
Insuficiencia cardíaca

Abel José Sánchez Alba y Arturo Cortés
Sánchez



1. Índice

- Descripción del dataset
- Exploración de variables
- Algoritmos de clasificación
- Clustering como preprocesamiento
- Reglas de asociación
- Conclusiones

Dataset

Contiene datos de los pacientes con insuficiencia cardíaca que fueron admitidos en el Instituto de Cardiología y el Hospital Alidado en Faisalabad, Pakistan durante abril-diciembre (2015).

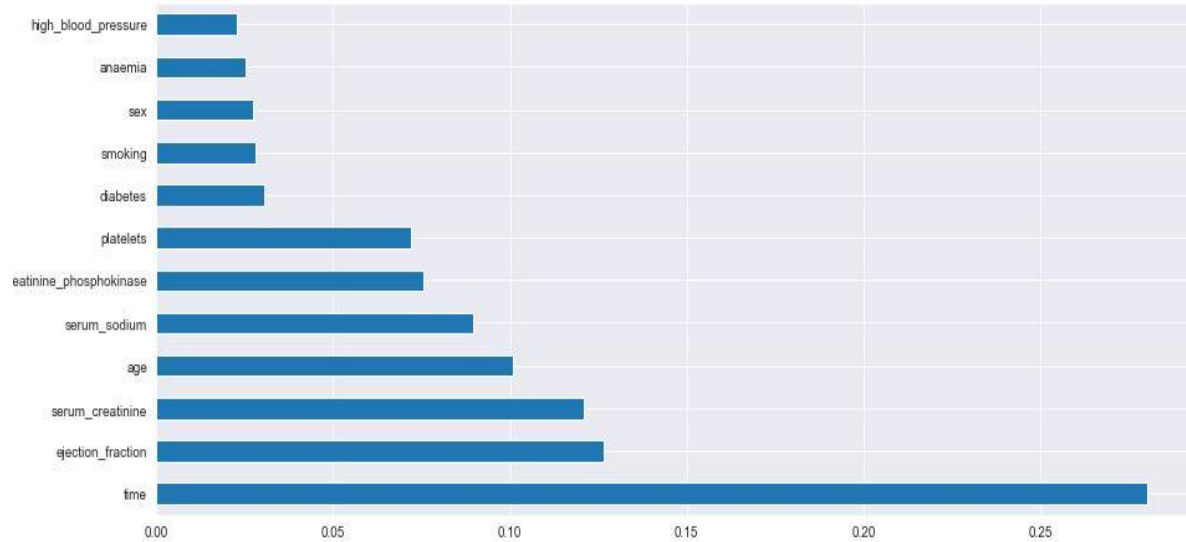
Dataset

- age
- anemia
- creatinine_phosphokinase
- diabetes
- ejection_fraction
- high_blood_pressure
- platelets
- serum_creatinine
- serum_sodium
- sex
- smoking
- time
- DEATH_EVENT

— Exploración de variables

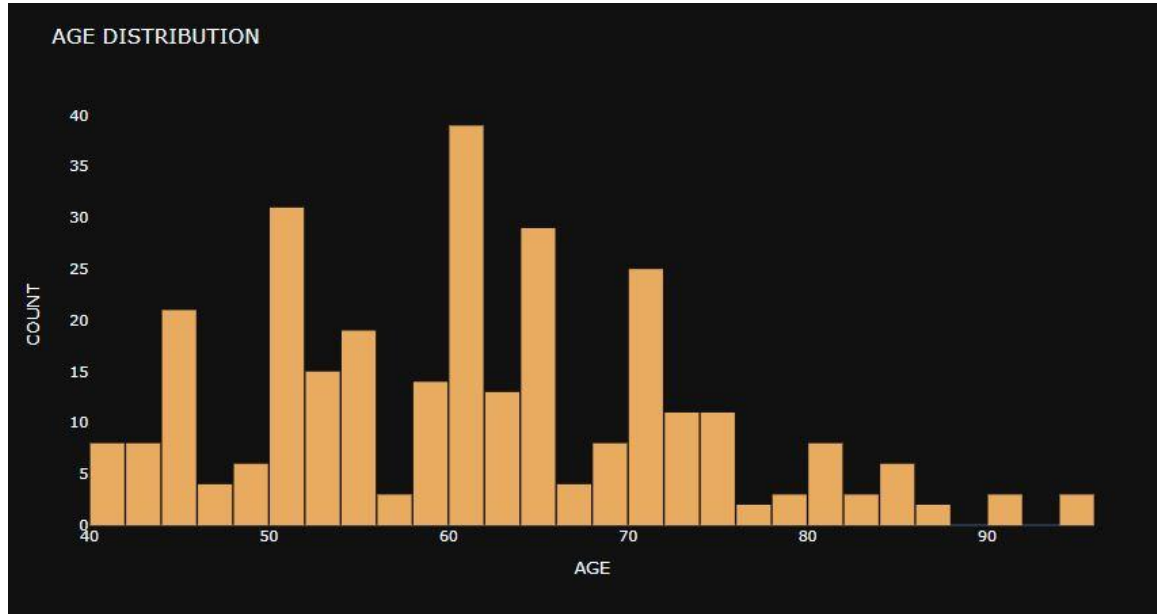
- Variable Indicadora DEATH_EVENT
- Variables relativas a análisis de sangre
- Variables binarias
- Variables normales

