

30 de septiembre del 2022



# Ataques Cardiacos

## Predicción

Coder House - **Data Science** - Comision 29730

Profesor: **Damian Dapuetto** - Tutor: **Miguel Garcia**

# TABLA DE CONTENIDOS

01

## PROBLEMÁTICA y OBJETIVO

Presentación de la  
problemática y  
objetivo del modelo

02

## FUENTE DATA SET

Descripción y  
procesamiento de  
los datos.

03

## EDA

Análisis Univariado,  
Bivariado y  
Multivariado.  
Hallazgos y  
Conclusiones

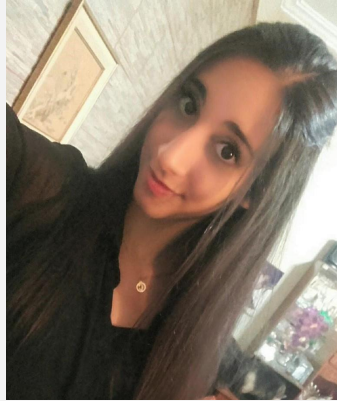
04

## ALGORITMO

Elección del  
Algoritmo,  
entrenamiento,  
Performance y  
Conclusiones



# OUR TEAM



CAROLINA  
ABUELO

Futura Psicóloga,  
entusiasta, ávida de  
aprender cosas nuevas



CAROLINA  
ORSE

Analista de Negocios, a  
favor del aprendizaje  
continuo

6699



“29 de Septiembre: Día Mundial  
del Corazón.

Se celebra con el propósito de  
concientizar sobre las  
enfermedades  
cardiovasculares, su  
prevención, control y  
tratamiento.”

# INTRODUCCION



- ❑ Las **enfermedades cardiovasculares** son aquellas que afectan al corazón y a todas las arterias del organismo.
- ❑ Su principal causa es la aterosclerosis, que es el depósito de placas de colesterol en el interior de las paredes de las arterias, provocando su obstrucción y comprometiendo la llegada de la sangre a órganos vitales como el corazón, el cerebro y el riñón.
- ❑ Por esta razón, la enfermedad arterial aterosclerótica es la principal causa del infarto agudo de miocardio (IAM), del accidente cerebrovascular (ACV) y de los aneurisma.

# 01

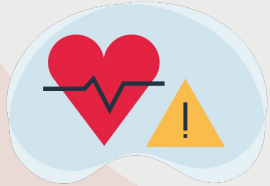
## PROBLEMÁTICA Y OBJETIVOS



# PROBLEMÁTICA

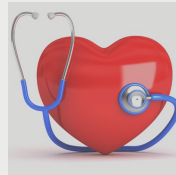
## ECV

En la actualidad, las enfermedades cardiovasculares (ECV) son las principales causas de muerte a nivel mundial.



## DIAGNOSTICO TARDIO

Los especialistas llegan al diagnóstico basados en estudios clínicos del paciente y en base a las experiencias personales, lo que aumenta el riesgo de errores y retrasa la implementación del tratamiento adecuado.



## ATAQUE CARDIACO

La OMS estimó en 2015 que el 31% de todas las muertes mundiales fueron por ECV.





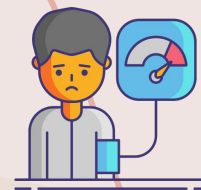
Antecedentes  
familiares



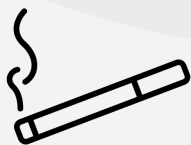
Sedentarismo



Estres



Elevada  
presión arterial



Tabaquismo



## FACTORES DE RIESGO



Obesidad



Elevado colesterol en  
sangre



**Investigación:** Identificar si existen patrones que puedan indicarnos cuando una persona sufrirá un ataque cardíaco.

## Hipótesis:

01	Los parámetros clínicos son predictores de mortalidad por ataque cardíaco
02	Las patologías clínicas (diabetes, anemia e hipertensión) son factores predictores de mortalidad por ataque cardíaco
03	El consumo de tabaco es un factor predictor de mortalidad por ataque cardíaco
04	La edad y el sexo son un factor predictor de mortalidad por ataque cardíaco

# OBJETIVO



Construir un modelo de **Machine Learning** utilizando **algoritmos de clasificación** para que especialistas médicos, aseguradoras y empresas de salud, den un diagnóstico más certero a la hora de prevenir un ataque cardíaco.

## Desafío

Realizar un **modelo predictivo** en base a un data set.

## Resultados

La métrica que definirá cuál es el mejor modelo es el **Accuracy**.

## Solucion

Aportar una **herramienta** que permita dar diagnósticos más certeros.

# 02

## FUENTE DE DATOS



# DATA SET

## DATA SET de kaggle.com

**Autores:** Davide Chicco, Giuseppe Jurman.

**Publicación:** *"Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone"* en la revista de investigación BMC Medical Informatics and Decision Making (2020).

