

Las 5 situaciones prácticas

Las 5 situaciones prácticas de la ciencia de datos

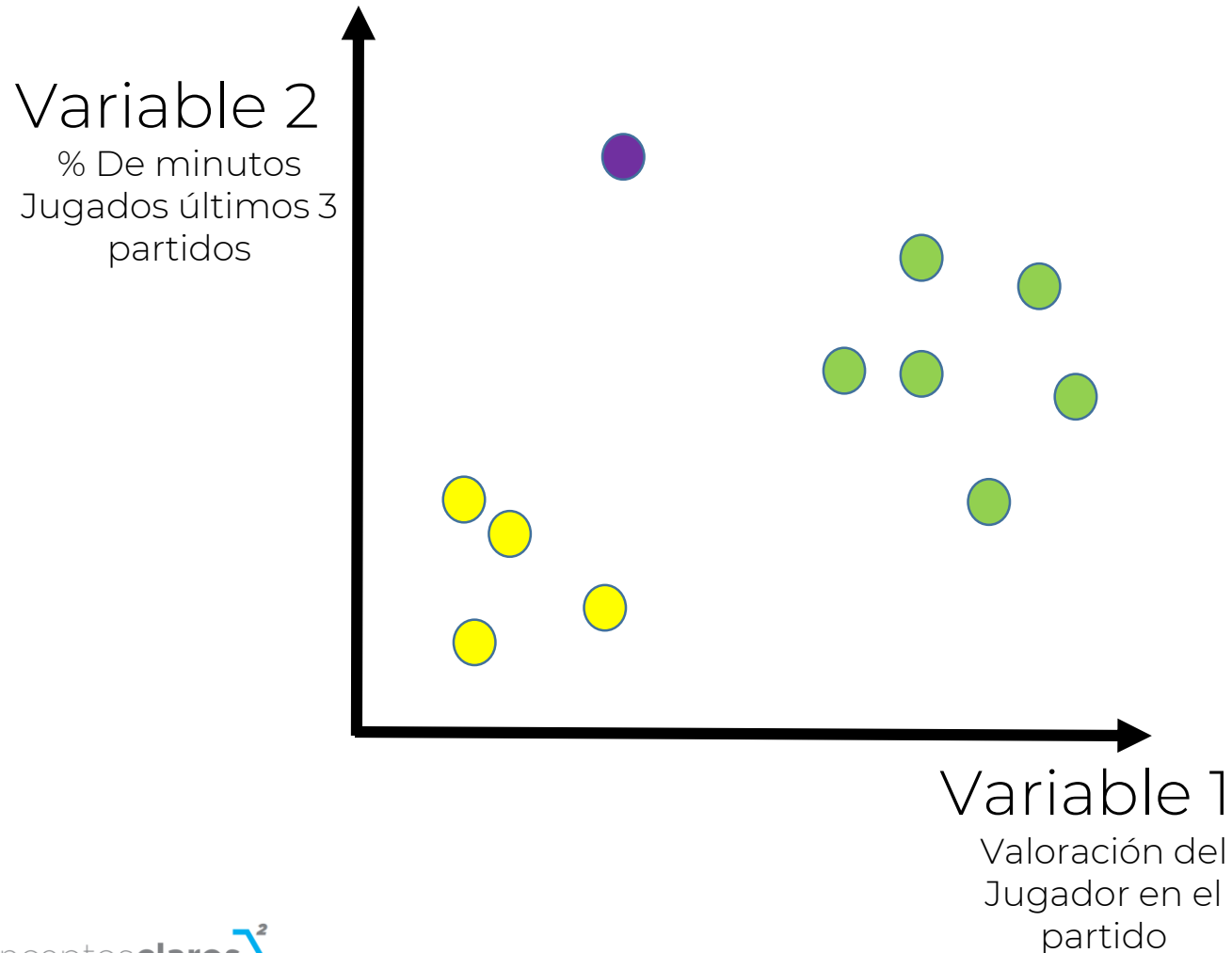
El mapa de aplicaciones prácticas de ciencia de datos



El mapa de aplicaciones prácticas de ciencia de datos



Clustering (no supervisado)



