



APRENDIZAJE AUTOMÁTICO CON R

Modelos de Segmentación
Eugenio Grant

Agenda

01

*AI &
Machine
Learning*

02

*Modelos
de
Clustering*

03

Código R

Artificial Intelligence

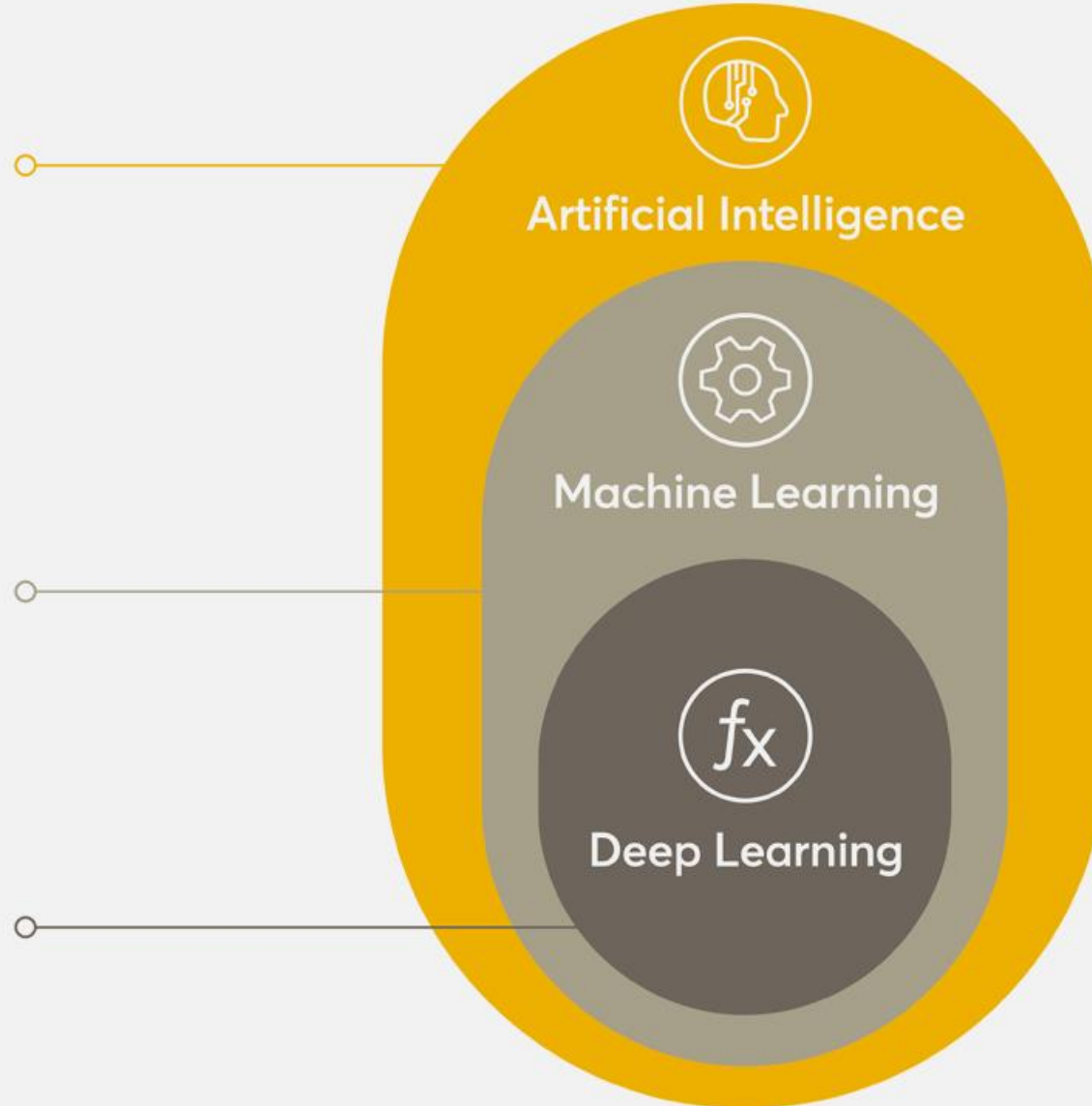
Any technique which enables computers to mimic human behaviour.

Machine Learning

Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

Deep Learning

Subset of ML which makes the computation of multi-layer neural networks feasible.



¿Qué es Aprendizaje Automático?

“Campo de estudio que le da a las computadoras la habilidad de ‘aprender’ sin ser explícitamente programadas para ello” – Arthur Samuel



¿Qué es Inteligencia Artificial?

La Inteligencia Artificial es la simulación de procesos de inteligencia humana por parte de máquinas, especialmente sistemas informáticos.



Hi, how can I help?



Uber



amazon



Artificial Intelligence



Engineering of making Intelligent Machines and Programs

Machine Learning



Ability to learn without being explicitly programmed

Deep Learning



Learning based on Deep Neural Network

1950's

1960's

1970's

1980's

1990's

2000's

2006's

2010's

2012's

2017's





Herramientas... de muchas



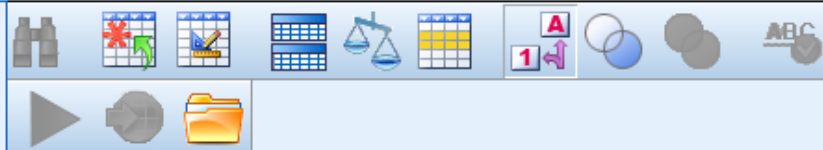
IBM

Create and train a machine learning model without coding

Explore the key capabilities of IBM SPSS Modeler and the New User Interface



- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify**
 - TwoStep Cluster...
 - K-Means Cluster...**
 - Hierarchical Cluster...
 - Tree...
 - Discriminant...
 - Nearest Neighbor...
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Quality Control
- ROC Curve...



Visible: 0 of 0 Variables

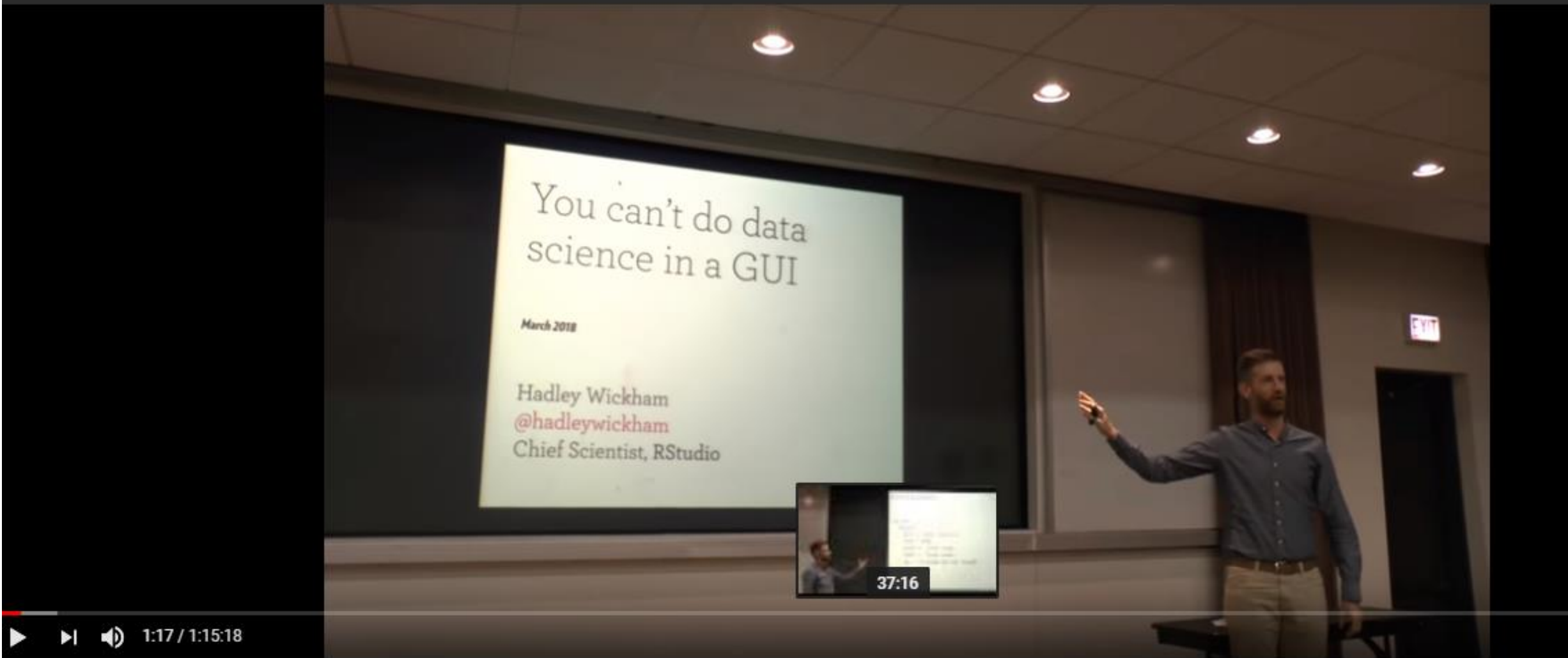
	var	var		var	var	var	var	var	var	var	var	var	var
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													

quick cluster VARS /missing=listwise /criteria=cluster(K) mxiter (100)
converge(0) /method=kmeans(noupdate) /save cluster (CLUSTER) /print initial.



