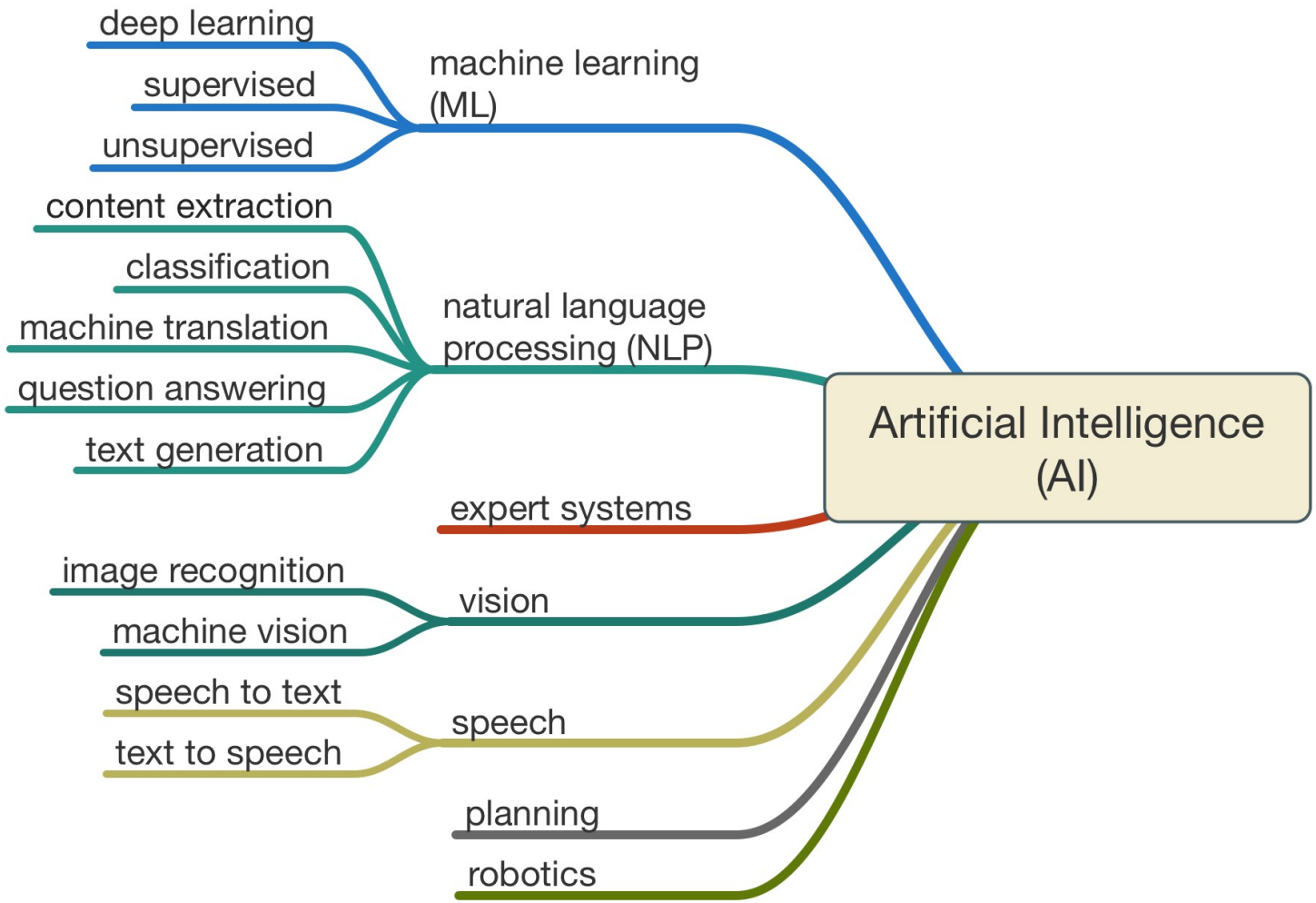


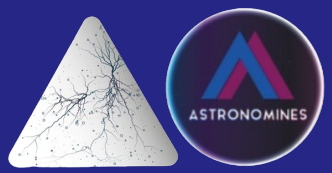
AI definition

Artificial Intelligence

Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems

Types of AI


Artificial Intelligence (AI)
machine learning (ML)
deep learning
supervised
unsupervised
natural language processing (NLP)
content extraction
classification
machine translation
question answering
text generation
expert systems
vision
image recognition
machine vision
speech
speech to text
text to speech
planning
robotics