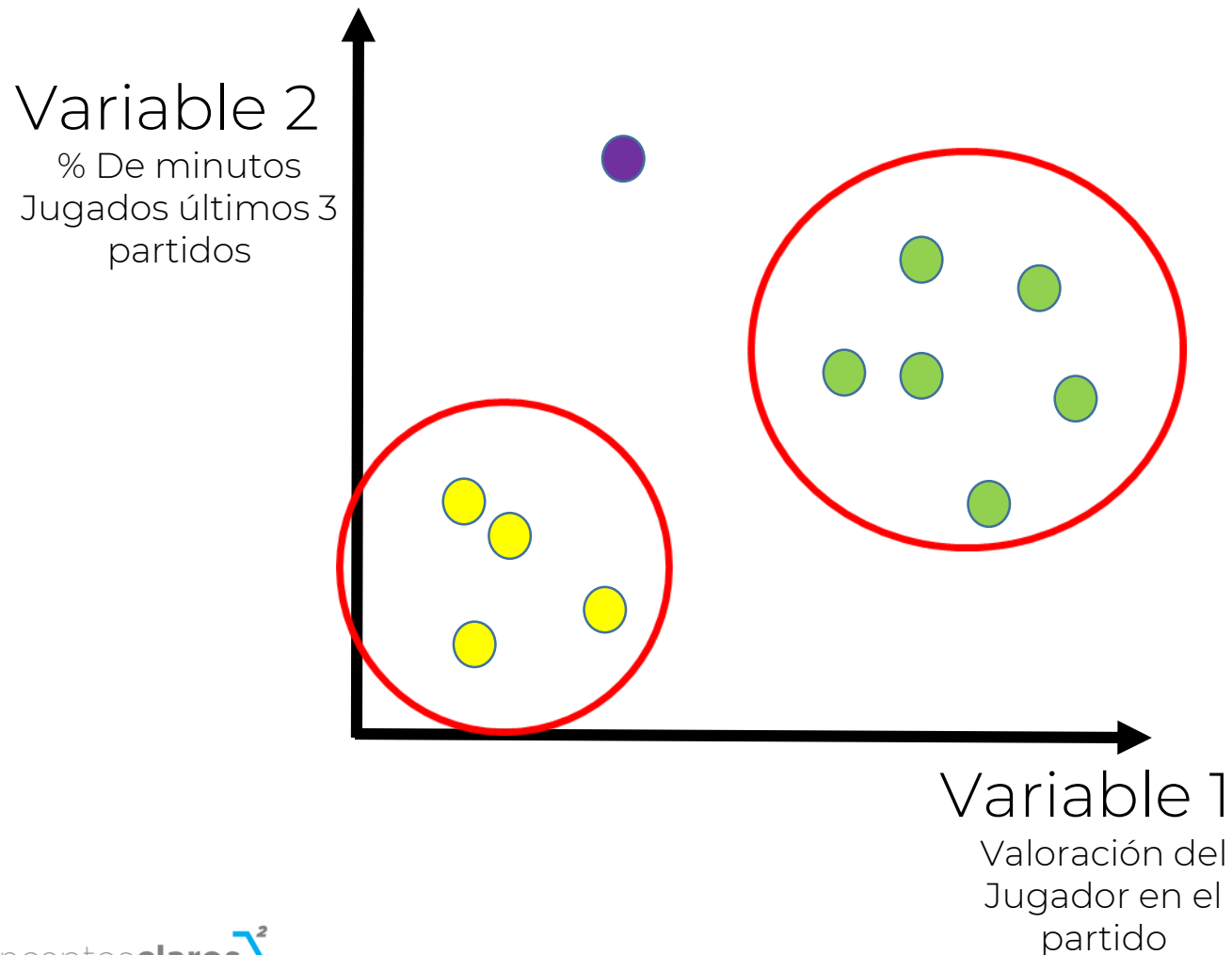
Objetivo:

A partir de características (features)

Construir grupos que sean lo suficientemente distintos

“Queremos crear 3 grupos de jugadores. Los malos, normales y buenos según el %de minutos jugados los últimos 3 partidos y su valoración”

Clustering (no supervisado)



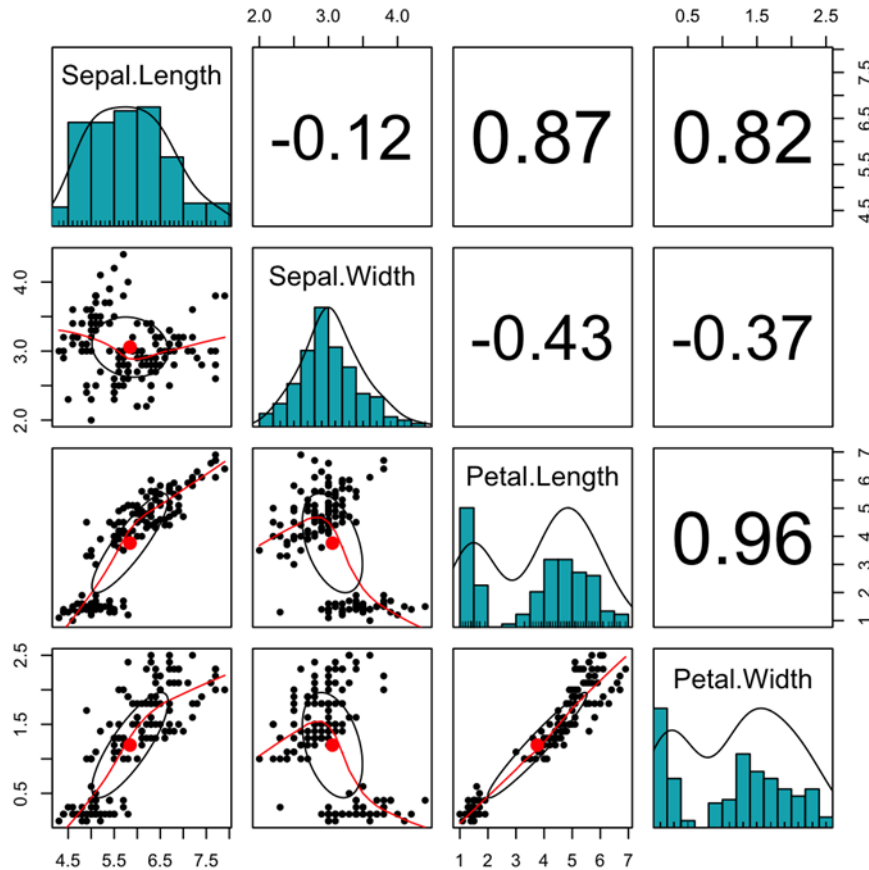
Objetivo:

A partir de características (features)

Construir grupos que sean lo suficientemente distintos

“Queremos crear 3 grupos de jugadores. Los malos, normales y buenos según el %de minutos jugados los últimos 3 partidos y su valoración”

Clustering (no supervisado)



Objetivo:

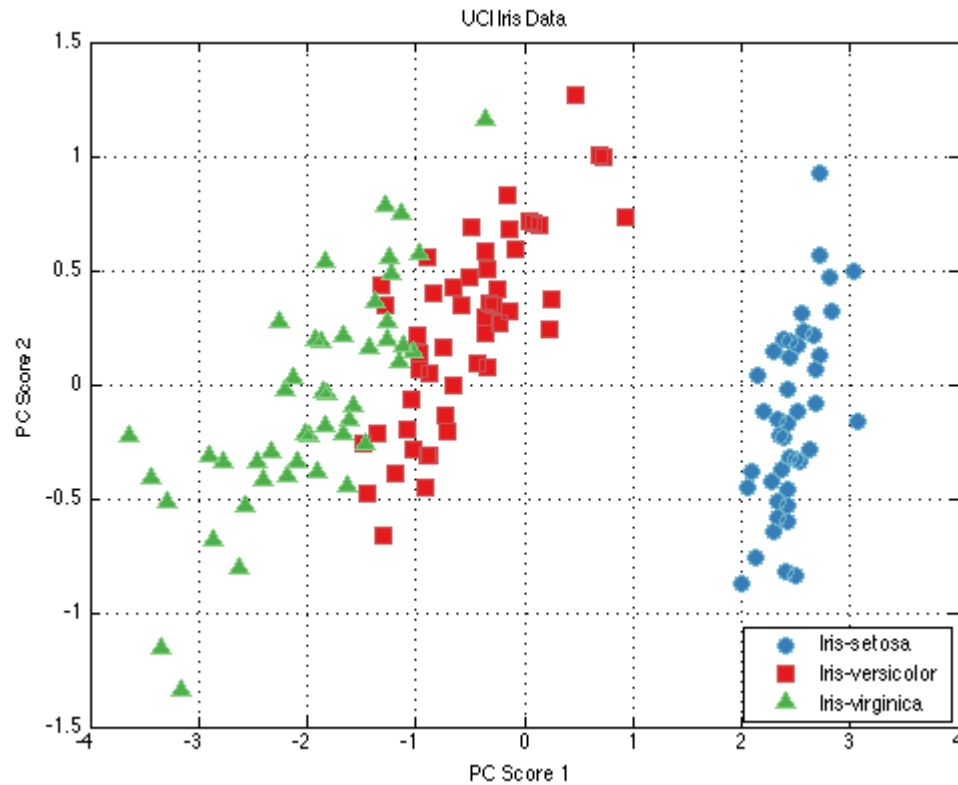
A partir de características (features)

Construir grupos que sean lo suficientemente distintos

Nos sirve para crear una nueva clase basada en las características

“A partir de las dimensiones de los pétalos y los sépalos poder dividir las observaciones en grupos según estas características”

Clustering (no supervisado)



Objetivo:

A partir de características (features)

Construir grupos que sean lo suficientemente distintos

Nos sirve para crear una nueva clase basada en las características

“A partir de las dimensiones de los pétalos y los sépalos poder dividir las observaciones en grupos según estas características”

