

# RIESGO DE ATAQUE CEREBRAL



# Peligro de derrame Cerebral

- Exploramos la salud cerebral de una muestra de 10000 individuos de una ciudad, a través de atributos relacionados con su entorno y salud, para poder así predecir futuros



Según la Organización Mundial de la Salud (OMS), el accidente cerebrovascular es la segunda causa de muerte a nivel mundial y es responsable de aproximadamente el 11% del total de muertes.

Y yo te pregunto a ti:

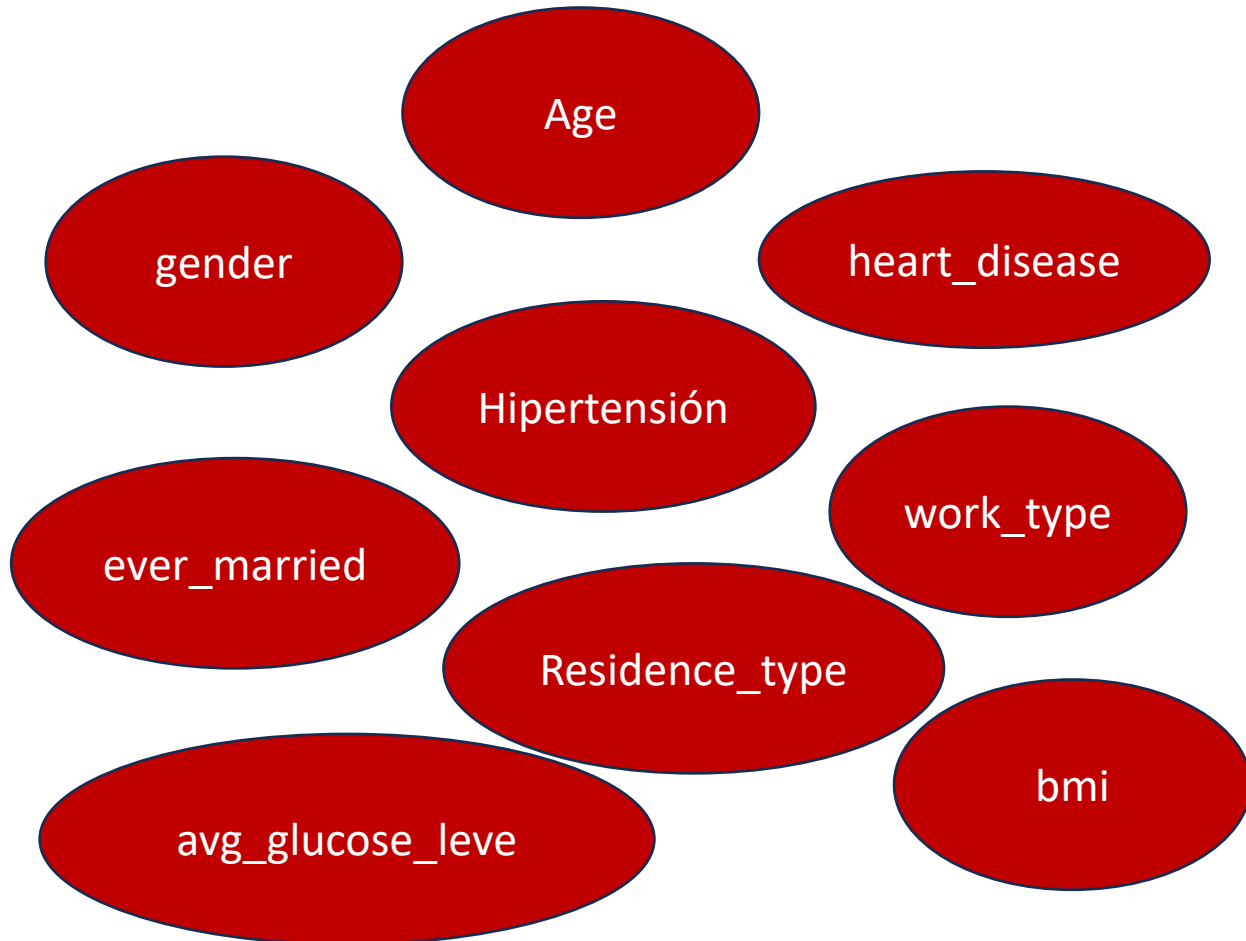
**¿Cuál crees que son los factores de mayor riesgo?**

