**Eduardo Mora González y Diego Sánchez De La Fuente**

**¿Cómo realizar la limpieza y análisis de datos?**

Diagrama

Descripción generada automáticamente

**PRACTICA 2**

**Asignatura:**

Tipología y ciclo de vida de los datos

**Bloque:**

PRA2: ¿Cómo realizar la limpieza y análisis de datos?

**Apellidos: Sánchez de la Fuente**

**Nombre: Diego**

**Apellidos: Mora González**

**Nombre: Eduardo**

**Fecha: Enero de 2023**

CONTENIDO

[**1.** **MOTIVACIÓN** 3](#_Toc123226279)

[**2.** **DESCRIPCION DATASET** 3](#_Toc123226280)

[**3.** **OBJETIVO** 4](#_Toc123226281)

[**4.** **INTEGRACIÓN Y SELECCIÓN** 4](#_Toc123226282)

[**5.** **LIMPIEZA DE LOS DATOS** 4](#_Toc123226283)

[**6.** **ANÁLISIS DE LOS DATOS** 4](#_Toc123226284)

[**7.** **REPRESENTACIÓN DE LOS RESULTADOS** 4](#_Toc123226285)

[**8.** **RESOLUCIÓN DEL PROBLEMA** 4](#_Toc123226286)

[**9.** **CÓDIGO** 4](#_Toc123226287)

[**REFERENCIAS** 4](#_Toc123226288)

# **MOTIVACIÓN**

En Europa, el paro cardiaco es una de las primeras causas de mortalidad y en España fallecen en torno a 100 personas al día por este suceso [1], esto representa aproximadamente el 31% de las muertes a nivel mundial.

# **DESCRIPCION DATASET**

Por la razón anterior, se han seleccionado un conjunto de datos [2] que contiene 12 características. El conjunto de datos ha sido elegido por las características que estos contienen, ya que son los parámetros típicos usados en los estudios de problemas del corazón, y es por eso por lo que tras el análisis de estos se puede sacar unas conclusiones bastantes interesantes.

Cómo se ha mencionado anteriormente tenemos 12 características distintas:

* **Age**: edad del paciente [años]
* **Sex**: sexo del paciente [M: Masculino, F: Femenino]
* **ChestPainType**: tipo de dolor de pecho [TA: angina típica, ATA: angina atípica, NAP: dolor no anginal, ASY: asintomático]
* **RestingBP**: presión arterial en reposo [mm Hg]
* **Cholesterol**: colesterol sérico [mm / dl]
* **FastingBS**: azúcar en sangre en ayunas [1: si BS en ayunas> 120 mg / dl, 0: en caso contrario]
* **RestingECG**: resultados del electrocardiograma en reposo [Normal: Normal, ST: con anomalía de la onda ST-T (inversiones de la onda T y / o elevación o depresión del ST> 0,05 mV), LVH: que muestra una hipertrofia ventricular izquierda probable o definitiva según los criterios de Estes]
* **MaxHR**: frecuencia cardíaca máxima alcanzada [Valor numérico entre 60 y 202]
* **ExerciseAngina**: angina inducida por el ejercicio [Y: Sí, N: No]
* **Oldpeak**: oldpeak = ST [Valor numérico medido en depresión]
* **ST\_Slope**: la pendiente del segmento ST del ejercicio pico [Up: uploping, Flat: flat, Down: downsloping]
* **HeartDisease**: clase de salida [1: enfermedad cardíaca, 0: Normal]

# **OBJETIVO**

Se puede decir que el objetivo buscado es predecir la posibilidad de que una persona tenga un alto riesgo de ser diagnosticado como un paciente cardíaco a través de las diversas características. Para llegar a al objetivo se tiene pensado realizar diversos métodos de análisis para así relacionar las diversas características para obtener unos parámetros finales y así concluir la posibilidad de que una persona tenga o no una enfermedad cardiaca.

# **INTEGRACIÓN Y SELECCIÓN**

Al tener un conjunto bastante completo, se ha decidió usar la 12 características de este para contestar a la pregunta de como predecir si una persona va a tener riesgo de ser diagnosticado como un paciente cardiaco.

# **LIMPIEZA DE LOS DATOS**

# **ANÁLISIS DE LOS DATOS**

# **REPRESENTACIÓN DE LOS RESULTADOS**

# **RESOLUCIÓN DEL PROBLEMA**

# **CÓDIGO**

# **REFERENCIAS**

1. [https://fundaciondelcorazon.com/prensa/notas-de-prensa/2900-solo-el-30-de espanoles-sabe-realizar-la-reanimacion-cardio-pulmonar-rcp-.html](https://fundaciondelcorazon.com/prensa/notas-de-prensa/2900-solo-el-30-de%20espanoles-sabe-realizar-la-reanimacion-cardio-pulmonar-rcp-.html)
2. <https://www.kaggle.com/fedesoriano/heart-failure-prediction?select=heart.csv>