ML Definition

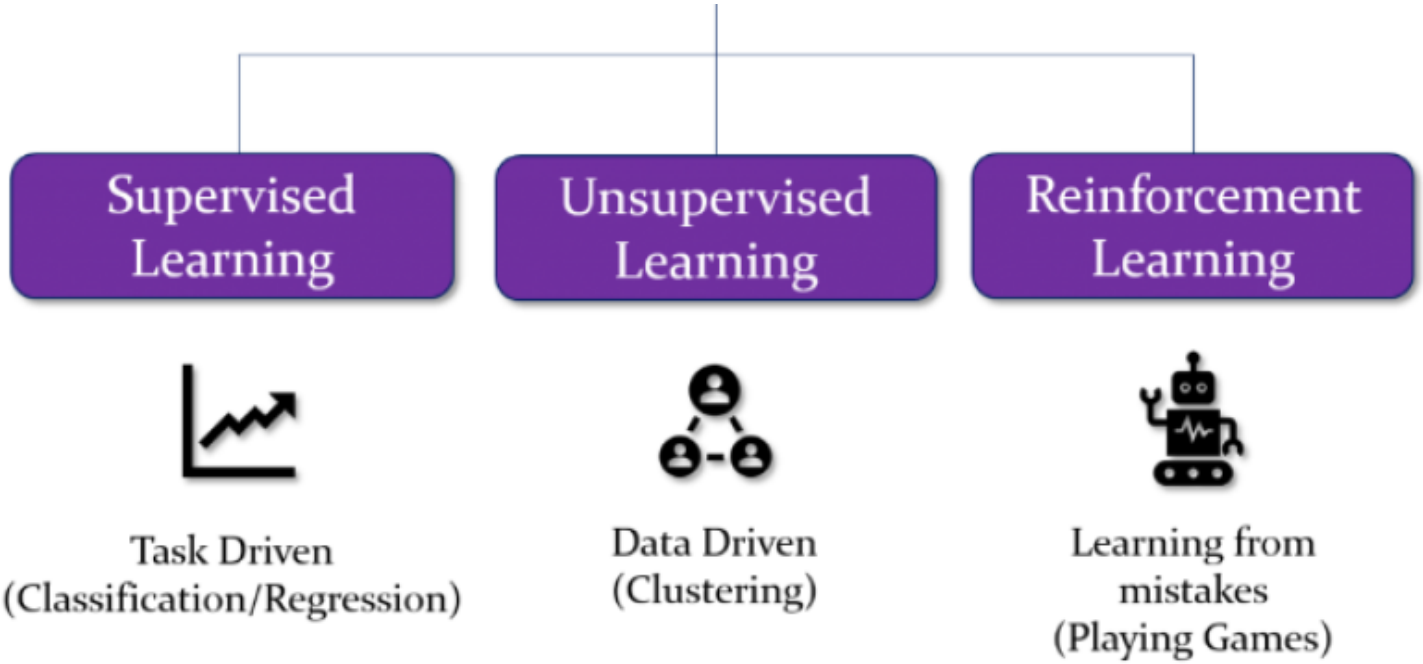
Machine Learning

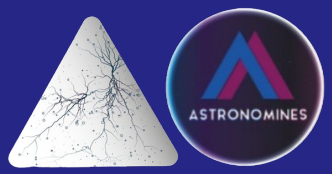
The process of making the machine able to learn from the data it got, thus without the intervention of human beings.

