

Ciudad de México a 4 de abril de 2022

Ejercicio de clase:

Grupo: 5CDM1

Equipo No. _____

Integrantes:

Bhumika Gupta, Aditya Rawat, Akshay Jain, Arpit Arora, Naresh Dhami. (2017). Analysis of Various Decision Tree Algorithms for Classification in Data Mining. *International Journal of Computer Applications* (0975–8887). Volume 163 – No 8, April 2017.

Leer el artículo mencionado y responder las siguientes preguntas:

Ejercicio No. 1

a) Describa la minería de datos

b) Explique las razones por las cuales se utilizan los árboles de decisión

c) ¿Cuál es la diferencia entre un árbol 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.

	A	B	C	D	E	F	G
1	precioCompra	costoManto	noPuertas	NoPasajeros	tamCajuela	Seguridad	evalcarro
2	vhigh	med	dos	cuatro	small	low	acc
3	vhigh	med	dos	tres	med	high	acc
4	high	med	dos	cuatro	big	med	acc
5	med	low	dos	more	med	high	acc
6	high	med	dos	more	big	med	acc
7	vhigh	med	tres	cuatro	small	low	acc
8	high	low	tres	dos	med	high	acc
9	vhigh	med	tres	cuatro	big	med	acc
10	med	med	cuatro	cuatro	small	high	acc
11	high	med	cuatro	cuatro	med	med	acc
12	vhigh	low	cuatro	more	big	med	acc
13	vhigh	med	cuatro	more	big	high	acc
14	med	med	5more	cuatro	small	high	acc
15	high	med	5more	cuatro	med	med	acc
16	vhigh	med	5more	cuatro	med	high	acc
17	vhigh	vhigh	dos	dos	small	low	unacc
18	vhigh	vhigh	dos	tres	small	med	unacc
19	vhigh	vhigh	dos	more	small	high	unacc
20	high	low	cuatro	tres	med	high	unacc
21	high	low	tres	dos	big	low	unacc
22							

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 (<i>Recall</i>)		
Tasa de positivos falsos		
Tasa de negativos falsos		
Especificidad		