

## Integrantes

Laura Herrera Giraldo  
Jaime Alexis Herrera Ruiz  
Arley Fuentes Arenales

### 1. Problema predictivo

Según el Centro para el Control y la Prevención de Enfermedades (CDC), las enfermedades cardíacas son una de las principales causas de muerte en los EE.UU. Aproximadamente la mitad de todos los estadounidenses tienen al menos uno de los tres factores de más riesgo clave de enfermedad cardíaca: presión arterial alta, colesterol alto y el tabaquismo. Además, a estos factores se incluyen también, la diabetes, la obesidad, el sedentarismo o la ingesta de alcohol.

A partir de los datos obtenidos, se pretende construir un modelo, a través de métodos de aprendizaje automático que permitan detectar patrones que puedan predecir la posibilidad de una persona de padecer enfermedades cardíacas.

### 2. Dataset

<https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>

El conjunto de datos proviene del CDC y hacen parte del Sistema de Vigilancia de Factores de Riesgo del Comportamiento (BRFSS), que realiza encuestas telefónicas anuales para recopilar datos sobre el estado de salud de los residentes de EE. UU.

Consta de 319795 filas y 18 columnas. La gran mayoría de las columnas son preguntas que se hacen a los encuestados sobre su estado de salud:

- **HeartDisease:** encuestados que alguna vez informaron haber tenido una enfermedad cardíaca coronaria (CHD) o un infarto de miocardio (IM).
- **BMI:** Índice de Masa Corporal (IMC).
- **Smoking:** ¿Ha fumado al menos 100 cigarrillos en toda su vida? (La respuesta Sí o No).
- **AlcoholDrinking:** bebedores empedernidos (hombres adultos que toman más de 14 tragos por semana y mujeres adultas que toman más de 7 tragos por semana)
- **Stroke:** (Alguna vez le dijeron) (usted tuvo) un accidente cerebrovascular?
- **PhysicalHealth:** Pensando en su salud física ¿durante cuántos días durante los últimos 30 días su salud física no fue buena? (0-30 días).
- **MentalHealth:** Pensando en su salud mental, ¿durante cuántos días durante los últimos 30 días su salud mental no fue buena? (0-30 días).
- **DiffWalking:** ¿Tiene serias dificultades para caminar o subir escaleras?
- **Sex:** ¿Eres hombre o mujer?
- **AgeCategory:** Categoría de edad de catorce niveles.
- **Race:** Valor de raza/etnicidad

- **Diabetic:** (Alguna vez le dijeron) (usted tenía) diabetes?
- **PhysicalActivity:** Adultos que informaron haber realizado actividad física o ejercicio durante los últimos 30 días además de su trabajo habitual.
- **GenHealth:** ¿Diría usted que, en general, su salud es...
- **SleepTime:** en promedio, ¿cuántas horas duermes en un período de 24 horas?
- **Asthma:** (Alguna vez le dijeron) (usted tenía) asma?
- **KidneyDisease:** sin incluir cálculos renales, infección de la vejiga o incontinencia, ¿alguna vez le dijeron que tenía una enfermedad renal?
- **SkinCancer:** (Alguna vez le dijeron) (usted tenía) cáncer de piel?

### 3. Métricas de desempeño

El modelo es de clasificación binaria, por ende, se utilizarían las siguientes métricas:

- Matriz de confusión
- Exactitud (ACC)
- Sensibilidad (TPR)
- Especificidad (TNR)
- Precisión
- Regresión logística
- Árbol de decisiones
- K-Vecinos más cercano (KNN).
- Máquina de vectores de soporte (SVC)

### 4. Primer criterio del desempeño deseable en producción

El porcentaje de acierto esperado de la predicción sobre el padecimiento de enfermedades cardíacas es de al menos de un 85% ya que el modelo se utilizará para dar sugerencias sobre estilos de vida saludable para disminuir riesgo de estas enfermedades.

### Fuentes

*HeartDiseases*. (2022, enero 28). <https://www.cdc.gov/nchs/fastats/heart-disease.htm>

*Infographics—Health, United States—Products*. (2020, enero 2).

<https://www.cdc.gov/nchs/hus/spotlight/2019-heart-disease-disparities.htm>

*Trends in Cancer and Heart Disease Death Rates Among Adults Aged 45–64: United States, 1999–2017*. (2019, mayo 22)

[https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68\\_05-508.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_05-508.pdf)

*Personal Key Indicators of Heart Disease*. (s. f.). Recuperado 4 de julio de 2022, de <https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>

*Binary Classification—Amazon Machine Learning*. (s. f.). Recuperado 4 de julio de 2022, de <https://docs.aws.amazon.com/machine-learning/latest/dg/binary-classification.html>