

PROCESS MINING

ASSIGNMENT

NO 1

Name : Zubair Afzal

Section : G

ROLL NO : 221-2591

- Input Variables:
 - Job level ✓
 - Employee Recognition ✗
 - Company size ✗
 - Company Reputation ✗
 - Martial Status ✗
 - Remote Work ✓
 - Performance Rating ✗
 - Work Life Balance ✓
 - Gender
 - Overtime

- Target Variable : Attrition

- Attrition:

Left = 22 , Stayed = 29 , Total = 51

$$E = \left[-\frac{22}{51} \log_2 \left(\frac{22}{51} \right) - \frac{29}{51} \log_2 \left(\frac{29}{51} \right) \right]$$

$$E = 0.9864$$

Entropy of the target variable "Attrition" is 0.9864.

①

• GENDER:

	Left	Stayed	Total
Male	15	12	27
Female	7	17	24

$$E_{\text{Male}} = \left[-\frac{15}{27} \log_2 \left(\frac{15}{27} \right) - \frac{12}{27} \log_2 \left(\frac{12}{27} \right) \right] \\ = 0.991$$

$$E_{\text{Female}} = \left[-\frac{7}{24} \log_2 \left(\frac{7}{24} \right) - \frac{17}{24} \log_2 \left(\frac{17}{24} \right) \right] \\ = 0.8709$$

$$W \cdot E = E_{\text{Male}} + E_{\text{Female}} \\ = \frac{27}{51} \times 0.991 + \frac{24}{51} \times 0.8709 \\ = 0.9345$$

$$\text{Info. Gain} = 0.9864 - \boxed{0.0519} \quad 0.9345$$

Info-Gain
 Gender = 0.0519

• Work Life balance:

	Left	Stayed	Total
Poor	2	6	8
Excellent	5	2	7
Good	16	11	27
Fair	5	10	15

$$\boxed{\text{Info-Gain} = 0.070}$$

• Job Level:

	Left	Stayed	Total
Mid	10	10	20
Entry	19	5	24
Senior	0	7	7

$$E = \frac{20}{51} \left[-\frac{10}{20} \log_2 \left(\frac{10}{20} \right) - \frac{10}{20} \log_2 \left(\frac{10}{20} \right) \right] + \frac{24}{51} \left[-\frac{19}{24} \log_2 \left(\frac{19}{24} \right) - \frac{5}{24} \log_2 \left(\frac{5}{24} \right) \right] + \frac{7}{51} [0]$$

$$E = \frac{20}{51} [1] + \frac{24}{51} [0.7383] + 0$$

Weighted Entropy = 0.7396

$$\boxed{\text{Info-Gain} = 0.2468}$$

• Remote Work:

No	Left	Stayed	Total
Yes	26	12	38
	3	10	13

$$E = \frac{38}{51} \left[-\frac{26}{38} \log_2 \left(\frac{26}{38} \right) - \frac{12}{38} \log_2 \left(\frac{12}{38} \right) \right] + \frac{13}{51} \left[-\frac{3}{13} \log_2 \left(\frac{3}{13} \right) - \frac{10}{13} \log_2 \left(\frac{10}{13} \right) \right]$$

$$\begin{aligned}
 E &= \frac{8}{51} \left[-\frac{2}{8} \log_2 \left(\frac{2}{8} \right) - \frac{6}{8} \log_2 \left(\frac{6}{8} \right) \right] + \frac{7}{51} \left[-\frac{5}{7} \log_2 \left(\frac{5}{7} \right) \right. \\
 &\quad \left. - \frac{2}{7} \log_2 \left(\frac{2}{7} \right) \right] + \frac{21}{51} \left[-\frac{10}{21} \log_2 \left(\frac{10}{21} \right) - \frac{11}{21} \log_2 \left(\frac{11}{21} \right) \right] \\
 &\quad + \frac{15}{51} \left[-\frac{5}{15} \log_2 \left(\frac{5}{15} \right) - \frac{10}{15} \log_2 \left(\frac{10}{15} \right) \right] \\
 &= \frac{8}{51} (0.8113) + \frac{7}{51} (0.8631) + \frac{21}{51} (0.9989) + \\
 &\quad \frac{15}{51} (0.9183)
 \end{aligned}$$