YouTube

Search



Hadley Wickham – “You can't do data science in a GUI”

Up next

AU

Diferencias entre R y Python

The screenshot shows the RStudio IDE. The top-left pane displays a script named 'Gotham-City-2018-Plan.Rmd' with R code for plotting assault data against urban population. The top-right pane shows the 'Environment' tab with a list of objects including '.gitignore', 'top100.txt.swp', 'Gotham-City-2018-Plan.Rmd', 'Gotham-City-2018-Plan.nb.html', 'crime2018.Rproj', and 'top100.txt'. The bottom-left pane shows the 'Console' with a list of installed packages. The bottom-right pane shows the 'Terminal' with a list of installed packages.

```
1 profvis::profvis({
2   plot(Assault ~ UrbanPop, data = USArrests)
3   lm(Assault ~ UrbanPop, data = USArrests)
4 })
5
```

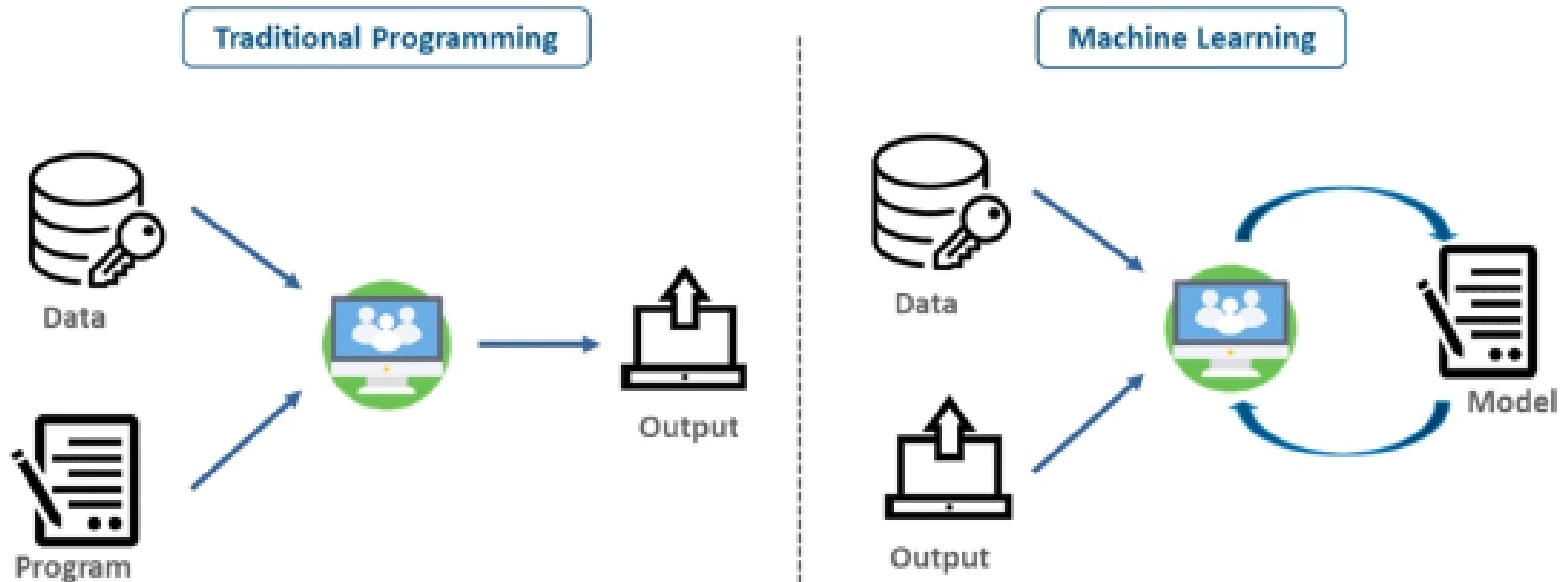
Name	Description	Version
assertthat	Easy Pre and Post Assertions	0.2.0
aws.s3	AWS S3 Client Package	0.3.3
aws.signature	Amazon Web Services Request Signatures	0.3.5
backports	Reimplementations of Functions Introduced Since R-3.0.0	1.1.0
base64enc	Tools for base64 encoding	0.1-3
BH	Boost C++ Header Files	1.62.0-1
binder	Parametrized Active Bindings	0.1
binderapp	An 'Rcpp' Interface to Active Bindings	0.2
bitops	Bitwise Operations	1.0-6
blogdown	Create Blogs and Websites with R Markdown	0.0.70
bookdown	Authoring Books and Technical Documents with R Markdown	0.4.5
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-19
brew	Templating Framework for Report Generation	1.0-6
broom	Convert Statistical Analysis Objects into Tidy Data Frames	0.4.2
caTools	Tools: moving window statistics, GIF, Base64, ROC AUC, etc.	1.17.1

The screenshot shows the JupyterLab IDE. The left pane displays a script with Python code using NumPy and SciPy for data analysis and plotting. The right pane shows the 'Environment' tab with a list of installed packages. The bottom-right pane shows two plots: a 3D surface plot and a 2D polar plot.

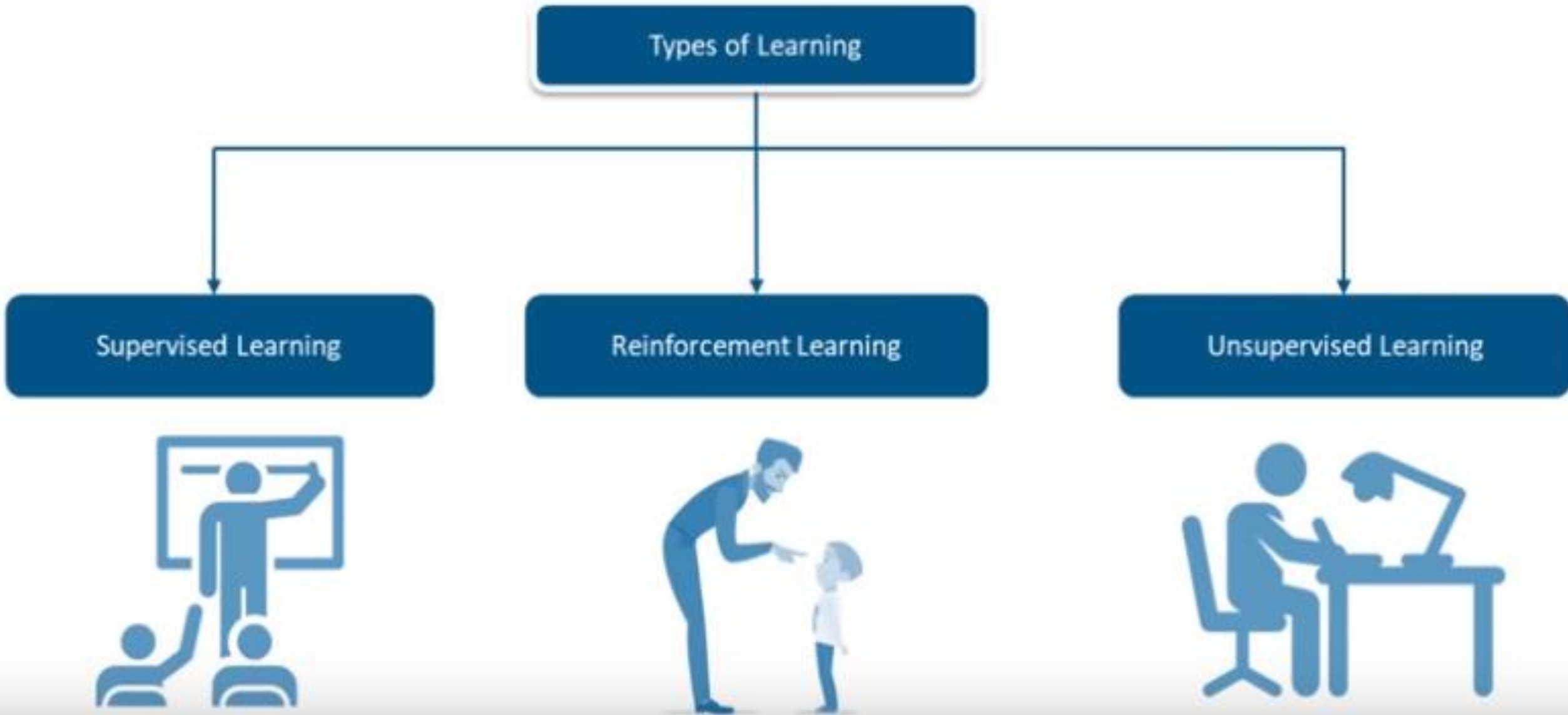
```
8 import numpy as np
9 from numpy import cos, linspace, pi, sin, random
10 from scipy.interpolate import splrep, splev
11 # Generate data for analysis
12
13 # Make ascending spiral in x space
14 t = linspace(0, 1.75 * 2 * pi, 200)
15
16 x = sin(t)
17 y = cos(t)
18 z = t
19
20 # Add noise
21 x += random.normal(scale=0.1, size=x.shape)
22 y += random.normal(scale=0.1, size=y.shape)
23 z += random.normal(scale=0.1, size=z.shape)
24
25 # Fit a 2D surface to the data
26
27 # Spline parameters
28 smoothness = 3.0 # Smoothness parameter
29 kparam = 2 # Spline order
30 nests = -1 # Estimate of number of knots needed (-1 = automatic)
31
32 # Find the knot points
33 knot_points = splprep([x, y, z], smoothness, kparam, nests=-1)
34
35 # Evaluate points, including interpolated points
36 xnew, ynew, znew = splev(linspace(0, 1, 400), knot_points)
37
38 # Plot results
39
40 # Plot results
41
42 # Plot results
43
44 # Plot results
45
46 # Plot results
47
48 # Plot results
49
50 # Plot results
51
```

Name	Type	Size	Value
Series	Series	20	Series object of type float64...
df	DataFrame	(1, 2)	Column name: knot1, knot2
fillname	str	1	Column name: knot1, knot2
list_test	list	2	[Series, Series]
znew	list	1	Series
r	float64	1	7.81180501010101
radius	float64	1	9.40010101010101
region	float64	1	9.40010101010101
znew	Series	1	Series object of type float64...
test_name	Series	1	Series object of type float64...