### Criterios de selección:

- No posee datos nulos
- No posee datos duplicados
- Sin variables categóricas
- Información balanceada
- Distribución normal
- Menos de 300 observaciones, permitiéndonos una carga rápida de las query, rapidez en el entrenamiento y optimizaciones



# VARIABLES

Variable	Type	Descripción
Edad	float	Edad del paciente
Sexo	integer	Sexo del paciente
Creatina Cinasa	integer	Nivel de la enzima CPK en sangre
Fraccion de eyeccion	integer	Porcentaje de sangre que bombea el corazón
Plaquetas	float	Cantidad de plaquetas en sangre
Diabetes	integer	Pacientes con la patología
Sodio en Sangre	integer	Cantidad de Sodio en Sangre

# VARIABLES

Anemia	integer	0.76
Hipertension	integer	0.76
Fumador	integer	Consumo de Tabaco del paciente
Tiempo	integer	Tiempo de Observacion del paciente
Creatina en sangre	float	Desecho generado por los músculos
Muerte	integer	Variable Target

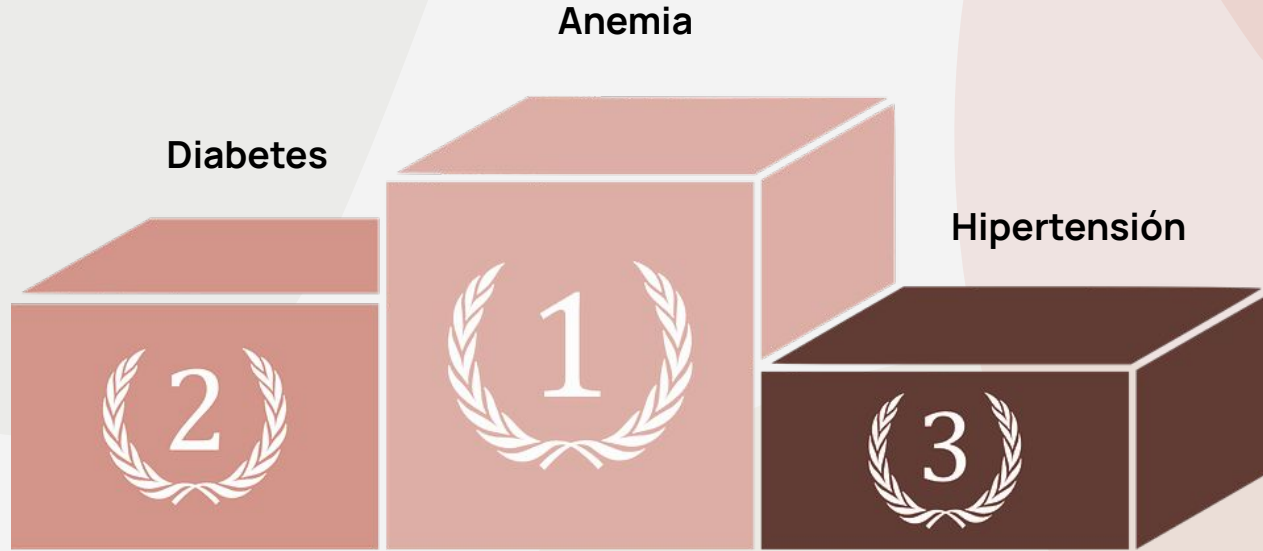
03



# EXPLORATORY DATA ANALYSIS (EDA)

# COMORBILIDADES

¿Las patologías clínicas son inductores de mortalidad por ataque cardíaco?

