

# Ejercicio 1

## Parte 1

a/b) Se setea el atributo “2do\_Ataque\_Corazon” como label ya que es el atributo que queremos predecir.

Row No.	Edad (integer) regular	Estado_civil (integer) regular	Sexo (integer) regular	Categoria_Peso (integer) regular	Colesterol (integer) regular	Manejo_stress (integer) regular	Trat_ansiedad (integer) regular	2do_Ataque_Corazon (polynomial) label
1	60	2	0	1	150	1	50	Si
2	69	2	1	1	170	0	60	Si
3	52	1	0	0	174	1	35	No
4	66	2	1	1	169	0	60	Si
5	70	3	0	1	237	0	65	Si
6	52	1	0	0	174	1	35	No
7	58	2	1	0	140	0	45	No
8	59	2	1	0	143	0	45	Si
9	60	2	0	0	139	0	45	No
10	51	1	1	0	174	1	40	No

c) Se verifica que son 2 clases Si/No

2do_Ataque_Corazon	Nominal	0	Least Si (68)	Most No (70)	Values No (70), Si (68)
--------------------	---------	---	------------------	-----------------	----------------------------

d) Se renombra el nodo a entrenamiento



## Parte 2

Row No.	Edad (integer) regular	Estado_civil (integer) regular	Sexo (integer) regular	Categoria_Peso (integer) regular	Colesterol (integer) regular	Manejo_stress (integer) regular	Trat_ansiedad (integer) regular
1	61	0	1	1	139	1	50
2	55	2	1	2	163	0	40
3	53	1	1	1	172	0	55
4	58	1	1	2	206	0	70

a) Porque los datos del training son integer.

## Parte 3

a) Rangos de evaluación

ExampleSet (Evaluación)		ExampleSet (Entrenamiento)				
Name	Type	Missing	Statistics			Filter
✓ Edad	Integer	0	Min 42	Max 81	Average 62.932	
✓ Estado_civil	Integer	0	Min 0	Max 3	Average 1.696	
✓ Sexo	Integer	0	Min 0	Max 1	Average 0.623	
✓ Categoria_Peso	Integer	0	Min 0	Max 2	Average 0.920	
✓ Colesterol	Integer	0	Min 122	Max 239	Average 178.265	
✓ Manejo_stress	Integer	0	Min 0	Max 1	Average 0.457	
✓ Trat_ansiedad	Integer	0	Min 35	Max 80	Average 55.435	


# Rangos de entrenamiento












ExampleSet (Evaluación)		ExampleSet (Entrenamiento)				
Name	Type	Missing	Statistics			Filter
Label ✓ 2do_Ataque_Corazon	Nominal	0	Least Si (68)	Most No (70)	Values No (70), Si (68)	
✓ Edad	Integer	0	Min 42	Max 81	Average 62.978	
✓ Estado_civil	Integer	0	Min 0	Max 3	Average 1.696	
✓ Sexo	Integer	0	Min 0	Max 1	Average 0.623	
✓ Categoria_Peso	Integer	0	Min 0	Max 2	Average 0.920	
✓ Colesterol	Integer	0	Min 122	Max 239	Average 177.391	
✓ Manejo_stress	Integer	0	Min 0	Max 1	Average 0.442	
✓ Trat_ansiedad	Integer	0	Min 35	Max 80	Average 55.435	

- Los rangos entre los datasets son iguales como se corrobora en las imágenes
- b) Si están. Necesitamos verificar para ver si tenemos que rellenar con algún otro atributo.
  - c) Podríamos detectar outliers

## Ejercicio 2

**Parameters** ×

 **Logistic Regression**

<b>solver</b>	AUTO	
<input type="checkbox"/> <b>reproducible</b>		
<input type="checkbox"/> <b>use regularization</b>		
<input checked="" type="checkbox"/> <b>standardize</b>		
<input type="checkbox"/> <b>non-negative coefficients</b>		
<input checked="" type="checkbox"/> <b>add intercept</b>		
<input checked="" type="checkbox"/> <b>compute p-values</b>		
<input checked="" type="checkbox"/> <b>remove collinear columns</b>		
<b>missing values handling</b>	MeanImputation	
<b>max iterations</b>	0	
<b>max runtime seconds</b>	0	

C-

solver: Selecciona el “solver” a utilizar

reproducible: Determina si el modelo es reproducible. Si está fijado, se puede definir el nivel de paralelismo manualmente.

use regularization: Determina si se desea utilizar regularización.