Paradigma Aprendizaje Automático



Tipos de Aprendizaje

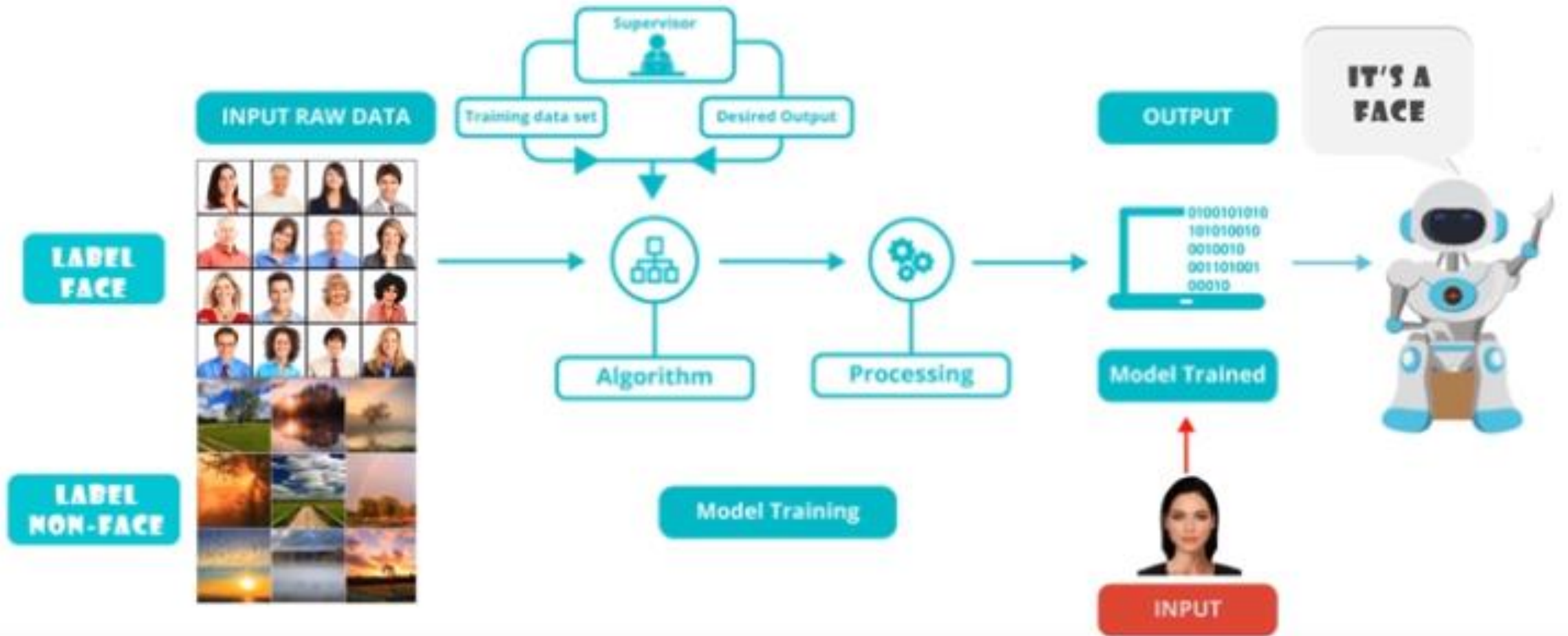


Aprendizaje Supervisado

En el aprendizaje supervisado se tiene una variable de entrada (x) y una variable de salida (Y) y se utiliza un algoritmo para aprender la función que asocia la variable de entrada con la de salida.

Se le llama Supervisado porque el proceso de un algoritmo aprendiendo de la data de entrenamiento se asocia con el de un profesor supervisando el proceso del aprendizaje.

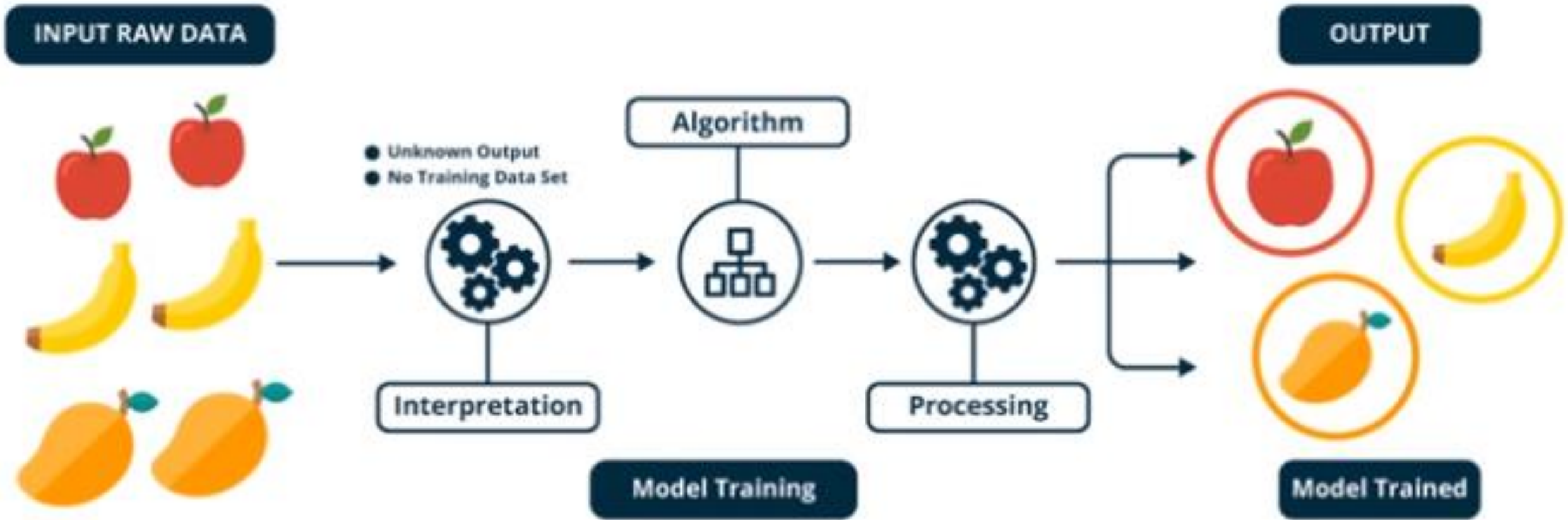
Aprendizaje Supervisado



Algoritmos Aprendizaje Supervisado

- Regresión Lineal
- Regresión Logística
- Árboles de Decisión
- Random Forest
- Clasificador Naive Bayes

Aprendizaje NO Supervisado



Aprendizaje NO Supervisado

En el aprendizaje NO supervisado se tienen datos no etiquetados y el algoritmo busca extraer inferencias por medio de análisis exploratorio.

Se le llama NO Supervisado porque es el equivalente de aprender sin un profesor que guíe el aprendizaje.

