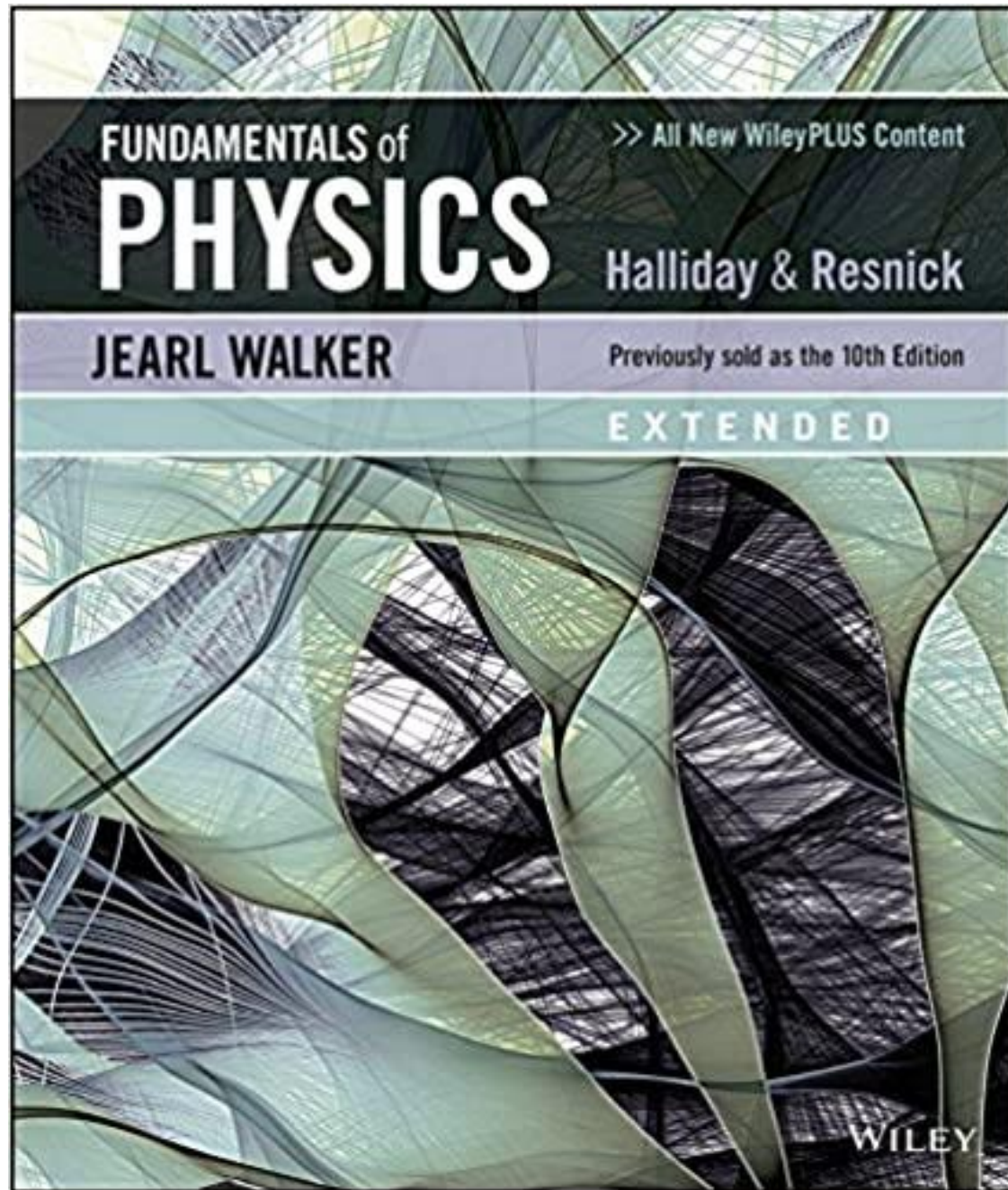


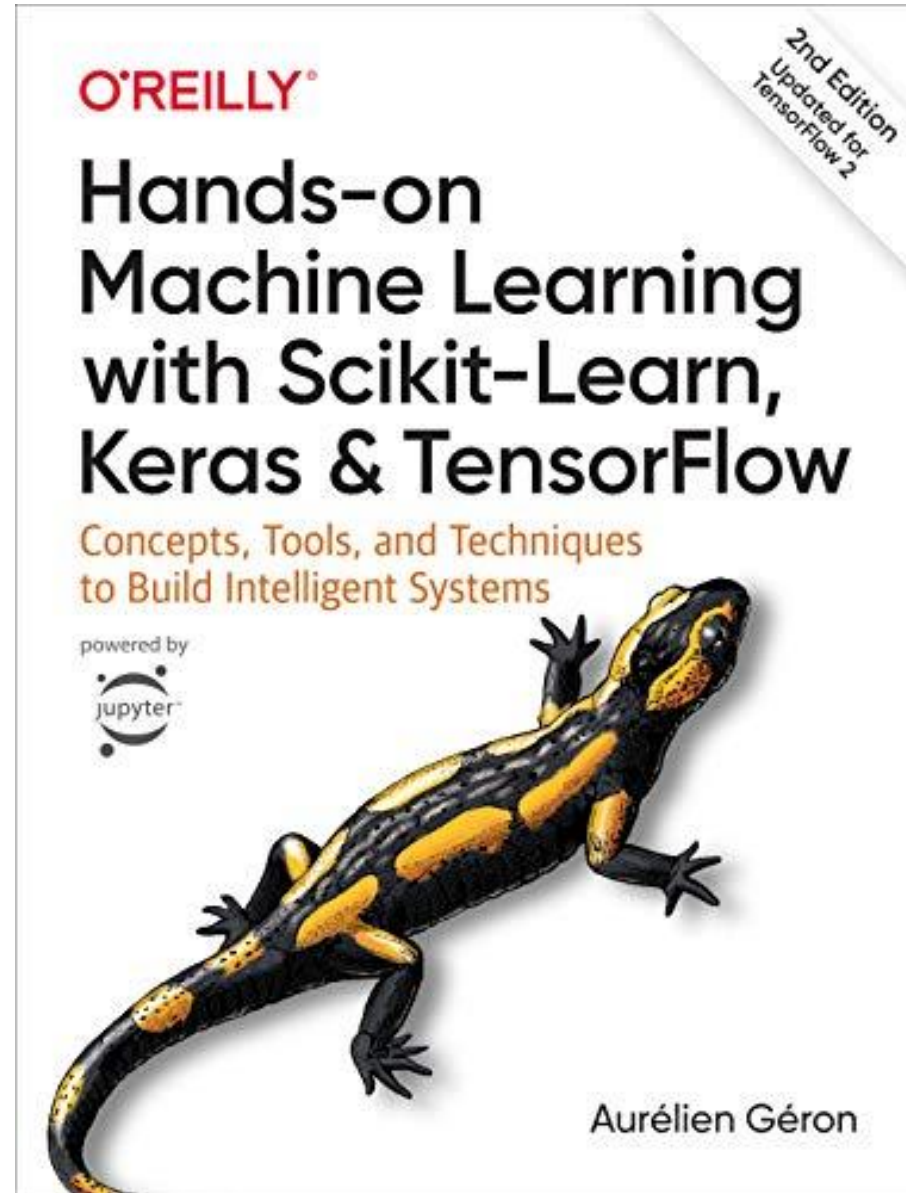
Supervised Learning

Unsupervised Learning

Supervised learning



Reading the book with the goal: **Learning ML**



Supervised learning

Paying attention with the goal: **Learning to paint**

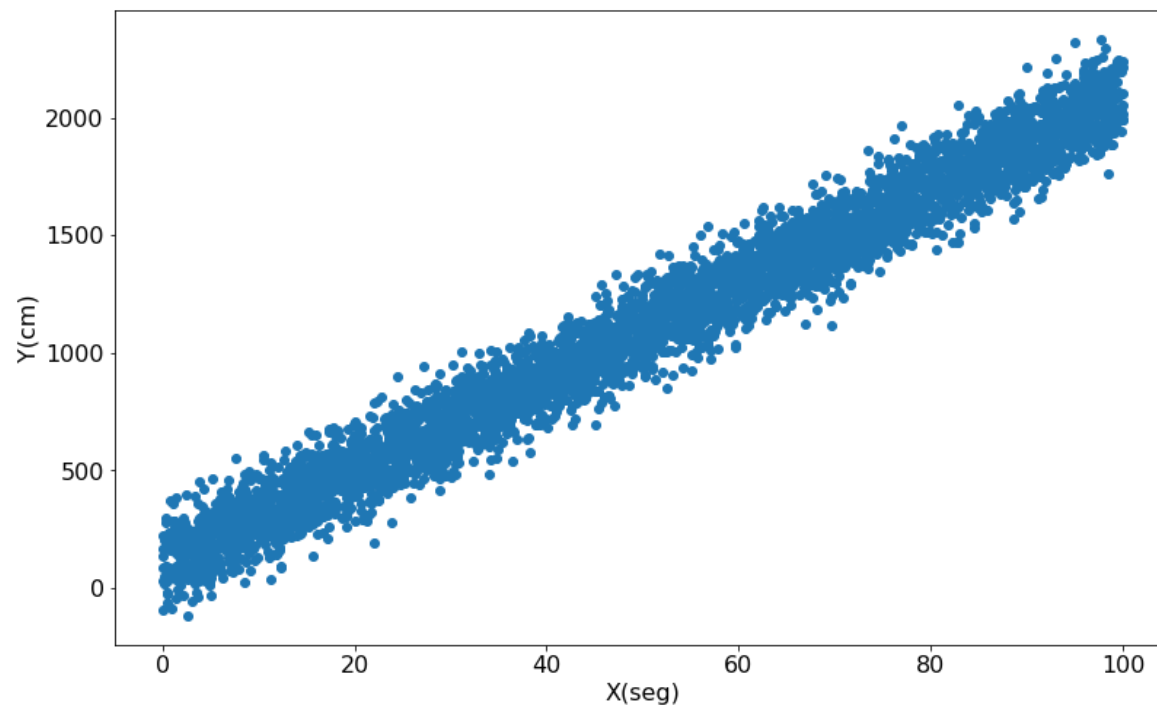


