Vidhi Kansala

CWID: 20029496

HW-04 – CART

Dataset:

	Occupation	Gender	Age.	Salary	Level (T)
4.1	Service	F	45	48k	L3
		M	25	25k	L1
	Mngment	M	33	35k	L2
	"	M	25	45k	L3
	"	F	35	65k	L4
	"	M	26	45k	L3
	"	F	45	~~65~~k	L4
	Sales	F	40	~~45~~k	L3
		M	30	40k	L2
	staff	F	50	40k	L2
		M	25	25k	L1

	P(j)	$-P_j \log P_j$
~~Level~~ L1	2/11	0.447
	3/11	0.511
	4/11	0.531
	2/11	0.447
		Σ 1.936.

	PL	PR.	P(j/tL) L1	L2	L3	L4
~~Split~~: Age <= 30	0.455	0.545	0.4			

Category	Split	PL	PR.	Level	P(j/tL)	P(j/tR)	2PL×PR
Age.	<= 30	0.45	0.545	L1	0.4	0.	0.496
				L2	0.2	0.3	
				L3	0.4	0.3	
				L4	0.0	0.3	
	~~35-40~~	0.273	0.727	L1	6	0.25	0.397
	<=40			L2	0.3	0.25	
				L3	0.3	0.375	
				L4	0.3	0.125	
	<=50	0.273	0.727	L1	0	0.25	0.397
				L2	0.3	025	
				L3	0.3	0.375	
				L4	0.3	0.125	
Occupation	~~Sales~~	0.273	0.727	L1	0.3	0.125	0.397
	Service			L2	0.3	0.250	
				L3	0.3	0.375	
				L4	0	0.25	
	Mngmnt	0.636		L1	0	0.286	0.463
		0.364		L2	0	0.429	
				L3	0.5	0.286	
				L4	0.5	0.0	
	Sales	0.182	0.818	L1	0	0.2	0.298
				L2	0.5	0.5	
				L3	0.5	0.3	
				L4	0	0.2.	

$\phi_L(s/t)$	$\phi_R(s/t)$	P(j)	$-P_j \log(P_j)$	$\Sigma -P_j \log P_j$	PL%	$(\Sigma -P_j \log P_j)$%
0.933	0.468	0.4	0.529	1.522	0.455	
		0.2	0.464			0.692
		0.4	0.529			
		0	0.			
0583	0.231	6	0	1.585	0.273	0.432
		0.3	0.528			
		0.3	0.528			
		0.3	0.528			
0.583	0.231	0	0	1.585	0.273	0.432
		03	0.528			
		0.3	0.528			
		0.3	0.528			
0.583	0.231	0.3	0.528	1.585	0.273	0.432
		03	0.528			
		0.3	0.528			
		0	0			
1.429	0.661	0	0	1	0.364	0.364
		0.	0			
		0.5	0.5			
		0.5	0.5			
0.889	0.264	0.0	0	1	0.182	0.182
		0.5	0.5			
		0.5	0.5			
		0	0			

Category	Split by	PL	PR	Level	P(j/tL)	P(j/tR)	2PLx PIR	q(s/t)	Φ(s/t)	P(j)	-P(j)logP(j)	ΣPjlogPj	PL%	(ΣPjlogPj)
	Staff	0.182	0.818	L1	0.5	0.11	0.298	1.33	0.397	0.5	0	1	0182	0.182
				L2	0.5	0.22				05	0.5			
				L3	0.5	0.44				0	0.5			
				L4	0	0.22				0	0			
Gender	Male	0.545	0.455	L1	0.3	0	0.496	0.933	0463	03	0.528	1.585	0.545	0.865
				L2	0.3	0.2				03	0.528			
				L3	0.3	0.4				0.3	0.523			
				L4	0	0.4				0	0			
	Female	0.455	0.545	L1	0.0	0.3	0.496	0.933	0.468	0.0	0.0	1.522	0.455	
				L2	0.2	0.3				0.2	0.464			0.692
				L3	0.4	0.3				0.4	0529			
				L4	0.4	0.				0.4	0529			

Optimal Split: services = Management