

Univ. Condori Quispe Diego Joel
12926609 L.P.

Tenemos el dataset de citas, donde tenemos las variables categoricas

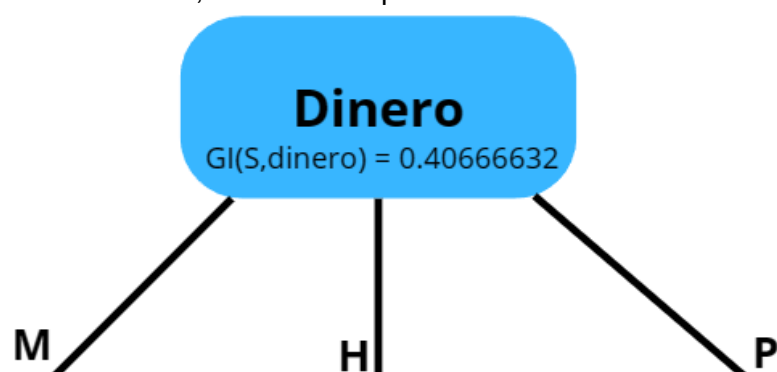
- genero(M: Masculino, F: Femenino)
- dinero (Harto, Medio, Poco)
- bañarse (S: Si, N: No)
- comprometido (S: Si, N: No)
- regalo (C: Chocolate, F: Flores, P: Peluche, L: Locion)
- cine (S: Si, N: No)
- clima (S: Sol, L: Lluvia, V: Ventarron, N: Nublado)
- Pareja (S: Si, N: No)

genero	dinero(Harto, medio, po	bañarse	compror	regalo (f	cine	clima(Sol,llu	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
M	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

Al sacar la entropia de cada variable categorica con respecto a la variable objetivo Pareja, como tambien la ganancia de informacion, tenemos:

TOTAL	SI	NO	Entropia		G.I.	
	14	8	6	H(S)	0.985228136	
Genero						GI(S,Genero)
M	7	3	4	H(S_M)	0.985228136	
F	7	5	2	H(S_F)	0.8631205686	
Dinero						GI(S,Dinero)
H	5	2	3	H(S_H)	0.9709505945	
M	4	1	3	H(S_M)	0.8112781245	
P	5	5	0	H(S_P)	0	
Bañarse						GI(S,Bañarse)
S	8	6	2	H(S_S)	0.8112781245	
N	6	2	4	H(S_N)	0.9182958341	
Comprometido						GI(S,Comprometido)
S	7	4	3	H(S_S)	0.985228136	
N	7	4	3	H(S_N)	0.985228136	
Regalos						GI(S,Regalos)
P	3	1	2	H(S_P)	0.9182958341	
F	3	3	0	H(S_F)	0	
C	5	3	2	H(S_C)	0.9709505945	
L	3	1	2	H(S_L)	0.9182958341	
Cine						GI(S,Cine)
S	8	4	4	H(S_S)	1	
N	6	4	2	H(S_N)	0.9182958341	
Clima						GI(S, Clima)
S	4	3	1	H(S_S)	0.8112781245	
L	3	3	0	H(S_L)	0	
N	4	1	3	H(S_N)	0.8112781245	
V	3	1	2	H(S_V)	0.9182958341	

Quien tiene mas ganancia de información es el dinero, ya que es el valor más informativo y relevante para tomar decisiones, será el nodo padre:



Ahora como usaremos el algoritmo ID3 para arboles de decision, por cada rama vamos a realizar el mismo proceso de búsqueda de la mayor ganancia de información considerando el hecho de que por cada rama, digamos M, en nuestro dataset inicial debemos tomar en cuenta todas las filas que tengan M de la columna dinero y eliminar la columna dinero y volver a sacar la entropía. Empezaremos por M:

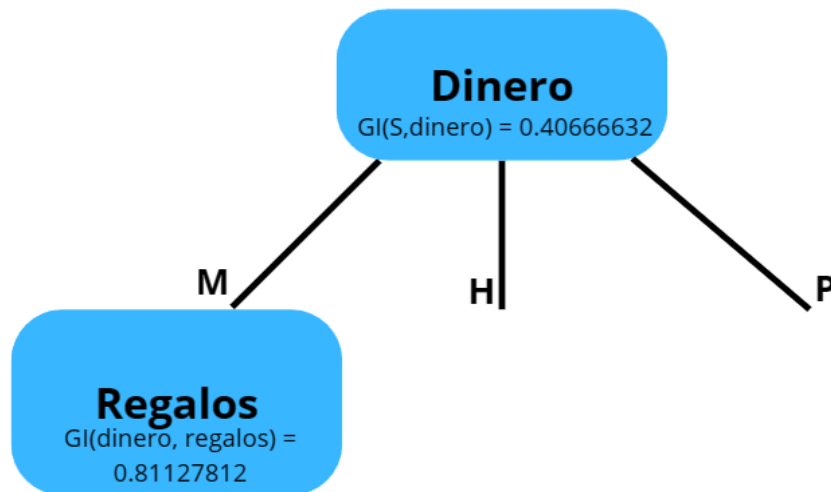
Rama M(Mucho):