# Variables de estudio:

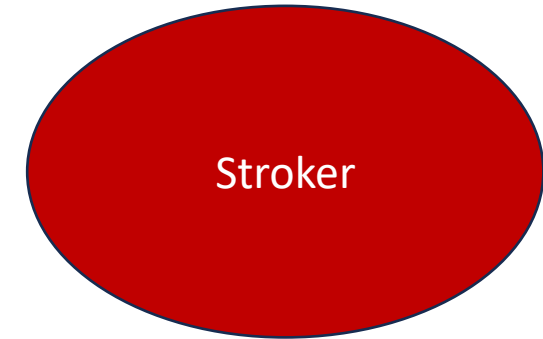
- 2 gender: "Masculino", "Femenino" u "Otro"
- 3 age: edad del paciente
- 4 hypertension: 0 si el paciente no tiene hipertensión, 1 si el paciente tiene hipertensión
- 5 heart\_disease: 0 si el paciente no tiene enfermedades cardíacas, 1 si el paciente tiene una enfermedad cardíaca.
- 6 ever\_married: "No" o "Sí"
- 7 work\_type: "Niños", "Trabajo gubernamental", "Nunca trabajó", "Privado" o "Autónomo"
- 8 Residence\_type: "Rural" o "Urbano"
- 9 avg\_glucose\_level: nivel promedio de glucosa en sangre
- 10 bmi: índice de masa corporal
- 11 smoking\_status: "Exfumador", "Nunca fumó", "Fuma" o "Desconocido"
- 12 stroke: 1 si el paciente tuvo un derrame cerebral o 0 si no lo tuvo

