

Primer Parcial INF - 354 “INTELIGENCIA ARTIFICIAL”

Nombre: Steve Brandom Nina Huacani

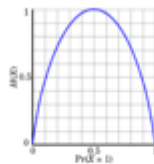
Pregunta 8. Del dataset de citas obtenida en la clase, realice el cálculo en Excel para obtener el árbol de decisión, grafique el árbol.

Se tiene el siguiente Dataset de citas realizado en clases

Dataset de citas 21 de septiembre, y= si o no tenemos pareja							
Genero	Dinero (Harto, medio, poco)	Bañarse	Con Compromi so	Regalo (Peluche, flores, Chocolates.	Cine	Clima (Sol, lluvioso, nublado, ventarrón)	Pareja
M	M	N	S	C	S	S	N
F	H	S	S	F	S	L	S
F	H	S	N	F	N	N	S
M	H	N	N	C	N	V	N
M	H	N	S	L	S	V	N
F	M	S	N	P	S	N	N
F	P	S	S	P	S	S	S
F	P	S	N	C	S	S	S
M	P	S	S	L	N	L	S
F	M	S	S	F	N	L	S
M	H	S	N	L	N	N	N
F	M	N	S	P	S	N	N
M	P	N	N	C	N	V	S
M	P	N	N	C	S	S	S

Tomaremos en cuenta las siguientes formulas

Entropia



$$-\sum_{i=1}^k p_i \log_2(p_i)$$

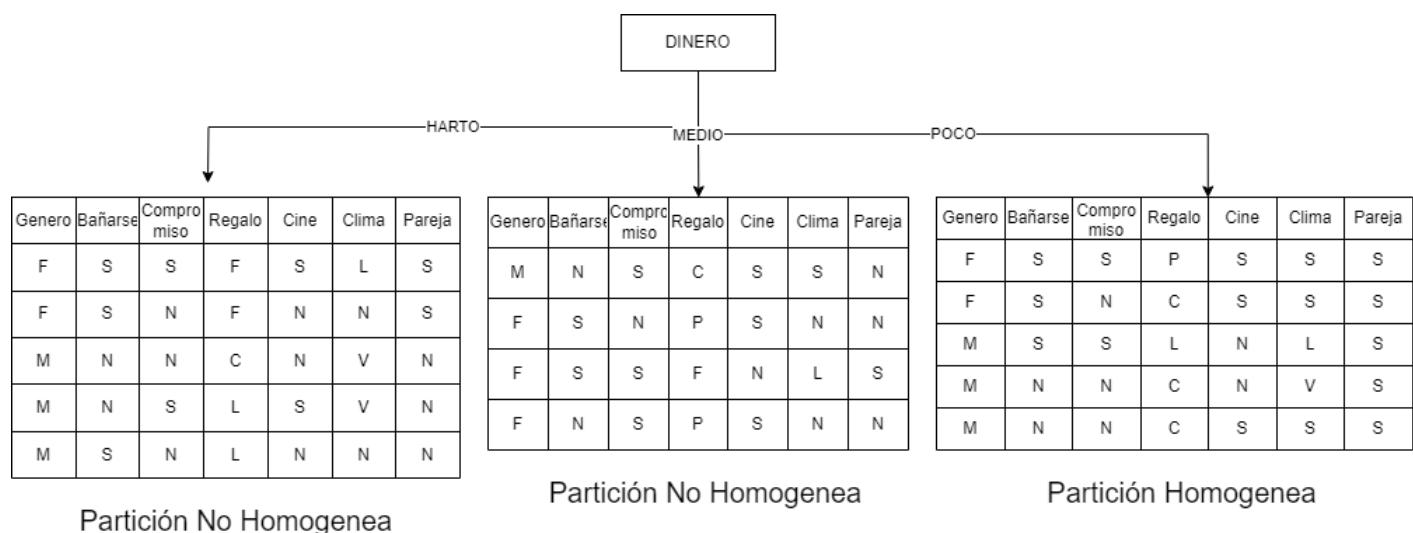
Ganancia de información

$$IG = I(Dp) - \sum_j \frac{N_j}{N_p} I(D_j)$$

Tenemos la siguiente tabla de entropías y de ganancia de información

VARIABLES	TOTAL	S	N	Entropia	Ganancia de informaci
TOTAL	14	8	6	0,9852281	
Genero					0,0610538
M	7	3	4	0,9852281	
F	7	5	2	0,8631206	
Dinero					0,4066663
H	5	2	3	0,9709506	
M	4	1	3	0,8112781	
P	5	5	0	0	
Bañarse					0,1280853
S	8	6	2	0,8112781	
N	6	2	4	0,9182958	
Compromiso					0
S	7	4	3	0,9852281	
N	7	4	3	0,9852281	
Regalo					0,2449047
P	3	1	2	0,9182958	
F	3	3	0	0	
C	5	3	2	0,9709506	
L	3	1	2	0,9182958	
Cine					0,0202442
S	8	4	4	1	
N	6	4	2	0,9182958	
Clima					0,324863
S	4	3	1	0,8112781	
L	3	3	0	0	
N	4	1	3	0,8112781	
V	3	1	2	0,9182958	