Esquema del clustering

VARIABLES
DE
ENTRADA

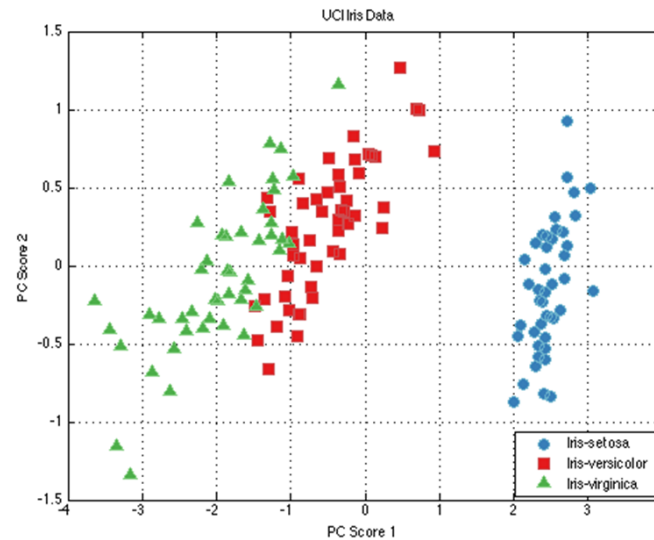


CLUSTERING



VARIABLE
DE
SALIDA

DIMENSIONES FLORES:
PETAL LENGTH
SEPAL LENGTH
PETAL WIDTH
SEPAL WIDTH



CLASE – ESPECIE DE FLOR

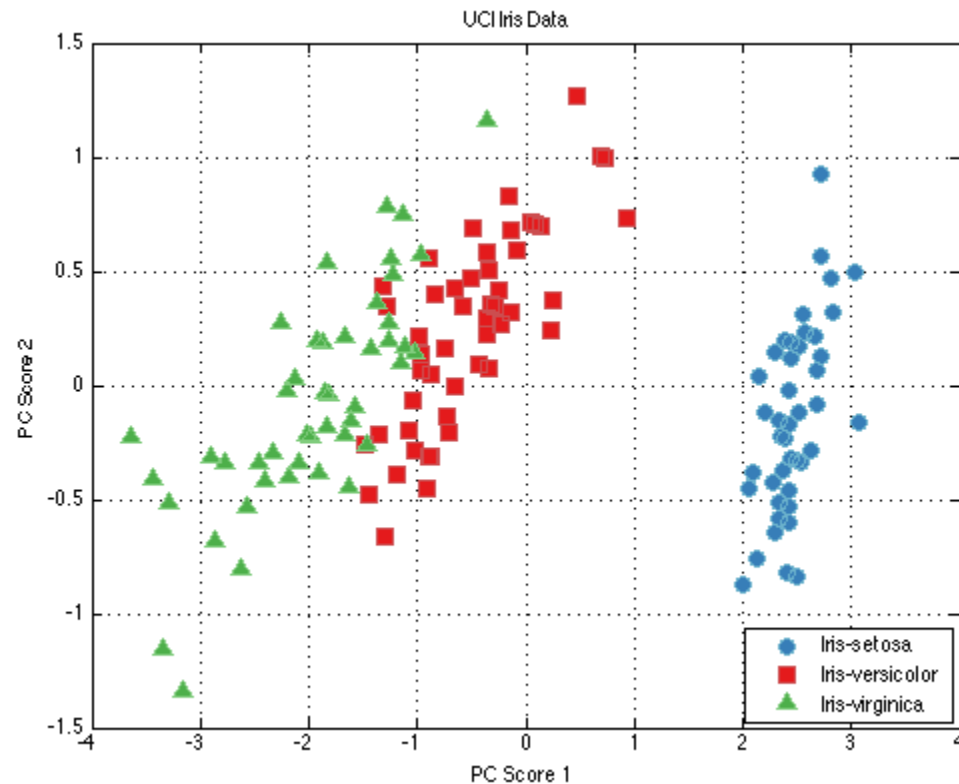
NO SUPERVISADO

CREO UNA NUEVA VARIABLE DE
SALIDA QUE ES UNA CLASE

El mapa de aplicaciones prácticas de ciencia de datos



Reducción dimensional (PCA...)



Objetivo:

A partir de características (features)

Reducir a nuevas variables con mayor importancia

- Nos sirve para poder dibujar en dos dimensiones
- Transformar los datos para calcular modelos predictivos posteriormente
- Reducir a menos variables

“A partir de las dimensiones de los pétalos y los sépalos queremos dibujar el % máximo de información en 2D”

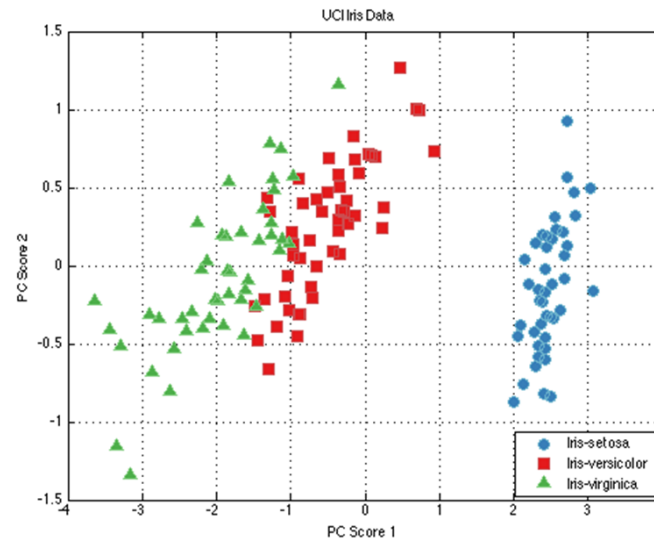
Esquema de la reducción dimensional

VARIABLES
DE
ENTRADA

REDUCCIÓN
DIMENSIONAL

VARIABLES
DE
SALIDA

DIMENSIONES FLORES:
PETAL LENGTH
SEPAL LENGTH
PETAL WIDTH
SEPAL WIDTH



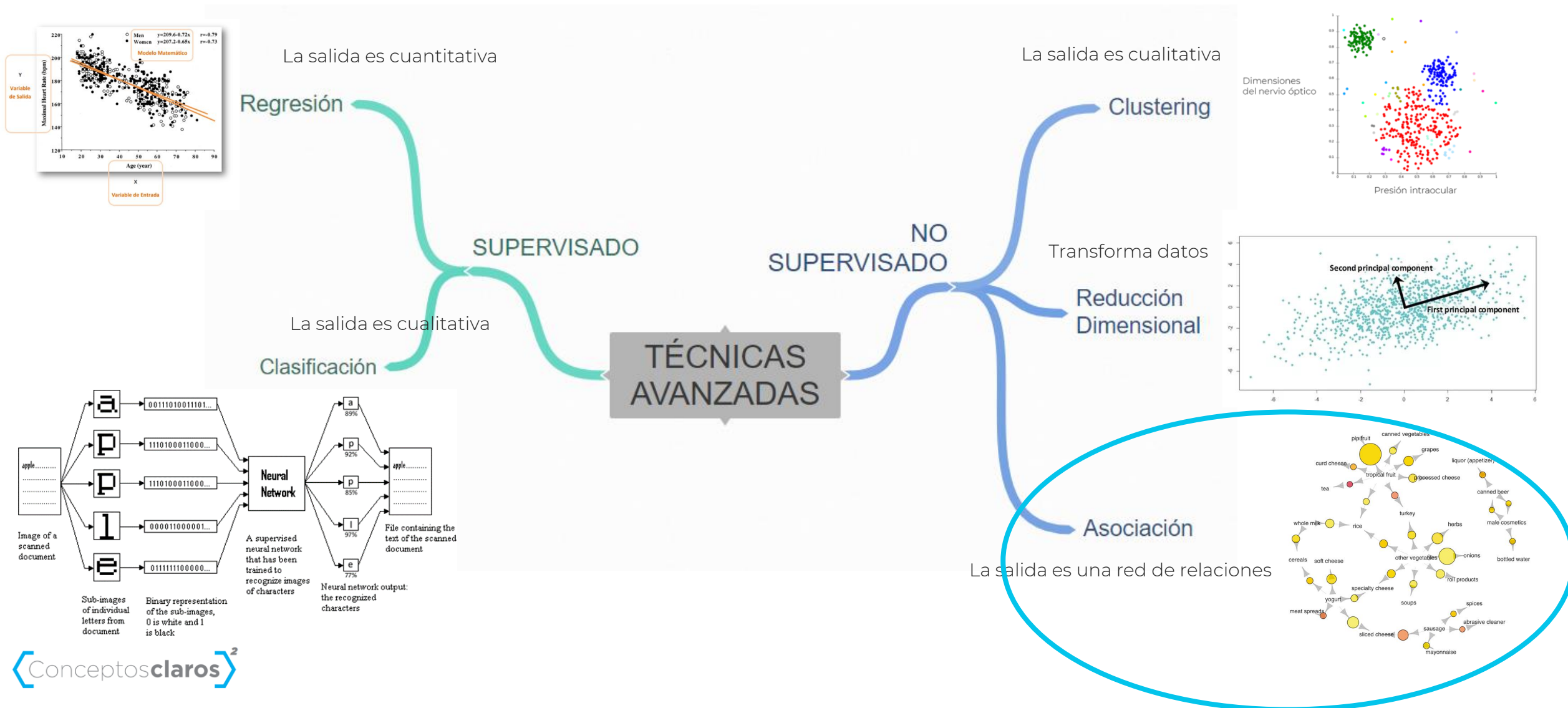
NUEVAS VARIABLES
ORDENADAS POR
INFORMACIÓN