Algoritmos Aprendizaje NO Supervisado

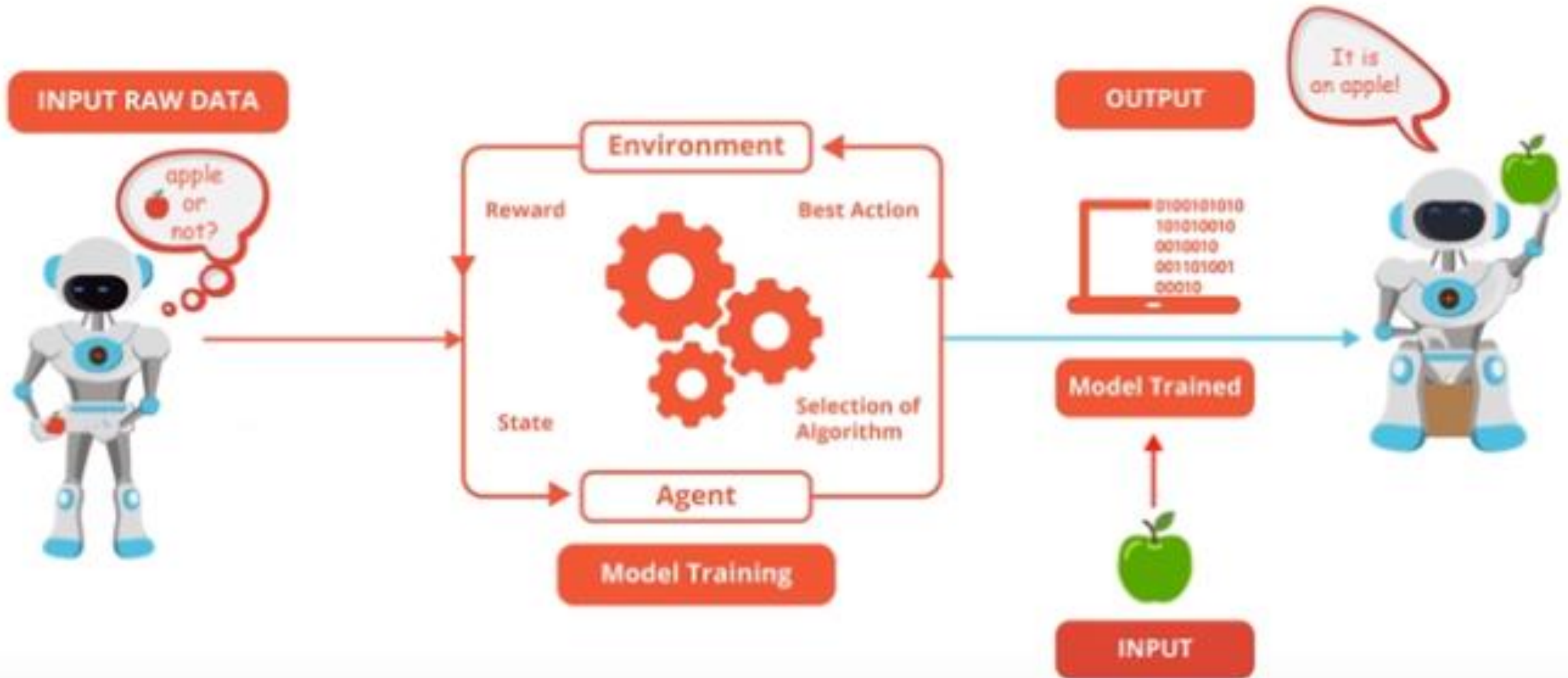


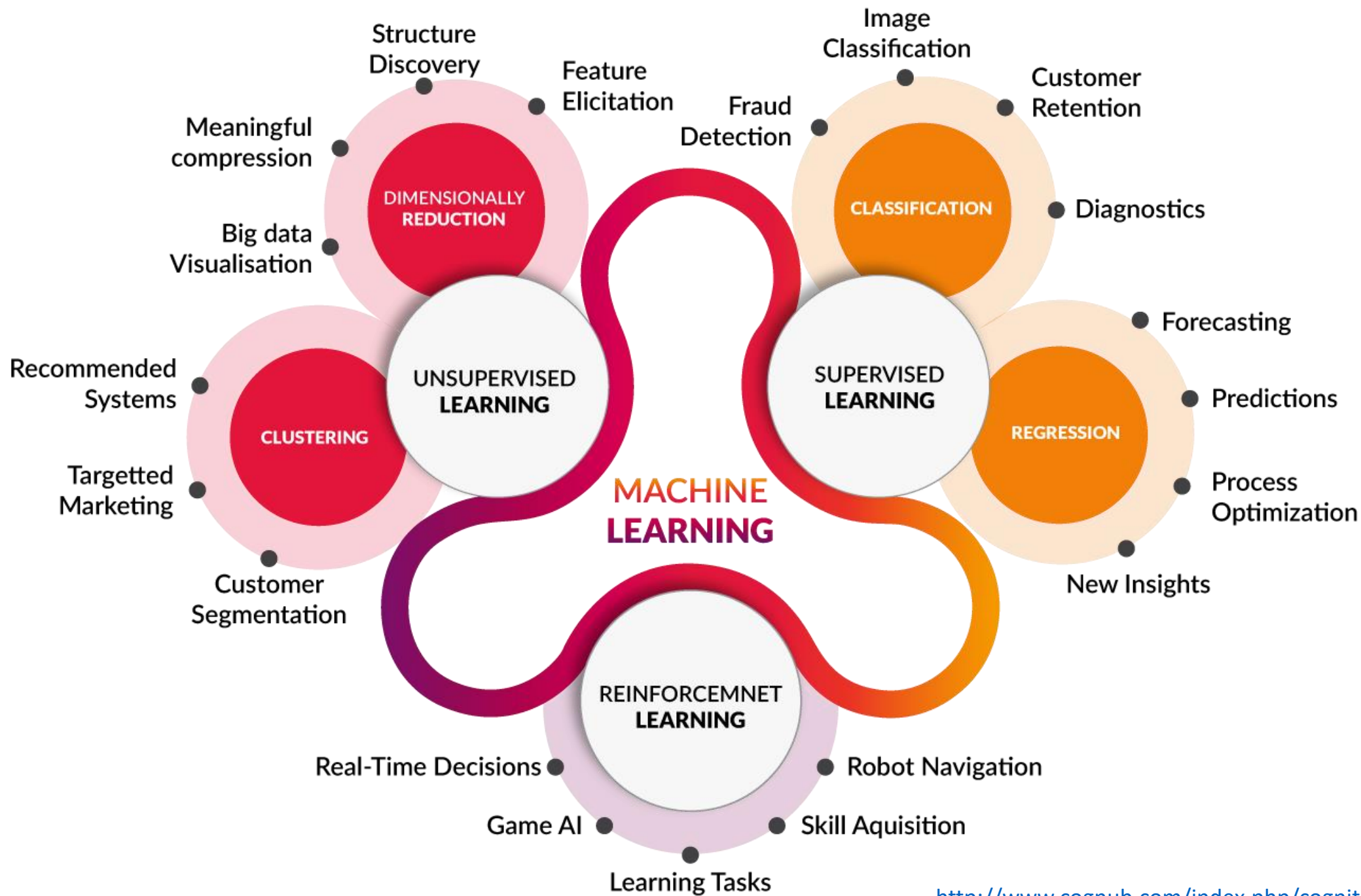
Clasificación



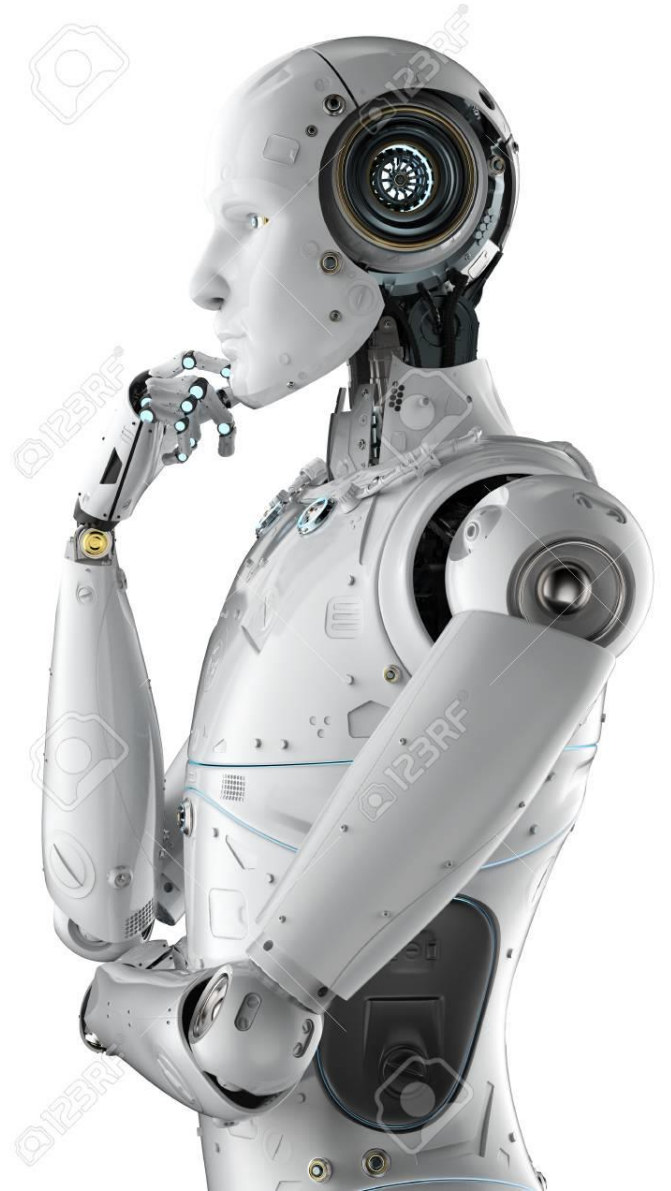
