

4.1 Construct a classification and regression tree to classify salary based on the other variables only one split level.

Name: Ali Abdullah Ahmad

CWID: 20031246

HW_04_CART

Category to split	Split	PL	PR	Level	P(j tL)	P(j tR)	2PL * PR	q(s t)	Φ(s t)
Age	Less than or equal to 30	0.455	0.545	1	0.400	0.000	0.496	0.933	0.463
				2	0.200	0.333			
				3	0.400	0.333			
				4	0.000	0.333			
	Less than or equal to 40	0.273	0.727	1	0.000	0.250	0.397	0.583	0.231
				2	0.333	0.250			
				3	0.333	0.375			
				4	0.333	0.125			
	Less than or equal to 50	0.273	0.727	1	0.000	0.250	0.397	0.583	0.231
				2	0.333	0.250			
				3	0.333	0.375			
				4	0.333	0.125			
Occupation	Service	0.273	0.727	1	0.333	0.125	0.397	0.583	0.231
				2	0.333	0.250			
				3	0.333	0.375			
				4	0.000	0.250			
	Management	0.364	0.636	1	0.000	0.286	0.463	1.429	0.661
				2	0.000	0.429			
				3	0.500	0.286			
				4	0.500	0.000			
	Sales	0.182	0.818	1	0.000	0.222	0.298	0.889	0.264
				2	0.500	0.222			
				3	0.500	0.333			
				4	0.000	0.222			
	Staff	0.182	0.818	1	0.500	0.111	0.298	1.333	0.397
				2	0.500	0.222			
				3	0.000	0.444			
				4	0.000	0.222			
Gender	Male	0.545	0.455	1	0.333	0.000	0.496	0.933	0.463
				2	0.333	0.200			
				3	0.333	0.400			
				4	0.000	0.400			
	Female	0.455	0.545	1	0.000	0.333	0.496	0.933	0.463
				2	0.200	0.333			
				3	0.400	0.333			
				4	0.400	0.000			

Occupation	Gender	Age	Salary (\$)	Level	Age group
Service	Female	45	48000	3	Less than or equal to 50
Service	Male	25	25000	1	Less than or equal to 30
Service	Male	33	35000	2	Less than or equal to 40
Management	Male	25	45000	3	Less than or equal to 30
Management	Female	35	65000	4	Less than or equal to 40
Management	Male	26	45000	3	Less than or equal to 30
Management	Female	45	70000	4	Less than or equal to 50
Sales	Female	40	50000	3	Less than or equal to 40
Sales	Male	30	40000	2	Less than or equal to 30
Staff	Female	50	40000	2	Less than or equal to 50
Staff	Male	25	25000	1	Less than or equal to 30

Total row 11

Answer – The first Classification and Regression Treen node on the basis of Salary would be Occupation since the highest $\phi(s|t)$ value fall under **Occupation = Management** ie of **0.661** among all splits.