

PREDECIR ENFERMEDADES DEL CORAZÓN EN SUDÁFRICA

ANGELICA MARIA CARDENAS MARTINEZ

CORPORACIÓN UNIVERSITARIA IBEROAMERICANA Y
BOGOTÁ INSTITUTE OF TECHNOLOGY

28/10/2021



TABLA DE CONTENIDOS



1. INTRODUCCION -----.
2. DESCRIPCIÓN DEL CASO -----.
3. OBJETIVOS DEL MODELO -----.
4. DESCRIPCIÓN DE LOS DATOS -----.
5. HALLAZGOS ENCONTRADOS POR EL EDA -----.
6. ALGORITMO ELEGIDO -----.
7. METRICAS DEL MODELO -----.
8. FUTURAS LINEAS -----.
9. CONCLUSIONES -----.

DESCRIPCIÓN DEL CASO

En una región de sudáfrica, los hombres de esa comunidad tienden a tener un alto riesgo de padecer enfermedades cardiacas y por este motivo se tomaron ciertos datos médicos, y con informacion se busca predecir si es posible de padecer la enfermedad o no .



OBJETIVOS DEL MODELO

Con este modelo de clasificación buscamos predecir si existe alguna enfermedad de corazón o no, para ello unos de los algoritmos que vamos a utilizar son máquinas de vector de soporte conocido como SVM.



DESCRIPCIÓN DE LOS DATOS

TEMATICA:

Una muestra retrospectiva de hombres en una región de alto riesgo de enfermedades cardíacas de el Cabo Occidental, Sudáfrica. Hay aproximadamente dos controles por caso de CHD. Muchos de los hombres con enfermedad coronaria positiva se han sometido a un tratamiento de reducción de la presión arterial y otros programas para reducir sus factores de riesgo después de su evento de enfermedad coronaria. En algunos casos, las mediciones se realizaron después de estos tratamientos. Estos datos se toman de un conjunto de datos más grande, descrito en Rousseauw et al, 1983, South African Medical Journal

VARIABLES:

Pasis : Es la presión arterial sistólica

Tabaco: La acumulación en Kg

Pbcoles: La proteína baja del colesterol

Adiposidad: sobre peso severo o morbido

Familia: Antecedentes familiares presente o ausente

Tipo: Comportamiento tipo-A


Obesidad

Alcohol: Consumo recurrente de alcohol

Edad

Efcad: Enfermedades cardíacas



 `data.head(10)`

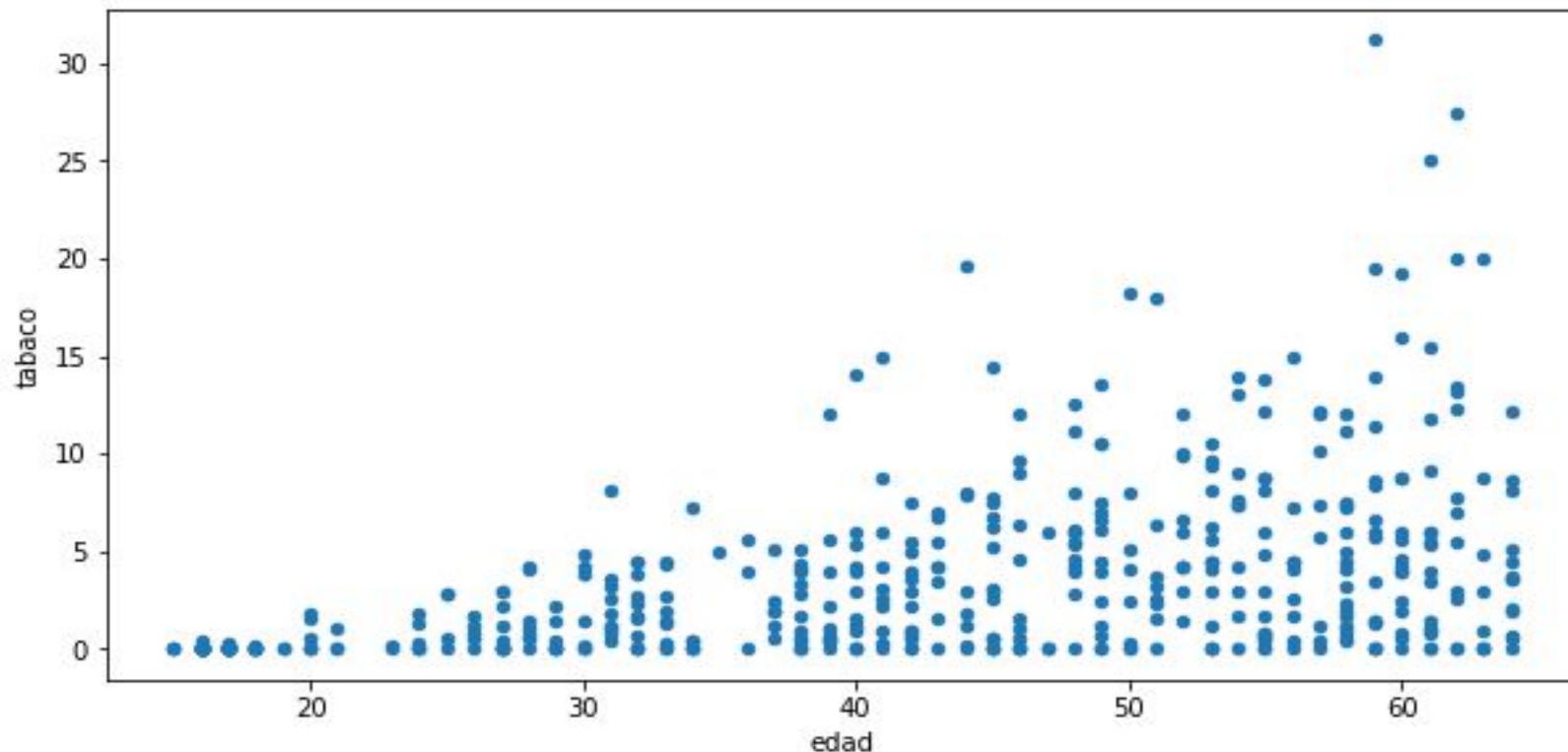
	pasis	tabaco	pbcoles	adiposidad	familia	tipo	obesidad	alcohol	edad	enfcard
0	160	12.00	5.73	23.11	1	49	25.30	97.20	52	2
1	144	0.01	4.41	28.61	2	55	28.87	2.06	63	2
2	118	0.08	3.48	32.28	1	52	29.14	3.81	46	1
3	170	7.50	6.41	38.03	1	51	31.99	24.26	58	2
4	134	13.60	3.50	27.78	1	60	25.99	57.34	49	2
5	132	6.20	6.47	36.21	1	62	30.77	14.14	45	1
6	142	4.05	3.38	16.20	2	59	20.81	2.62	38	1
7	114	4.08	4.59	14.60	1	62	23.11	6.72	58	2
8	114	0.00	3.83	19.40	1	49	24.86	2.49	29	1
9	132	0.00	5.80	30.96	1	69	30.11	0.00	53	2