Paying attention with the goal: **Producing gold**

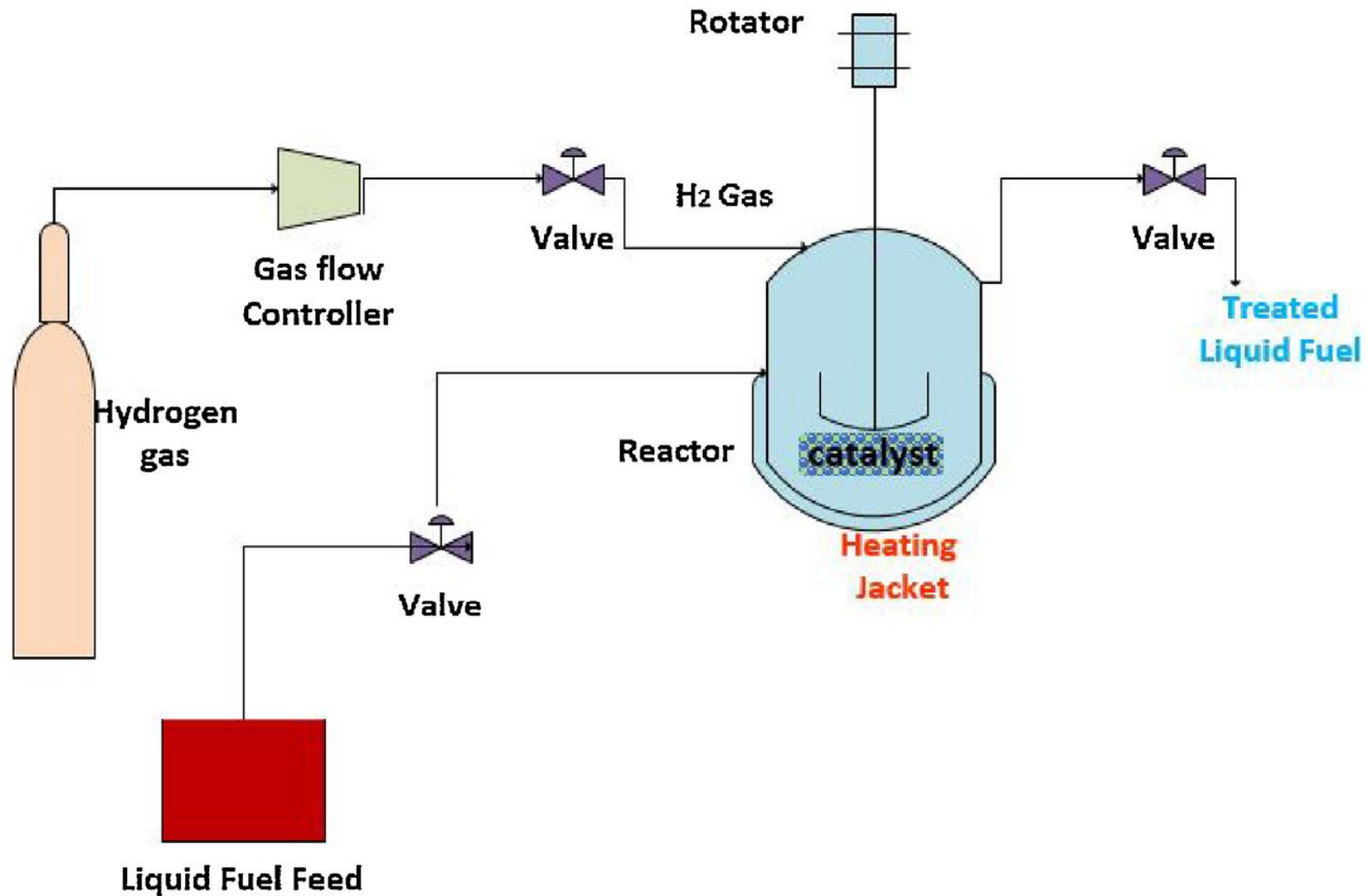
Supervised learning

Supervised Learning

Phenomena under study: Objective \rightarrow distance

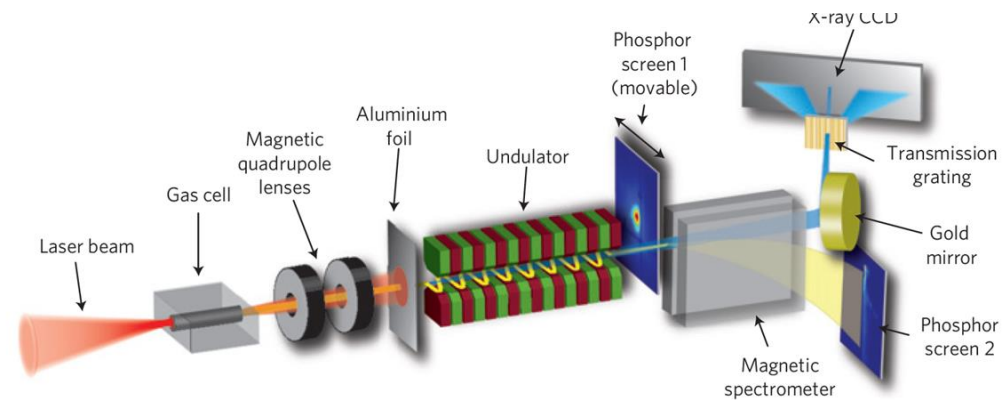
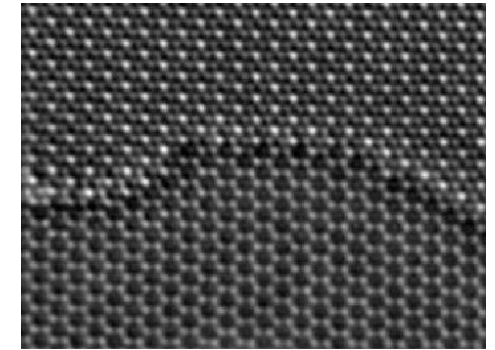


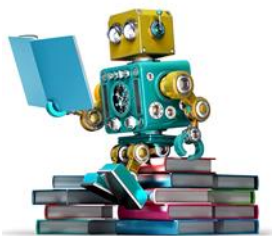
