Heart Disease

PORTAFOLIOS DE EQUIPO
Caso de estudio 1

Descripción
del Contexto

Enfermedades cardíacas

Principal causa de muerte en todo el mundo

17,5 millones

Muertes anuales

Fuente: OMS

Enfermedades cardíacas

Principales factores de riesgo

Colesterol	Hipertensión	Obesidad	Mala alimentación
Inactividad física	Consumo alcohol	Consumo de tabaco	Familiares que las padezcan

Enfermedades cardíacas

Las 10 más comunes

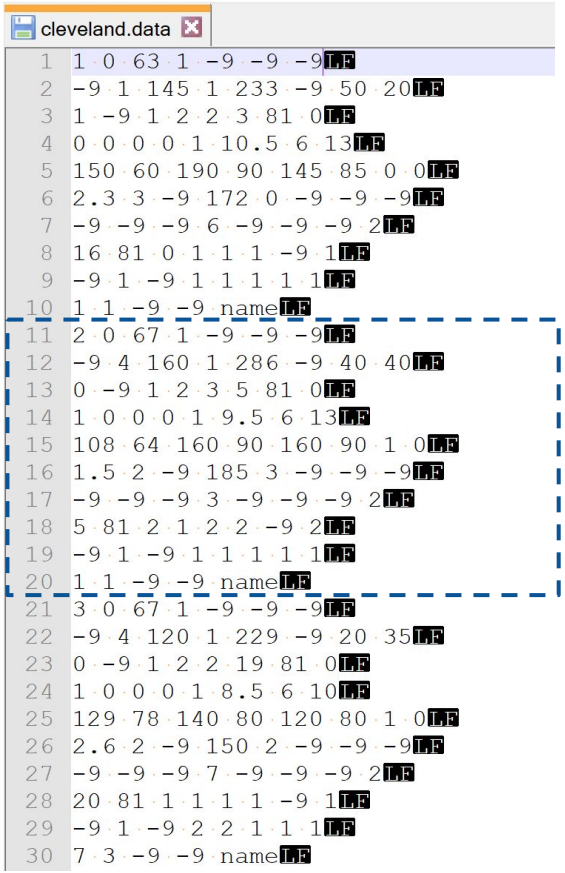
1. Infarto
2. Cardiopatía isquémica
3. Insuficiencia cardíaca
4. Muerte súbita
5. Miocardiopatías
6. Miocardiopatía dilatada
7. Miocardiopatía hipertrófica
8. Miocardiopatía restrictiva
9. Valvulopatías
10. Valvulopatía mitral

Preprocesamiento de los datasets

Preprocesamiento

1

Analizar los datasets e **identificar** cada instancia



Preprocesamiento

2

Agregar **encabezados** de atributos y ordenar **instancias por filas**

cleveland.data

```
ekgday ekgyr dig prop nitr pro diuretic proto thaldur thaltime met thalach thalrest tpeakbps tpeakbpd dummy
-9 -9 -9-9 -9 -9 6 -9 -9 -9 216 81 0 1 1 1 -9 1-9 1 -9 1 1 1 1 11 1 -9 -9 name
-9 -9 -9-9 -9 -9 3 -9 -9 -9 25 81 2 1 2 2 -9 2-9 1 -9 1 1 1 1 11 1 -9 -9 name
-9 -9 -9-9 -9 -9 7 -9 -9 -9 220 81 1 1 1 1 -9 1-9 1 -9 2 2 1 1 17 3 -9 -9 name
 -9 -9-9 -9 -9 3 -9 -9 -9 24 81 0 1 1 1 -9 1-9 1 -9 1 1 1 1 11 1 -9 -9 name
-9-9 -9 -9 3 -9 -9 -9 218 81 0 1 1 1 -9 1-9 1 -9 1 1 1 1 11 1 -9 -9 name
 -9 -9 -9-9 -9 -9 3 -9 -9 -9 310 81 0 1 1 1 -9 1-9 1 -9 1 1 1 1 11 1 -9 -9 name
-9 -9-9 -9 -9 3 -9 -9 -9 22 81 3 1 2 1 -9 1-9 2 -9 2 1 1 1 17 1 -9 -9 name
-9 -9-9 -9 -9 3 -9 -9 -9 721 81 0 1 1 1 -9 1-9 1 -9 1 1 1 1 11 1 -9 -9 name
 -9-9 -9 -9 7 -9 -9 -9 73 81 2 1 2 1 -9 1-9 1 -9 2 1 1 1 67 2 -9 -9 name
-9 -9 -9-9 -9 -9 7 -9 -9 -9 76 81 1 1 1 1 -9 1-9 1 -9 2 1 1 1 11 1 -9 -9 name
 -9 -9 -9-9 -9 -9 6 -9 -9 -9 71 81 0 1 1 1 -9 1-9 1 -9 1 1 1 1 11 1 -9 -9 name
```

