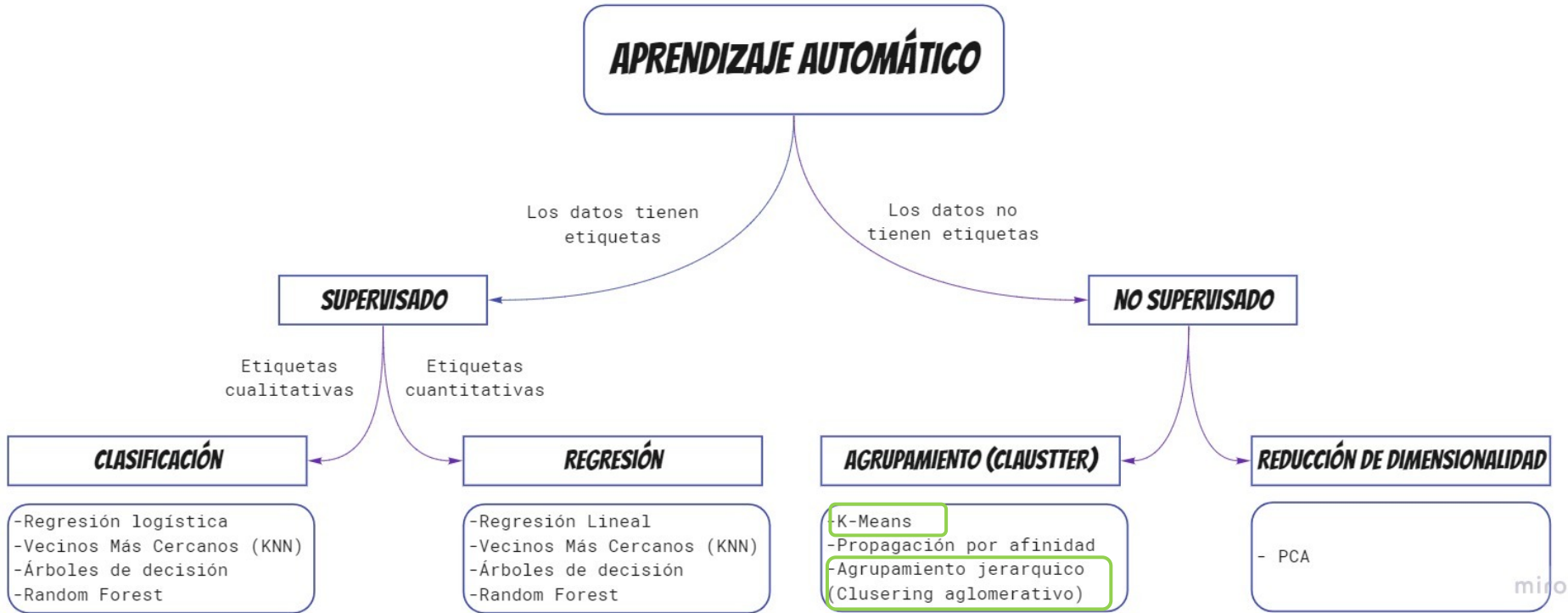


# Ciencia de Datos

- Módulo 3

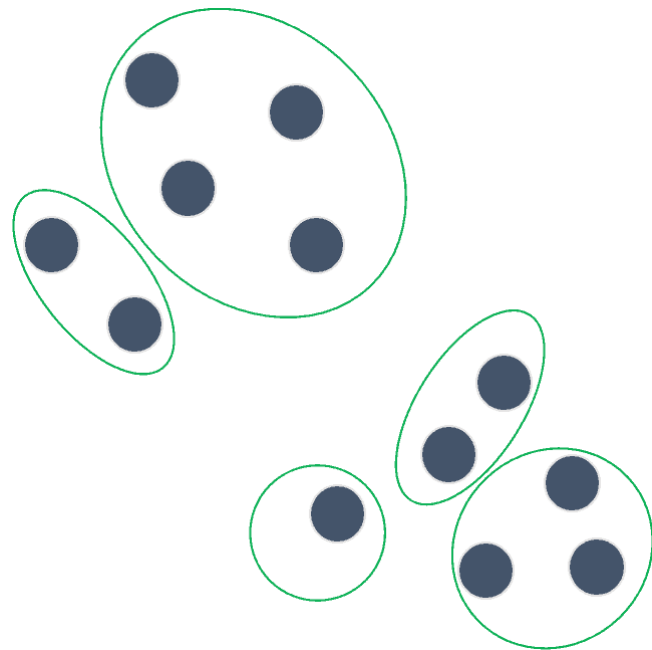
## Clusters





# Clusters

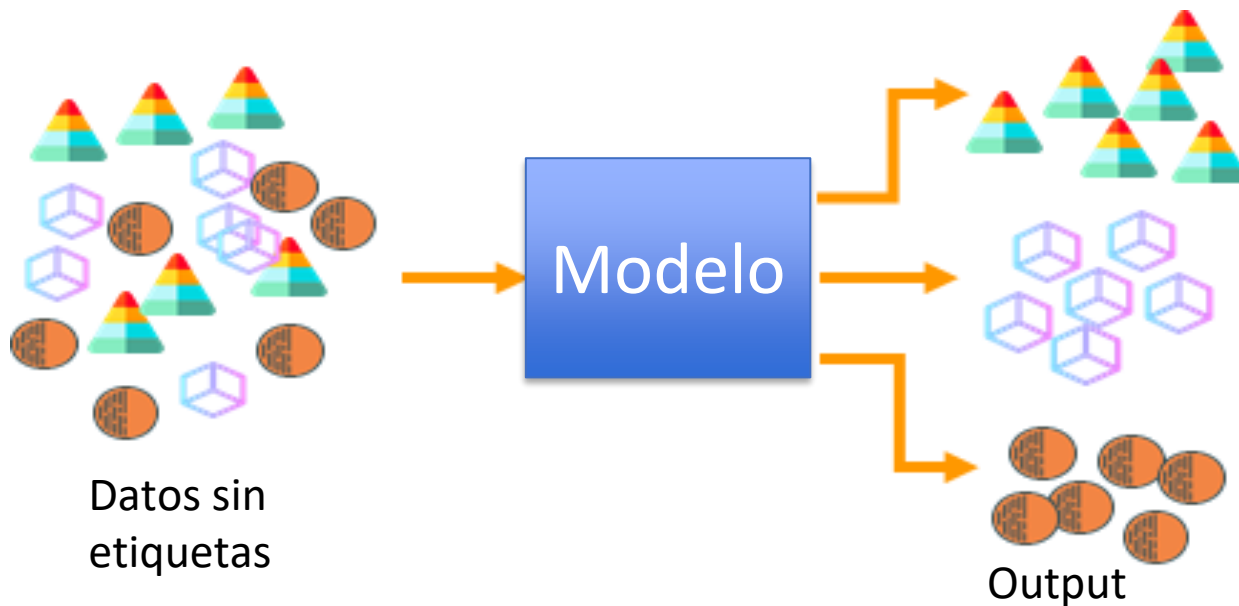
Uno de los objetivos más comunes del aprendizaje automático, más específicamente del aprendizaje no supervisado, es agrupar los datos por características similares. A estos grupos los llamamos Clusters.