Reducción de Dimensiones

Reinforcement Learning

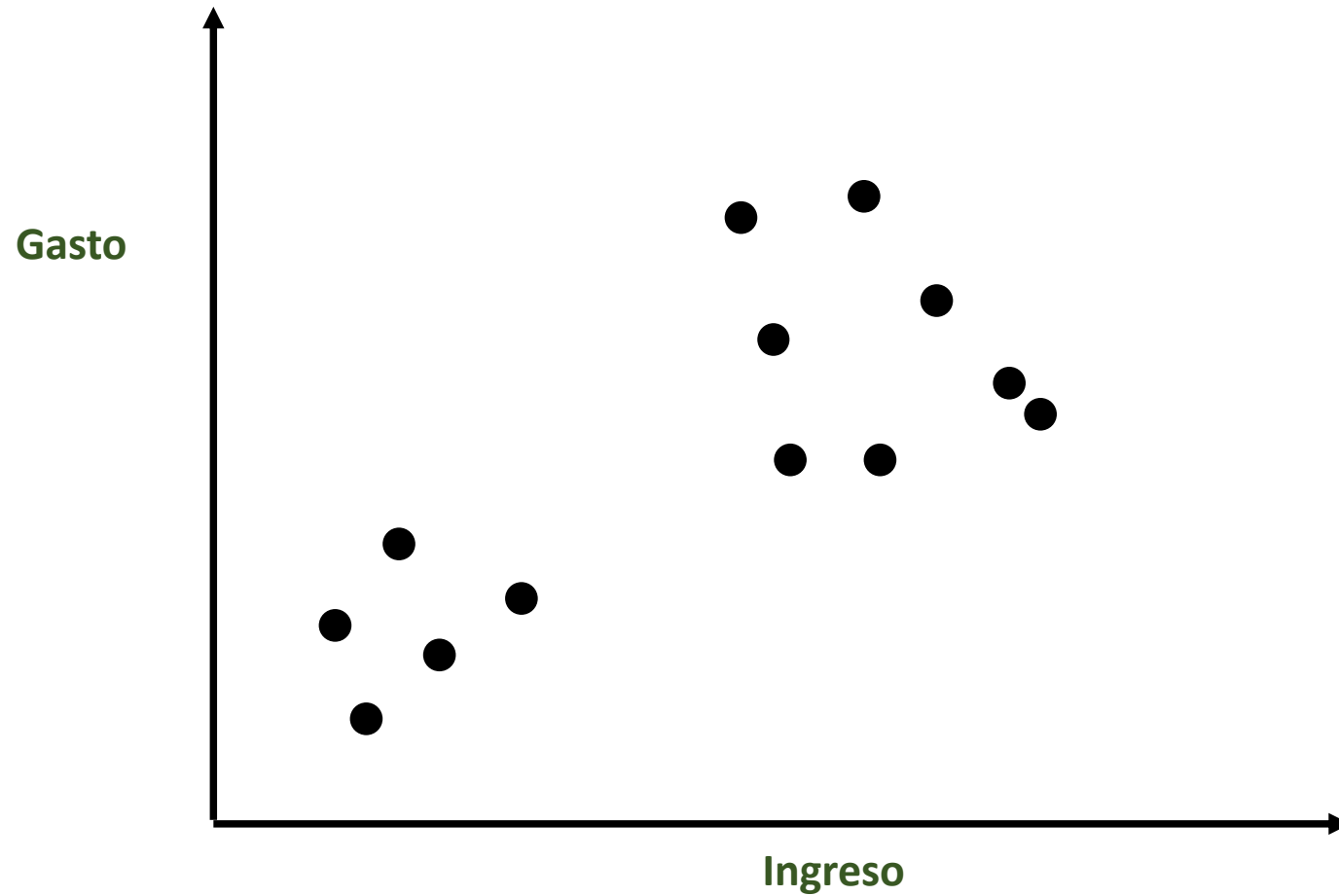




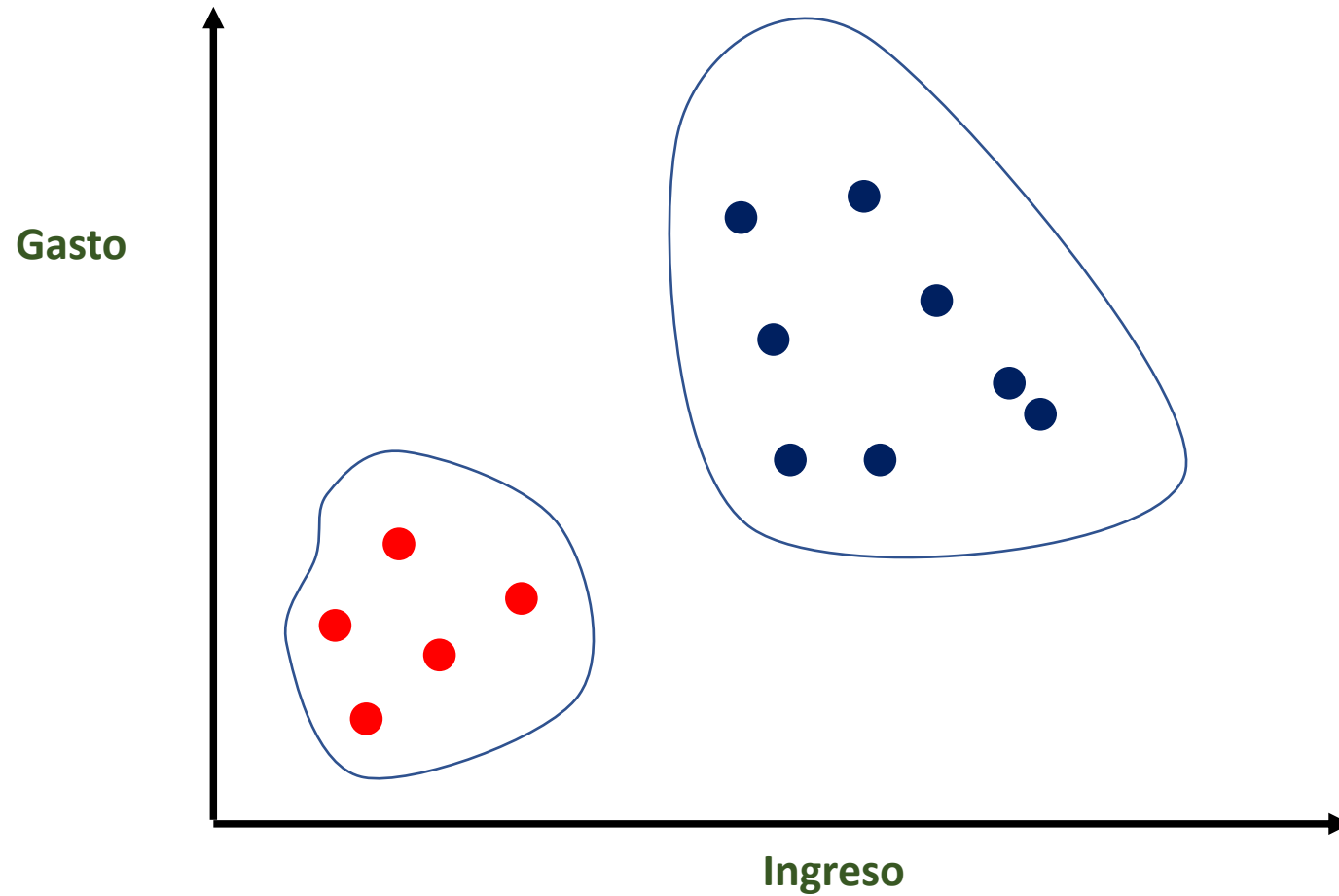
¿Puede pensar una máquina?



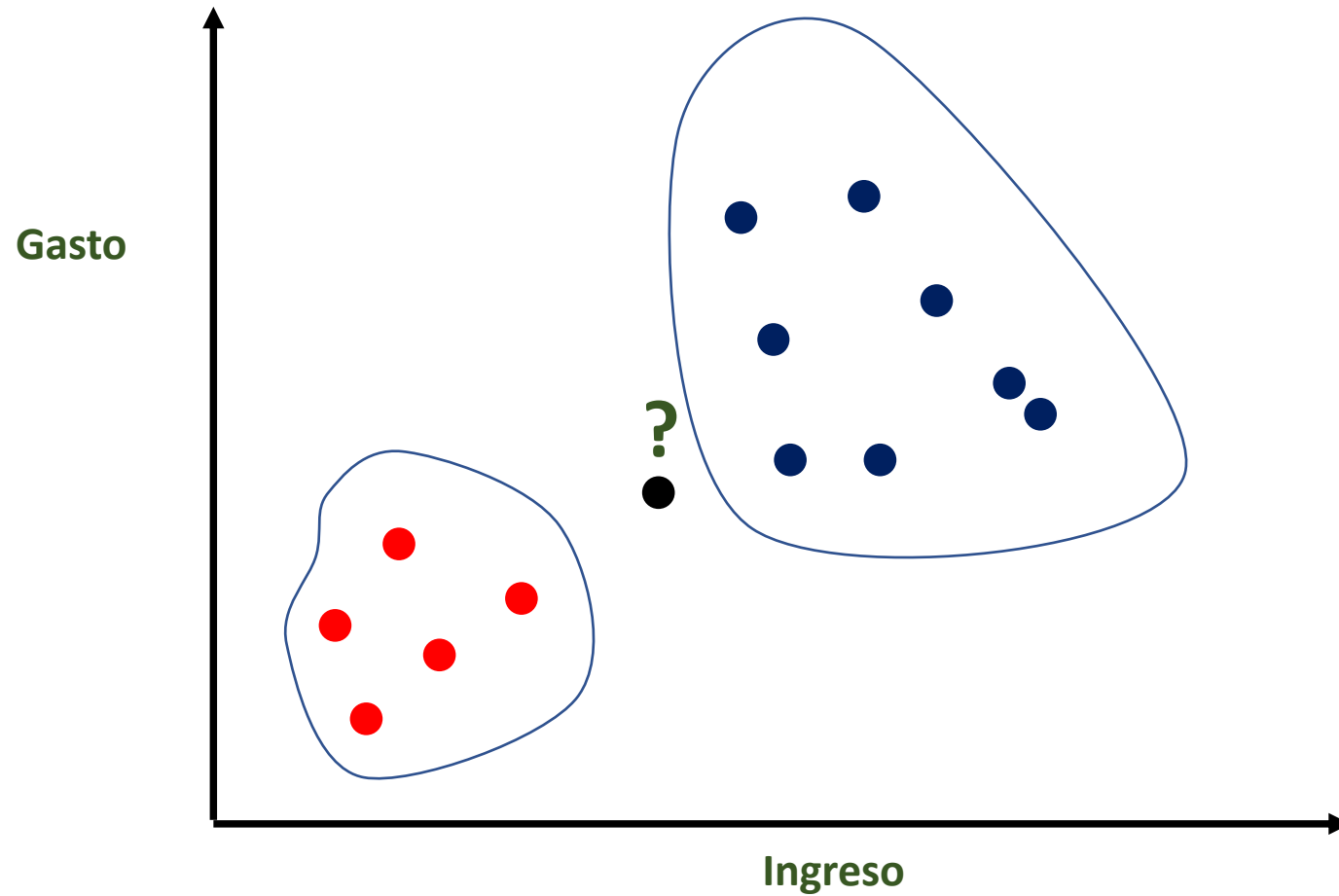
Ejemplo Centros Comerciales



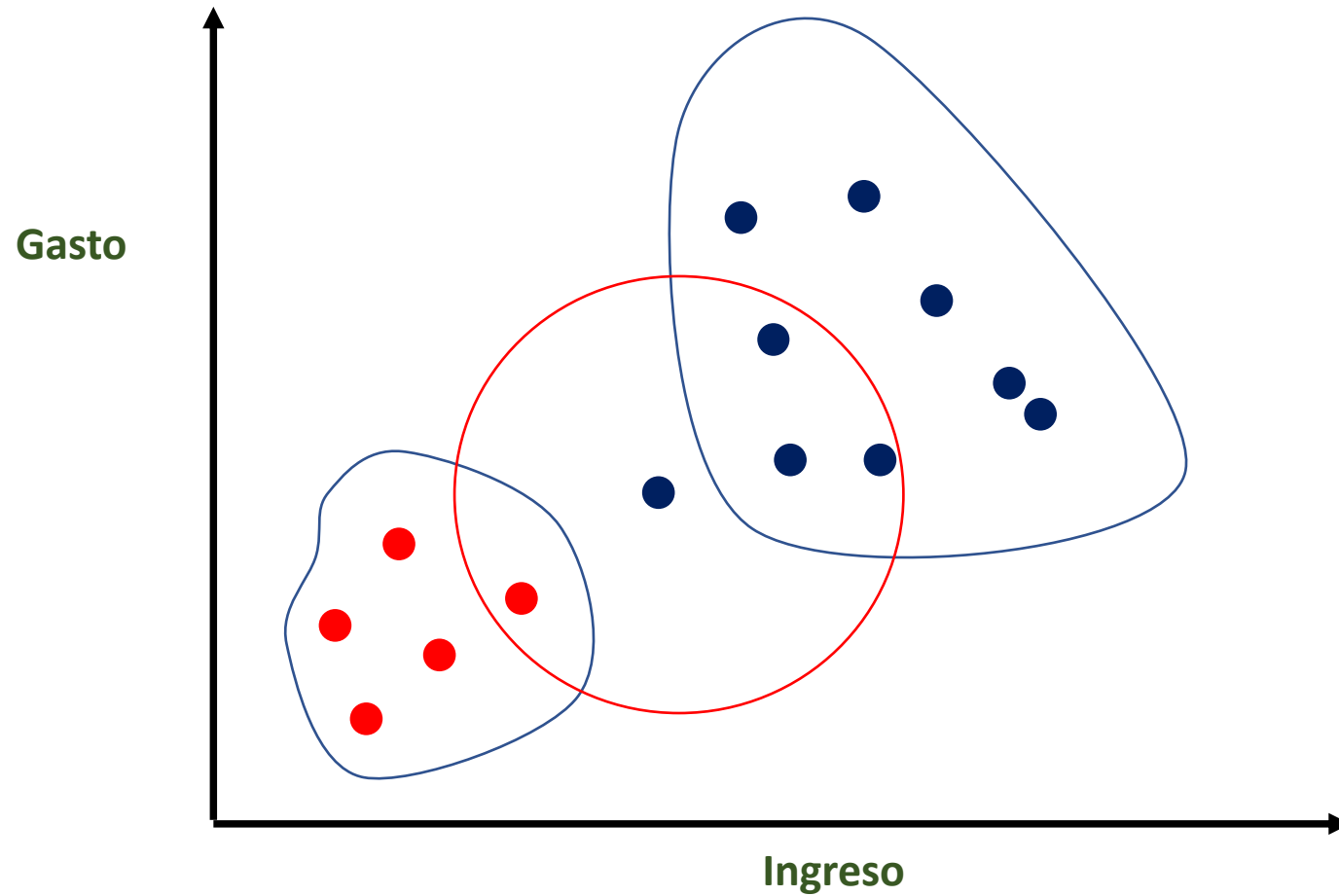
Ejemplo Centros Comerciales



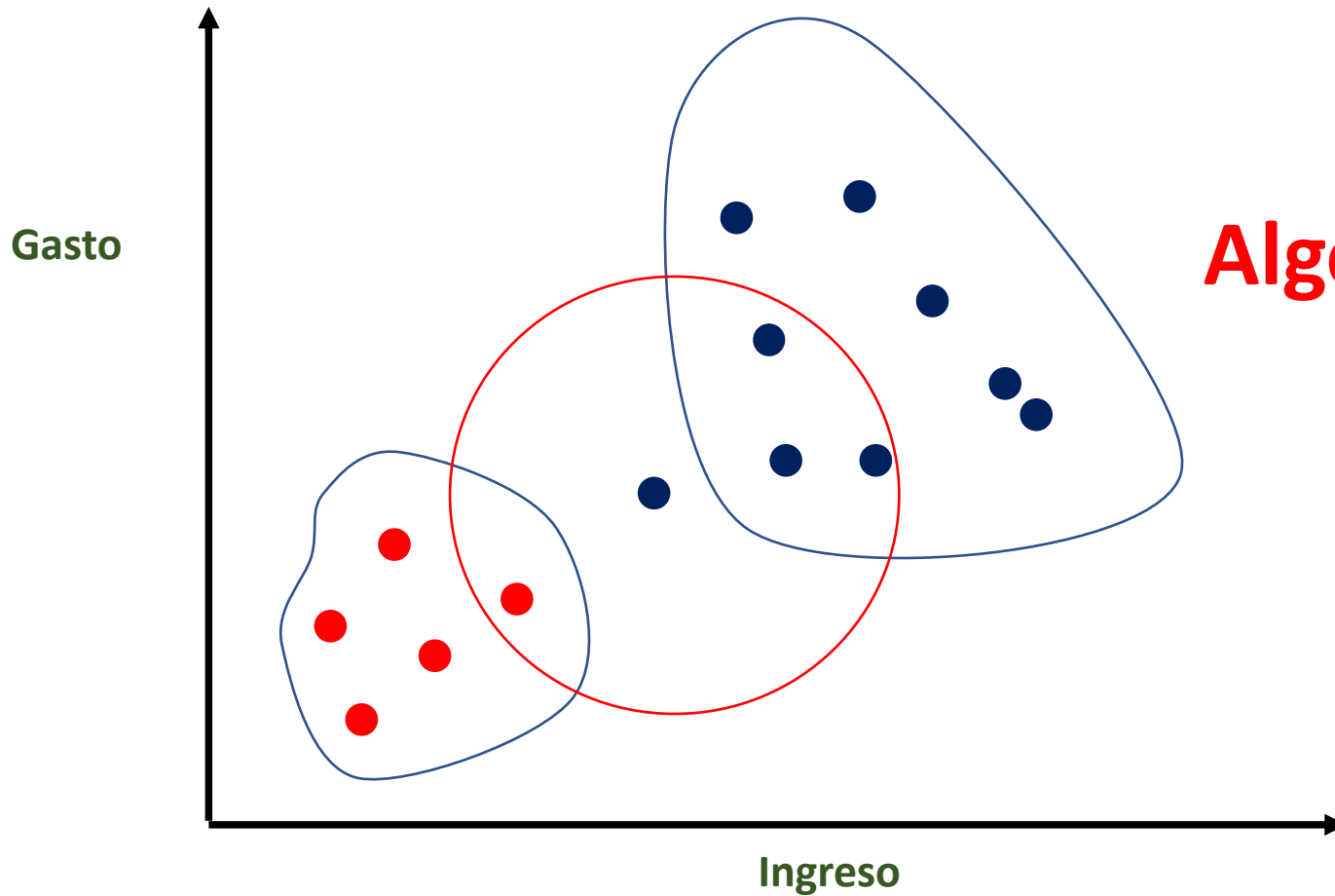
Ejemplo Centros Comerciales



Ejemplo Centros Comerciales



Ejemplo Centros Comerciales



**Clasificación:
Algoritmo k-Nearest**

Modelos de Clustering

Cluster + Factoextra R Packages

Partitioning Clustering

K-means

PAM

CLARA

Cluster + Dendextend + Factoextra R Packages

Hierarchical Clustering

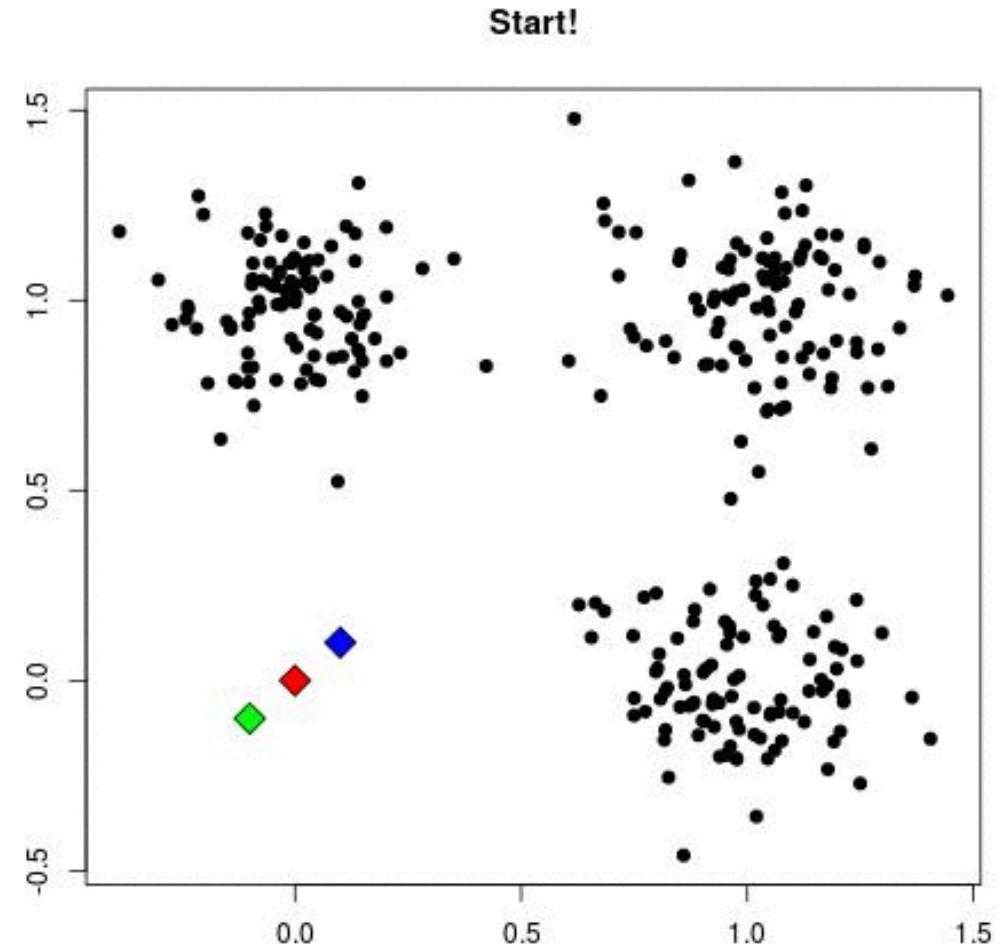
Agglomerative