Extra Trees



A partir de la variable indicadora de las plaquetas de una persona, estas aportan más información al algoritmo que variables como la presión arterial o el sexo.

Distribución de edad

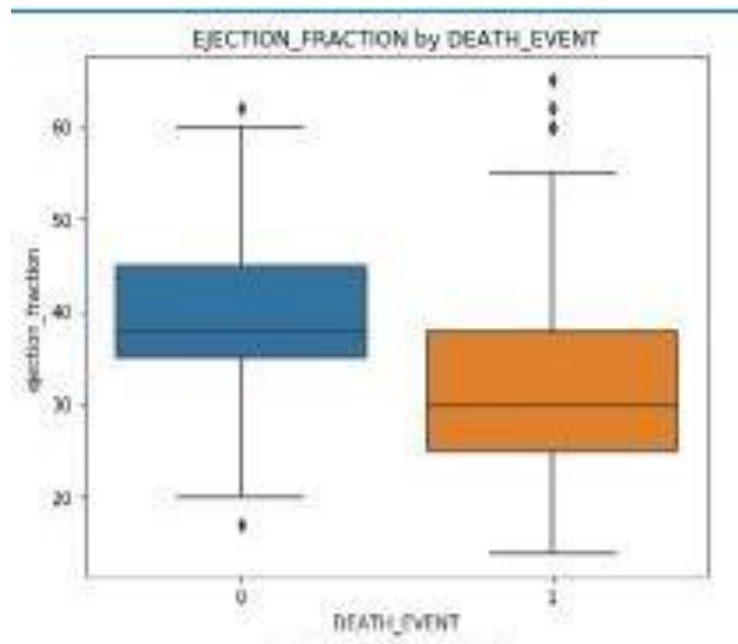
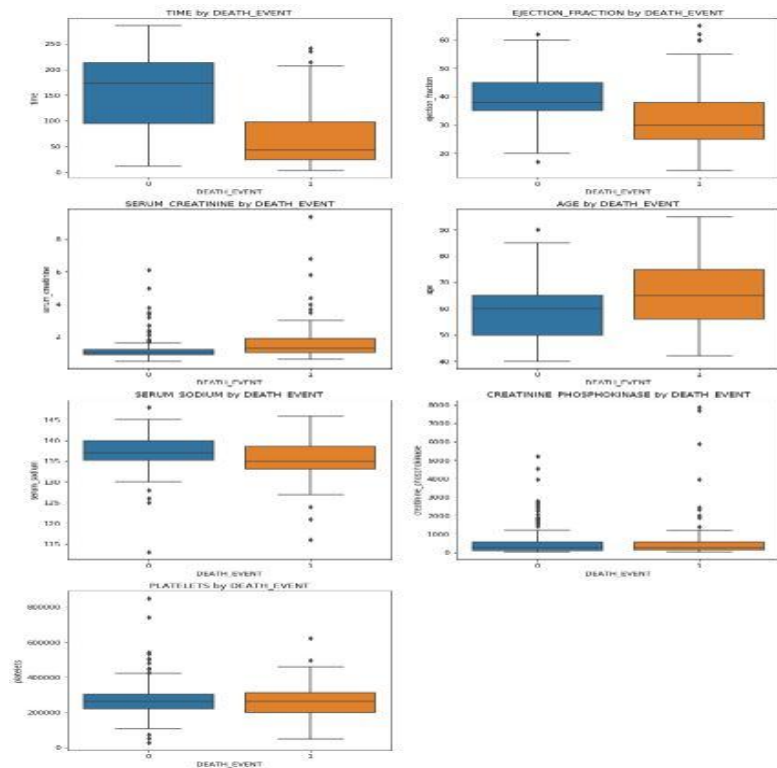