PC1
PC2
PC3
PC4

NO SUPERVISADO

NUEVAS VARIABLES ORDENADAS
POR INFORMACIÓN

El mapa de aplicaciones prácticas de ciencia de datos

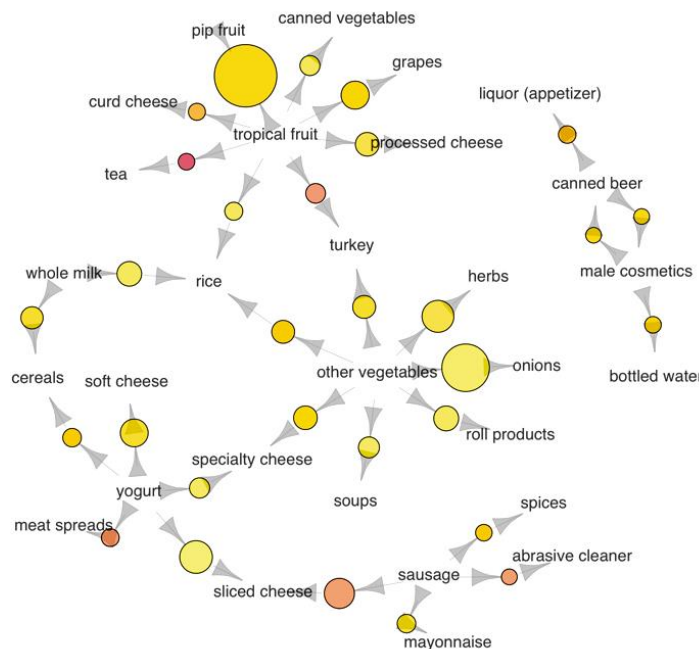


Asociación (no supervisado)

$$\text{Support } \{\text{🍎}\} = \frac{4}{8}$$

Transaction 1	🍎 🍺 🥛 🍗
Transaction 2	🍎 🍺 🥛
Transaction 3	🍎 🍺
Transaction 4	🍎 🍏
Transaction 5	🍼 🍺 🥛 🍗
Transaction 6	🍼 🍺 🥛
Transaction 7	🍼 🍺
Transaction 8	🍼 🍏

Table 1. Example Transactions



Objetivo:

A partir de un histórico de datos podemos observar asociaciones entre variables.

“Según el histórico de compras en supermercados se ha observado que muchos de los consumidores compran salsichas y queso en lonchas”