Hidrodesulfuration



Temperature, Pressure, Hydrogen Flow rate, Inlet sulfur concentration,
equipment operating days ...

Output sulfur concentration

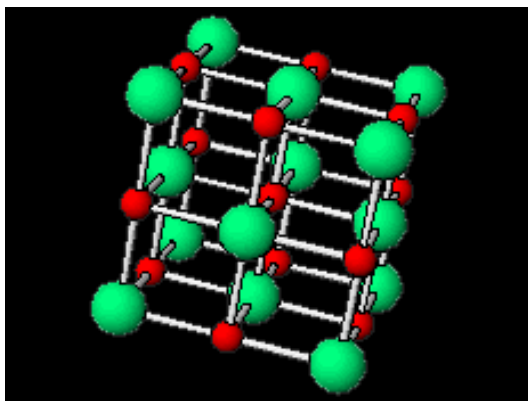
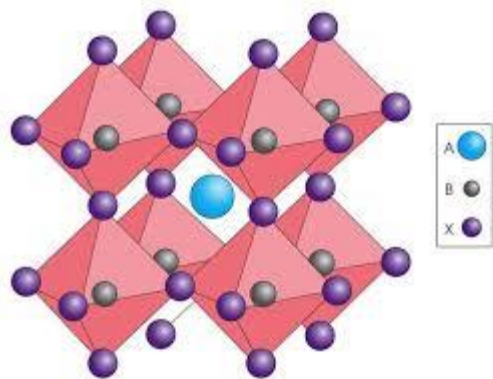
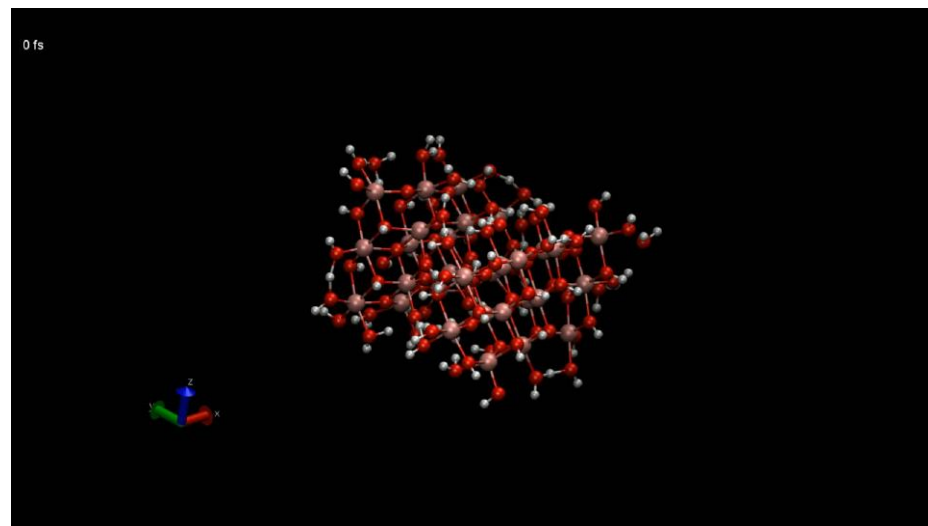
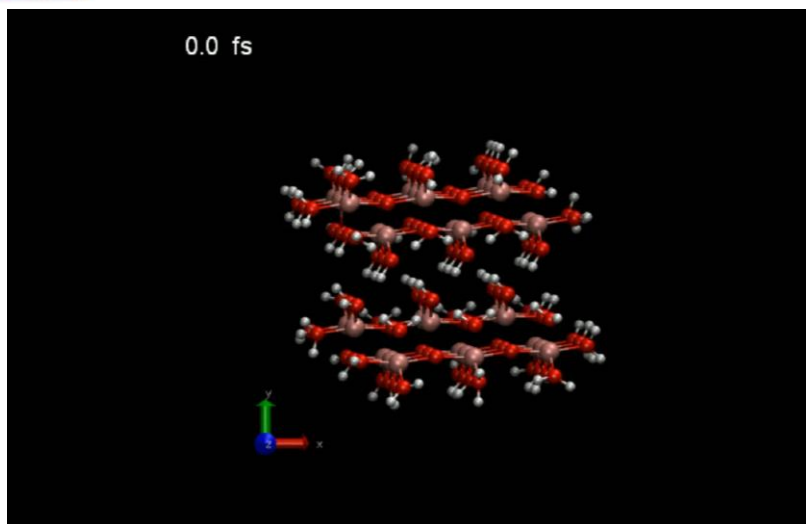