De dinero	M(mucho)					
genero	bañarse	compromet	regalo (Pelucine		clima(Sol,llu	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

Sacamos las entalpías y luego la ganancia de información, la entropía de conjunto de datos $H(S)$ ya no será la de parejas, sino la de dinero $H(\text{dinero})$, por lo cual, por cada variable categórica tendremos que sacar la Ganancia de información $GI(\text{dinero, variable_categorica})$, ya que dinero es el nodo padre.

TOTAL	SI	NO	Entropia			
	4	1	3	$H(\text{Dinero})$	0.8112781245	
Genero					$GI(\text{Dinero, Genero})$	0.12255624
M	1	0	1	$H(S_M)$	0	
F	3	1	2	$H(S_F)$	0.9182958341	
Bañarse					$GI(\text{Dinero, Bañarse})$	0.31127812
S	2	1	1	$H(S_S)$	1	
N	2	0	2	$H(S_N)$	0	
Comprometido					$GI(\text{Dinero, Comprometido})$	0.12255624
S	3	1	2	$H(S_S)$	0.9182958341	
N	1	0	1	$H(S_N)$	0	
Regalos					$GI(\text{Dinero, Regalos})$	0.81127812
P	2	0	2	$H(S_P)$	0	
F	1	1	0	$H(S_F)$	0	
C	1	0	1	$H(S_C)$	0	
L	0	0	0	$H(S_L)$	0	
Cine					$GI(\text{Dinero, Cine})$	0.81127812
S	3	0	3	$H(S_S)$	0	
N	1	1	0	$H(S_N)$	0	
Clima					$GI(\text{Dinero, Clima})$	0.81127812
S	1	0	1	$H(S_S)$	0	
L	1	1	0	$H(S_L)$	0	
N	2	0	2	$H(S_N)$	0	
V	0	0	0	$H(S_V)$	0	

Podemos tomar entre Regalos, Cine, clima ya que tienen la mayor cantidad de entropía, de todos modos llegaríamos a lo mismo, por lo que tomaremos regalos



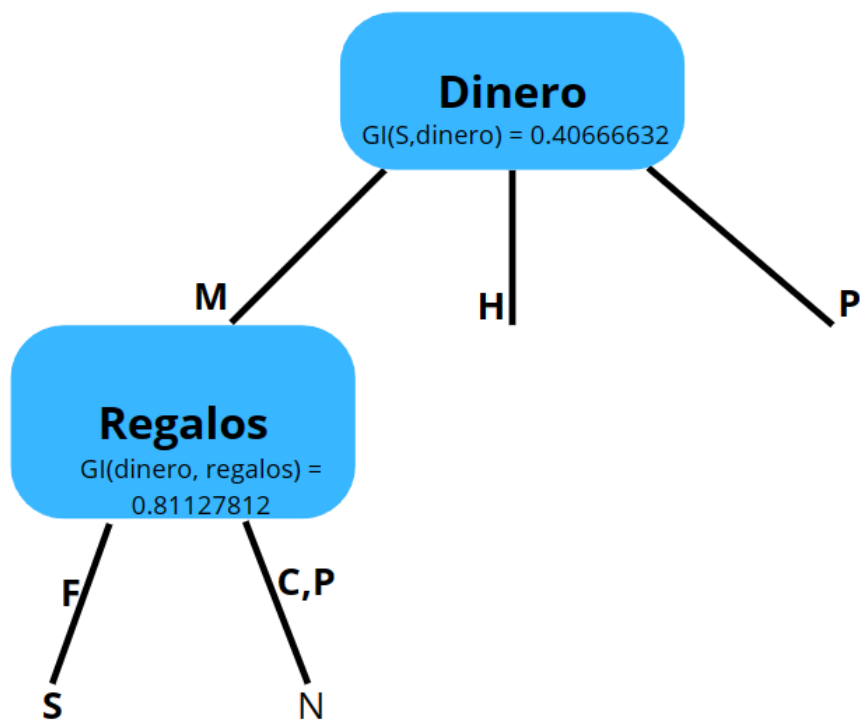
Antes de eliminar las filas y volver a sacar la ganancia de información para determinar las hojas, notemos de la columna regaló con respecto a la variable objetivo que es Pareja:

regalo	Pareja
C	N
P	N
F	S
P	N

Para todos los casos ya está definido un único valor, es decir tanto para C, como para P el resultado es N, para P solo su resultado es N, y para F solo es S, entonces simplificando:

Regalo	Pareja
C,P	N
F	S

Por tanto en nuestro árbol:



Ahora nos centraremos en la rama H (Harto), así que solo tomaremos en cuenta las filas que tengan H en la columna Dinero y eliminamos la columna dinero:

De Dinero	H(Harto)					
genero	bañarse	compromet	regalo (Pelu	cine	clima(Sol,llu	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

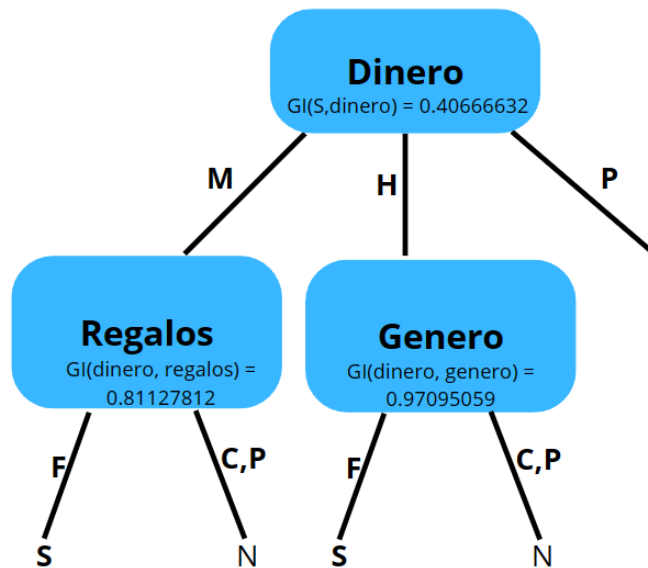
Sacamos la entropía de cada columna y la ganancia de información:

TOTAL		SI	NO	Entropia			
	5	2	3	H(Dinero)	0.9709505945		
Genero						GI(Dinero, Genero)	0.97095059
M	3	0	3	H(S_M)	0		
F	2	2	0	H(S_F)	0		
Bañarse						GI(Dinero, Bañarse)	0.41997309
S	3	2	1	H(S_S)	0.9182958341		
N	2	0	2	H(S_N)	0		
Comprometido						GI(Dinero, Comprometido)	0.01997309
S	2	1	1	H(S_S)	1		
N	3	1	2	H(S_N)	0.9182958341		
Regalos						GI(Dinero, Regalos)	0.97095059
P	0	0	0	H(S_P)	0		
F	2	2	0	H(S_F)	0		
C	1	0	1	H(S_C)	0		
L	2	0	2	H(S_L)	0		
Cine						GI(Dinero, Cine)	0.97095059
S	2	1	1	H(S_S)	1		
N	3	1	2	H(S_N)	0.9182958341		
Clima						GI(Dinero, Clima)	0.57095059
S	0	0	0	H(S_S)	0		
L	1	1	0	H(S_L)	0		
N	2	1	1	H(S_N)	1		
V	2	0	2	H(S_V)	0		

Podemos elegir, entre cine, genero, regalos ya que tienen la mayor cantidad de entropía, pero a género a comparación de cine, regalos no tendremos que seguir buscando más hojas, de tomar uno de estos a posteriori implica realizar una poda del árbol, así que tomaremos genero, pero hay un detalle en genero, para sus valores F y M son únicos con respecto a la variable objetivo, es el motivo por el cual se lo elijo:

genero	Pareja	genero	Pareja
F	S	F	S
F	S	M	N
M	N		
M	N		
M	N		

Así que nuestro árbol quedaría:



Ahora para la rama P (poco), tomamos las filas que tengan el valor P en la columna dinero y luego eliminamos la columna dinero, debemos notar algo en este caso, para todos los casos, todos nos llevaran a S así que no se sacaran la ganancia de información ya que todos nos llevan al mismo resultado:

De dinero	P(poco)						
genero	dinero(Harto, medio, poco)	bañarse	compromet	regalo (Pelucine		clima(Sol,llu	Pareja
F	P	S	S	P	S	S	S
M	P	S	N	C	S	S	S
M	P	S	S	L	N	L	S
M	P	N	N	C	N	V	S
M	P	N	N	C	S	S	S

Por tanto nuestro árbol finalmente quedaría de la siguiente forma:

