

Nombre: Lopez Chavez Pablo

Carrera: Ing en ciencias de la Computación

Materia: Inteligencia Artificial 1

Dataset: <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction?select=test.csv>

Colab: Para editar

https://colab.research.google.com/drive/1wvh91WVjl81_WpxMCl1luCSRmFu4AcOT?usp=sharing

#id identificador de cada registro

#gender Genero

#Customer type Tipo de cliente. Indica si el pasajero es un cliente frecuente o no

#Age La edad del pasajero.

#Type of Travel Tipo de viaje. Describe si el viaje es por negocios (Business travel) o por placer/personal (Personal Travel).

#Class: La clase en la que viaja el pasajero, que podría ser "Business", "Eco" (Económica) o "Eco Plus" (Económica Plus).

Flight Distance: La distancia del vuelo en millas.

Inflight wifi service: Calificación del servicio de wifi a bordo, donde una calificación más alta indica un mejor servicio.

Departure/Arrival time convenient: Calificación de la conveniencia del horario de salida y llegada, donde una calificación más alta indica mayor conveniencia.

Ease of Online booking: Calificación de la facilidad para reservar en línea, donde una calificación más alta indica mayor facilidad.

Gate location: Calificación de la ubicación de la puerta de embarque, donde una calificación más alta indica que la ubicación es más conveniente.

Food and drink: Calificación de la calidad de la comida y bebida durante el vuelo.

Online boarding: Calificación del proceso de embarque en línea.

Seat comfort: Calificación de la comodidad del asiento.

Inflight entertainment: Calificación del entretenimiento a bordo (películas, música, etc.).

On-board service: Calificación del servicio en el avión (atención de la tripulación, etc.).

Leg room service: Calificación del espacio para las piernas.

Baggage handling: Calificación del manejo del equipaje (entrega, cuidado, etc.).

Checkin service: Calificación del servicio de check-in.

Inflight service: Calificación general del servicio a bordo.

Cleanliness: Calificación de la limpieza en el avión.

Departure Delay in Minutes: Minutos de retraso en la salida del vuelo.

Arrival Delay in Minutes: Minutos de retraso en la llegada del vuelo.

Satisfaction: Nivel de satisfacción del pasajero con la experiencia general del vuelo. Puede ser una categoría como "Satisfied" (Satisfecho) o "Dissatisfied" (Insatisfecho).

Nombre: airline passenger satisfaction

Temática: satisfaction de pasajeros en vuelos

Primera parte

Primer bloque

Leemos el data set y Codificar la columna 'Gender' usando LabelEncoder, Ajusta el codificador a la columna 'Gender' y luego la transforma,

reemplazando los valores categóricos (como 'Male' y 'Female') por valores numéricos (por ejemplo, 'Male' puede ser 1 y 'Female' puede ser 0), Codificar la columna 'Customer Type' usando LabelEncoder

Se aplica el mismo procedimiento a la columna 'Customer Type'. Si la columna tiene valores como 'Loyal Customer' y 'Disloyal Customer',

estos se convertirán en números, por ejemplo, 0 y 1.

Codificar la columna 'Type of Travel' usando LabelEncoder

valores numericos, Convertir 'satisfaction' a 1 para 'satisfied' y 0 para 'neutral or dissatisfaction', Reorganizar las columnas para poner 'Class' al final, Guardar el dataset preprocesado en un nuevo archivo CSV, usando drive

Segundo Bloque

Cargamos de nuevo Eliminar la primera fila que contiene los encabezados de las columnas, Convertir la columna 'Class' en valores numéricos, Eliminar las filas que contienen valores faltantes, Convertir el DataFrame a una matriz NumPy si es necesario,

Tercer Bloque

La entrada es de 23 elemento contando con x0, 3 etiquetas (1, 2 o 3), los datos de entrenamiento el 80% para x desde las 0 hasta las 20714 Todas menos la ultima para y solo la ultima, desde las 20714 hasta..

Bloque Sigmoid

la funcion que nos ayuda a transformar cualquier valor entre 0 y 1 para modelar probabilidades, ya que la salida puede interpretarse como una probabilidad.

Bloque lrCostFuncion

algoritmo de optimización para ajustar los parámetros de un modelo de regresión logística durante el entrenamiento. Ayuda a encontrar los valores óptimos de los parámetros que minimizan el costo total, teniendo en cuenta la regularización para evitar sobreajuste

OnevsAll

predicciones en un problema de clasificación multiclase

realiza el producto punto entre X y la transpuesta de all_theta. all_theta tiene forma (num_labels, n + 1), donde n es el número de características en X más 1 para el sesgo