# VARIABLES OBJETIVO

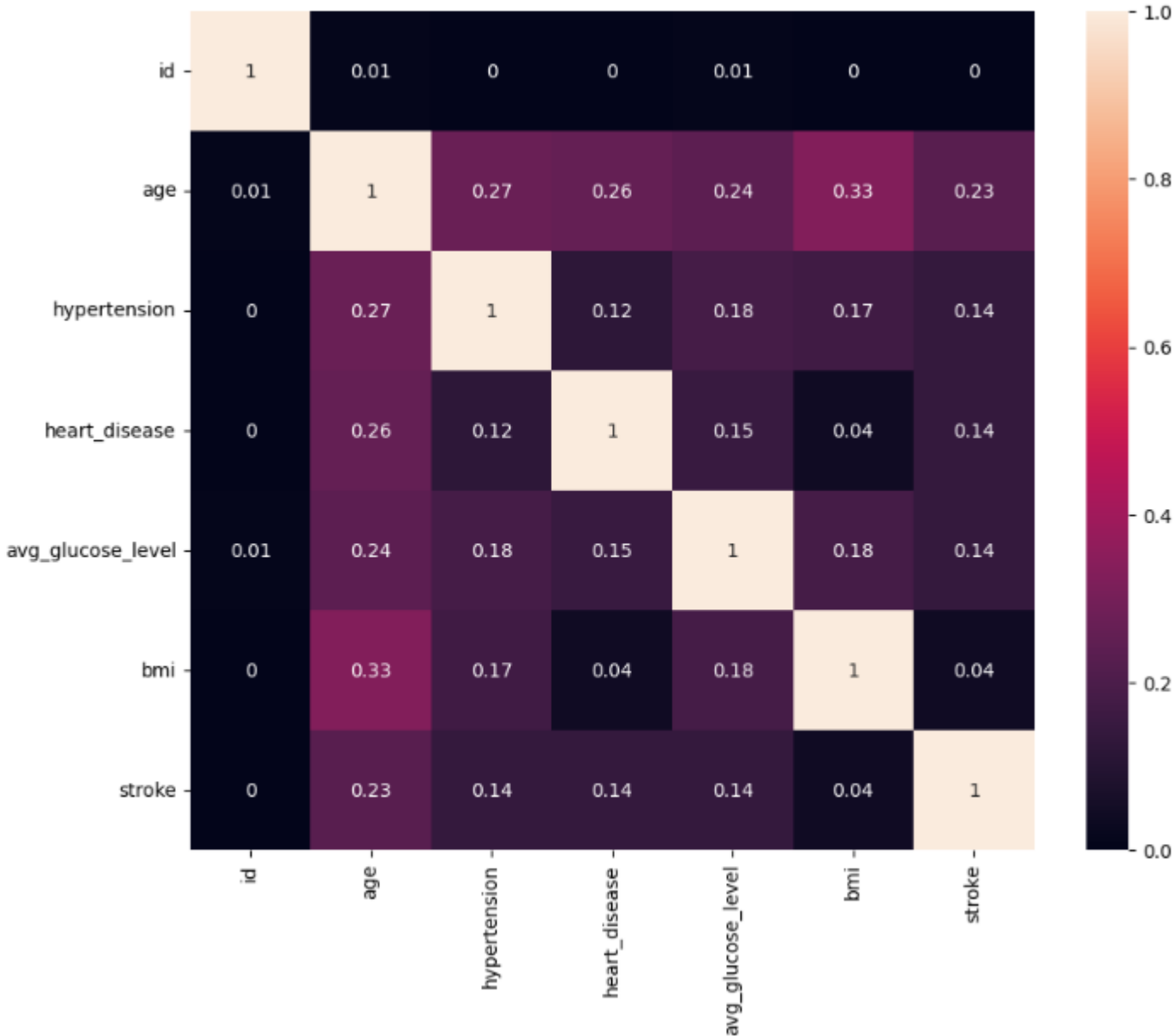
## VARIABLES PREDICTORAS



## VARIABLES OBJETIVO



# Exploración visual 01

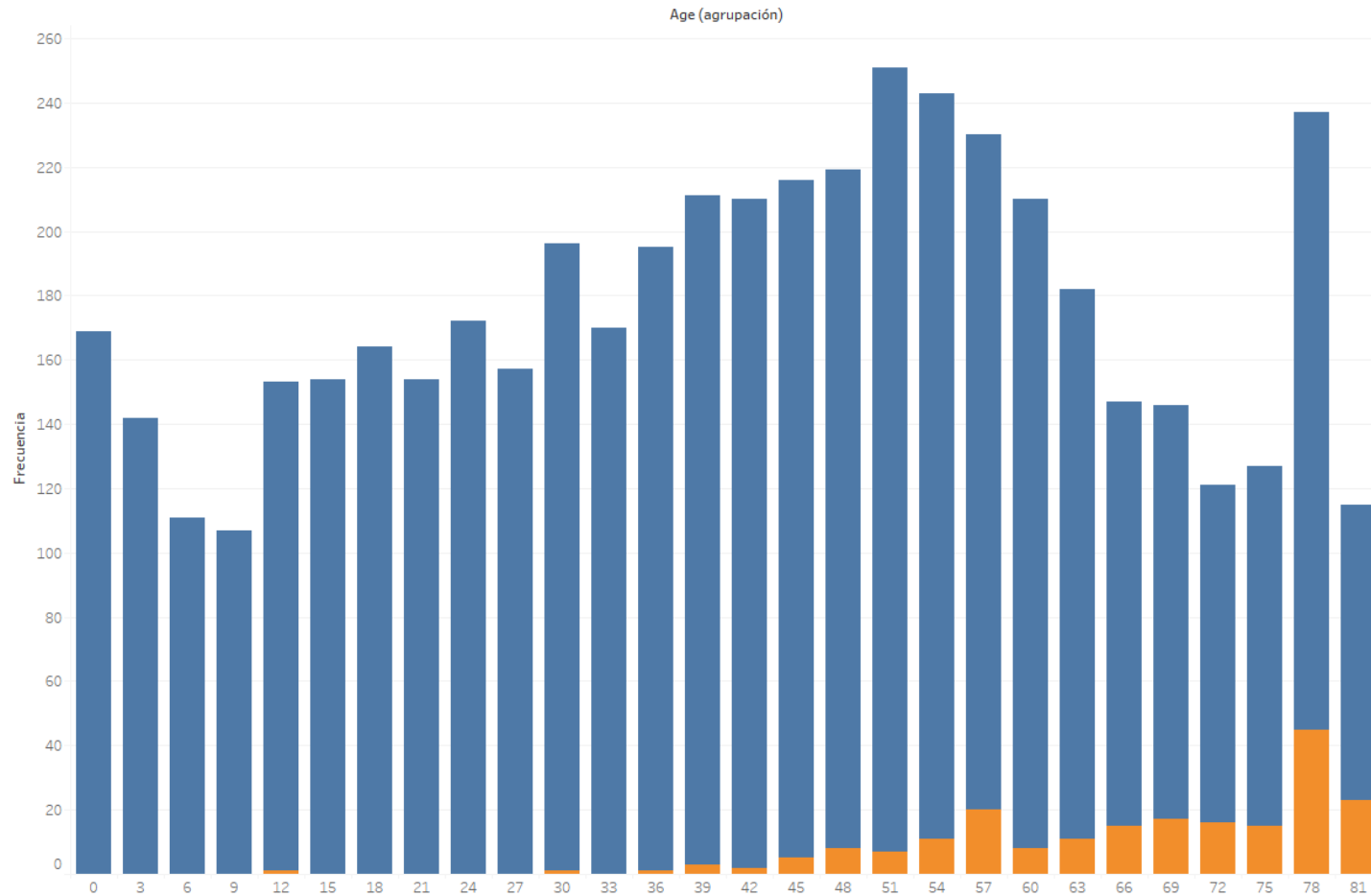


## Matriz de correlación:

A simple vista no se pueden ver fuertes correlaciones entre variables

La relación más relevante: 0.33 entre age-bmi.