Selección

Se tendrá en cuenta el conjunto de variables que consta de:

1. Plaquetas
2. creatinine_phosphokinase
3. serum_sodium
4. edad
5. serum creatinine
6. ejection fraction
7. time

BoxPlots

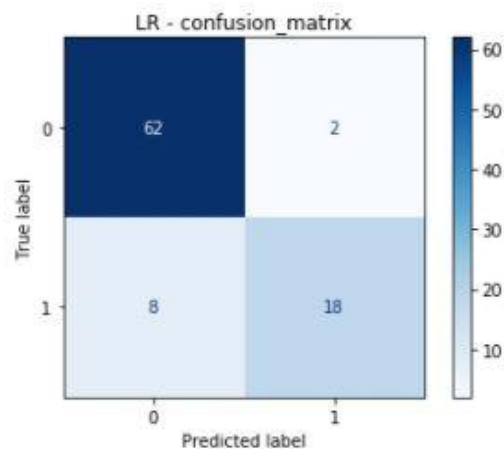
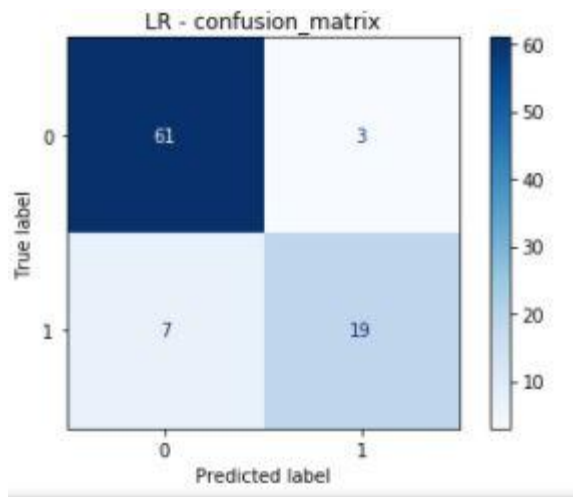


Algoritmos de clasificación Estudiados

Algoritmos basados en:

1. Árboles, random forest
2. Gradient Boost, Extra trees
3. Naïve-Bayes
4. LDA, QDA
5. KNN

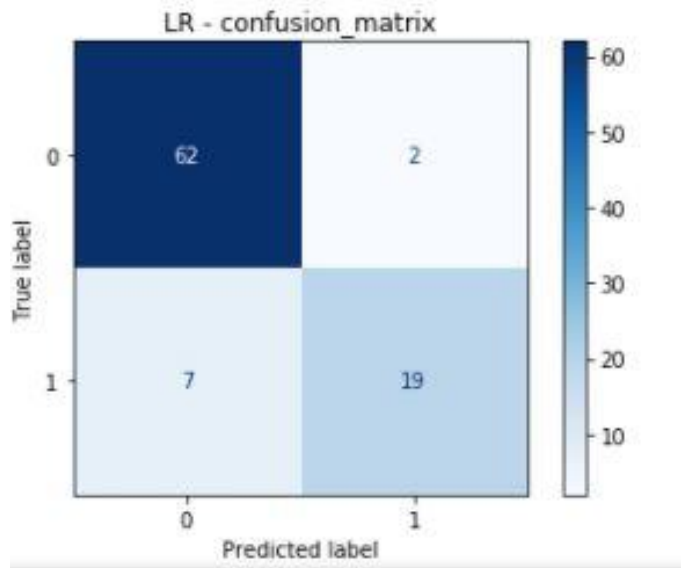
Resultados de los algoritmos



```
{'time': 0.3932150228678265, 'ejection_fraction':  
0.16295587180998997, 'serum_creatinine': 0.18757721131664665,  
'age': 0.1085802869865146, 'serum_sodium': 0.06858438342128573,  
'creatinine_phosphokinase': 0.044669383573307596, 'platelets':  
0.034417840024428936}
```