- *Compute*
- *Visualize*
- *Interpret*
- *Compare*

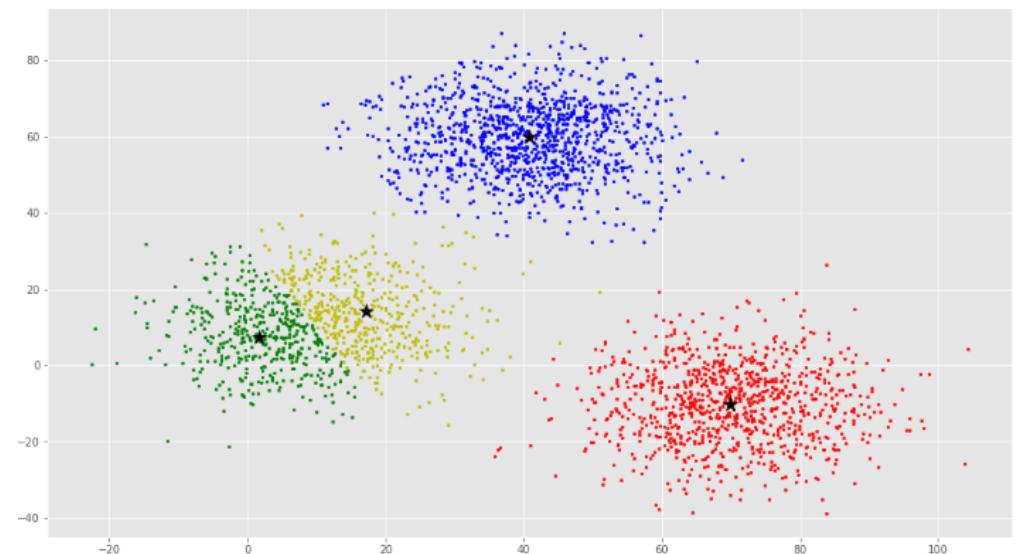
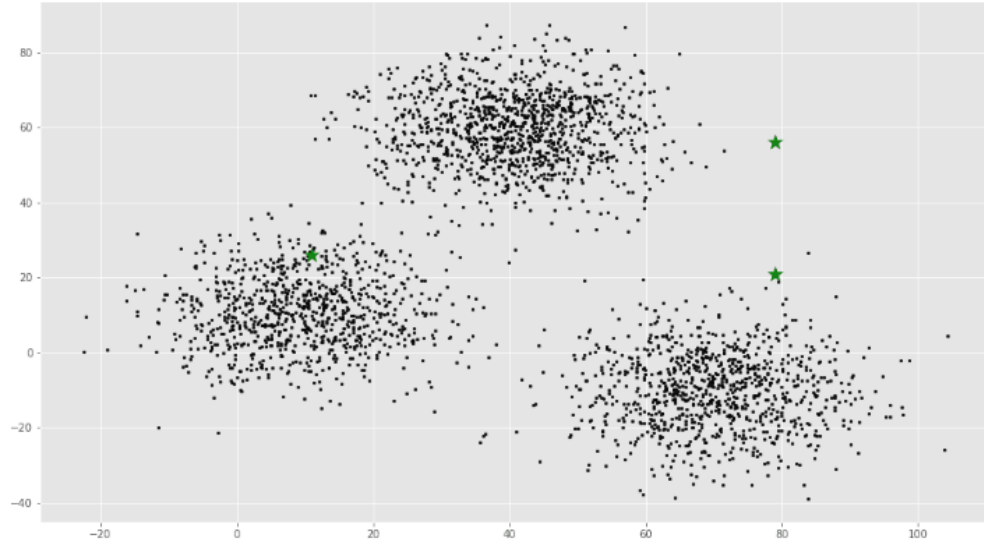
Divisive

Algoritmo k-Medias

- **Paso 1** - Seleccione K puntos aleatorios de inicio llamados centroides
- **Paso 2** - Asigne cada elemento al cluster más cercano basado en la distancia a cada centroide.
- **Step 3** - Genere un centroide nuevo tomando el promedio de los puntos asignados a cada cluster.
- **Step 4** - Repita los pasos 2 y 3 hasta que ninguna de las asignaciones a los clusters cambie.