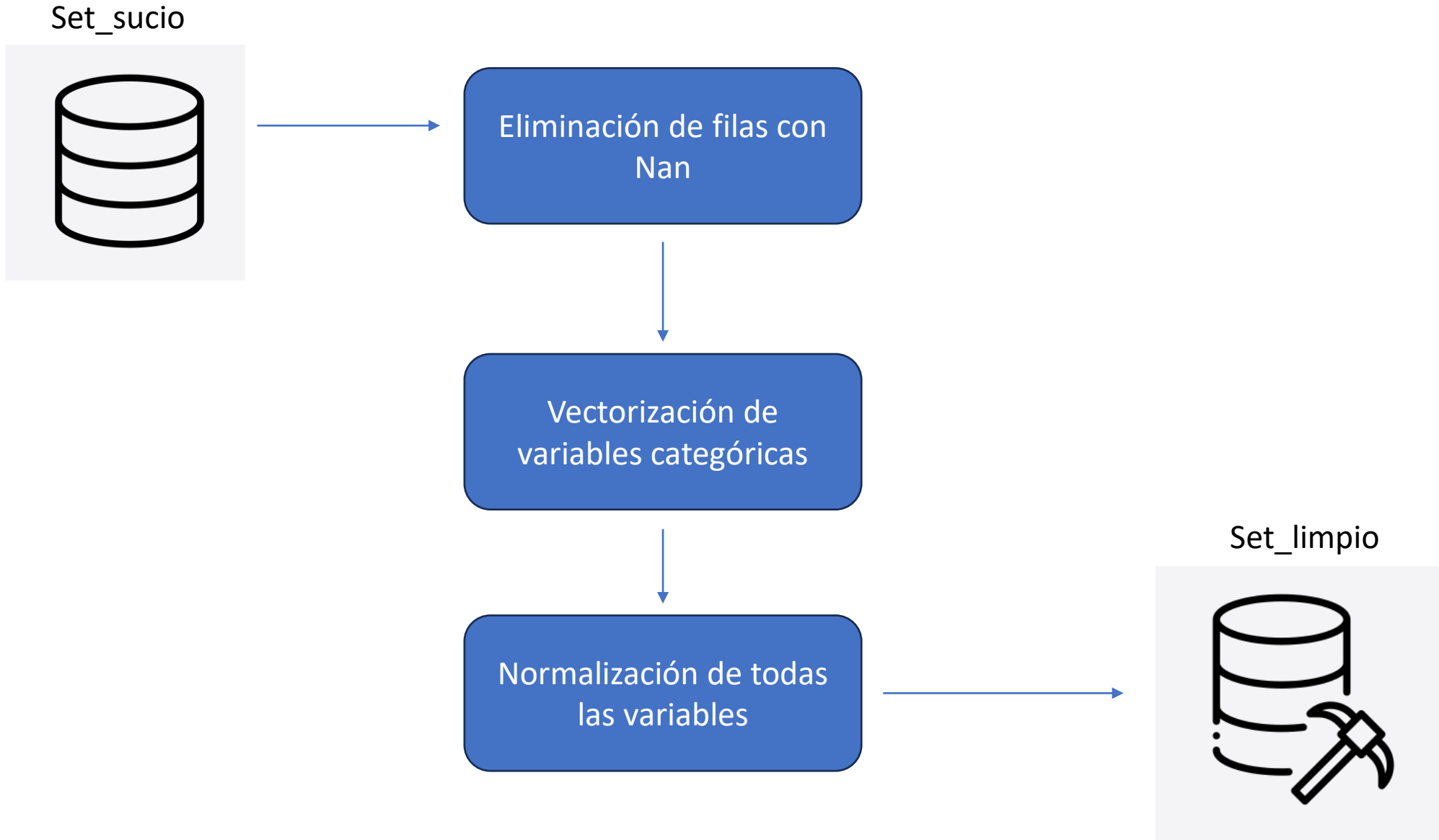
# Exploración visual 02



Sesgo en aquellas personas que mayor edad

azul: Distribución de los pacientes sin riesgo  
naranja: Distribución de pacientes con riesgo

