

MARÍA LUZ PÉREZ
SAURA

MARTÍN PIERANGELI

MARCELO RAPONI

JAVIER CEFERINO
RODRÍGUEZ

MACHINE LEARNING
UNSAM 2021

COGNITO

HEART FAILURE PREDICTION



DATASET

heart_failure_clinical_records_dataset.csv

299 PACIENTES 12 FEATURES 1 TARGET

BINARIAS

anaemia: Disminución de glóbulos rojos o hemoglobina

sex: Mujer u hombre

smoking: Si el paciente fuma

diabetes: Si el paciente tiene diabetes

high_blood_pressure: Si el paciente tiene hipertensión

CONTINUAS

age: Edad del paciente (años)

creatinine_phosphokinase: Nivel de enzimas CPK en sangre (mcg/L)

ejection_fraction: Porcentaje de sangre eyectada del corazón por contracción (%)

platelets: Plaquetas en sangre (kiloplatelets/mL)

serum_creatinine: Nivel de creatinina sérica en sangre (mg/dL)

serum_sodium: Nivel de sodio sérica en sangre (mEq/L)

time: Periodo de seguimiento (días)

TARGET

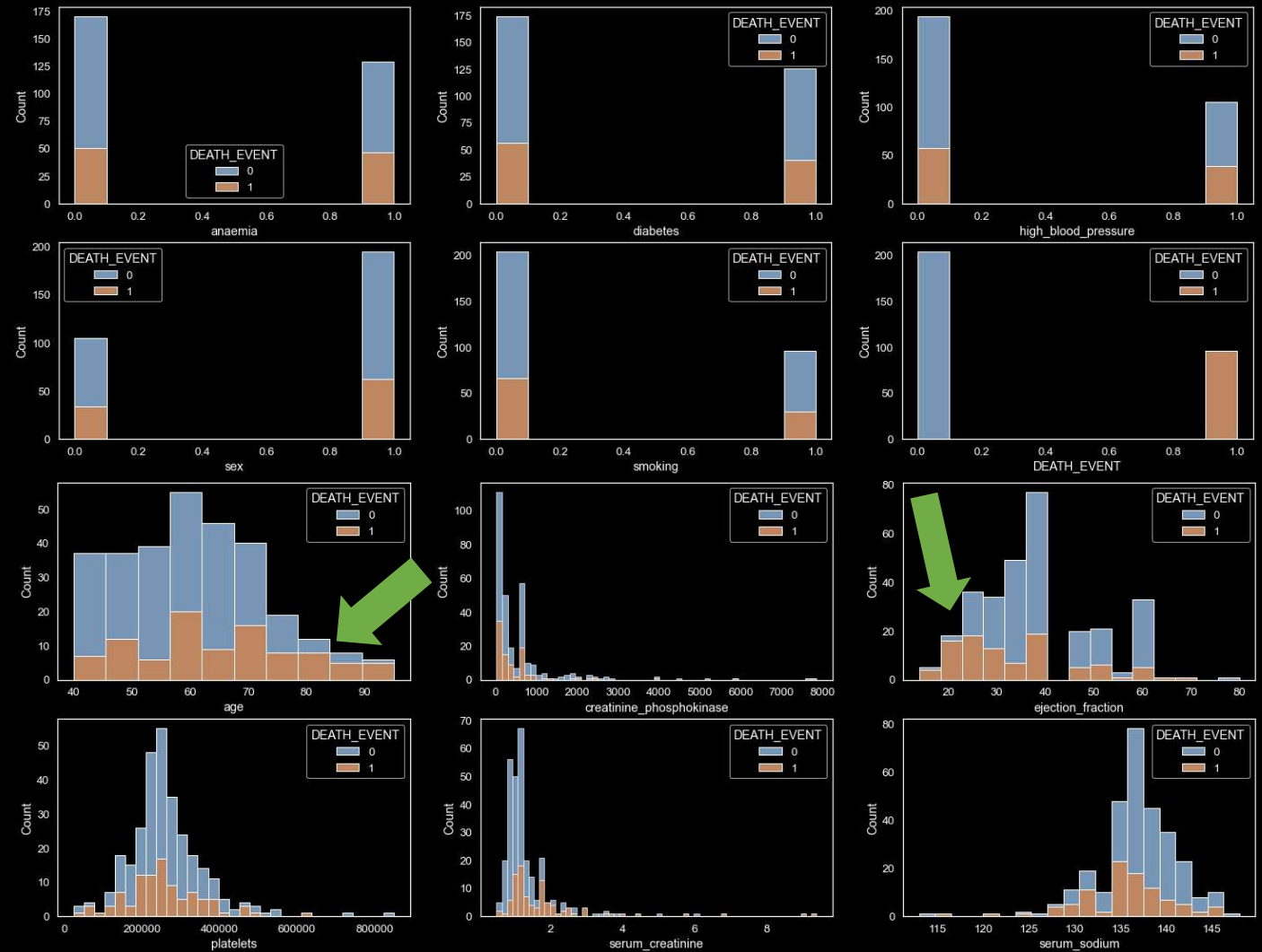
DEATH_EVENT: Si el paciente falleció durante el periodo de seguimiento

Disponible en:

<https://www.kaggle.com/andrewmvd/heart-failure-clinical-data>

DATA ANALYSIS

ESTADÍSTICA DE DATOS



DATA ANALYSIS

ESTADÍSTICA DE DATOS

ALGUNAS OBSERVACIONES:

`creatinine_phosphokinase.mean()` = 581.8 vs. 120mcg/L como valor máximo normal

- Valores altos en pacientes con AC o Pericarditis posterior a un AC

`ejection_fraction.mean()` = 38 vs. 41% como mínimo crítico

- La disminución puede deberse a un AC o Alta Presión

`serum_creatinine.mean()` = 1.4 vs. 1.3(H) y 1.38 vs. 1.1mg/dL(M) ó como máximo normal

- Valores altos indican problemas renales, muy relacionado con la presión y el corazón

`serum_sodium.mean()` = 136.7 dentro de los valores normales 135-140mEq/L

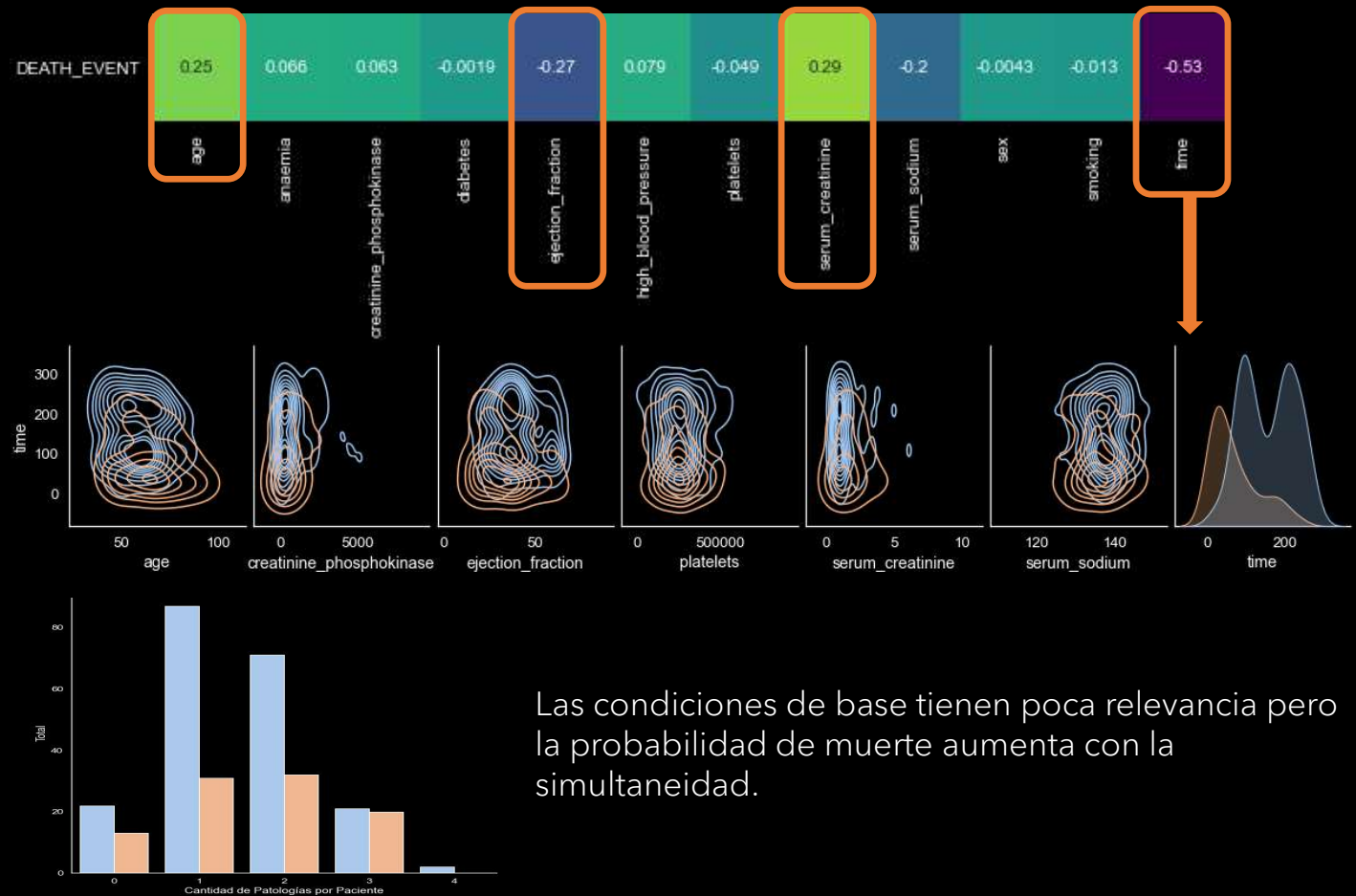
`platelets.mean()` = 98k vs. 150k/mL como mínimo normal

En general, los valores están por fuera de lo normal lo que indica que los pacientes sufren definitivamente complicaciones relacionadas.

DATA ANALYSIS

ESTADÍSTICA DE DATOS

Lo anterior se condice con la matriz de correlación:



PLAN DE TRABAJO

PROBLEMA DE CLASIFICACIÓN

Proponemos aplicar algún modelo de clasificación e inferir si el paciente muere o no (con un intervalo de confianza).

LOGISTIC REGRESSION
NAIVE BAYES
DECISION TREE
SUPER VECTOR MACHINES

SELECCIÓN DE FEATURES MÁS SIGNIFICATIVOS

Buscaremos, dependiendo la conveniencia de nuestro modelo, que y cuantos features son realmente necesarios y si es necesario o no aplicar algún método de reducción de grado como PCA.