Esquema de las asociación

VARIABLES
DE
ENTRADA

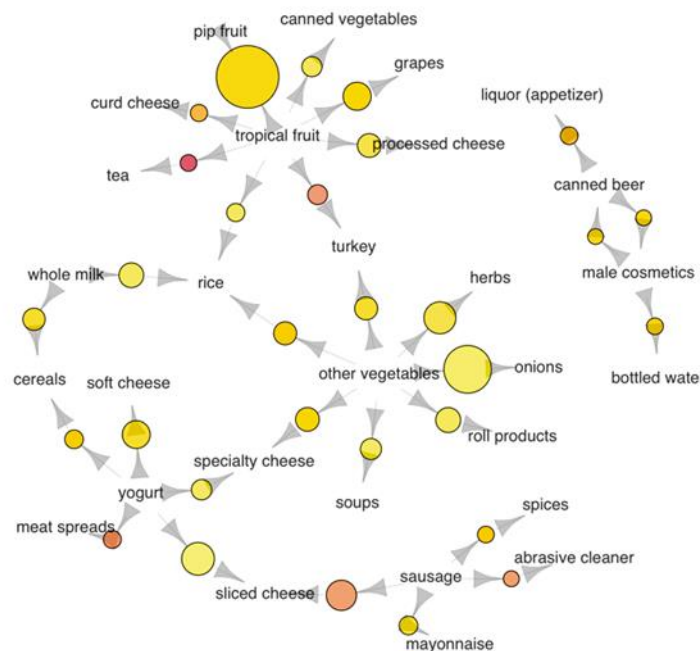


ASOCIACIÓN



RELACIONES
ENTRE LAS
ENTRADAS

Históricos de Transacciones
en el supermercado

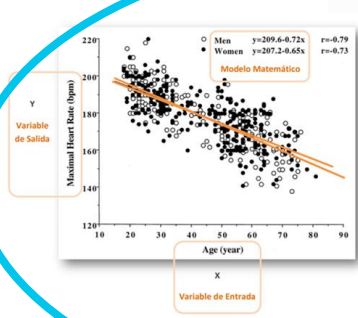


RELACIONES - PATRONES

NO SUPERVISADO

CREO RELACIONES ENTRE LAS
ENTRADAS

El mapa de aplicaciones prácticas de ciencia de datos

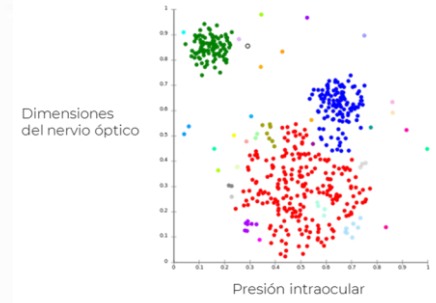


La salida es cuantitativa

Regresión

La salida es cualitativa

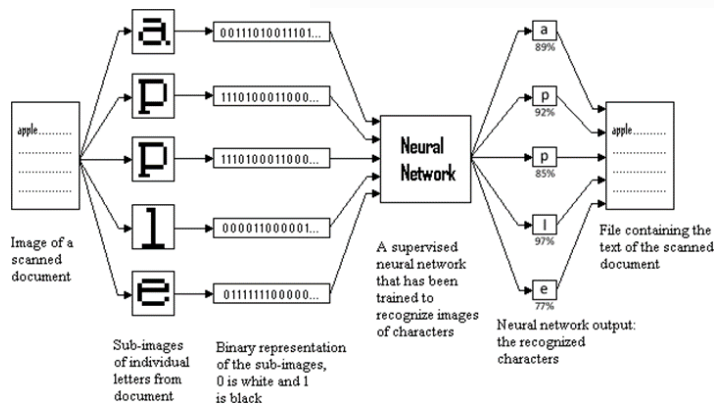
Clustering



SUPERVISADO

La salida es cualitativa

Clasificación

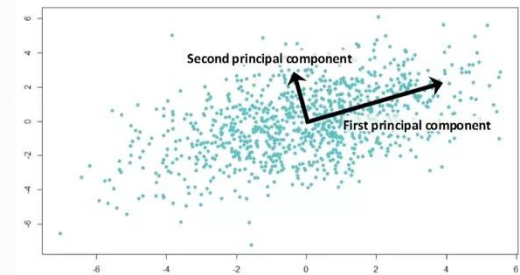


TÉCNICAS AVANZADAS

NO SUPERVISADO

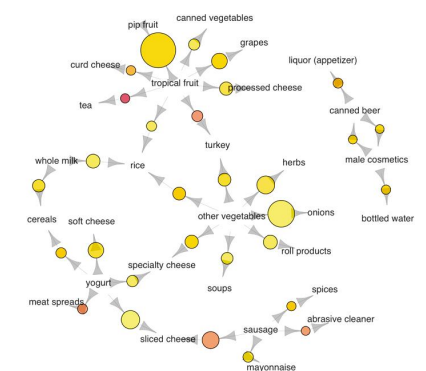
Transforma datos

Reducción Dimensional

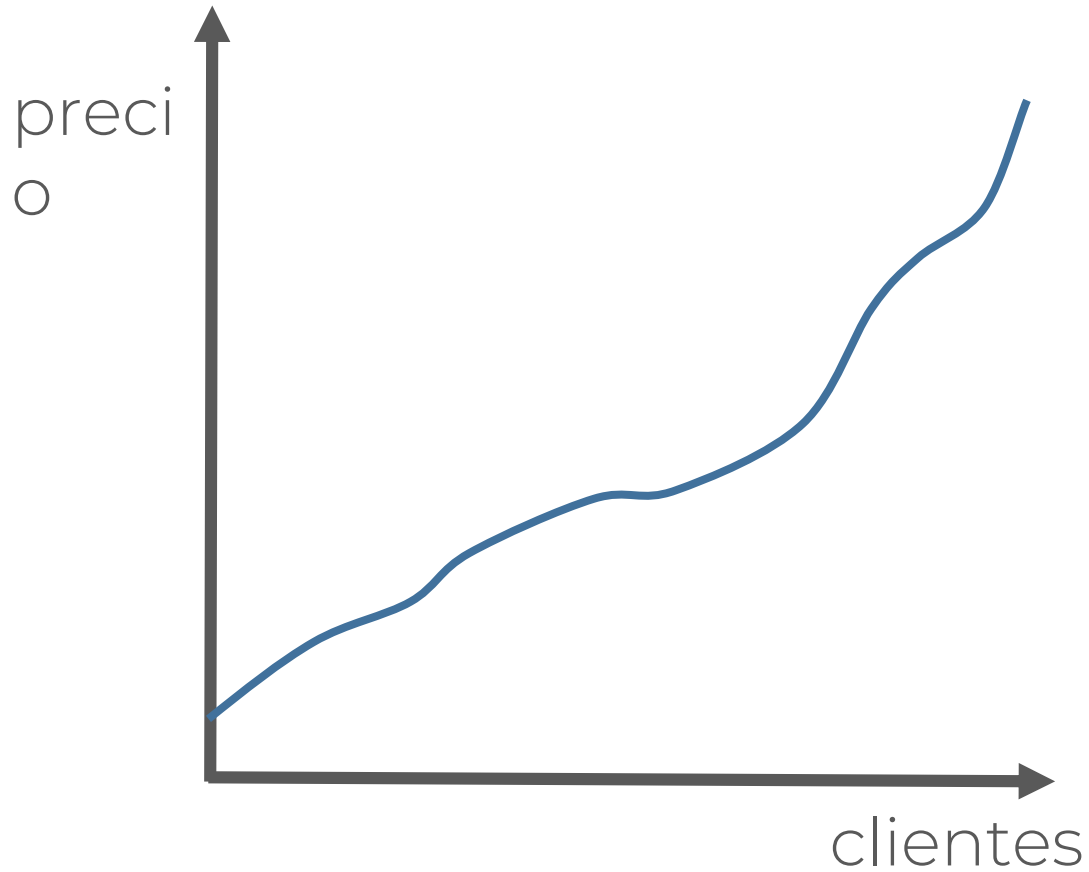


Asociación

La salida es una red de relaciones



Regresión (supervisado)