Weighted Entropy = 0.9269

Info-Gain = 0.0595

• Performance Rating:

	Left	Stayed	Total
Average	15	18	33
High	5	3	8
Below Average	2	5	7
Low	0	3	3

$$\begin{aligned}
 E &= \frac{33}{51} \left[-\frac{15}{33} \log_2 \left(\frac{15}{33} \right) - \frac{18}{33} \log_2 \left(\frac{18}{33} \right) \right] + \frac{8}{51} \left[-\frac{5}{8} \log_2 \left(\frac{5}{8} \right) \right. \\
 &\quad \left. - \frac{3}{8} \log_2 \left(\frac{3}{8} \right) \right] + \frac{7}{51} \left[-\frac{2}{7} \log_2 \left(\frac{2}{7} \right) - \frac{5}{7} \log_2 \left(\frac{5}{7} \right) \right] + \frac{3}{51} [0] \\
 &= \frac{33}{51} [0.990] + \frac{8}{51} [0.9544] + \frac{7}{51} [0.8631] + 0
 \end{aligned}$$

Weighted Entropy = 0.9114

$$E = \frac{38}{51} [0.8997] + \frac{13}{51} [0.7793]$$

$$= 0.8691$$

Weighted Entropy = 0.8691

Info. Gain = 0.1173

Overtime:

	Left	Stayed	Total
No	22	13	35
Yes	7	9	16

$$E = \frac{35}{51} \left[-\frac{22}{35} \log_2 \left(\frac{22}{35} \right) - \frac{13}{35} \log_2 \left(\frac{13}{35} \right) \right] + \frac{16}{51} \left[-\frac{7}{16} \log_2 \left(\frac{7}{16} \right) - \frac{9}{16} \log_2 \left(\frac{9}{16} \right) \right]$$

$$E = \frac{35}{51} [0.9518] + \frac{16}{51} [0.9864]$$

Weighted = 0.9634
Entropy

Info. Gain = 0.0230

Employee Recognition:

Medium	Left	Stayed	Total
Low	10	4	14
High	10	12	22
Very high	6	6	12
	3	0	3

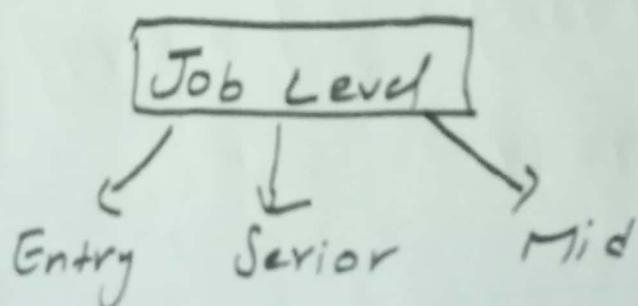
$$E = -\frac{14}{51} \left[-\frac{10}{14} \log_2 \left(\frac{10}{14} \right) - \frac{4}{14} \log_2 \left(\frac{4}{14} \right) \right] + \frac{22}{51} \left[-\frac{10}{22} \log_2 \left(\frac{10}{22} \right) - \frac{12}{22} \log_2 \left(\frac{12}{22} \right) \right] + \frac{12}{51} [1] + \frac{3}{51} [0]$$

$$E = \frac{14}{51} [0.8631] + -\frac{4}{14} [0.9940] + \frac{12}{51} + 0$$

Weighted Entropy = 0.9010

Info. Gain = 0.0853

⇒ **JOB LEVEL** has the highest Information gain, so it is **Root Node**



② ENTRY SPLIT

- Remote Work

	Left	Stayed	Total
No	16	3	19
Yes	3	2	5

$$\begin{aligned}
 E &= \frac{19}{57} \left[-\frac{16}{19} \log_2 \left(\frac{16}{19} \right) - \frac{3}{19} \log_2 \left(\frac{3}{19} \right) \right] + \\
 &\quad \frac{5}{57} \left[-\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) \right] \\
 &= \frac{19}{57} [0.6292] + \frac{5}{57} [0.9710]
 \end{aligned}$$