Clustering como Preprocesamiento

Se ha ejecutado el algoritmo Kmeans como preprocesamiento del conjunto de datos



```
{'time': 0.4852948824630747, 'ejection_fraction':  
0.13216178583589114, 'serum_creatinine':  
0.0888216901567347, 'age': 0.07905896267487082,  
'serum_sodium': 0.11133771588231858,  
'creatinine_phosphokinase': 0.05241916588566074,  
'platelets': 0.03641304347826086, 'cluster':  
0.014492753623188404}
```

Reglas de asociación

Algoritmos usados:

- Apriori
 - Hotspot
 - PredictiveApriori
 - FilteredAsociator
 - Tetrius
 - Association rule learner
-

Reglas de asociación

serum_creatinine=05-139 time=228-259 28 ==> DEATH_EVENT=no 28 acc:(0.99354)
ejection_fraction=33-40 time=228-259 25 ==> DEATH_EVENT=no 25 acc:(0.9932)
serum_creatinine=05-139 time=172-200 23 ==> DEATH_EVENT=no 23 acc:(0.9929)
creatinine_phosphokinase=23-806 serum_creatinine=05-139 time=200-228 23 ==> DEATH_EVENT=no 23 acc:(0.9929)
age=50-55 creatinine_phosphokinase=23-806 high_blood_pressure=no 22 ==> DEATH_EVENT=no 22 acc:(0.99272)
age=50-55 creatinine_phosphokinase=23-806 serum_creatinine=05-139 22 ==> DEATH_EVENT=no 22 acc:(0.99272)
anaemia=no creatinine_phosphokinase=23-806 time=228-259 21 ==> DEATH_EVENT=no 21 acc:(0.99252)
age=50-55 creatinine_phosphokinase=23-806 sex=hombre 20 ==> DEATH_EVENT=no 20 acc:(0.99228)
age=60-65 ejection_fraction=33-40 19 ==> DEATH_EVENT=no 19 acc:(0.99202)
age=50-55 anaemia=no creatinine_phosphokinase=23-806 18 ==> DEATH_EVENT=no 18 acc:(0.99171)
ejection_fraction=33-40 time=200-228 16 ==> DEATH_EVENT=no 16 acc:(0.99094)
anaemia=no high_blood_pressure=no smoking=no time=228-259 16 ==> DEATH_EVENT=no 16 acc:(0.99094)
platelets=190080-272570 time=200-228 15 ==> DEATH_EVENT=no 15 acc:(0.99044)
platelets=190080-272570 time=4-32 15 ==> DEATH_EVENT=si 15 acc:(0.99044)
age=50-55 creatinine_phosphokinase=23-806 platelets=190080-272570 14 ==> DEATH_EVENT=no 14 acc:(0.98985)
anaemia=si time=200-228 13 ==> DEATH_EVENT=no 13 acc:(0.98913)
creatinine_phosphokinase=23-806 ejection_fraction=33-40 time=4-32 13 ==> DEATH_EVENT=si 13 acc:(0.98913)
ejection_fraction=33-40 time=172-200 13 ==> DEATH_EVENT=no 13 acc:(0.98913)
age=40-45 anaemia=no platelets=190080-272570 11 ==> DEATH_EVENT=no 11 acc:(0.98714)
age=50-55 creatinine_phosphokinase=23-806 smoking=si 11 ==> DEATH_EVENT=no 11 acc:(0.98714)
diabetes=no time=116-144 10 ==> DEATH_EVENT=no 10 acc:(0.98573)

Reglas de asociación

- platelets=190080-272570 time=200-228 15 ==> DEATH_EVENT=no 15 acc:(0.99044)
 - platelets=190080-272570 time=4-32 15 ==> DEATH_EVENT=si 15 acc:(0.99044)
 - creatinine_phosphokinase=23-806 ejection_fraction=33-40 time=4-32 13 ==>
DEATH_EVENT=si 13 acc:(0.98913)
-

Conclusiones

- Los algoritmos basados en la estadística son igual de efectivos para el problema que los algoritmos basados en árboles.
- Los valores fisiológicos del paciente más importantes son la 'ejection_fraction' y la 'serum_creatinine' ambos relacionados con la actividad física.
- El agrupamiento ayuda en la clasificación.
- Los algoritmos de reglas de asociación han sido capaces de identificar la importancia de la variable 'time', pero dicha variable no puede ser usada para nuevos pacientes.

FIN