Atomic distributions



Energy, Force



New compounds

High Energy Physics

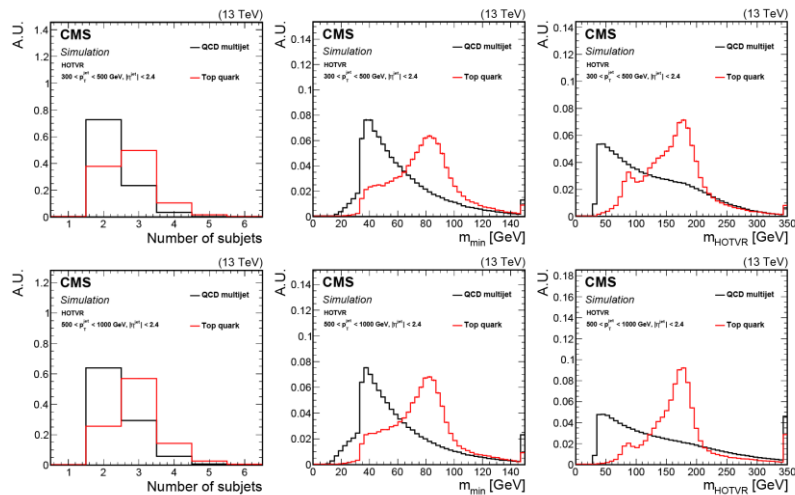
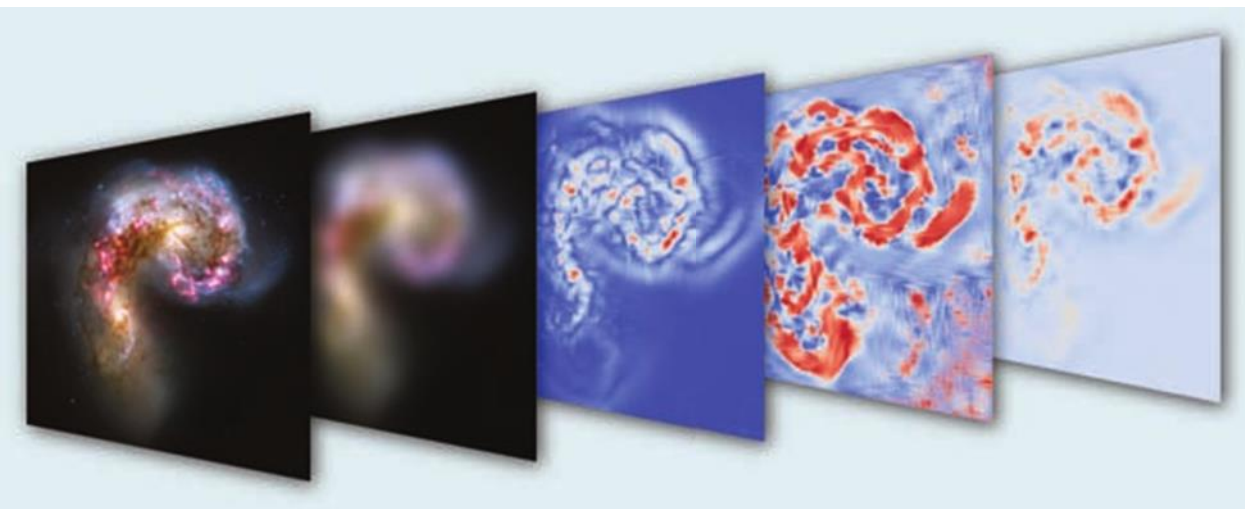


Figure 4. Shape comparison of the main variables of the HOTVR algorithm for signal and background jets, in two different regions of the jet p_T as displayed in the plots.

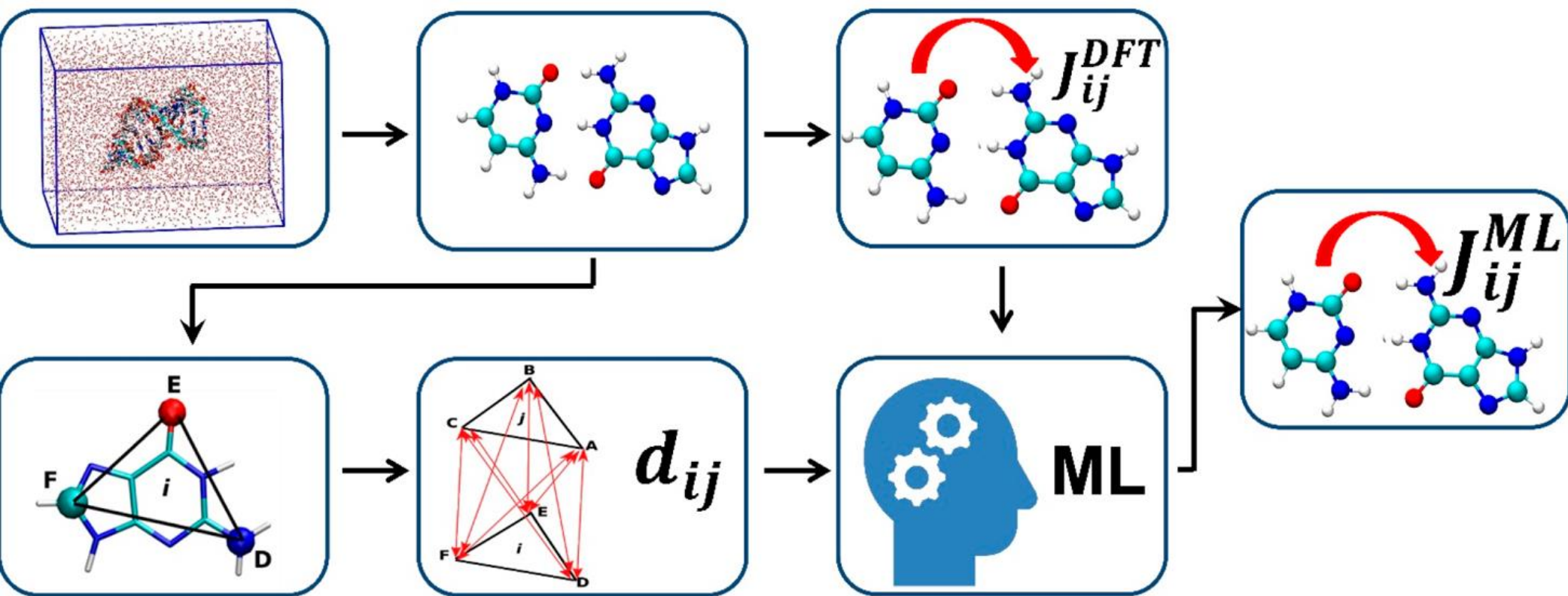
Detecting of heavy, energetic, Hadronically decaying particles