Preprocesamiento

3 Importar los datasets a rapidminer

A

cleveland
282 instancias
75 atributos

C

long-beach-va
200 instancias
75 atributos

B

hungarian
294 instancias
75 atributos

D

switzerland
123 instancias
75 atributos

899 instancias, **75** atributos

Preprocesamiento

4 Eliminar atributos por cantidad de **missing values**

- ❌ **pncaden** (100%)
- ❌ **restwm** (96,7%)
- ❌ **exeref** (99,8%)
- ❌ **om2** (63,6%)
- ❌ **dm** (89,4%)
- ❌ **exerwm** (99,4%)
- ❌ **restef** (96,9%)
- ❌ **ramus** (63,1%)
- ❌ **famhist** (46,9%)
- ❌ **ca** (66,7%)
- ❌ **diag** (62,1%)
- ❌ **thal** (53,1%)

---

12 atributos eliminados

Preprocesamiento

5 Eliminar atributos **irrelevantes** (1/2)

- ❌ id
- ❌ name
- ❌ ccf
- ❌ thalach
- ❌ tpeakbpd
- ❌ ekgmo
- ❌ ekgday
- ❌ ekgyr
- ❌ thalrest
- ❌ thaltime
- ❌ dig
- ❌ prop
- ❌ nitr
- ❌ met
- ❌ tpeakbps
- ❌ pro
- ❌ diuretic
- ❌ proto
- ❌ thaldur
- ❌ oldpeak

Preprocesamiento

5 Eliminar atributos **irrelevantes** (2/2)

- ❌ slope
- ❌ rldv5
- ❌ painloc
- ❌ rldv5e
- ❌ restckm
- ❌ cday
- ❌ exerckm
- ❌ cmo
- ❌ cyr

---

29 atributos eliminados

Preprocesamiento

6 Eliminar atributos por **falta de información**

❌ thalsev	❌ earlobe	❌ lvx2	❌ lvx4
❌ thalpul	❌ lvx1	❌ lvx3	❌ lvf
❌ cathef	❌ junk	❌ dummy	❌ xhypo

---

12 atributos eliminados

Preprocesamiento

7 Generación del atributo **smoke_gen**

SI cigs **>** 0 **||** years **>** 0 **||** smoke **==** 1 **1**

SINO SI !missing(cigs) **||** !missing(years) **||** !missing(smoke) **0**

SINO Vecino **más cercano**

❌ cigs ❌ years ❌ smoke ➕ smoke_gen

Atributos **Utilizados**

Total : 20

- ✓ age
- ✓ sex
- ✓ painexer
- ✓ relres
- ✓ cp
- ✓ trestbps
- ✓ htn
- ✓ chol
- ✓ fbs
- ✓ restecg
- ✓ trestbpd
- ✓ exang
- ✓ lmt
- ✓ ladprox
- ✓ laddist
- ✓ cxmain
- ✓ om1
- ✓ rcadist
- ✓ smoke_gen
- ✓ rcaprox

Preprocesamiento

8 Imputación de missing values

K - NN

---

con K = 1

9 Normalización de los datos

Transformación Z

Resultado
de la regresión

Resultado de la regresión

Confianza

87,98%

+/- 2,55%

Preguntas?

Muchas gracias!