Procedemos a seleccionar el mejor atributo, en este caso el atributo con el que tenemos mayor Ganancia de Información es “Dinero”, procedemos a seleccionarlo como atributo de decisión para el nodo raíz así obtenemos lo siguiente:



Como se puede observar no todas las particiones son homogéneas, de manera que será necesario iterar el proceso con los atributos que todavía no se han tratado

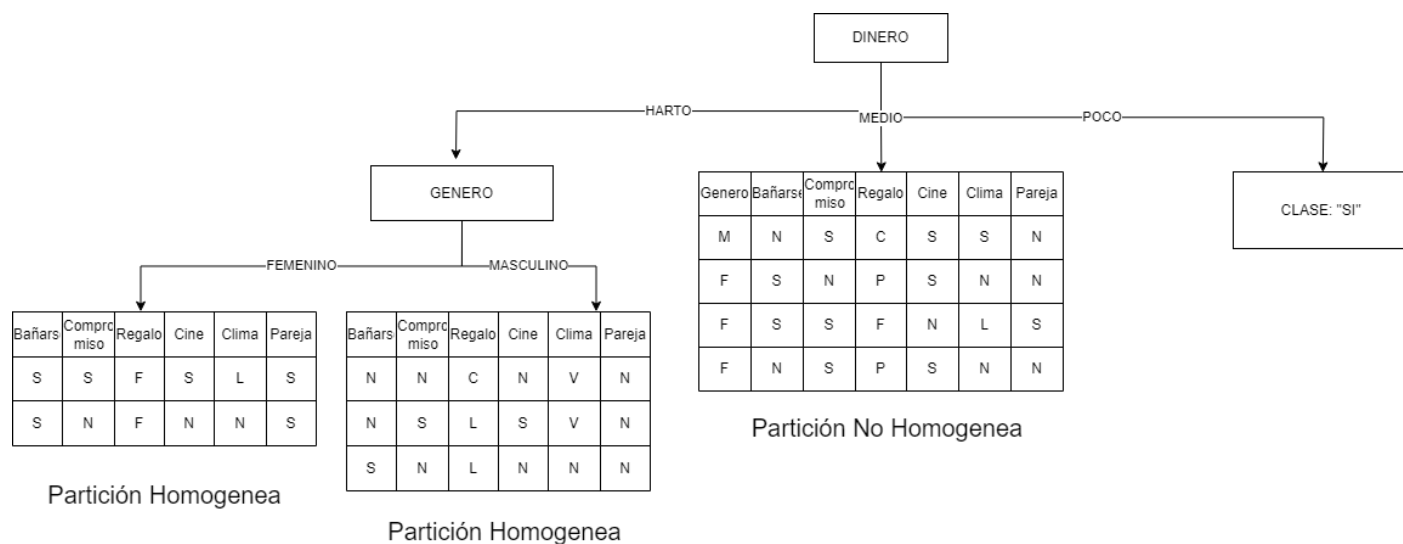
Comenzamos con el valor de “Harto Dinero” tenemos las siguientes tablas

Genero	Bañarse	Con Compromiso	Regalo (Peluche, flores, Chocolate)	Cine	Clima (Sol, lluvia o, nublado, ventarrón)	Pareja
F	S	S	F	S	L	S
F	S	N	F	N	N	S
M	N	N	C	N	V	N
M	N	S	L	S	V	N
M	S	N	L	N	N	N

VARIABLES	TOTAL	S	N	Entropia	Ganancia de informacion
TOTAL	5	2	3	0,970950594	
Genero					0,970950594
M	3	2	0	0	
F	2	0	3	0	
Bañarse					0,419973094
S	3	2	1	0,918295834	
N	2	0	2	0	
Compromiso					0,019973094
S	2	1	1	1	
N	3	1	2	0,918295834	
Regalo					0,970950594
P	0	0	0	0	
F	2	2	0	0	
C	1	0	1	0	
L	2	0	2	0	
Cine					0,019973094
S	2	1	1	1	
N	3	1	2	0,918295834	
Clima					0,570950594
S	0	0	0	0	
L	1	1	0	0	
N	2	1	1	1	
V	2	0	2	0	