Ejemplo K-Medias

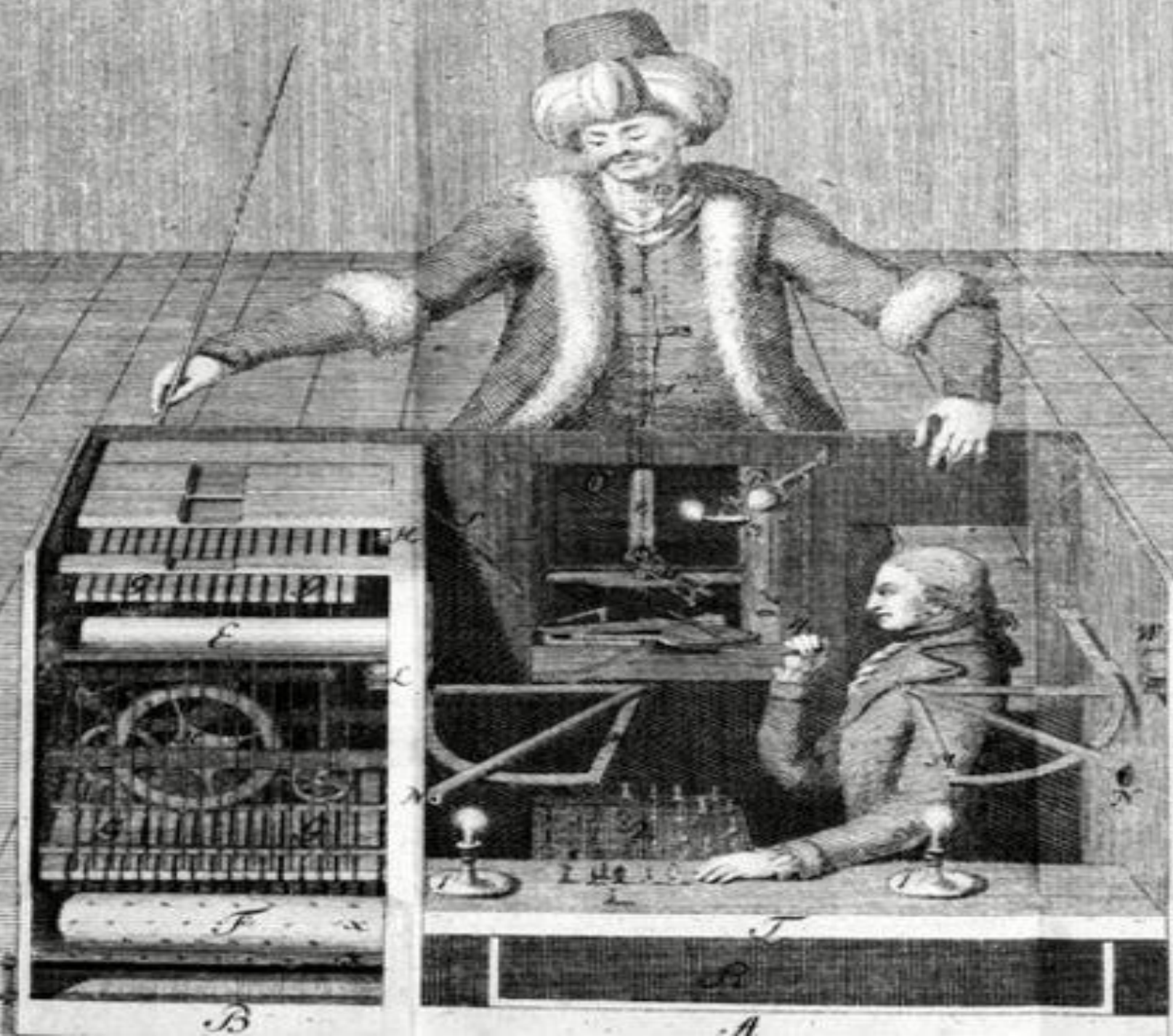




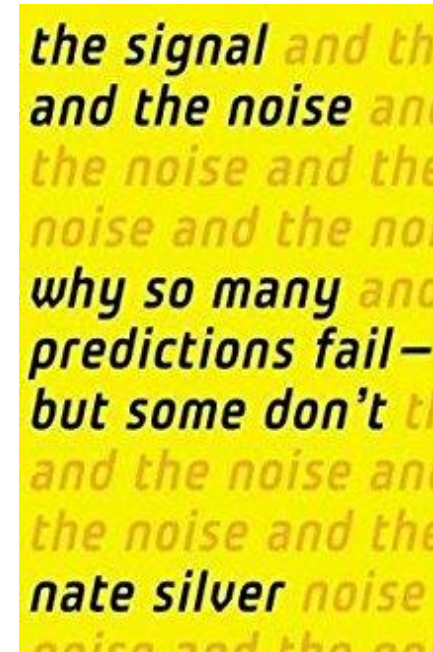
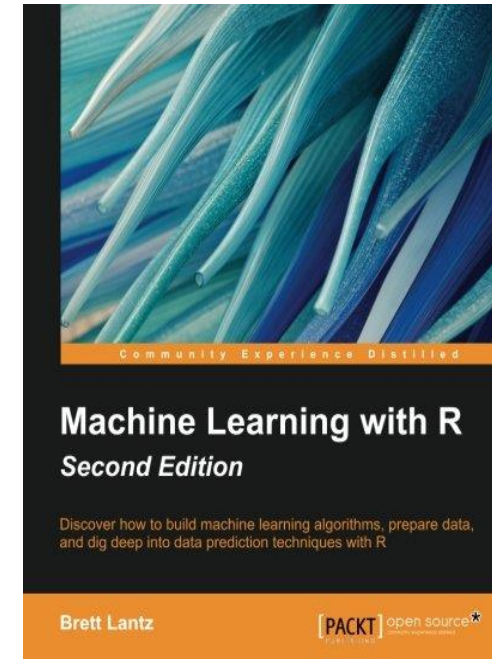
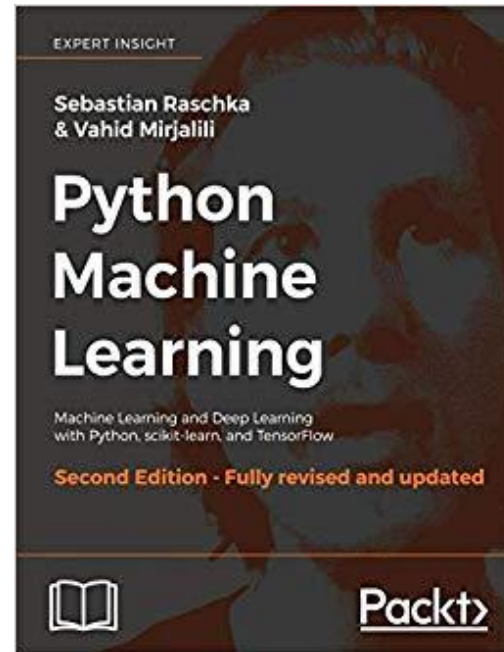
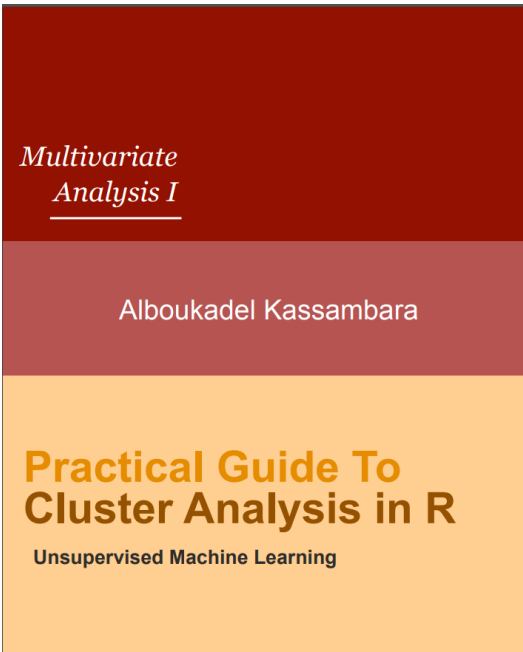
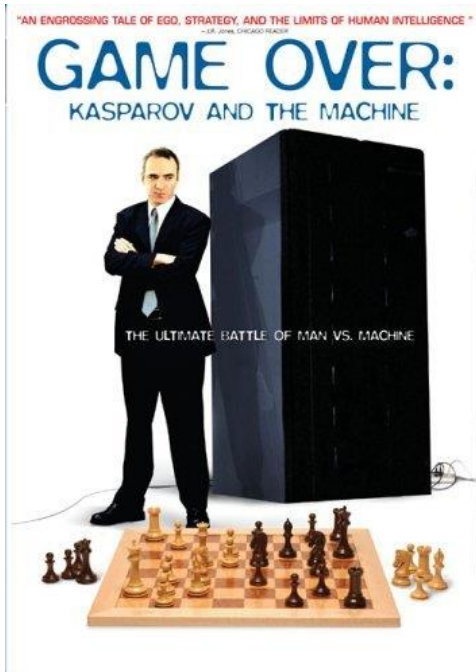


Tabulación de Resultados


		Total	Segmento 1	Segmento 2	Segmento 3	Segmento 4
		%	%	%	%	%
Sexo	Total	100.0	100.0	100.0	100.0	100.0
	Hombre	47.3	43.0	48.9	46.9	41.5
	Mujer	52.7	57.0	51.1	53.1	58.5
Edad	Total	100.0	100.0	100.0	100.0	100.0
	21-29	29.5	16.6	19.0	30.7	31.5
	30-39	22.7	16.8	24.1	23.2	25.1
	40-49	13.9	37.2	21.4	9.0	8.2
NSE	Total	100.0	100.0	100.0	100.0	100.0
	Alto	8.0	1.7	4.6	9.7	4.2
	Medio	46.3	24.5	42.3	49.2	48.0
	Bajo	45.6	73.8	53.0	41.1	47.8



Material



¿Preguntas?



APRENDIZAJE AUTOMÁTICO CON R

Modelos de Segmentación

Eugenio Grant

LinkedIn: <https://www.linkedin.com/in/eugeniogrant/>
Correo: eugenio.grant@gmail.com