**Examen final DAT-245**

**NOMBRE:** KEVIN EDSON ZAPANA CABRERA  
**CI:** 12895010

**DESCRIPCION DEL TITANIC DATASET**

Acerca de este archivo

El hundimiento del Titanic es uno de los naufragios más infames de la historia.

El 15 de abril de 1912, durante su viaje inaugural, el RMS Titanic, considerado por muchos como un barco “insumergible”, se hundió tras chocar con un iceberg. Desafortunadamente, no había suficientes botes salvavidas para todos a bordo, lo que provocó la muerte de 1502 de los 2224 pasajeros y tripulantes.

Si bien hubo un elemento de suerte involucrado en la supervivencia, parece que algunos grupos de personas tenían más probabilidades de sobrevivir que otros.

En este desafío, le pedimos que construya un modelo predictivo que responda a la pregunta: “¿Qué tipo de personas tenían más probabilidades de sobrevivir?” utilizando datos de los pasajeros que se detallara a continuación

**COLUMNAS   
LA CLASE DEL DATASET ES:**

**Clase (Survived):**

Descripción: Es la variable objetivo del dataset, indicando si un pasajero sobrevivió o no al desastre del Titanic.

**Valores:** Binarios.

0: No sobrevivió.

1: Sobrevivió.

**Importancia:** Es el enfoque principal para tareas de predicción y clasificación supervisada.

**PassengerId (Entero):**

Identificador único asignado a cada pasajero.

Función principal: Diferenciar a cada pasajero en el dataset.

Ejemplo de valores: 1, 2, 3, ...

**Survived (Clase, Binaria):**

Indica si el pasajero sobrevivió o no al naufragio.

Valores posibles:

0: No sobrevivió.

1: Sobrevivió.

**Pclass (Entero Categórico):**

Clase del boleto comprado.

Representa el nivel socioeconómico del pasajero.

Valores posibles:

1: Primera clase.

2: Segunda clase.

3: Tercera clase.

**Name (Texto):**

Contiene los nombres completos de los pasajeros.

Incluye el apellido, un título social (e.g., Mr., Mrs.), y en algunos casos nombres adicionales.

**Sex (Entero Binario):**

Sexo del pasajero.

Valores posibles:

0: Masculino.

1: Femenino.

**Age (Continua):**

Edad del pasajero en años.

Rango: 0.42 (bebé) a 80.0 años.

Ejemplo de características:

Mediana: 28 años.

Promedio: 29.7 años.

Desviación estándar: 14.5 años.

**SibSp (Entero):**

Número de hermanos o cónyuges que acompañaban al pasajero.

Rango: 0 a 8.

**Parch (Entero):**

Número de padres o hijos que acompañaban al pasajero.

Rango: 0 a 6.

**Ticket (Texto):**

Contiene el número o código del boleto de cada pasajero.

Puede incluir solo números, solo letras o una combinación de ambos.

**Fare (Continua):**

Precio pagado por el boleto.

Valores normalizados para una mejor representación.

Rango original: 0 a 512.33 (moneda de la época).

**Cabin (Texto):**

Indica el número de cabina asignada al pasajero.

Algunos valores son letras y números combinados (ejemplo: "C85", "B28"), otros solo letras.

Los valores presentes reflejan el nivel y ubicación de la cabina dentro del barco.

**Embarked (texto)**

Puerto desde el que el pasajero abordó el Titanic.

Valores posibles:

C: Cherbourg.

Q: Queenstown.

S: Southampton.

**DESCRIPCIÓN DEL OBJETIVO DE INVESTIGACIÓN A PARTIR DEL DATASET**

El objetivo de esta investigación es desarrollar un modelo predictivo que permita estimar la probabilidad de supervivencia de los pasajeros del Titanic basándose en las características disponibles en el dataset, como edad, sexo, clase de cabina, tarifa, entre otras. A través de un análisis de los datos, se busca identificar patrones y relaciones que expliquen por qué ciertos pasajeros sobrevivieron y otros no. Utilizando técnicas de aprendizaje supervisado, como algoritmos de clasificación, el modelo será capaz de predecir la clase objetivo (supervivencia) de nuevos pasajeros, ayudando a comprender los factores que influyeron en las probabilidades de supervivencia en este evento histórico.