Continuamos con el desarrollo del árbol

Podemos elegir entre “Genero” y “Regalo” por que son los atributos que nos ofrecen mayor ganancia de información



Aun tenemos una partición no homogénea por lo que será necesario iterar nuevamente para continuar con los demás atributos

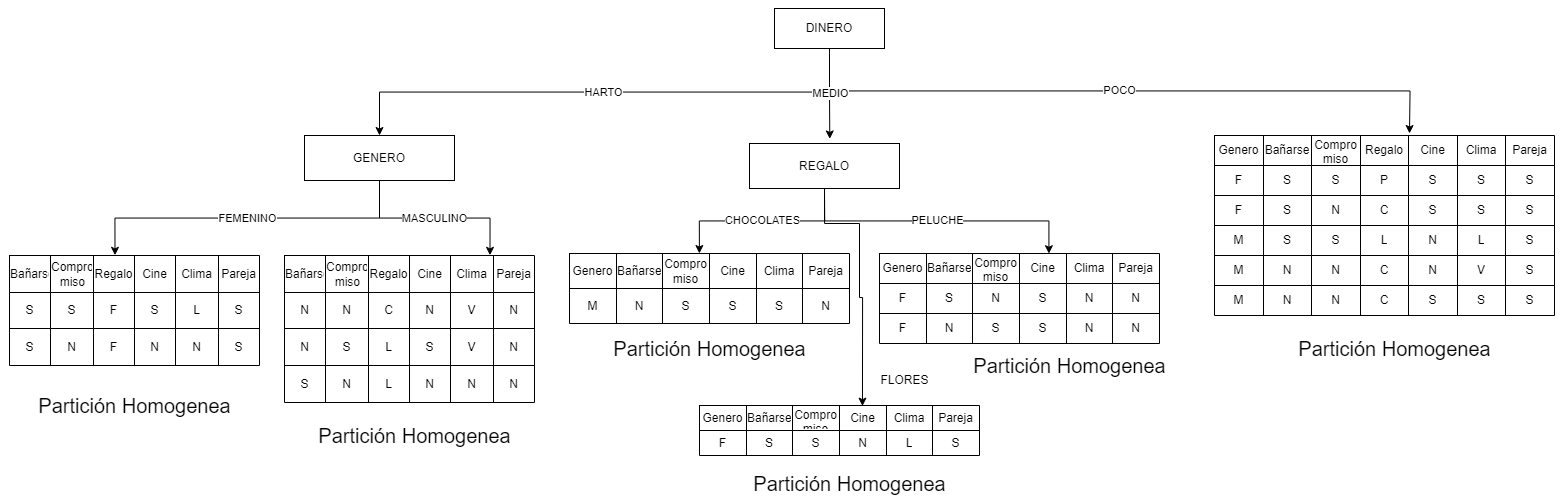
Comenzamos con el valor de "Dinero Medio" tenemos las siguientes tablas

Genero	Bañarse	Con Compromiso	Regalo (Peluche, flores, Chocolate s. Locion)	Cine	Clima (Sol, lluvias o, nublado, ventarron)	Pareja
M	N	S	C	S	S	N
F	S	N	P	S	N	N
F	S	S	F	N	L	S
F	N	S	P	S	N	N

VARIABLES	TOTAL	S	N	Entropia	Ganancia de informacion
TOTAL	4	1	3	0,811278124	
Genero					0,122556249
M	1	0	1	0	
F	3	1	2	0,918295834	
Bañarse					0,311278124
S	2	1	1	1	
N	2	0	2	0	
Compromiso					0,122556249
S	3	1	2	0,918295834	
N	1	0	1	0	
Regalo					0,811278124
P	2	0	2	0	
F	1	1	0	0	
C	1	0	1	0	
L	0	0	0	0	
Cine					0,811278124
S	3	0	3	0	
N	1	1	0	0	
Clima					0,811278124
S	1	0	1	0	
L	1	1	0	0	
N	2	0	2	0	
V	0	0	2	0	

Continuamos con el desarrollo del árbol

Podemos elegir entre “Cine”, “Regalo” y “Clima” por qué son los atributos que nos ofrecen mayor ganancia de información, escogeremos “Regalo”



ARBOL DE DECISION ALGORITMO ID3