Learning From Experience

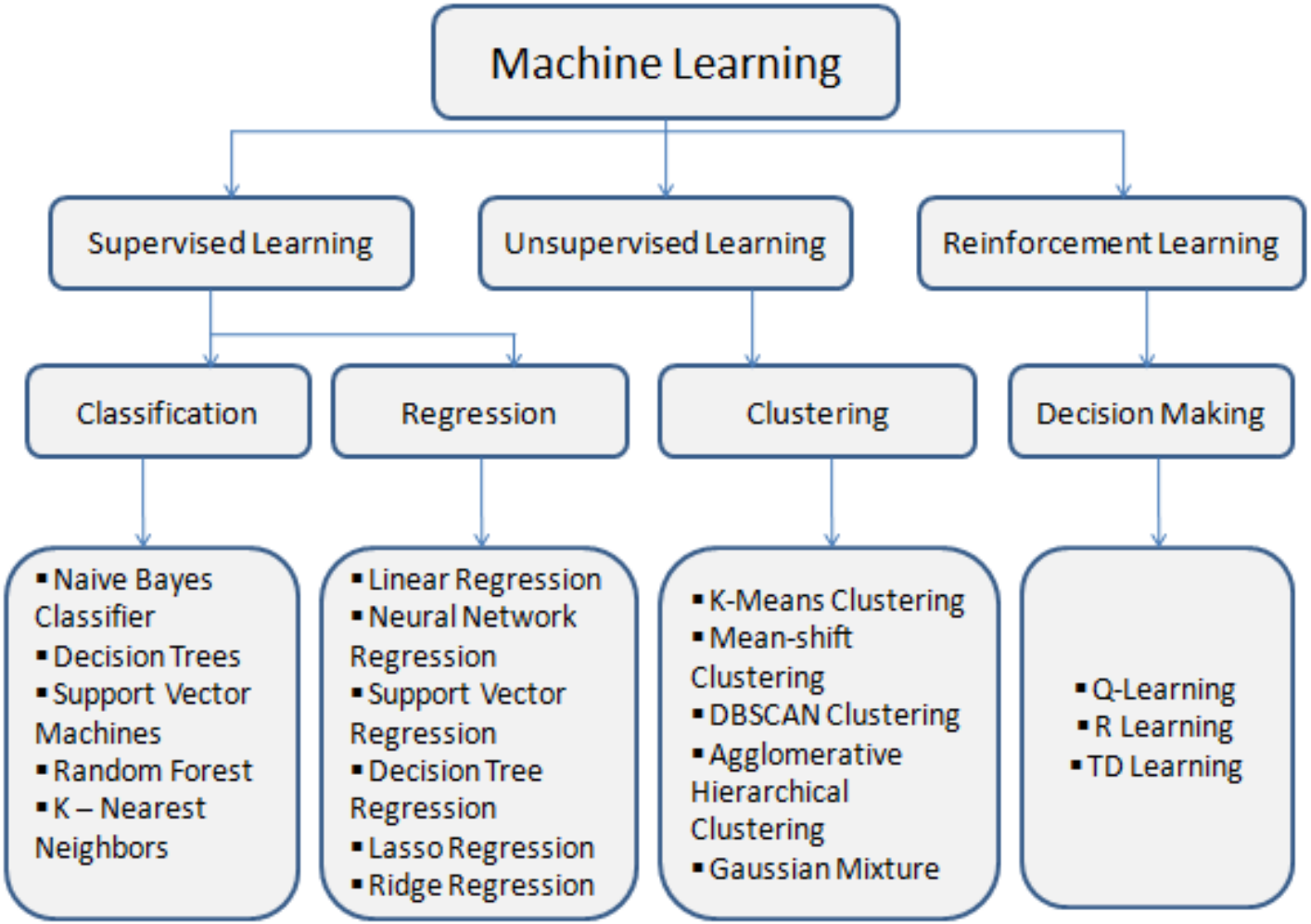
Data from the past + Algorithm $\rightarrow$ Find module (parameters) $\rightarrow$ Prediction of the future