standardize: Estandariza las columnas numéricas antes del proceso mediante Z-index.

non-negative coefficients: Impide que hayan coeficientes negativos.

add intercept

compute p-values

remove linear columns

missing values handling

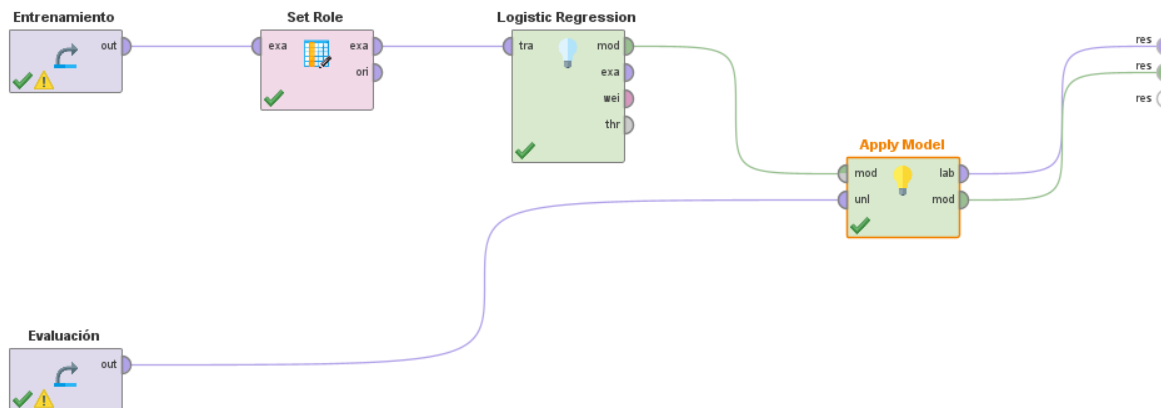
max iterations

max runtime seconds

#### 4. Coeficientes del modelo

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Edad	-0.119	-0.937	0.078	-1.528	0.127
Estado_civil	-1.278	-1.055	0.531	-2.409	0.016
Sexo	-0.215	-0.104	0.851	-0.253	0.801
Categoria_Peso	-4.056	-3.102	0.976	-4.154	0.000
Colesterol	-0.009	-0.279	0.015	-0.587	0.557
Manejo_stress	0.071	0.035	0.949	0.075	0.941
Trat_ansiedad	0.054	0.663	0.065	0.818	0.413
Intercept	12.419	0.342	4.690	2.648	0.008

#### Proceso generado



#### 8. Estadísticas de aplicar el modelo

Row No.	prediction(2...	confidence(Si)	confidence(No)	Edad	Estado_civil	Sexo	Categoria_P...	Colesterol	Manejo_stre...	Trat_ansied...
1	No	0.082	0.918	61	0	1	1	139	1	50
2	Si	0.986	0.014	55	2	1	2	163	0	40
3	No	0.118	0.882	53	1	1	1	172	0	55
4	Si	0.894	0.106	58	1	1	2	206	0	70
5	Si	0.582	0.418	62	2	1	1	148	1	50
6	No	0.011	0.989	70	1	0	0	172	0	60
7	No	0.004	0.996	52	1	0	0	171	1	35
8	No	0.086	0.914	50	1	1	1	172	0	55
9	Si	0.662	0.338	67	2	1	1	172	0	60
10	No	0.312	0.688	62	1	1	1	166	1	50
11	Si	0.993	0.007	66	2	1	2	220	0	60
12	No	0.015	0.985	56	2	1	0	141	0	45
13	Si	0.992	0.008	77	2	1	2	181	1	80
14	Si	0.582	0.418	64	2	1	1	174	0	60
15	Si	0.713	0.287	67	2	1	1	146	1	50
16	Si	0.834	0.166	62	3	1	1	171	0	55
17	Si	0.797	0.203	68	2	1	1	239	0	60
18	No	0.001	0.999	48	1	0	0	175	0	60
19	No	0.007	0.993	67	1	0	0	168	0	60

Se generan los atributos:

- Confidence(Si)
- Confidence(No)
- Prediction(2do\_Ataque\_Corazon)

Prediction prediction(2do_Ataque_Corazo...	Polynomial	0	Least No (340)	Most Si (350)	Values Si (350), No (340)
Confidence_Si confidence(Si)	Real	0	Min 0.001	Max 0.998	Average 0.493
Confidence_No confidence(No)	Real	0	Min 0.002	Max 0.999	Average 0.507

