Etapa 2 Diseño de sistemas de aprendizaje automatico

Nombre del estudiante:  
Jhon Edison Betancurt Lora

Grupo:

**Análisis de datos\_202016908A\_1394**

Tutor

Sandra Milena Patiño

Universidad Nacional Abierta y a Distancia-UNAD

Escuelas de Ciencias Básicas Tecnologías e Ingeniería

Programa de Ingeniería de sistemas

Quimbaya – Septiembre del 2023

**Regresión logística**

**Código**

import pandas as pd

from sklearn.linear\_model import LogisticRegression

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import confusion\_matrix, accuracy\_score, recall\_score, precision\_score, f1\_score

from sklearn.preprocessing import StandardScaler

reg\_logistica = pd.read\_csv("framingham.csv") // carga el archivo

reg\_logistica.dropna(inplace=True) // elimina filas NaN

bins = [-1, 0, 10, 20, 100] // aquí defino los contenedores

labels = ["No fumador", "Fumador ligero", "Fumador moderado", "Fumador intenso"]

reg\_logistica['cigsPerDay'] = pd.cut(reg\_logistica['cigsPerDay'], bins=bins, labels=labels)

X = reg\_logistica.drop('cigsPerDay', axis=1)

y = reg\_logistica['cigsPerDay']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

//Para quitar el error de poca iteraciones me recomendaron usar esta función

scaler = StandardScaler()

X\_train\_scaled = scaler.fit\_transform(X\_train)

X\_test\_scaled = scaler.transform(X\_test)

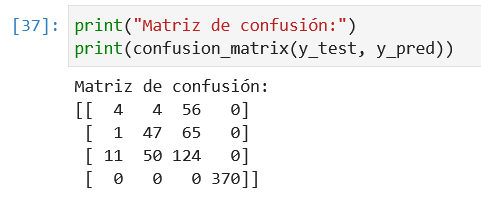
model = LogisticRegression(max\_iter=1000)

model.fit(X\_train\_scaled, y\_train)

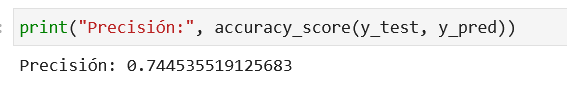
y\_pred = model.predict(X\_test\_scaled)

**Análisis**

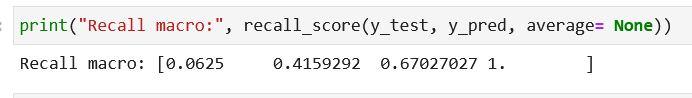
Como es un metodo de clasificación se puede analizar con lo siguiente



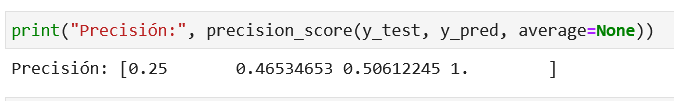
Aquí se puede ver que que el modelo esta sesgado hacia la Fumador intenso, las otras clases el modelo no fue tan preciso porque por ejemplo en Fumador moderado se obtuvieron 124 verdaderos positivos pero hay 11 + 50 casos incorrectos.



La precisión alcanzo un 74% aproximadamente lo que indica que el modelo no es perfecto pero pues funciona



Se puede ver que en el recall para cada clase que se definio varia demasiado, el Fumador intenso si tiene exactamente un 1 pero las demás clases tienen valores muy bajos como por ejemplo la clase No fumadores, esto diría que tendría problemas en detectar esta clase.



Esta precisión es igual al recall y tiene el mismo comportamiento, ya que esta varia mucho entre ellos.