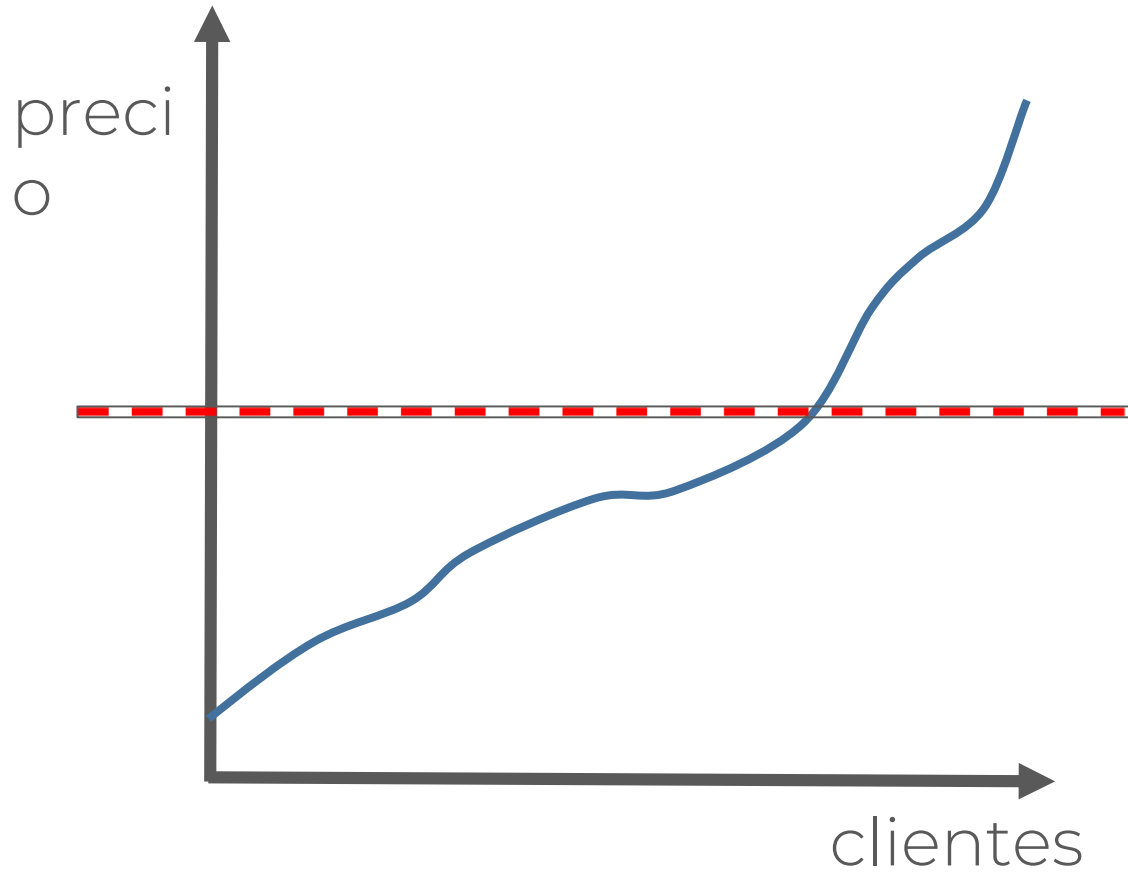


Objetivo:

A partir de una serie de variables predictoras (cuantitativas o cualitativas) poder predecir una característica cuantitativa de la forma más fiable posible, cuantificando el error

“A partir de las características de un cliente poder predecir los ingresos que te puede aportar”

Regresión (supervisado)



Objetivo:

A partir de una serie de variables predictoras (cuantitativas o cualitativas) poder predecir una característica cuantitativa de la forma más fiable posible, cuantificando el error

“A partir de las características de un cliente poder predecir los ingresos que te puede aportar”

