CORAZÓN A RITMO SANO

LAGOS MARTINEZ J.A. | MUÑOZ BARRIENTOS R.
RODRIGUEZ GUERRERO L.V. | DOMINGUEZ VICTORINO C.O.

UANL
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

INTRODUCCIÓN

El corazón es uno de los órganos más importantes ya que bombea sangre y suministra oxígeno y nutrientes a todo el cuerpo. Como cumple con estas funciones tan cruciales su bienestar es de suma importancia. Según la OMS, las enfermedades cardiovasculares son actualmente la principal causa de fallecimiento en todo el mundo representando un 31% de todas las muertes registradas.

Una herramienta que explore datos sobre la salud de un paciente y datos sobre su corazón podría ayudar a reducir el tiempo en el que a éste se le brinde un diagnóstico y así iniciar tratamiento correspondiente lo más pronto posible, incrementando sus posibilidades de recuperación e incluso prevenir la muerte.

OBJETIVOS

Predecir el diagnóstico sobre la enfermedad de corazón según los resultados de los exámenes de salud y datos biométricos del paciente.

Objetivos secundarios.

- Identificar datos de mayor influencia al momento de realizar un diagnóstico.
- Encontrar si existen algún grupo de edad o sexo sea más propenso a presentar problemas cardiacos.

METODOLOGÍA

La base de datos cuenta con 296 registros y 13 variaables más una columna 'Objetivo' que nos indica si el paciente fue diagnosticado con una enfermedad del corazón.
Dado que el objetivo es predicir el diagnóstico se utilizó el modelo regresión logística (1) para obtener los resultados.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (1)$$

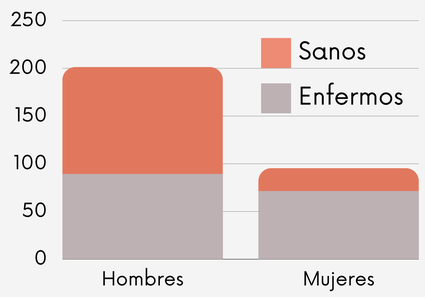

Fig. 1: Comparación diagnóstico por sexo.

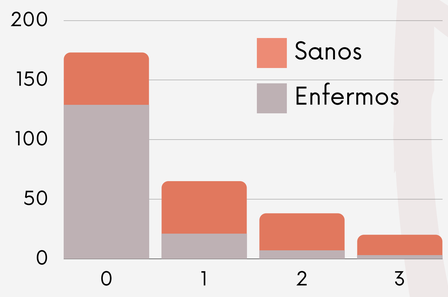

Fig. 2: Comparación diagnóstico por número de vasos sanguíneos observados.

En la Fig. 1 se observa que la proporción de mujeres que padecen de una enfermedad del corazón es mucho mayor a las que no.
En la Fig. 2 se ven los resultados de una fluorscopia, exámen donde se cuenta el número de vasos sanguíneos visibles. La gráfica indica que entre menor es este valor, mayor es la proporción de pacientes diagnosticdos con una enfermedad del corazón.

RECURSOS

Librerias:

- Pandas
- Numpy
- Seaborn
- Matplotlib
- Statsmodels

Base de Datos	Notebook

RESULTADOS

Utilizando solo el 80% de los registros para entrenar al modelo se llegó a la siguiente ecuación del modelo (2).

$$\begin{aligned}\log\left(\frac{p}{1-p}\right) = {} & 0.02(Edad) - 1.2468(Sexo) - 0.0179(PresArtRep) - 0.003(Col) \\ & +0.2851(NivAzuAyu) + 0.4598(Electro) + 0.0320(MaxRitCard) \\ & -0.7146(Angina) - 0.5304(MinElectro) + 0.7205(Prendiente) \\ & -1.3805(NumVasSang) - 1.0168(thal)\end{aligned} \qquad (2)$$

Donde p es la probabilidad de que el paciente padezca de una enfermedad del corazón. Si la expresión (3) es mayor a 0 significa que p>50% y al perfil se le asigna el diagnóstico de enfermedad del corazón.

$$\log\left(\frac{p}{1-p}\right) \qquad (3)$$

El 20% restante de los datos se utilizó para comparar las predicciones del modelo con los valores reales del diagnóstico.

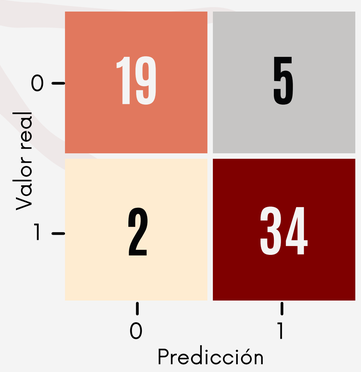

Fig. 3: Matriz de Confusión Regresión Logística

Con este modelo se logró predecir correctamente el diagnóstico de 53 de los 60 registros de prueba por lo que se cuenta con una precisión de 88.33%.

CONCLUSIONES

Existen varios factores que influyen al momento de realizar un diagnóstico sobre la salud del corazón. El análisis realizado indica que la variable con mayor influencia es el número de vasos sanguíneos observados en la fluoroscopia ya que tiene el coeficiente más alejado de cero. También se nota que los hombres tienden a padecer menos de enfermedades del corazón en comparación a las mujeres.

Se espera en un futuro mejorar la precisión de la predicción, una opción es revisar que la existencia de multicolinealidad entre variables independientes y en caso de que exista removerlas del modelo.