Types of ML

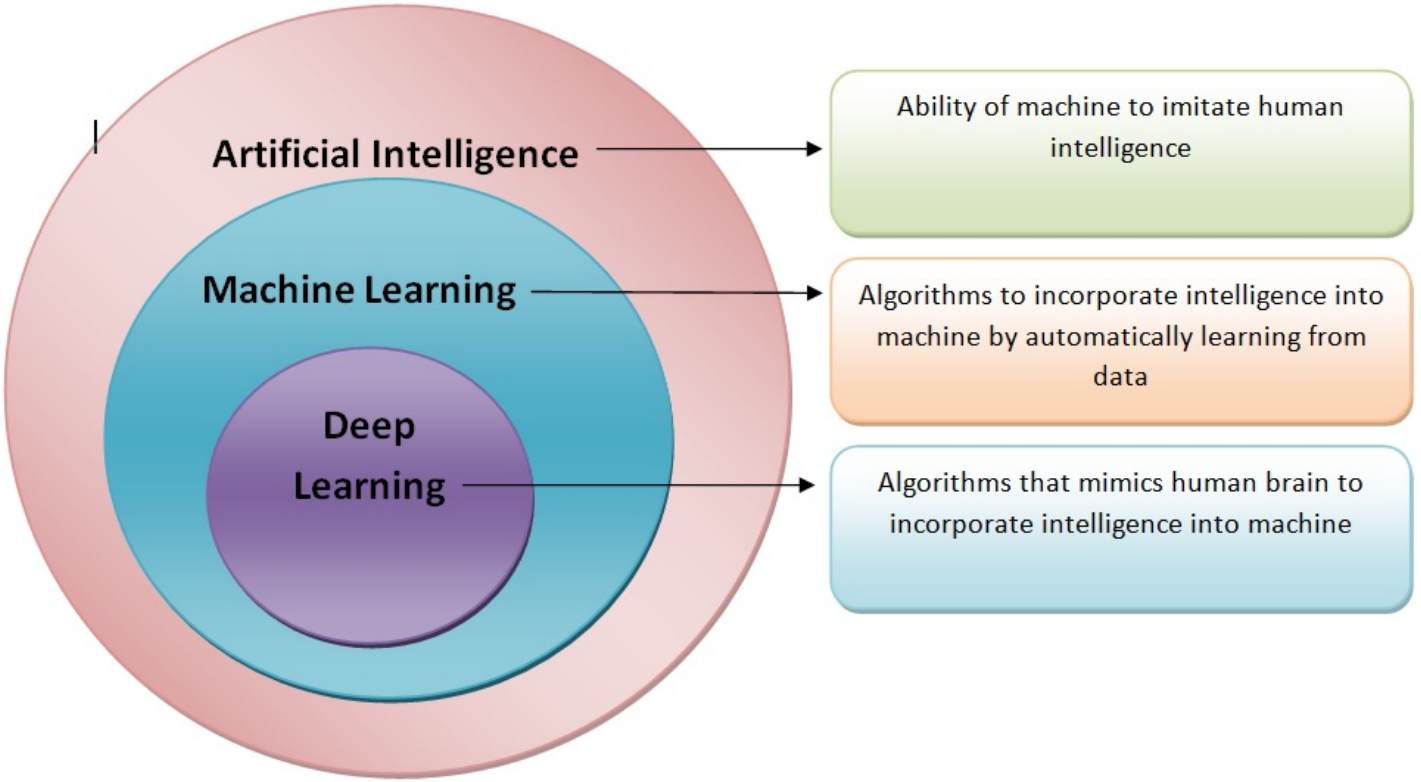

Supervised Learning
Unsupervised Learning
Reinforcement Learning
Task Driven (Classification/Regression)
Data Driven (Clustering)
Learning from mistakes (Playing Games)

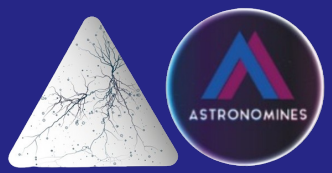


Types of ML


Machine Learning
Supervised Learning
Unsupervised Learning
Reinforcement Learning
Classification
Regression
Clustering
Decision Making
▪ Naive Bayes Classifier
▪ Decision Trees
▪ Support Vector Machines
▪ Random Forest
▪ K – Nearest Neighbors
▪ Linear Regression
▪ Neural Network Regression
▪ Support Vector Regression
▪ Decision Tree Regression
▪ Lasso Regression
▪ Ridge Regression
▪ K-Means Clustering
▪ Mean-shift Clustering
▪ DBSCAN Clustering
▪ Agglomerative Hierarchical Clustering
▪ Gaussian Mixture
▪ Q-Learning
▪ R Learning
▪ TD Learning

AI & ML & DL


Artificial Intelligence
Machine Learning
Deep Learning
Ability of machine to imitate human intelligence
Algorithms to incorporate intelligence into machine by automatically learning from data
Algorithms that mimics human brain to incorporate intelligence into machine



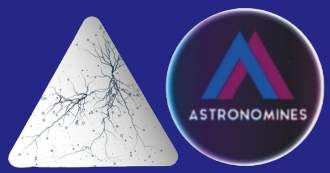
Algorithm

An Algorithm

A finite sequence of rigorous instructions, typically used to solve a class of specific problems or to perform a computation

Algorithm 1 Check Even or Odd

```
1: procedure CheckEvenOrOdd(n) ▷ Input: Integer n
2:     if n is divisible by 2 then
3:         Output: "Even"
4:     else
5:         Output: "Odd"
6:     end if
7: end procedure
```



DATA & DATASet

DATA

Facts and statistics collected together for reference or analysis.

Data set

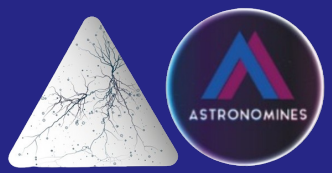
A data set (or dataset) is a collection of data.

Entity

An entity is something that exists and can be distinctly identified. For example, in a database of employees, each employee is an entity.

