<u>Ejercicio de clase:</u>	
Grupo: 5CDM1	Equipo No.
Integrantes:	
Analysis of Various Decision Tr	y Jain, Arpit Arora, Naresh Dhami. (2017). ee Algorithms for Classification in Data omputer Applications (0975–8887). Volume
Leer el artículo mencionado y responde	r las siguientes preguntas:
Ejercicio No. 1	
a) Describa la minería de datos	
b) Explique las razones por las cua	les se utilizan los árboles de decisión
c) ¿Cuál es la diferencia entre un ár	bol de clasificación y un árbol de regresión?

Ejercicio No. 2

Considerando los algoritmos: ID3, C4.5, CART y Random Forest. Realice un cuadro comparativo que considere los siguientes aspectos: descripción del algoritmo, criterio de partición que utiliza, si utiliza poda o no, tipo de datos que utiliza, ventajas y desventajas.

Criterio	ID3	C4.5	CART	Random Forest
Descripción				
Criterio partición				
Poda				
Tipos de datos que utiliza				

Criterio	ID3	C4.5	CART	Random Forest
Ventajas				
Desventajas				

Ejercicio 3

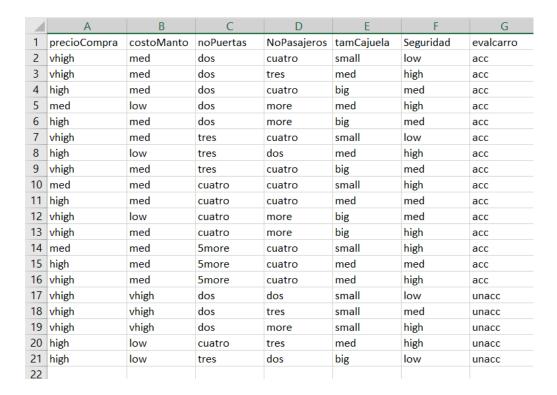
Considerando los siguientes criterios de selección de atributo para particionamiento: *Entropy (Information Gain), Gain Ratio* and *Gini Index*. Realice una descripción con sus propias palabras de cada uno de ellos.

Criterio	Descripción
Entropy (Information Gain)	
Gain Ratio	
Gini Index	

Ejercicio 4

Considerando el ejercicio de evaluación del artículo (laptop), aplique el proceso de cálculo de medidas de evaluación al conjunto de datos de carros (accesible, no accesible).

Realizar el proceso de los cinco (*visto en clase*) pasos de los cálculos para la elección del atributo de particionamiento.



Ejercicio 5

Plantee un conjunto de datos, con 15 registros y 5 atributos, cuyo atributo objetivo sea dicotómico y aplique las actividades que realizó en el ejercicio número 4 de esta guía.

Ejercicio 6

Suponga que tiene la siguiente matriz de confusión de la evaluación de carros.

Original/Predicción	Accesible	No accesible
Accesible	40	5
No accesible	2	3
	42	8

Calcule y explique las diversas medidas que puede generar con base en los datos de la matriz de confusión.

Medida	Cálculo	Explicación
Negativo verdadero		
Positivo falso		
Negativo falso		
Positivo verdadero		
Tasa de exactitud		

Tasa de error	
Precisión	
Sensibilidad (Recall)	
Tasa de positivos falsos	
Tasa de negativos falsos	
Especificidad	