Fuente: <https://www.iartificial.net/clustering-agrupamiento-kmeans-ejemplos-en-python/>



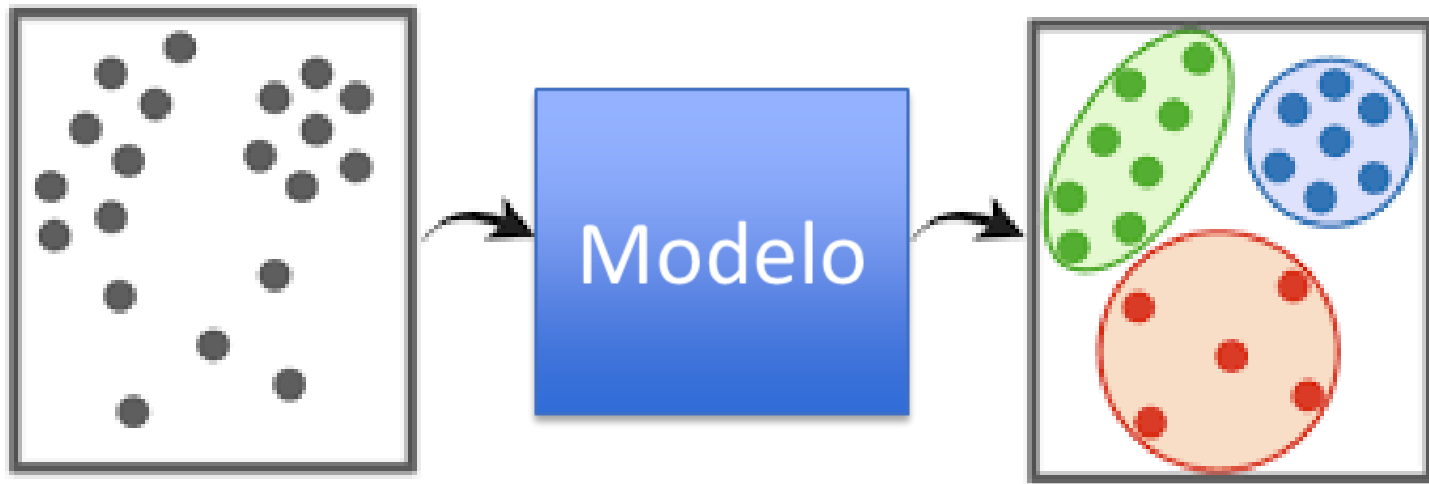
# Aprendizaje No supervisado



Fuente: <https://es.clariba.com/machine-learning-for-business>



# Aprendizaje No supervisado



Fuente: <https://es.clariba.com/machine-learning-for-business>

# Clusters

Los usos más comunes de esta técnica son:

- Segmentación de clientes
- Agrupamiento de productos u objetos.



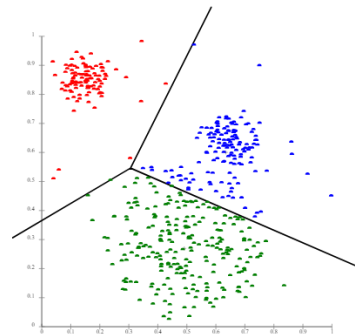
Fuente: <http://itm.ucam.edu/noticias/los-clusters-como-motor-de-desarrollo-empresarial>



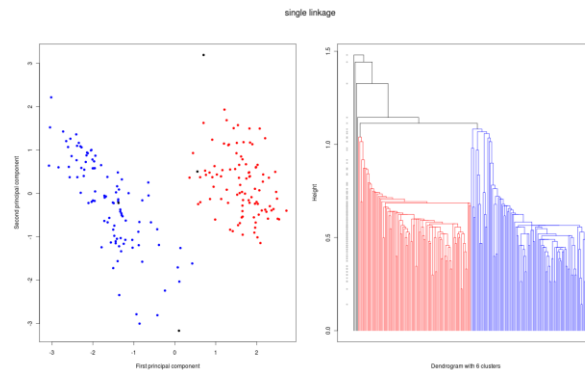
# Clusters

Existen muchos algoritmos de generación de clusters, entre los más populares están:

- Promedio K (K-means)
- Aglomeración jerárquica



Promedio K



Agrupamiento jerárquico

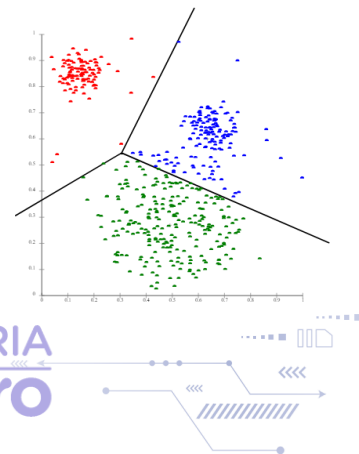


## Clusters - Promedio K

Este modelo genera una cantidad de clusters predefinida, y trata de buscar los K puntos (llamados centroides) tales que, minimicen la distancia de cada dato al centroide más cercano.

Este es un modelo con un entrenamiento iterativo en el que cada paso:

- Se calcula a qué cluster pertenece cada dato (A que centroide está más cerca cada dato).
- Se computa la media de todos los datos en cada cluster.
- Se actualiza el centroide con la media obtenida.





## Clusters - Promedio K

---

Visualizaciones:

<http://tech.nitoyon.com/en/blog/2013/11/07/k-means/>

<http://shabal.in/visuals/kmeans/3.html>

<https://www.naftaliharris.com/blog/visualizing-k-means-clustering/>

Más información:

<https://www.youtube.com/watch?v=4b5d3muPQmA>

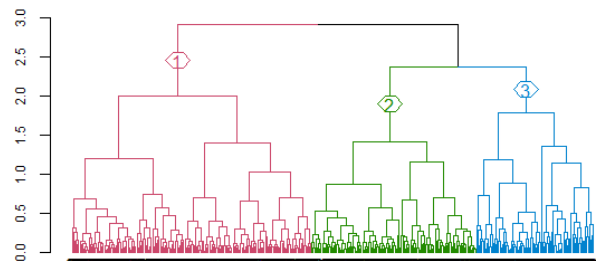


# Clusters - Aglomeración jerárquica

Este modelo va generando cluster unificando los datos más cercanos, dando como resultado una jerarquía de grupos. para esto:

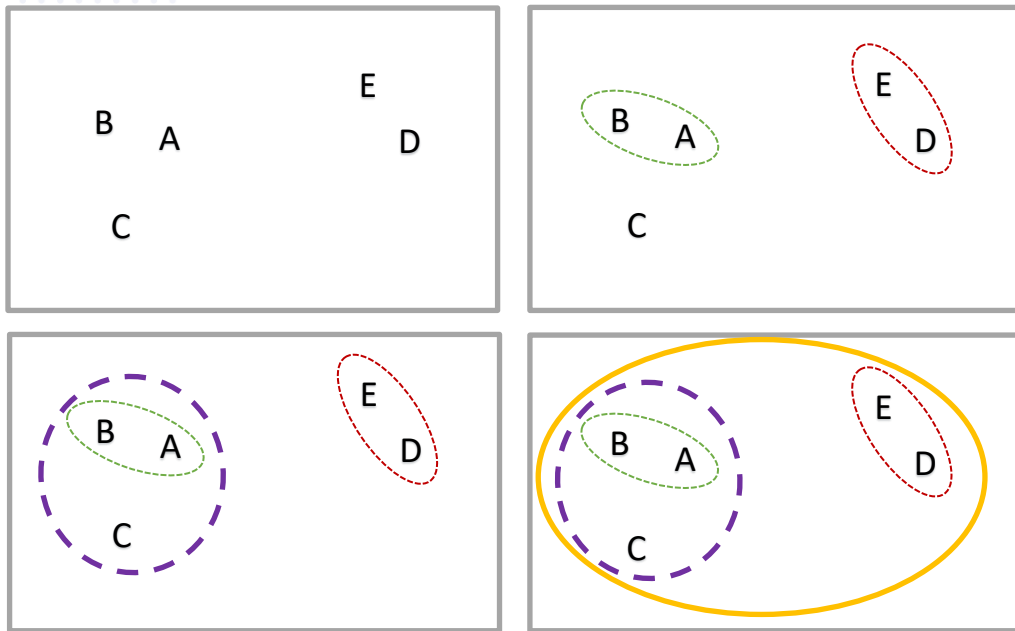
1. Calcula la distancia entre todos los datos.
2. Selecciona los dos datos con la menor distancia.
3. Esos dos datos se unen.
4. Cuando se unen dos datos, esos datos ya no son tomados en cuenta. En su lugar se considera un dato nuevo cuya posición se puede computar de diferentes formas siendo la más común el promedio de ambos datos.
5. Se repite el proceso hasta que solo queda un dato unificando a todo el conjunto.

Esto produce lo se conoce como dendograma.

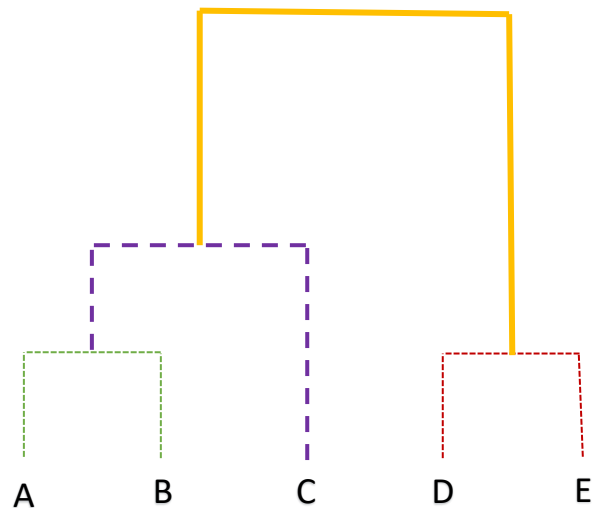


Fuente: [https://www.researchgate.net/figure/Dendograma-para-3-grupos-con-el-metodo-Ward\\_fig1\\_331462041](https://www.researchgate.net/figure/Dendograma-para-3-grupos-con-el-metodo-Ward_fig1_331462041)

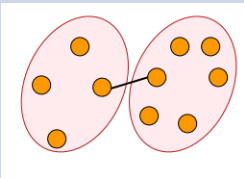
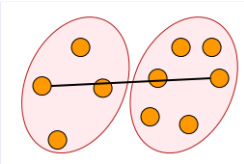
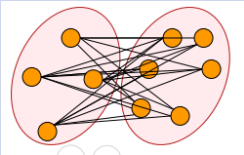
# Clusters - Aglomeración jerárquica



Dendrograma



# Clusters - Aglomeración jerárquica

Tipo de linkage	Descripción	Imagen
<b>Single</b>	Calcula todos los pares de distancias entre los miembros del cluster A y el cluster B y utiliza la mínima.	
<b>Completo</b>	Calcula todas los pares de distancias entre los miembros del cluster A y el cluster B y utiliza la máxima	
<b>Average</b>	Calcula todas los pares de distancias entre los miembros del cluster A y el cluster B y utiliza el promedio de todas	
<b>Ward</b>	Calcula la diferencia en la varianza total generada al aglomerar los diferentes clusters y busca la mínima	