Attributes

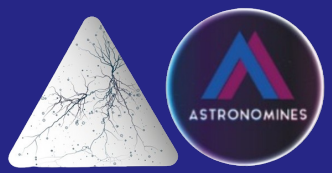
An attribute is a characteristic of an entity. For instance, in the context of an employee entity, attributes might include name, employee ID, age, and salary



Examples : Tabular Data

Titanic Data Set

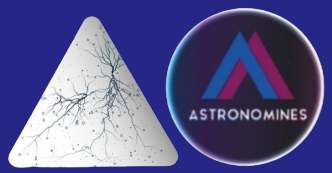
PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...
887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q



Examples : Tabular Data

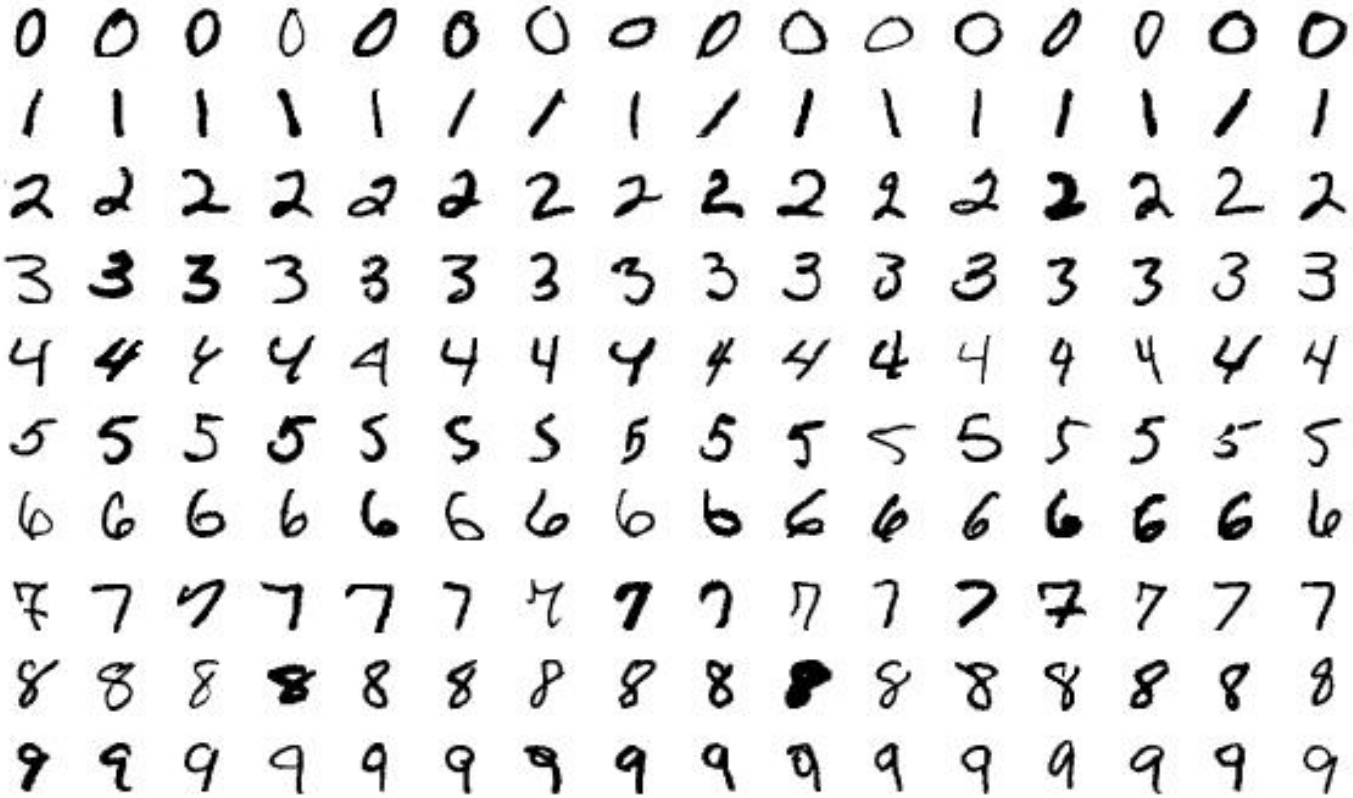
Features (Input Variables) The attributes that the model uses to make predictions or classifications

Target (Output Variables) The variable that the model is trying to predict or classify. It's what you want your machine learning model to learn and make predictions about



