🎓 PROYECTO DE INTELIGENCIA ARTIFICIAL

Predicción de enfermedades cardíacas mediante regresión logística
Dataset: https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset

🧠 Contexto

Las enfermedades cardiovasculares son una de las principales causas de mortalidad en el mundo. El *Framingham Heart Study* es un estudio longitudinal iniciado en 1948 en Massachusetts (EE. UU.) para identificar factores que contribuyen al desarrollo de enfermedades del corazón.
A partir de los datos recopilados de miles de pacientes, es posible aplicar **modelos de Inteligencia Artificial** para predecir la probabilidad de desarrollar una enfermedad cardíaca, con el fin de apoyar la prevención y la toma de decisiones médicas.
Este proyecto busca que los estudiantes **apliquen la regresión logística** para modelar la probabilidad de padecer una enfermedad cardíaca en función de variables de riesgo como la edad, presión arterial, colesterol, tabaquismo y diabetes.

🎯 Objetivo general

Desarrollar un modelo de **regresión logística binaria** que prediga la probabilidad de padecer enfermedad cardíaca en los próximos 10 años, utilizando el conjunto de datos *Framingham Heart Study* y aplicando técnicas de aprendizaje supervisado.

🎯 Objetivos específicos

1. Analizar las variables del dataset e identificar su relación con la aparición de enfermedad cardíaca.
2. Implementar un modelo de regresión logística con Python y bibliotecas de IA.
3. Evaluar el modelo mediante métricas de clasificación (Accuracy, Precision, Recall, F1, AUC).
4. Interpretar los coeficientes del modelo para identificar los factores de riesgo más relevantes.
5. Elaborar una aplicación web, móvil o de escritorio en la que se pruebe el modelo.

🧩 Descripción del dataset

El dataset contiene información de **4 240 registros** y **15 atributos**, incluyendo una variable objetivo-binaria que indica si una persona ha desarrollado o no enfermedad cardíaca en un periodo de 10 años.

Variable	Descripción	Tipo
male	1 = hombre, 0 = mujer	Categórica
age	Edad del paciente	Numérica
education	Nivel educativo	Categórica
currentSmoker	1 = fumador actual	Categórica
cigsPerDay	Cigarrillos por día	Numérica
BPMeds	1 = usa medicación para presión arterial	Categórica
prevalentStroke	1 = ha sufrido un accidente cerebrovascular	Categórica
prevalentHyp	1 = tiene hipertensión	Categórica
diabetes	1 = diabético	Categórica
totChol	Colesterol total (mg/dL)	Numérica
sysBP	Presión sistólica (mmHg)	Numérica

Variable	Descripción	Tipo
diaBP	Presión diastólica (mmHg)	Numérica
BMI	Índice de masa corporal	Numérica
heartRate	Frecuencia cardíaca (lpm)	Numérica
glucose	Nivel de glucosa (mg/dL)	Numérica
TenYearCHD	Variable objetivo: 1 = presenta enfermedad cardíaca a 10 años	Binaria

⚙️ Herramientas

- **Lenguaje:** Python 3.8+
- **Librerías:** pandas, numpy, matplotlib, seaborn, scikit-learn
- **Entorno:** Jupyter Notebook / Google Colab / VS Code
- **Dataset:** Framingham Heart Study Dataset – Kaggle

🧠 Interpretación esperada

- **Coeficientes positivos:** aumentan la probabilidad de desarrollar enfermedad cardíaca (por ejemplo, edad, presión arterial, tabaquismo).
- **Coeficientes negativos:** reducen la probabilidad.
- **AUC (Área bajo la curva ROC):** mide la capacidad del modelo para distinguir entre las dos clases (0 = sin riesgo, 1 = con riesgo). Un AUC cercano a 1 indica un modelo excelente.
- **Precision, Recall y F1-score:** permiten analizar equilibrio entre falsos positivos y falsos negativos.

📊 Entregables

1. **Notebook o script (.ipynb o .py)** con código comentado y resultados.
2. **Diapositiva que contenga un Informe técnico** con los apartados:
    - Introducción y contexto del problema
    - Metodología (pasos de limpieza, división, modelado, evaluación)
    - Resultados y análisis (incluyendo gráficos y coeficientes)
    - Conclusiones y recomendaciones
3. **Presentación oral (10 minutos)** con resumen de resultados y visualizaciones.

📜 Rúbrica de evaluación

Criterio	Excelente (10)	Bueno (8)	Regular (6)	Deficiente (4)
Análisis del problema	Claramente definido, contextualizado y con respaldo teórico	Bien definido pero con pocos detalles	Poca claridad	No presenta
Preparación de datos	Limpieza, normalización y análisis exploratorio correctos	Parcialmente adecuados	Falta procesamiento	No realiza limpieza
Implementación técnica	Código funcional, documentado y correctamente estructurado	Código funcional con errores menores	Parcialmente funcional	No ejecuta
Evaluación y resultados	Incluye métricas, curva ROC y análisis de variables	Resultados correctos pero con poca interpretación	Métricas incompletas	Ausente
Informe y presentación	Claros, con gráficas, interpretación y conclusiones	Claros pero con pocos detalles	Incompletos	Deficientes