Astronomy, Galaxies

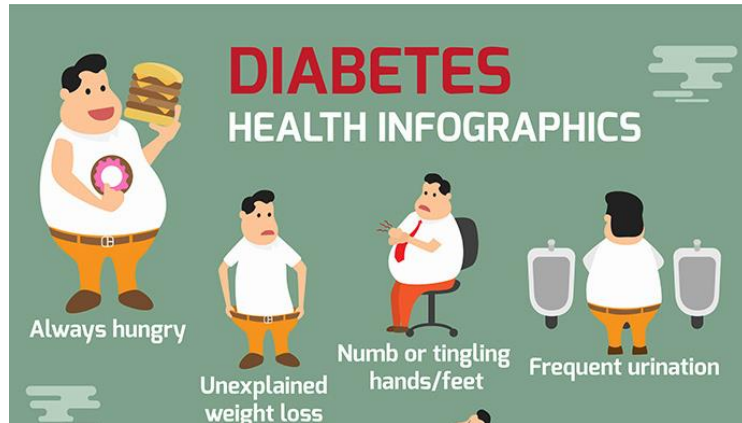


Galaxy inner structure

Machine Learning Prediction of **Electronic Coupling between the Guanine Bases of DNA**



Diabetes mortality

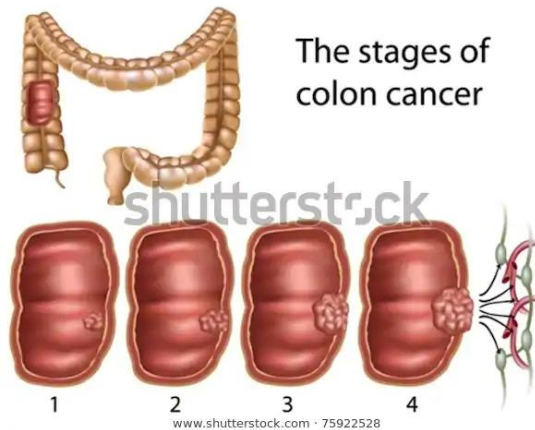
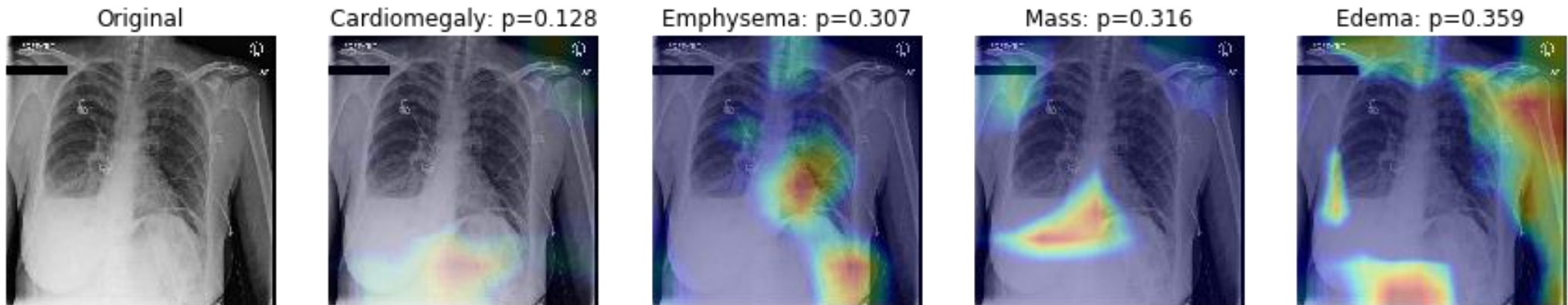


Medical prognosis:

Mortality after one year

Medical Diagnosis:

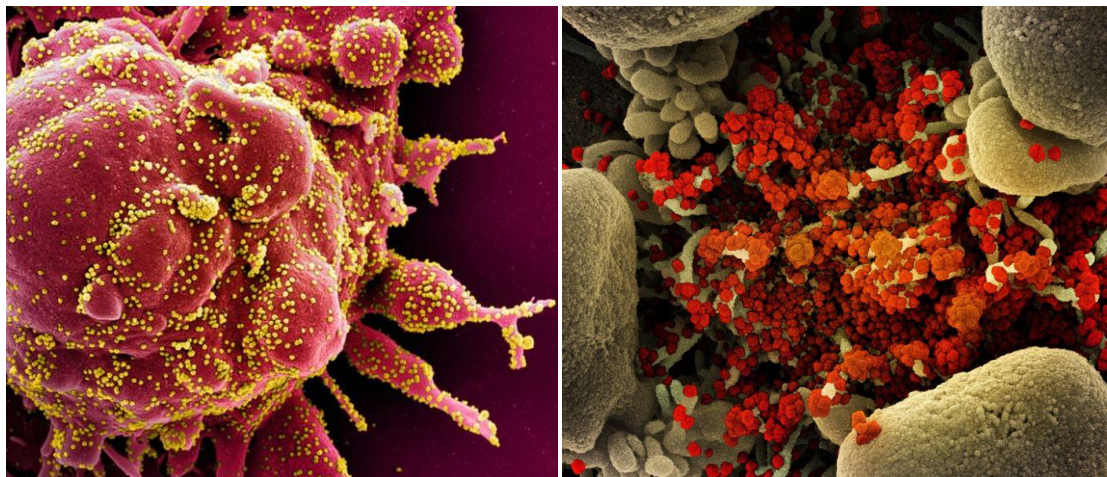
Pulmonary Diseases



Medical treatment:

Effect of drugs combination

Covid-19



Drug discovery and
Vaccine development

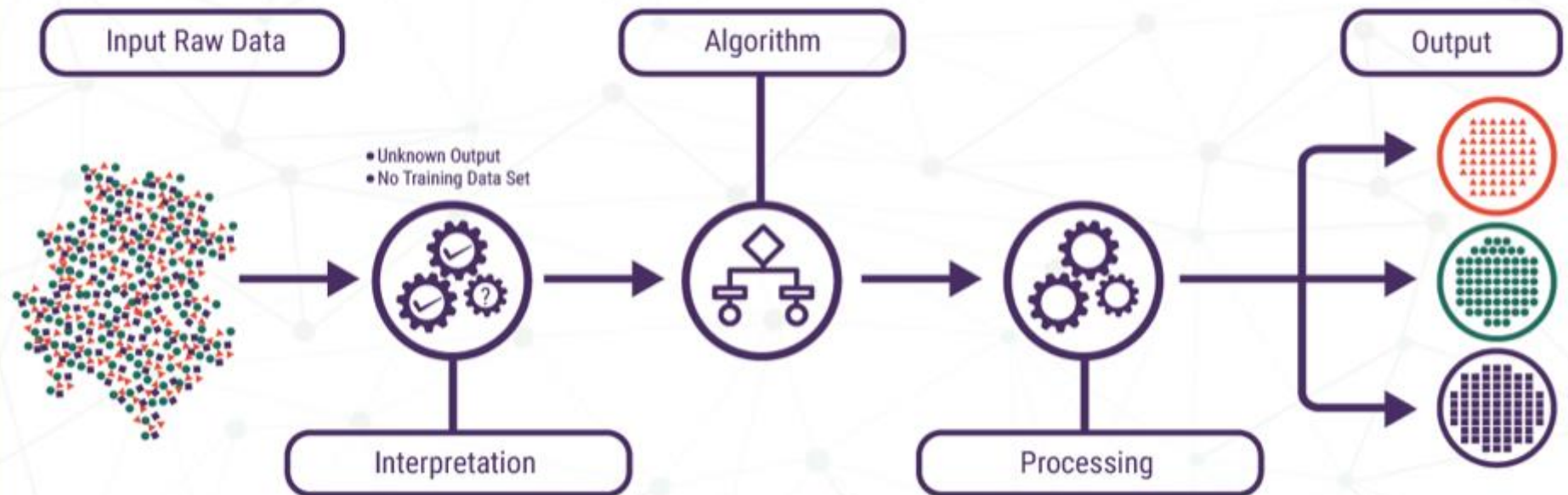
Automated extraction of **chemical synthesis actions**
from experimental procedures

[Nature communications, 11, 3601 \(2020\)](#)

Thematic analysis of 18 years of physics education
research conference proceedings using natural
language processing

[Phys. Rev. Phys. Education Research 16, 010142 \(2020\)](#)

UNSUPERVISED LEARNING



Alcaldías CDMX



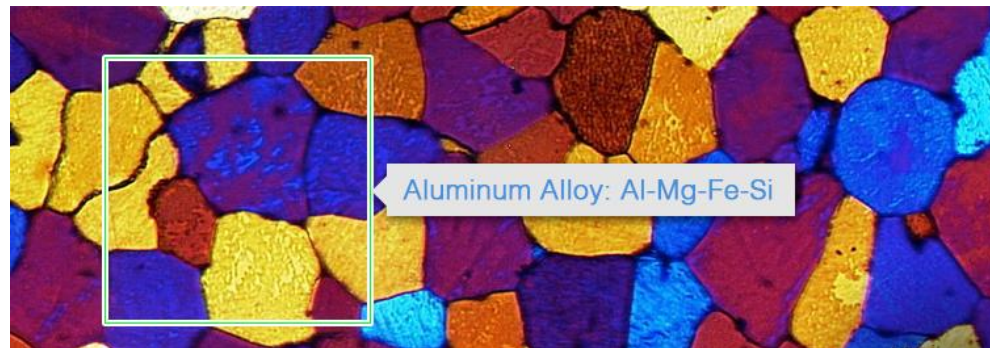
Fuente: Cuentas Rápidas Instituto Electoral de la CDMX



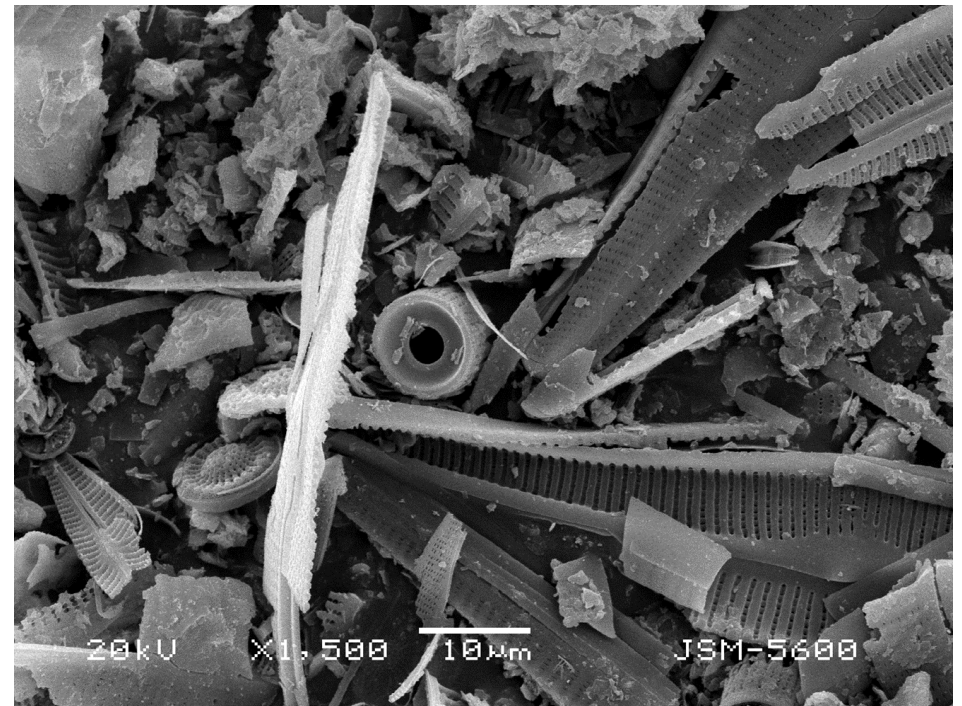
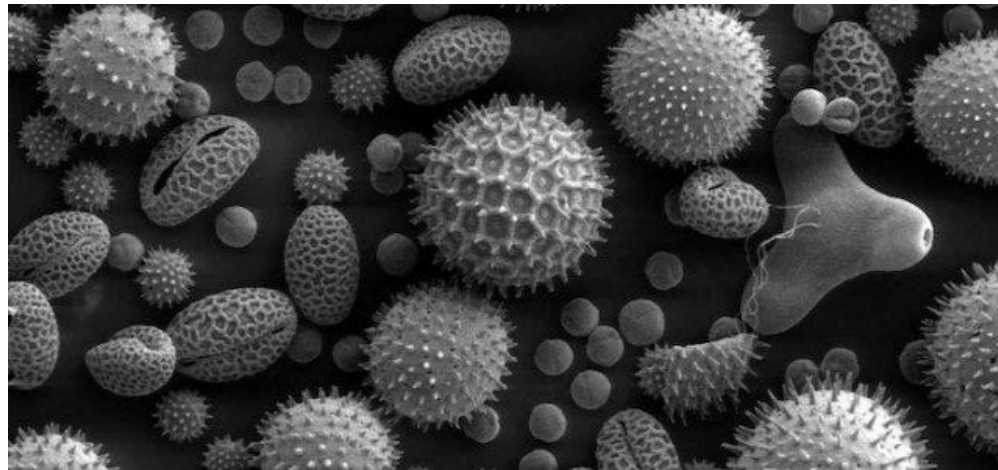
1	2	5	9	7	6	3	5	0	8
4	5	8	6	9	3	2	9	7	2
3	3	3	9	5	0	3	2	3	0
1	1	4	0	2	1	5	3	3	6
8	6	2	0	4	0	4	5	3	9
8	5	4	2	2	7	1	6	0	9
1	7	0	3	9	1	2	0	7	7
2	6	5	1	6	4	2	2	2	9
4	4	4	2	0	6	9	4	8	3
1	5	0	3	4	6	8	2	5	1