# TAREAS DE PRE PROCESADO



**D A T A S E T**

**Training Dataset**

**Testing Dataset**

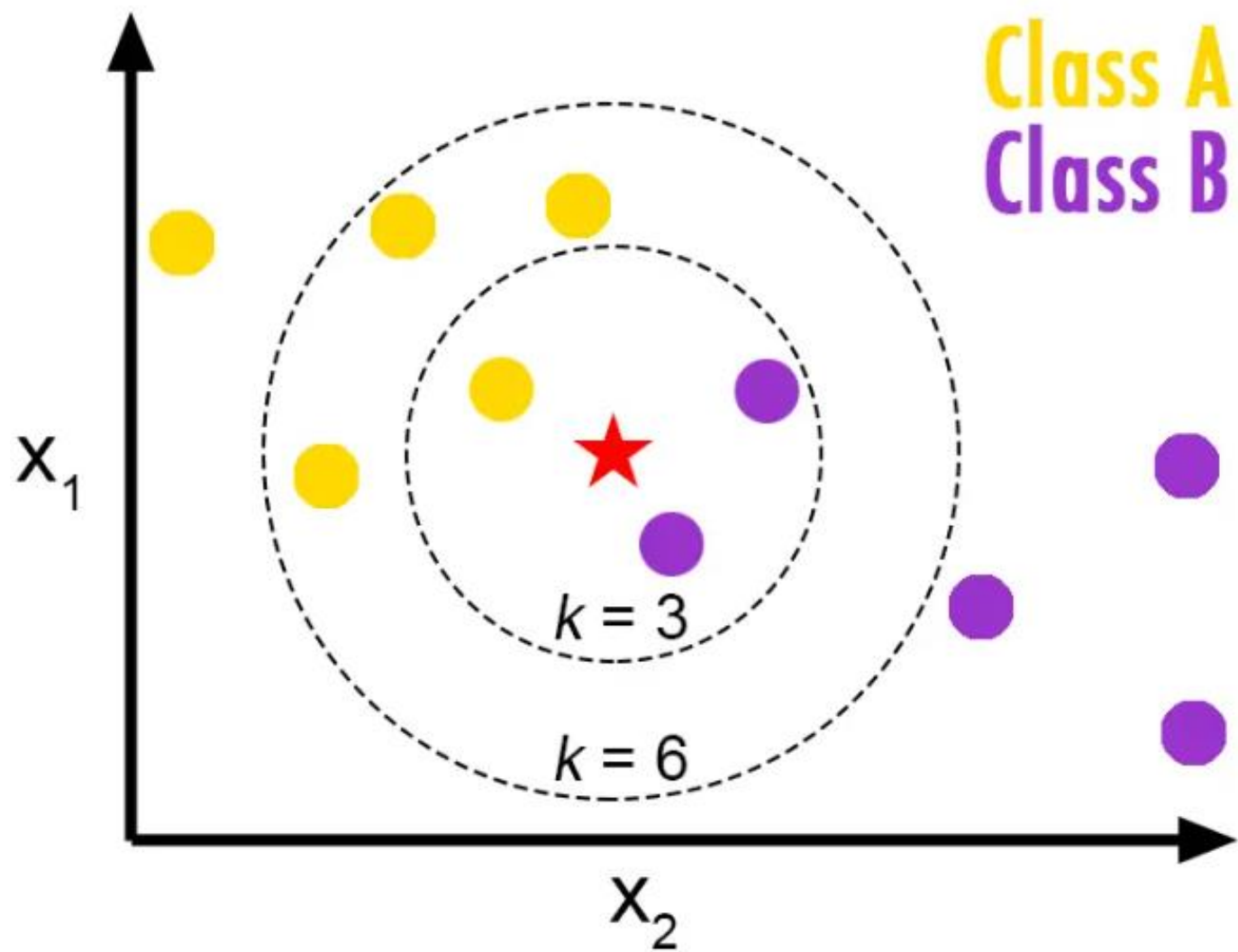


**75%**

**25%**



Modelo a  
utilizar: KNN



# Aplicamos ciclo for y la herramienta SequentialFeatureSelector:

Determinamos las mejores variables para las distintas cantidades (1 hasta 8).

**Mejores 1 características seleccionadas:** ('work\_type\_Private',)

**Mejores 2 características seleccionadas:** ('work\_type\_Private', 'smoking\_status\_formerly smoked')

**Mejores 3 características seleccionadas:** ('hypertension', 'work\_type\_Private', 'smoking\_status\_formerly smoked')

**Mejores 4 características seleccionadas:** ('hypertension', 'gender\_Male', 'work\_type\_Private', 'smoking\_status\_formerly smoked')