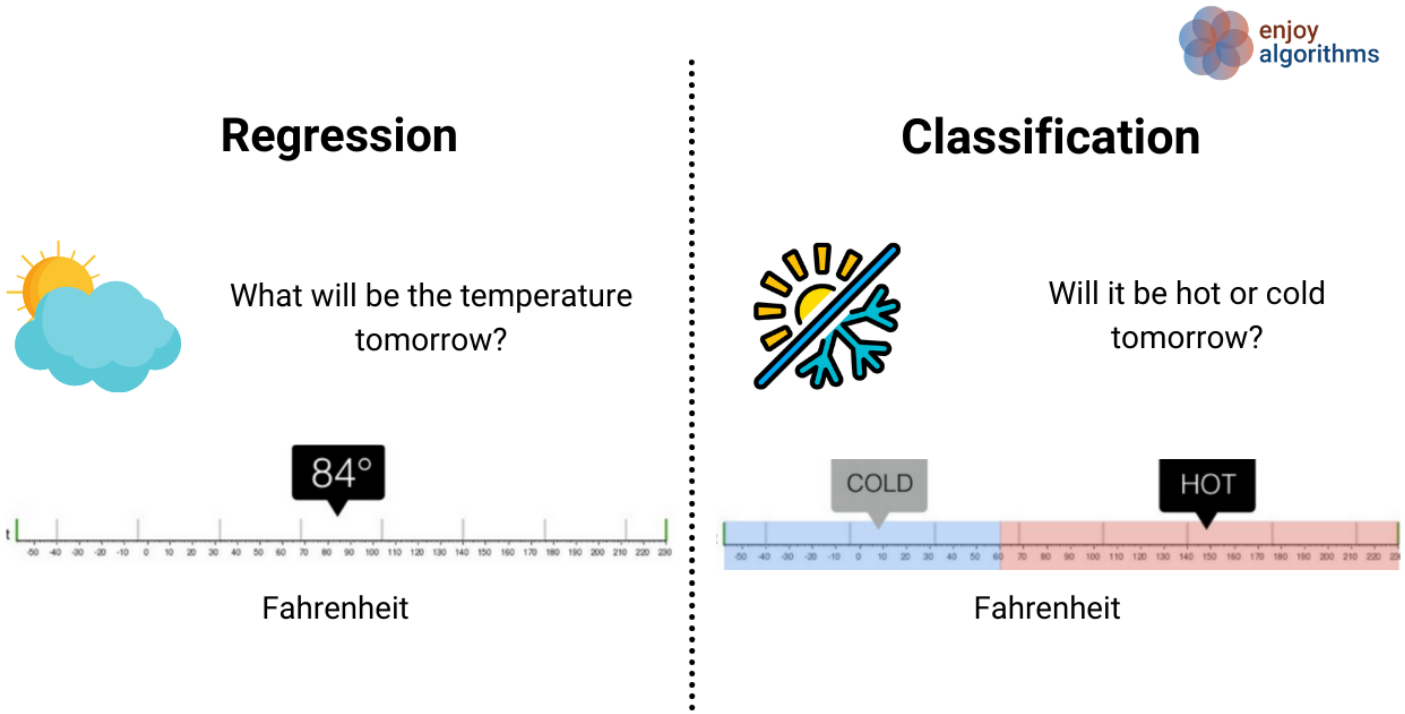
Exempels : Images Data

MNIST Data Set

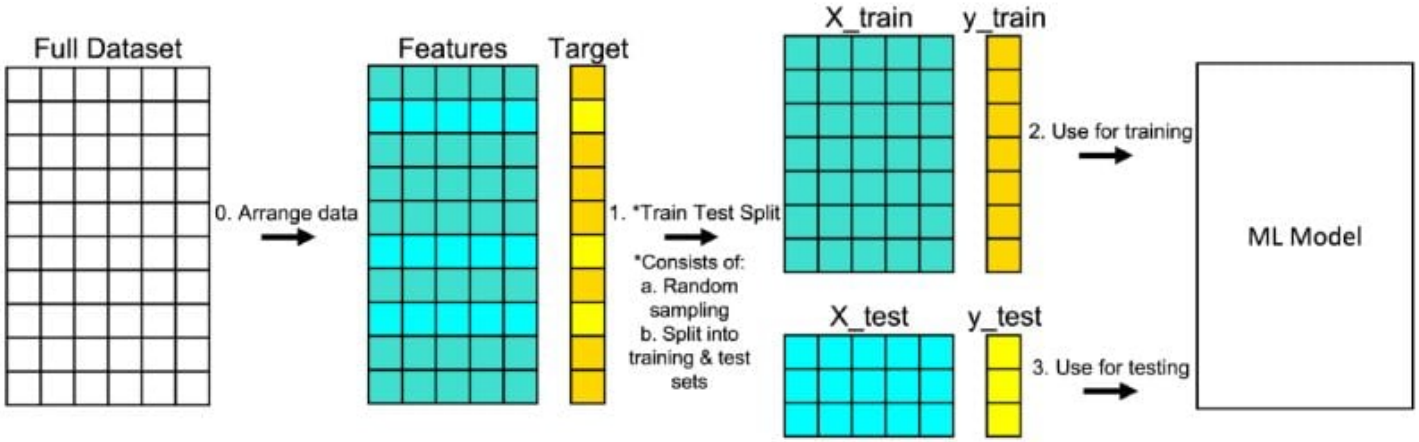


Objectif : Classification vs Regression

Is it a Classification or Regression task ?