Weighted Entropy = 0.7064

Info. Gain = 0.03778

• Company Size:

	Left	Stayed	Total
Medium			
Large	13	2	15
Small	4	0	4
	2	8	5

$$\begin{aligned}
 E &= \frac{15}{57} \left[-\frac{13}{15} \log_2 \left(\frac{13}{15} \right) - \frac{2}{15} \log_2 \left(\frac{2}{15} \right) \right] + \\
 &\quad \frac{4}{57} [0] + \frac{5}{57} \left[-\frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} \log_2 \left(\frac{3}{5} \right) \right] \\
 &= \frac{15}{57} [0.5665] + 0 + \frac{5}{57} [0.9710]
 \end{aligned}$$

Weighted Entropy = 0.5563

Info. Gain = 0.1819

• Work life balance:

	Left	Stayed	Total
Good	8	3	11
Excellent	2	1	3
Poor	2	0	2

$$\begin{aligned}
 E &= \frac{11}{57} \left[-\frac{8}{11} \log_2 \left(\frac{8}{11} \right) - \frac{3}{11} \log_2 \left(\frac{3}{11} \right) \right] + \frac{3}{57} \\
 &\quad \left[-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right] + 0 \\
 &= \frac{11}{57} [0.8454] + \frac{3}{57} [0.5436] + 0
 \end{aligned}$$

Weighted Entropy = 0.6834

Info. Gain = 0.0549

• Gender

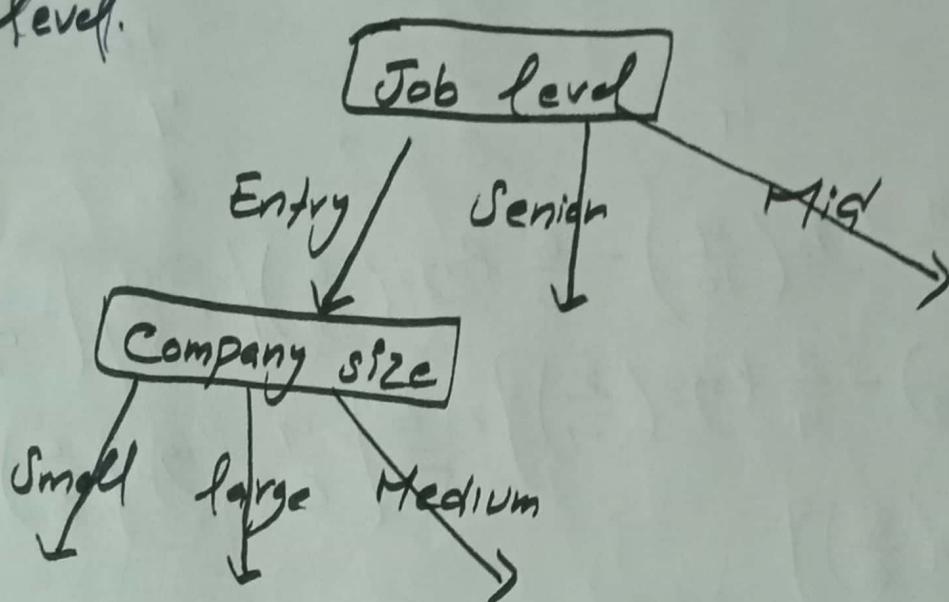
	Left	Stayed	Total
Female	12	1	13
Male	7	4	11

$$\begin{aligned}
 E &= \frac{13}{57} \left[-\frac{12}{13} \log_2 \left(\frac{12}{13} \right) - \frac{1}{13} \log_2 \left(\frac{1}{13} \right) \right] + \\
 &\quad \frac{11}{57} \left[-\frac{7}{11} \log_2 \left(\frac{7}{11} \right) - \frac{4}{11} \log_2 \left(\frac{4}{11} \right) \right] \\
 &= \frac{13}{57} [0.3912] = \frac{11}{57} [0.9457]
 \end{aligned}$$

Weighted Entropy = 0.6959

Info. Gain = 0.0929

→ Company size has high information gain,
so it is now root node for entry Job
level.