## Ejercicio 3

### Parte 1

La fila para el paciente indica con 91.8% de seguridad que no tendrá un segundo ataque al corazón por lo que el Doctor García puede estar seguro

Row No.	prediction(2...	confidence(Si)	confidence(No)	Edad	Estado_civil	Sexo	Categoria_Peso	Colesterol	Manejo_stre...	Trat_ansied...
1	No	0.082	0.918	61	0	1	1	139	1	50

### Parte 2

Para la fila en cuestión la decisión del Doctor García debería ser tomar acciones y tratar al paciente pues tiene grandes posibilidades de sufrir un segundo ataque (99.3%)

Row No.	prediction(2...	confidence(Si)	confidence(No)	Edad	Estado_civil	Sexo	Categoria_...	Colesterol	Manejo_str...	Trat
11	Si	0.993	0.007	66	2	1	2	220	0	60

### Parte 3

#### ¿Cómo usar la predicción en un caso de consultoría o desarrollo de un sistema real?

Debe indicarse la confianza para que el doctor en cuestión lo tenga en cuenta, y no solo un Si/No ya que esto puede inducir a falsos negativos y positivos.

#### ¿Cuántos pacientes tienen predicción de ataque cardíaco?

350 Pacientes tienen predicción positiva usando 0.5 como parámetro diferenciador de confianza.

Prediction prediction(2do_Ataque_Corazo...	Polynomial	0	Least No (340)	Most Si (350)	Values Si (350), No (340)
---	------------	---	-------------------	------------------	------------------------------

#### Tener en cuenta los niveles de confianza

Variando los niveles de confianza se pueden obtener distintos resultados. Se recomienda tener un umbral a favor de asumir un positivo, ya que el costo de un falso negativo es bastante superior al de un falso positivo.

#### ¿Cómo podríamos en RM analizar la performance global del modelo?

Podríamos destinar cierta parte del dataset a testing, y utilizar cross validation.

## Ejercicio 4

6. Analizar cómo se utiliza el método de Logistic Regression y documentarlo (alternativas, parámetros). Entregar el documento.

La clase `LogisticRegression` proviene del módulo `sklearn.linear_model`.

Su constructor tiene los siguientes parámetros:

- `penalty`
- `dual`
- `tol`
- `C`
- `fit_intercept`
- `intercept_scaling`
- `class_weight`
- `random_state`
- `solver`
- `max_iter`: Máximo número de iteraciones que tiene el algoritmo para terminar de entrenar la IA.
- `multi_class`
- `verbose`
- `warm_start`
- `n_jobs`
- `l1_ratio`

Para setear el solver utilizamos el siguiente parametro:

```
solver : {'lbfgs', 'liblinear', 'newton-cg', 'newton-cholesky', 'sag', 'saga'}, default='lbfgs'
```

## Atributos

Nombre	Descripción	Tipo
Age	Edad	numeric
Job	Tipo de trabajo. Puede ser "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student",  "blue-collar", "self-employed", "retired", "technician", "services")	categorical
marital	Estado Civil. Puede ser "married", "divorced", "single"; note: "divorced" means divorced or widowed	categorical
education	educación. Puede ser "unknown", "secondary", "primary", "tertiary"	categorical
default	Tiene crédito en default. Si o no	binary
balance	Promedio balance anual en euros	Numeric
housing	Tiene un préstamo de vivienda. Si o no	Binary
loan	Tiene un préstamo personal	Binary
contact	tipo de comunicación de contacto. Puede ser "unknown", "telephone", "cellular"	Categorico
day	último día de contacto del mes	Numeric
month	último mes de contacto del mes	Categorico
duration	Duracion de ultimo contacto en segundos	Numeric
campaign	número de contactos realizados durante esta campaña y para este cliente	Numérico

pdays	número de días que pasaron después de que el cliente fue contactado por última vez desde una campaña anterior. -1 significa que no fue contactado previamente	Numérico
previous	Número de contactos antes de esta campaña	Numeric
poutcome	Resultado de la campaña de marketing previa	Categorico: "unknown", "other", "failure", "success"
emp_var_rate	Variacion del empleo con una frecuencia cuatrimestral	Real
cons_price_idx	monthly average consumer price index	Real
cons_conf_idx	monthly average consumer confidence index	Real
euribor3m	daily three month Euribor rate	Real
nr_employed	quarterly average of the total number of employed	Real

La variable objetivo es y que es si el cliente se ha suscrito a un deposito a plazo. Es binario (si o no).