

CERTIFICACION CIENCIA DE DATOS

Examen 2

1. Utiliza el dataset de la diabetes para crear un modelo de regresión logística que te indique bajo que criterios una persona tiene riesgos de tener diabetes.
 - (a) Realiza una validación del modelo y explica el porque de tu razonamiento.
 - (b) Mediante una matriz de confusión da criterios para evaluar el modelo construido.
 - (c) Crea un modelo de regresión para que según el modelo contruido si una persona tiene un índice de glucosa superior a 150 y una edad superior a los 50 años nos pueda indicar ¿cuál es la probabilidad de tener diabetes?
2. Utiliza el dataset de Real State (que nos indica los precios de los inmuebles en Sao Paulo) para construir al menos un modelo de regresión lineal, argumenta acerca de las dificultades de la construcción
 - (a) Califica los modelos que se han construido y explica con que modelo te quedarías.
 - (b) Construye un modelo para estimar el precio de alquiler de un apartamento de 2 dormitorios, 1 suite, 2 baños con al menos una plaza de parking.
 - (c) Mejora el modelo anterior dedicandote a explorar solamente 2 zonas de la ciudad.
 - (d) Gracias al error medio da un estimación del precio máximo que se podría esperar en el alquiler de un apartamento de las características mencionadas en las zonas exploradas en el punto anterior.
3. Utiliza el dataset Customer para segmentar una población según los parámetros del dataset. Este dataset tiene la finalidad de calificar el valor de una propiedad en función de las características de las personas que la habitan.
 - (a) Da argumentos para la elección de la cantidad de clusters.
 - (b) Realiza un clustering alternativo al realizado y argumenta si los resultados son similares.
 - (c) Con el dataset PropertyValuation valida si las elecciones anteriores tienen sentido.

BUENA SUERTE!!!!