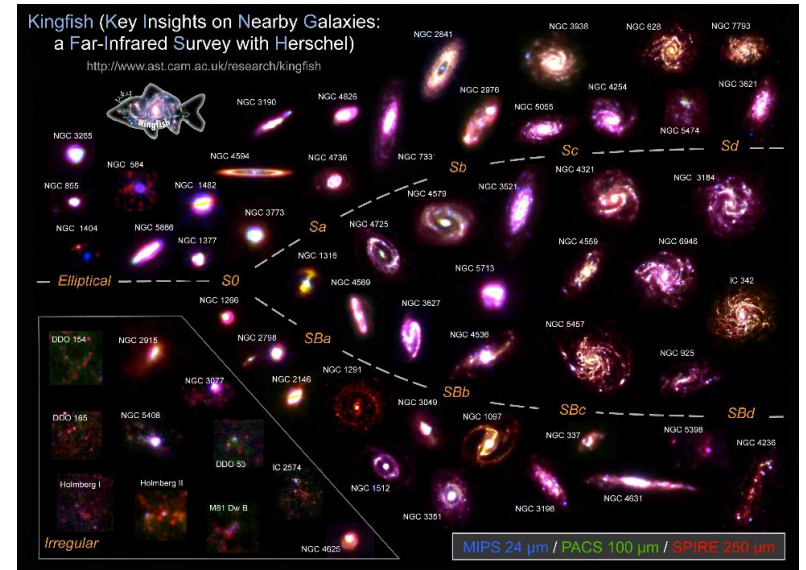
Digits clustering

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

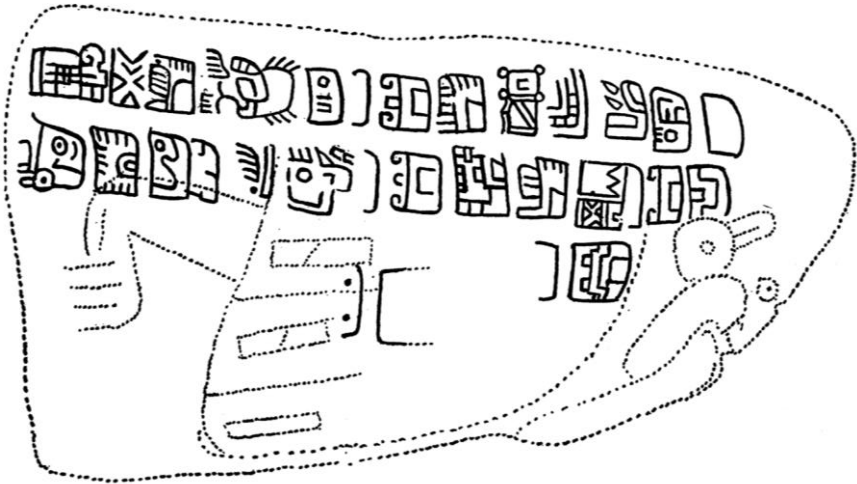


Microstructure clustering





Mexican Writings Epigraphy



K-means

SOME METHODS FOR CLASSIFICATION AND ANALYSIS OF MULTIVARIATE OBSERVATIONS

J.M_{AC}QUEEN (1987)

The main purpose of this paper is to describe a process for partitioning an N dimensional population into k sets on the basis of a sample. The process, which is called '**k-means**,' appears to give partitions which are reasonably efficient in the sense of within-class variance.

The way kmeans algorithm works is as follows:

Specify number of clusters K .

Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.

Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

Compute the sum of the squared distance between data points and all centroids.

Assign each data point to the closest cluster (centroid).

Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

The used metric to detect the clusters is:

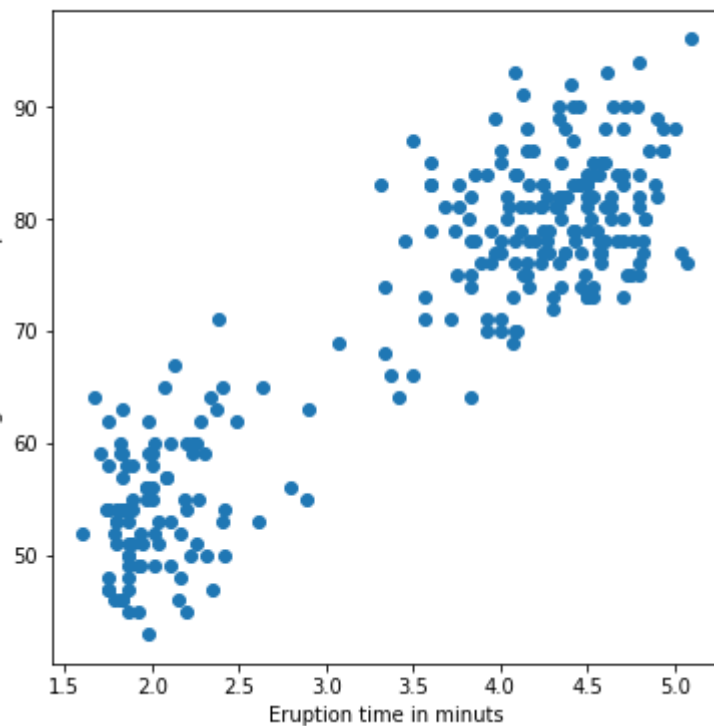
$$J = \sum_{i=1}^m \sum_{k=1}^K w_{ik} \|x^i - \mu_k\|^2$$

where $w_{ik} = 1$ for data point x_i if it belongs to cluster k ; otherwise, $w_{ik}=0$. Also, μ_k is the centroid of x_i 's cluster.