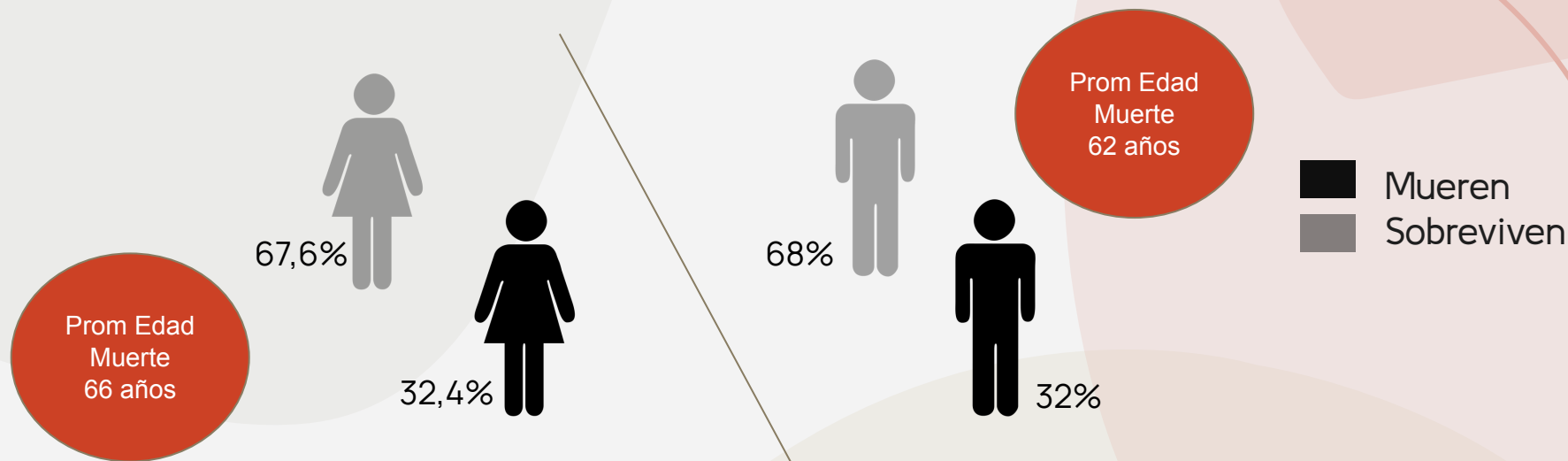


- ❑ En un análisis más detallado de la muestra de fallecidos, observamos que aquellos que padecen anemia, tienen una mayor probabilidad de sufrir un ataque cardíaco, ya que el 45% de los fallecidos eran anémicos.
- ❑ La diabetes queda en un segundo puesto ya que un 40% de los fallecidos eran diabéticos.
- ❑ La hipertensión en el tercer puesto, ya que el 39% de los fallecidos poseían dicha patología.



# GENERO

¿El género es un factor determinante de la mortalidad por ataque cardíaco?



Al separar la muestra por población, hombre/mujer y compararlos con su propio género, la diferencia entre hombres y mujeres que padecen un ataque cardíaco es poco significativa.

**Podemos concluir que el sexo no es un factor determinante a la hora de padecer un ataque cardíaco.**

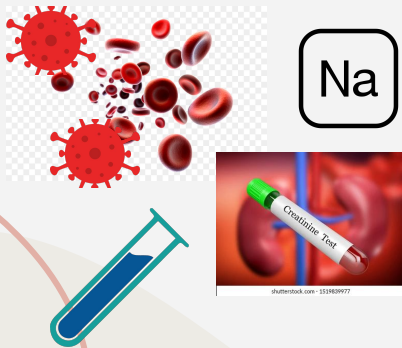
# OTRAS HIPÓTESIS ANALIZADAS

## ¿El consumo de tabaco es un factor predictor de mortalidad por ataque cardíaco?

Podemos inferir que el consumo de tabaco no es un factor determinante a la hora de predecir pacientes con posibles ataques cardíacos, ya que de los pacientes que fallecieron por enfermedades cardiovasculares **sólo un tercio** eran fumadores.