③ Mid split:

- Remote Work:

No	Left	Stayed	Total
Yes	10	4	14
No	0	6	6

$$E = \frac{14}{51} \left[-\frac{10}{14} \log_2 \left(\frac{10}{14} \right) - \frac{4}{14} \log_2 \left(\frac{4}{14} \right) \right] + \frac{6}{51} [0]$$

$$E = \frac{14}{51} [0.8631] + 0$$

$$\text{Weighted Entropy} = 0.6042$$

$$\text{Info-Gain} = 0.3958$$

• Company Size:

	Left	Stayed	Total
Medium	4	6	10
Large	2	2	4
Small	4	2	6

$$\begin{aligned}
 E &= \frac{10}{51} \left[-\frac{4}{10} \log_2 \left(\frac{4}{10} \right) - \frac{6}{10} \log_2 \left(\frac{6}{10} \right) \right] + \frac{4}{51} \left[-\frac{2}{4} \log_2 \left(\frac{2}{4} \right) - \frac{2}{4} \log_2 \left(\frac{2}{4} \right) \right] + \frac{6}{51} \left[-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right] \\
 &= \frac{10}{51} [0.9710] + \frac{4}{51} [0] + \frac{6}{51} [0.9183]
 \end{aligned}$$

Weight Entropy = 0.9610

Info-Gain = 0.0390

• Work Life Balance:

	Left	Stayed	Total
Poor	4	0	4
Excellent	0	2	2
Good	3	4	7
Fair	3	4	7

$$\begin{aligned}
 E &= \frac{4}{51} \left[-\cancel{\frac{4}{4} \log_2 0} \right] + \frac{2}{51} [0] + \frac{2}{51} \left[-\frac{3}{7} \log_2 \left(\frac{3}{7} \right) - \frac{4}{7} \log_2 \left(\frac{4}{7} \right) \right] \\
 &\quad + \frac{7}{51} \left[-\frac{3}{7} \log_2 \left(\frac{3}{7} \right) - \frac{4}{7} \log_2 \left(\frac{4}{7} \right) \right]
 \end{aligned}$$

$$= 0 + 0 + \frac{2}{51} [0.9852] + \frac{2}{51} [0.9852]$$

Weighted Entropy = 0.6897

[Info. Gain = 0.3103]

• Gender:

	Male	Left	Stayed	Total
Female	5	5	8	13
			2	
				13

$$E = -\frac{13}{51} \left[-\frac{5}{13} \log_2 \left(\frac{5}{13} \right) - \frac{8}{13} \log_2 \left(\frac{8}{13} \right) \right] + \frac{2}{51} \left[-\frac{5}{7} \log_2 \left(\frac{5}{7} \right) - \frac{2}{2} \log_2 \left(\frac{2}{7} \right) \right]$$