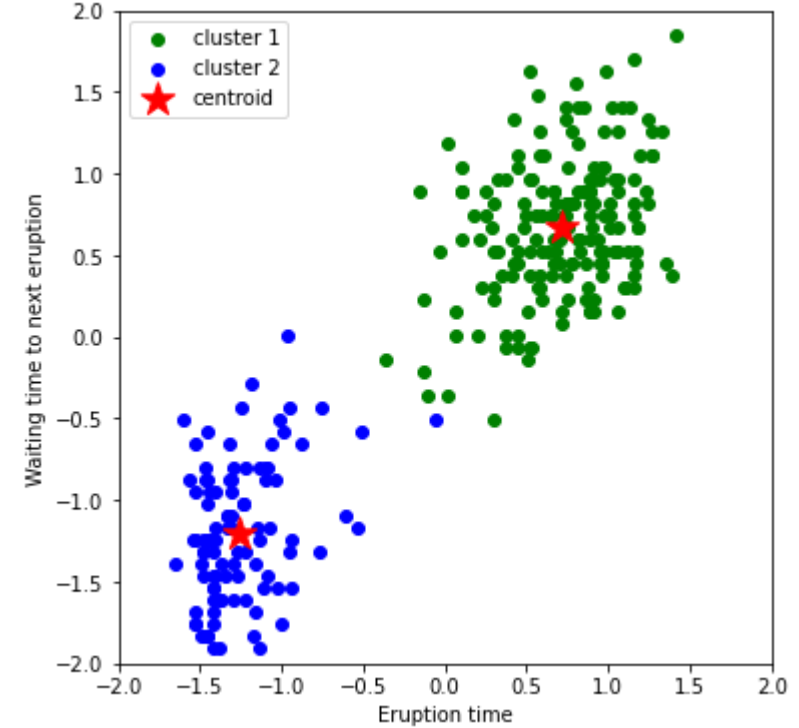
Geyser's Eruptions



Visualization of raw data



Visualization of clustered data



	eruptions	waiting
0	3.600	79
1	1.800	54
2	3.333	74
3	2.283	62
4	4.533	85

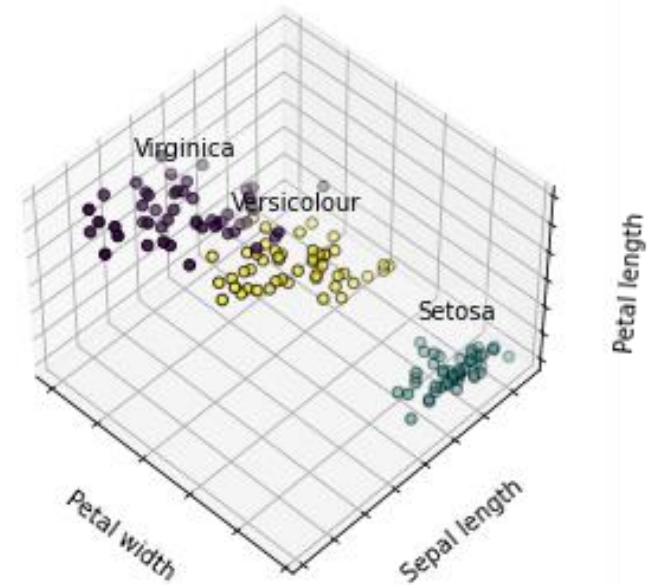
Iris Clustering



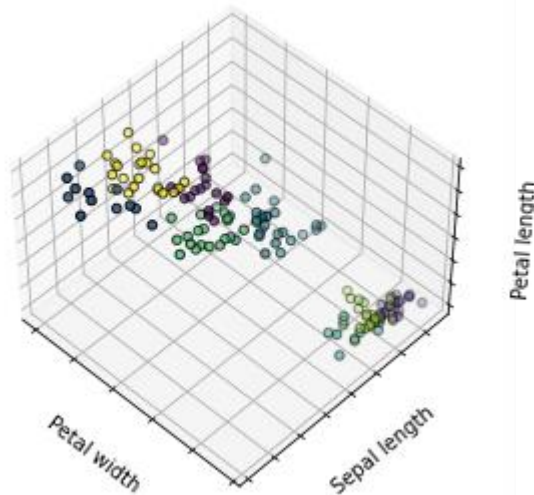
Sepal length width petal length width

```
[ [5.1 3.5 1.4 0.2]
  [4.9 3.  1.4 0.2]
  [4.7 3.2 1.3 0.2]
  [4.6 3.1 1.5 0.2]
  [5.  3.6 1.4 0.2]]
```

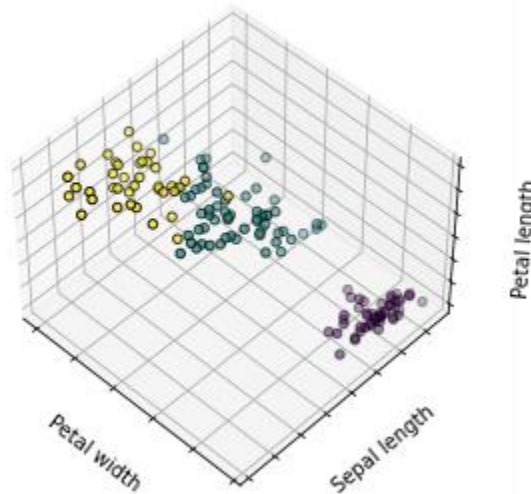
Ground Truth



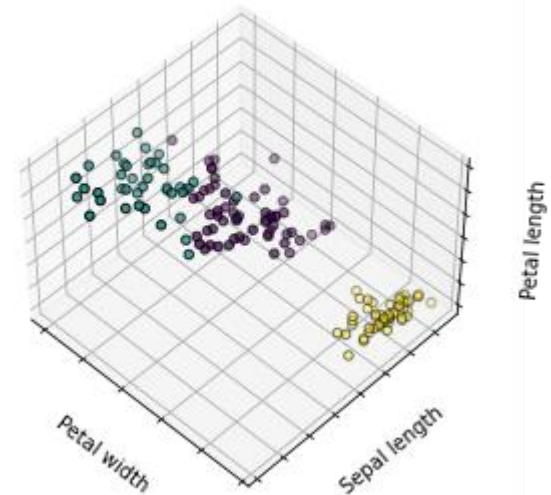
8 clusters



3 clusters

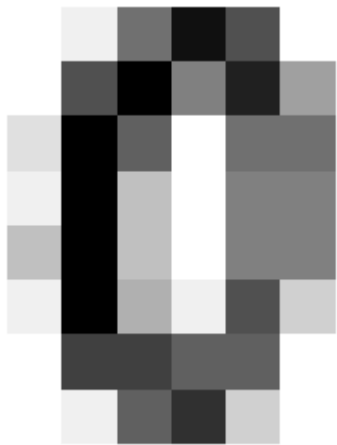


3 clusters, bad initialization



Digits clustering

1	2	5	9	7	6	3	5	0	8
4	5	8	6	9	3	2	9	7	2
3	3	3	9	5	0	3	2	3	0
1	1	4	0	2	1	5	3	3	6
8	6	2	0	4	0	4	5	3	9
9	5	4	2	2	7	1	6	0	9
1	7	0	3	9	1	2	0	7	7
2	6	5	1	6	4	2	2	2	9
4	4	4	2	0	6	9	4	8	3
1	5	0	3	4	6	8	2	5	1

[illegible][illegible][illegible][illegible]