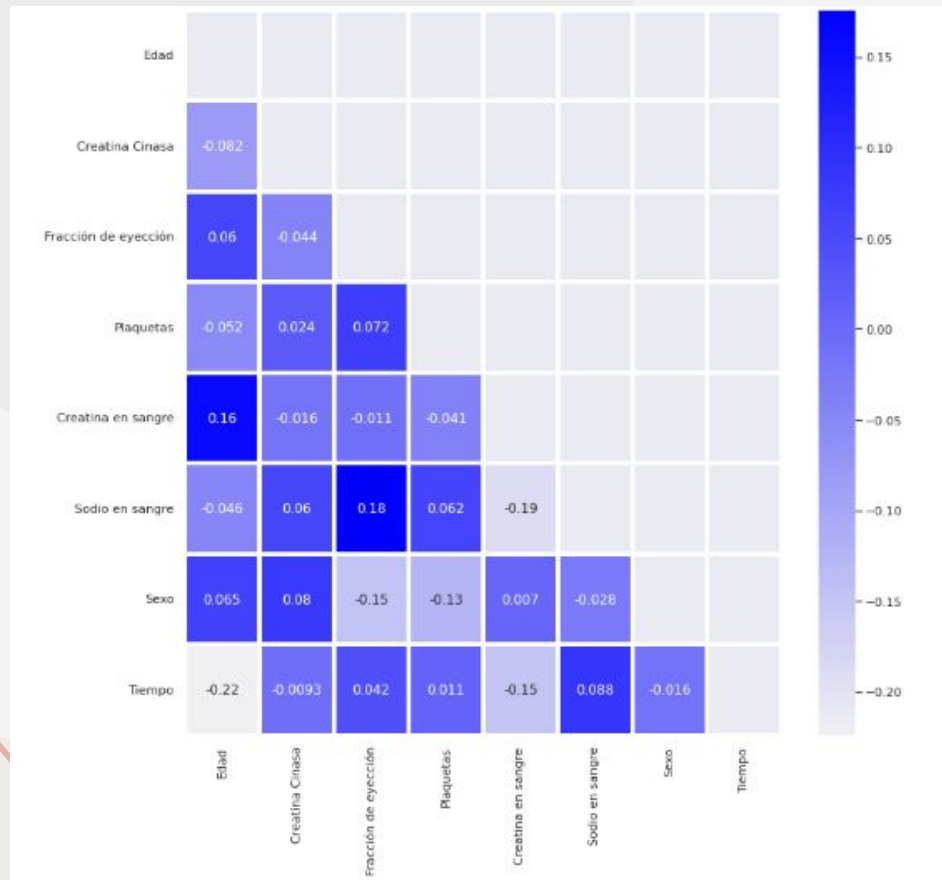


## ¿Los parámetros clínicos son predictores de mortalidad por ataque cardíaco?



- ❑ En base a las observaciones, podemos concluir que aquellos individuos que fallecieron por ataques cardíacos presentan valores más altos de cinasa y creatina en sangre afectando en un porcentaje mayor a las mujeres.
- ❑ El sodio presenta el fenómeno de hiponatremia.
- ❑ Las plaquetas en ambos géneros no presentan variación frente a la variable target.

# MATRIZ DE CORRELACIÓN



- La **creatina cinasa** y el **tiempo** son las variables con mayor correlación.
- La variable que más se relaciona con la posibilidad de tener un ataque cardíaco es la **creatina cinasa**.

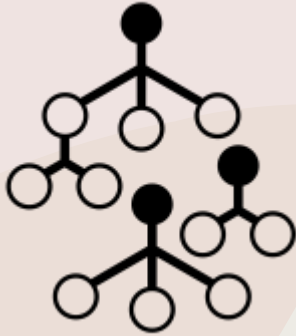
# 04

## ALGORITMO

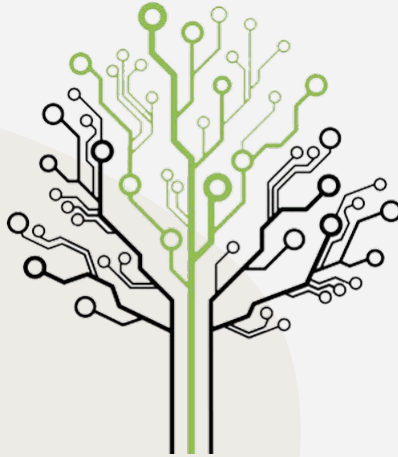


# MODELOS ML ELEGIDOS

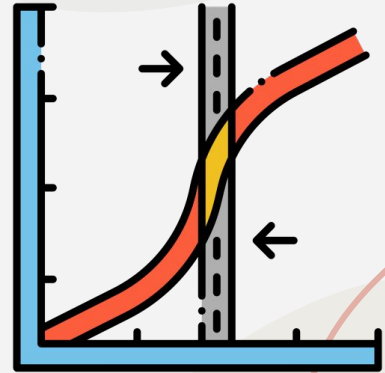
## 1) Random Forest Classifier



## 2) Decision Tree



## 3) Logistic Regression



# METRICAS



1°

RECALL

Porcentaje de clasificación correcta de los verdaderos positivos.

2°

PRECISION

Porcentaje de clasificación correcta de los falsos negativos.

3°

ACCURACY

Porcentaje de aciertos sobre el total.

# INDICADORES DE DESEMPEÑO DE LOS MODELOS

## MÉTRICAS DE LOS MODELOS SIN OPTIMIZAR

	Precision	Recall	Accuracy
Regresión logística	0.61	1.00	0.63
Random Forest Classifier	0.65	0.91	0.66
Árbol de Decisión	0.64	0.77	0.61

# MÉTRICAS DE LOS MODELOS OPTIMIZADOS

	Accuracy		Accuracy
Regression logistica	63%		72%
<b>Random Forest Classifier</b>	66%	→	76%
Arbol de Decision	61%		74%

El modelo elegido es **Random Forest Classifier**

*Dicho modelo arroja el mejor resultado de la métrica seleccionada (accuracy) tanto antes como después de ser optimizado, obteniendo 66,7% previa a la optimización y un 76,13% tras las mejoras*





# Gracias



Podes buscarnos en nuestras redes profesionales:



[www.linkedin.com/in/carolinaabuelo](https://www.linkedin.com/in/carolinaabuelo)

[www.linkedin.com/in/carolina-orse-7b9579143](https://www.linkedin.com/in/carolina-orse-7b9579143)