How to evaluate our model ?



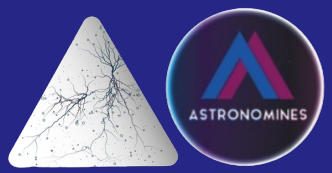
Error

Difference between a single actual value and a single predicted value

Lost

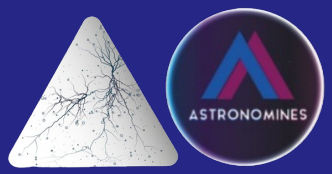
The average error over training data

Does this task really required ML ?



When to Use Machine Learning?

- There exists a pattern that can be learned, leading to potential performance improvement.
- No straightforward programmable solution is available.
- Sufficient data is available for the model to learn from.



Decision Trees for Automated Identification of Cosmic-Ray Hits in Hubble

Publications of the Astronomical Society of the Pacific
107: 279–288, 1995 March

Decision Trees for Automated Identification of Cosmic-Ray Hits in *Hubble Space Telescope* Images

STEVEN SALZBERG[1]
Department of Computer Science, Johns Hopkins University, Baltimore, Maryland 21218
Electronic mail: salzberg@cs.jhu.edu

RUPALI CHANDAR
Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218
Electronic mail: rupali@jhufos.pha.jhu.edu

HOLLAND FORD
Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218
Electronic mail: ford@stsci.edu

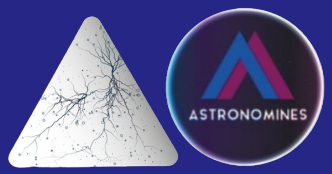
SREERAMA K. MURTHY
Department of Computer Science, Johns Hopkins University, Baltimore, Maryland 21218
Electronic mail: murthy@cs.jhu.edu

RICHARD WHITE
Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218
Electronic mail: rlw@stsci.edu

Received 1994 April 18; accepted 1994 November 30

ABSTRACT. We have developed several algorithms for classifying objects in astronomical images. These algorithms have been used to label stars, galaxies, cosmic rays, plate defects, and other types of objects in sky surveys and other image databases. Our primary goal has been to develop techniques that classify with high accuracy, in order to ensure that celestial objects are not stored in the wrong catalogs. In addition, classification time must be fast due to the large number of classifications and to future needs for on-line classification systems. This paper reports on our results from using decision-tree classifiers to identify cosmic-ray hits in *Hubble Space Telescope* images. This method produces classifiers with over 95% accuracy using data from a single, unpaired image. Our experiments indicate that this accuracy will get even higher if methods for eliminating background noise improve.

Salzberg et al. 1995



k-Nearest Neighbors for automated classification of celestial objects

Science in China Series G: Physics, Mechanics & Astronomy

© 2008 SCIENCE IN CHINA PRESS

Springer-Verlag

www.scichina.com
phys.scichina.com
www.springerlink.com

k-Nearest Neighbors for automated classification of celestial objects

LI LiLi[1,2,3†], ZHANG YanXia[1] & ZHAO YongHeng[1]

[1] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China;
[2] Department of Physics, Hebei Normal University, Shijiazhuang 050016, China;
[3] Weishanlu Middle School, Tianjin 300222, China