HALLAZGOS ENCONTRADOS POR EL EDA

✓
0s

```
[12] data.plot(x='edad',y='tabaco',kind='scatter',figsize=(10,5))
```

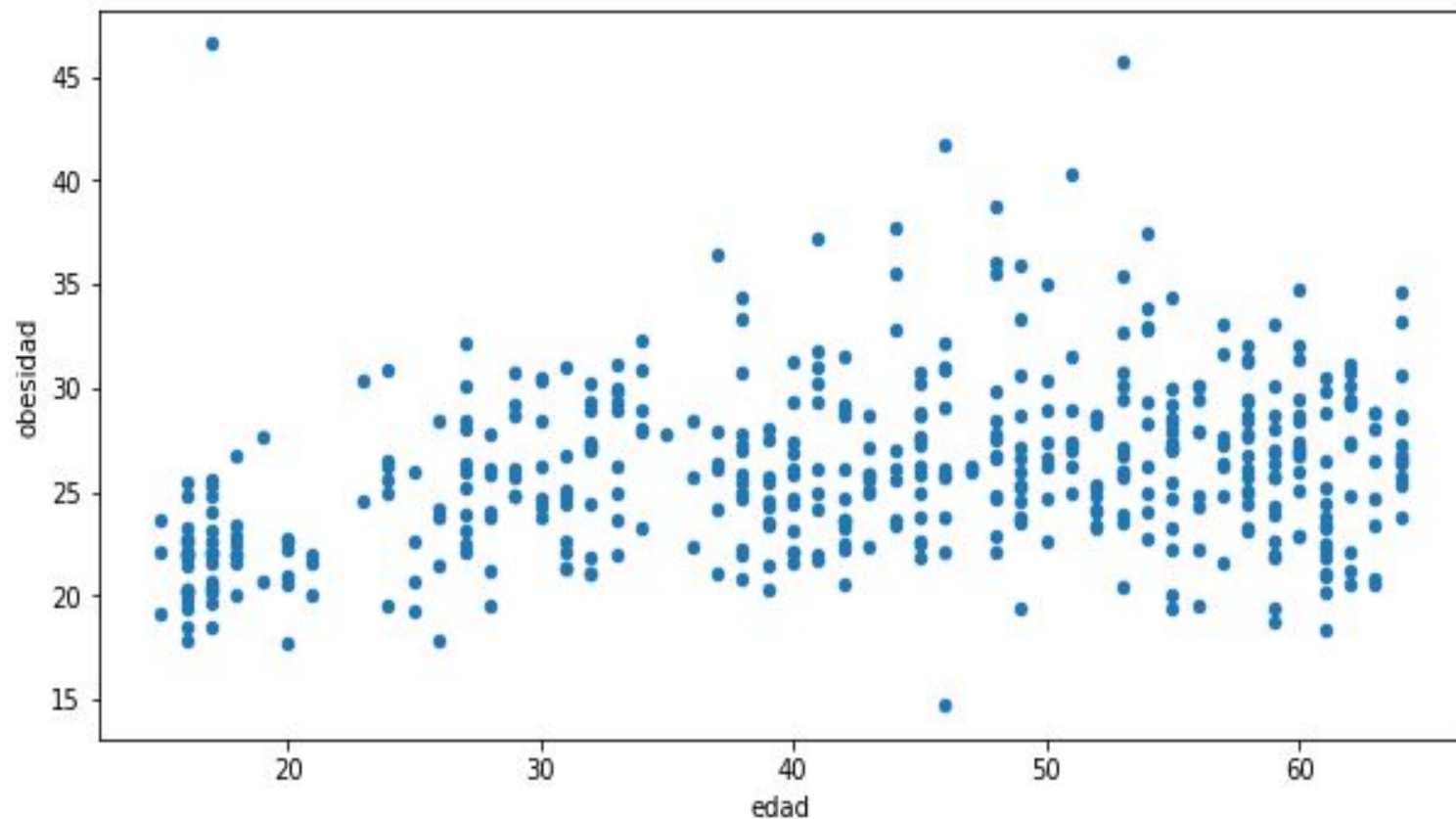
<matplotlib.axes._subplots.AxesSubplot at 0x7f5452f40850>





```
data.plot(x='edad',y='obesidad',kind='scatter',figsize=(10,5))
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f54527dfa10>



data.describe()

	pasis	tabaco	pbcoles	adiposidad	familia	tipo	obesidad	alcohol	edad	enfcard
count	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000	462.000000
mean	138.326840	3.635649	4.740325	25.406732	1.584416	53.103896	26.044113	17.044394	42.816017	1.346320
std	20.496317	4.593024	2.070909	7.780699	0.493357	9.817534	4.213680	24.481059	14.608956	0.476313
min	101.000000	0.000000	0.980000	6.740000	1.000000	13.000000	14.700000	0.000000	15.000000	1.000000
25%	124.000000	0.052500	3.282500	19.775000	1.000000	47.000000	22.985000	0.510000	31.000000	1.000000
50%	134.000000	2.000000	4.340000	26.115000	2.000000	53.000000	25.805000	7.510000	45.000000	1.000000
75%	148.000000	5.500000	5.790000	31.227500	2.000000	60.000000	28.497500	23.892500	55.000000	2.000000
max	218.000000	31.200000	15.330000	42.490000	2.000000	78.000000	46.580000	147.190000	64.000000	2.000000



0 s



```
data['adiposidad'].describe()
```

```
count      462.000000
```

```
mean       25.406732
```

```
std         7.780699
```

```
min         6.740000
```

```
25%        19.775000
```

```
50%        26.115000
```

```
75%        31.227500
```

```
max         42.490000
```

```
Name: adiposidad, dtype: float64
```

ALGORITMO ELEGIDO

El algoritmo a utilizar, que como se indicó anteriormente será el de Máquinas Vectores de Soporte, junto a un kernel lineal y las librerías que utilizare son las siguientes:

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score,
precision_score
```



MÉTRICAS DE DESEMPEÑO DEL MODELOS



- Matriz de confusión o error
- Precisión
- exactitud

FUTURAS LINEAS

Se podría hacer unas mejoras al cálculo del modelo, puede ser probar con varios algoritmos de clasificación y ver si se mejora los datos obtenidos o también seleccionando las características más significativa y verificar los resultados.



CONCLUSIONES

Los valores obtenidos es el de 67 datos correctos y 26 datos incorrectos, con esta información podemos decir que el modelo no acertó correctamente un gran número de datos



MUCHAS GRACIAS