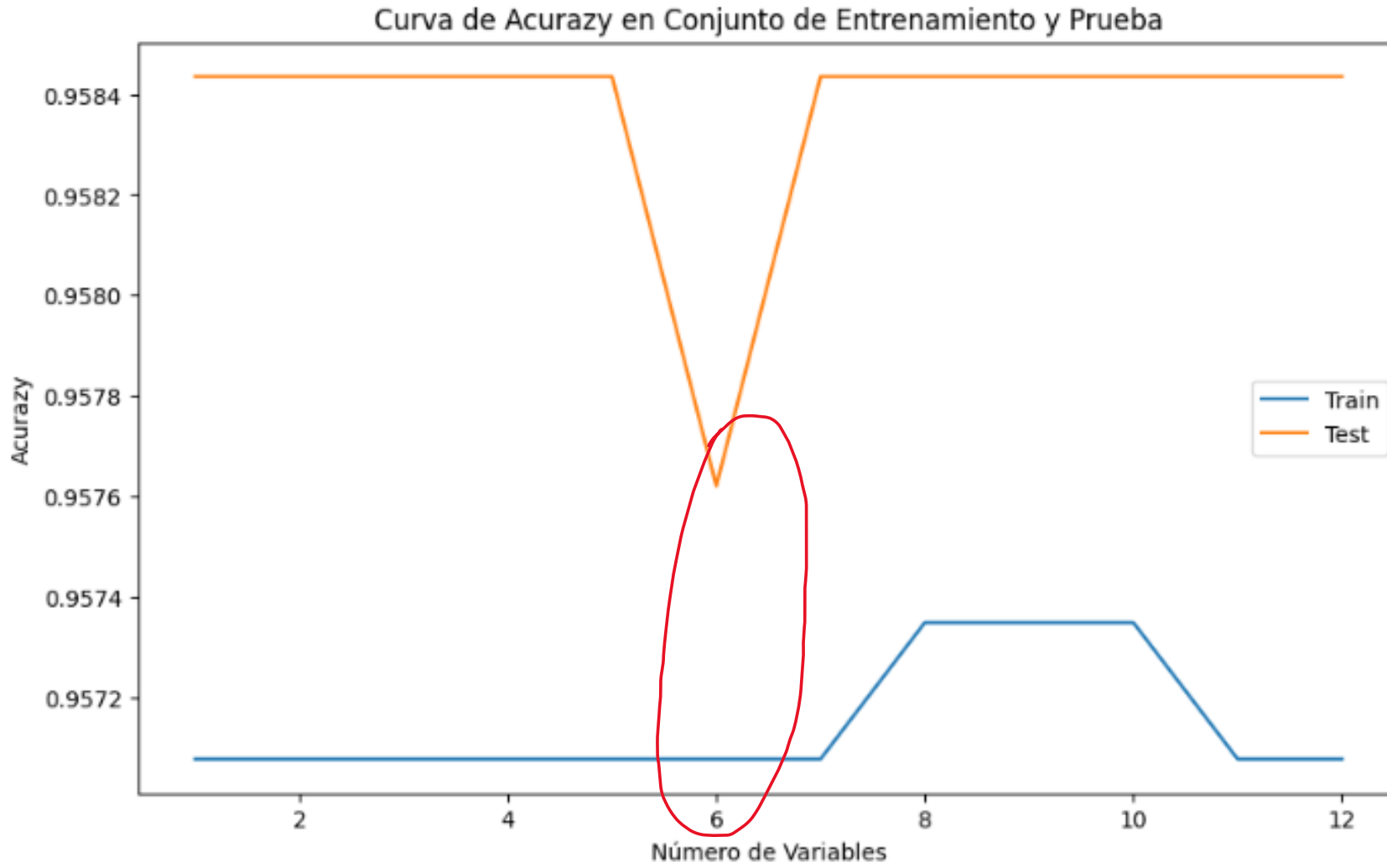
**Mejores 5 características seleccionadas:** ('hypertension', 'gender\_Male', 'ever\_married\_Yes', 'work\_type\_Private', 'smoking\_status\_formerly smoked')

**Mejores 6 características seleccionadas:** ('hypertension', 'bmi', 'gender\_Male', 'ever\_married\_Yes', 'work\_type\_Private', 'smoking\_status\_formerly smoked')

**Mejores 7 características seleccionadas:** ('hypertension', 'bmi', 'gender\_Male', 'ever\_married\_Yes', 'work\_type\_Private', 'work\_type\_Self-employed', 'smoking\_status\_formerly smoked')

**Mejores 8 características seleccionadas:** ('hypertension', 'heart\_disease', 'bmi', 'gender\_Male', 'ever\_married\_Yes', 'work\_type\_Private', 'work\_type\_Self-employed', 'smoking\_status\_formerly smoked')

# PUNTO DE EQUILIBRIO



Determinando los distintos modelos para las distintas cantidades de variables, podemos obtener la curva de train y la curve de test. Con esto podemos encontrar el punto de equilibrio entre over y under.

Como se aprecia en la gráfica, el punto de equilibrio está en 6 variables