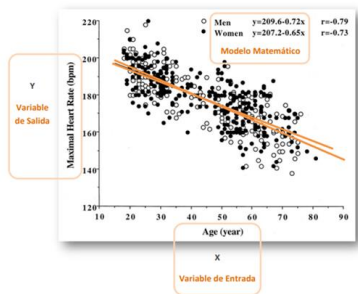
Esquema de la regresión



NO SUPERVISADO

CREO UNA NUEVA VARIABLE DE SALIDA QUE ES UNA CLASE

El mapa de aplicaciones prácticas de ciencia de datos

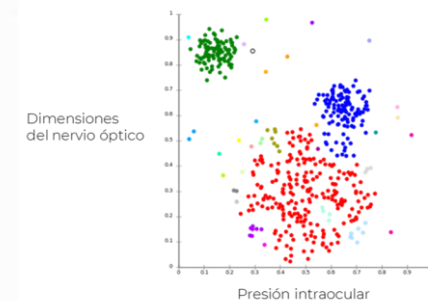


La salida es cuantitativa

Regresión

La salida es cualitativa

Clustering

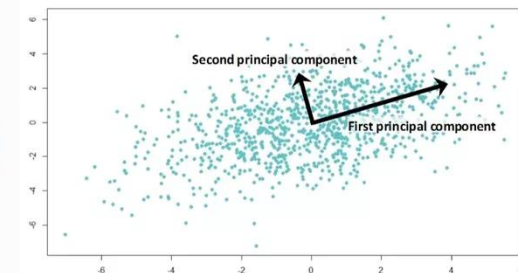


SUPERVISADO

NO SUPERVISADO

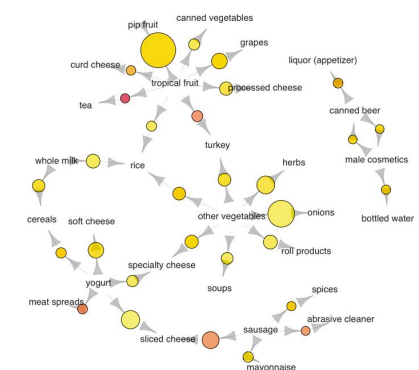
Transforma datos

Reducción Dimensional



Asociación

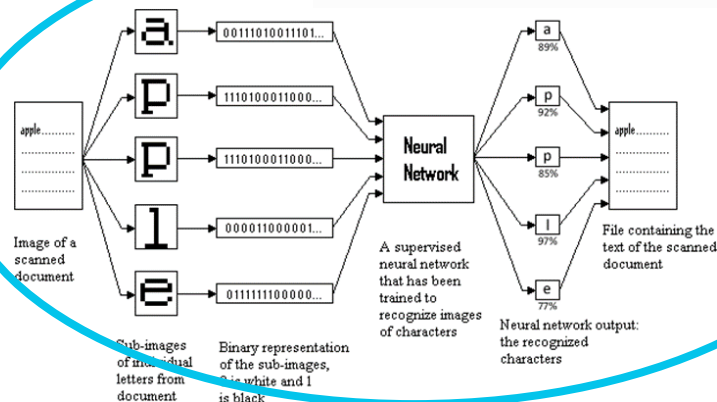
La salida es una red de relaciones



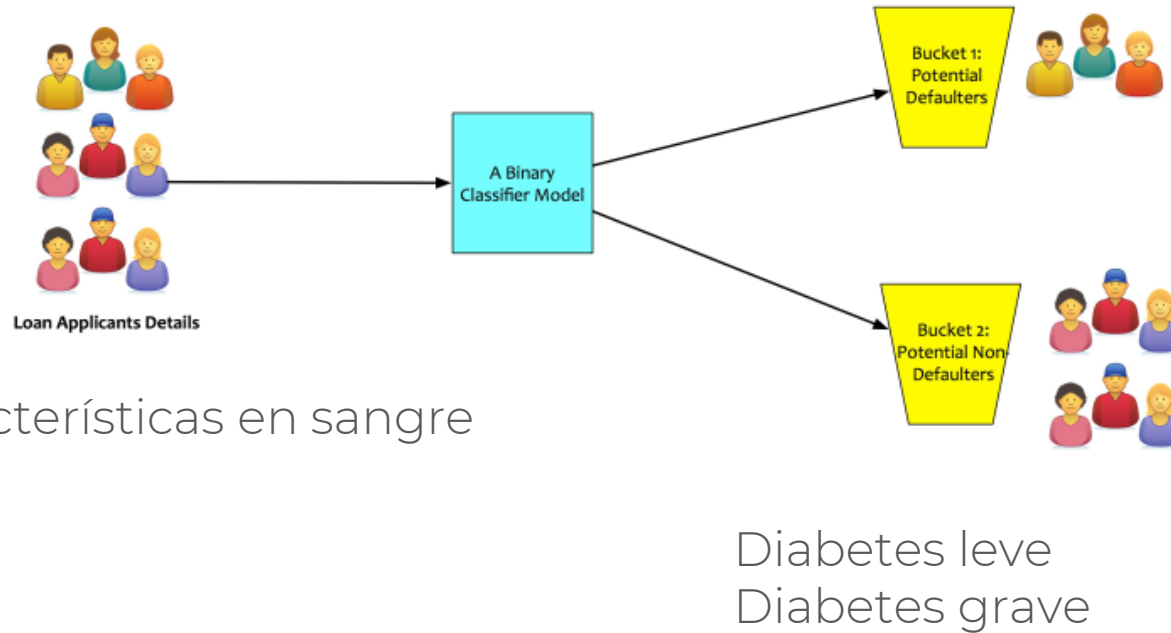
La salida es cualitativa

Clasificación

TÉCNICAS AVANZADAS



Clasificación (supervisado)



Características en sangre

Objetivo:

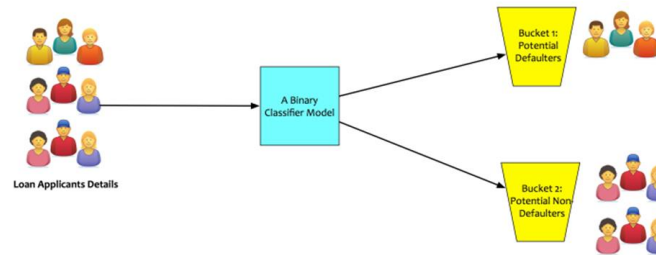
A partir de una serie de variables predictoras (cuantitativas o cualitativas) poder predecir el grupo de pertenencia

“A partir de características en sangre poder predecir si tenemos riesgo bajo o alto de padecer diabetes”

Esquema de la clasificación



Peso
Edad
Nivel de glucosa medio
...



Diabetes:
Leve
Grave

NO SUPERVISADO

CREO UNA NUEVA VARIABLE DE SALIDA QUE ES UNA CLASE

Take away

El resumen de la lección

Lo más importante de la lección

- Las 5 situaciones en la práctica más comunes son:
 - Clustering
 - Reducción dimensional
 - Asociación
 - Regresión
 - Clasificación

Tú turno

Un pequeño ejercicio para empezar

Tú turno

- Descarga la hoja de trabajo que te he preparado
- Según el problema que tienes en mente con tus datos, piensa qué tipo de situación es más interesante para tu caso.
- Dale una pensada 😊