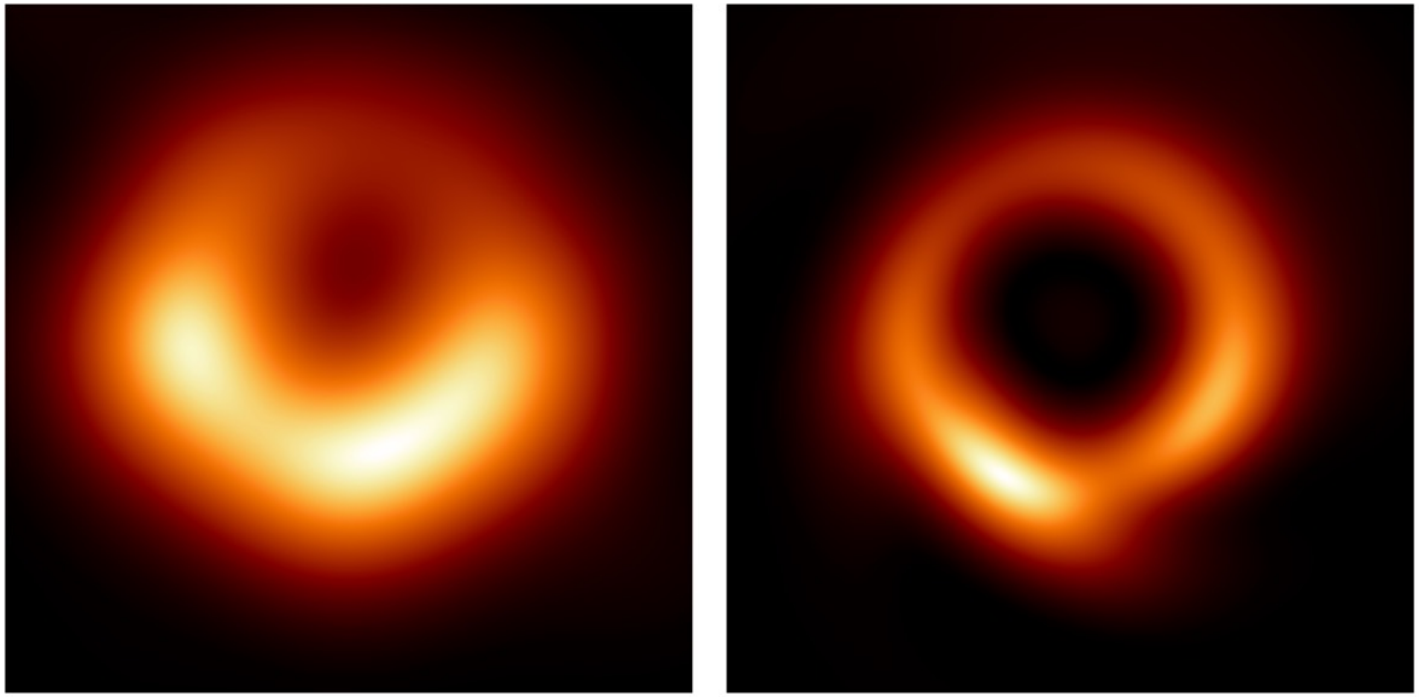
The nearest neighbors (NNs) classifiers, especially the *k*-Nearest Neighbors (*k*NNs) algorithm, are among the simplest and yet most efficient classification rules and widely used in practice. It is a nonparametric method of pattern recognition. In this paper, *k*-Nearest Neighbors, one of the most commonly used machine learning methods, work in automatic classification of multi-wavelength astronomical objects. Through the experiment, we conclude that the running speed of the *k*NN classier is rather fast and the classification accuracy is up to 97.73%. As a result, it is efficient and applicable to discriminate active objects from stars and normal galaxies with this method. The classifiers trained by the *k*NN method can be used to solve the automated classification problem faced by astronomy and the virtual observatory (VO).

k-Nearest Neighbors | data analysis | classification | astronomical catalogues

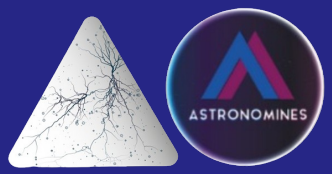
Li, Zhang, and Zhao 2023

First Over Black Hole Image Reconstruction

Supermassive black hole M87



"First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole" 2019

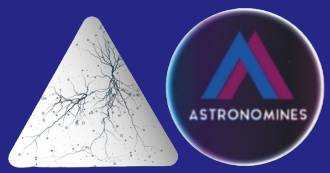


References I

"First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole" (Apr. 2019). In: *The Astrophysical Journal Letters*. DOI: 10.3847/2041-8213/ab0c57.

Li, LiLi, YanXia Zhang, and YongHeng Zhao (2023). "k-Nearest Neighbors for Automated Classification of Celestial Objects". In: *National Astronomical Observatories, Chinese Academy of Sciences*. URL: https://doi.org/10.1234/jastro.2023.xxx.x.

Salzberg, Steven et al. (Mar. 1995). "Decision Trees for Automated Identification of Cosmic-Ray Hits in Hubble Space Images". In: *Publications of the Astronomical Society of the Pacific*. DOI: 10.1086/133592. eprint: https://iopscience.iop.org/article/10.1086/133592.



UNLOCKING AI
Essentiel Concepts for ML Mystery

MAJDOUB Adnane

EAIC Co-Training Head

Second-year AI Engineering Student

ENSIAS AI Club x Astronomies

December 13, 2023