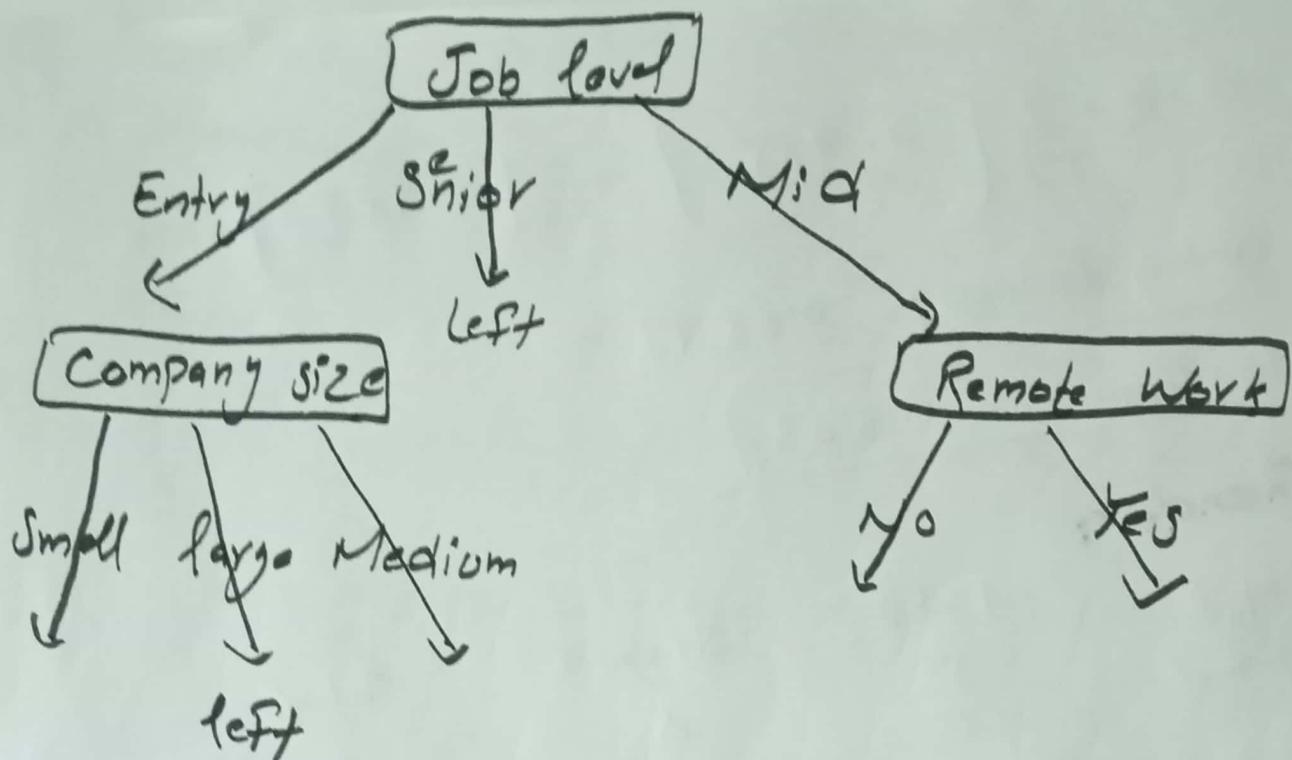
$$= \frac{13}{51} [0.9612] + \frac{2}{51} (0.8631)$$

Weighted Entropy = 0.9269

[Info. Gain = 0.0731]

\Rightarrow Remote work has high information gain so it is root node now for mid job level ~~not split~~

- Work life balance:



④ Company Size Small Split:

- Remote work:

	Left	Stayed	Total
Yes	0	4	4
No	6	3	9

$$\begin{aligned}
 E &= \frac{4}{51} [-0] + \frac{9}{51} \left[-\frac{6}{9} \log_2 \left(\frac{6}{9} \right) - \frac{3}{9} \log_2 \left(\frac{3}{9} \right) \right] \\
 &= 0 + \frac{9}{51} [0.9183]
 \end{aligned}$$

Weighted Entropy = 0.6357

Info-Gain = 0.3600

• Work life balance:

	Left	Stayed	Total
Fair	2	1	3
Good	2	4	6
Excellent	0	2	2
Poor	2	0	2

$$\begin{aligned}
 E &= \frac{3}{51} \left[-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right] + \frac{6}{51} \left[-\frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right. \\
 &\quad \left. - \frac{4}{6} \log_2 \left(\frac{4}{6} \right) \right] + \frac{2}{51} [0] + \frac{2}{51} [0] \\
 &= \frac{3}{51} [0.9183] + \frac{6}{51} [0.9183] + 0 + 0
 \end{aligned}$$

Weighted Entropy = 0.6337

Info-Gain = 0.3600

• Gender

	Left	Stayed	Total
Male	3	6	9
Female	3	1	4

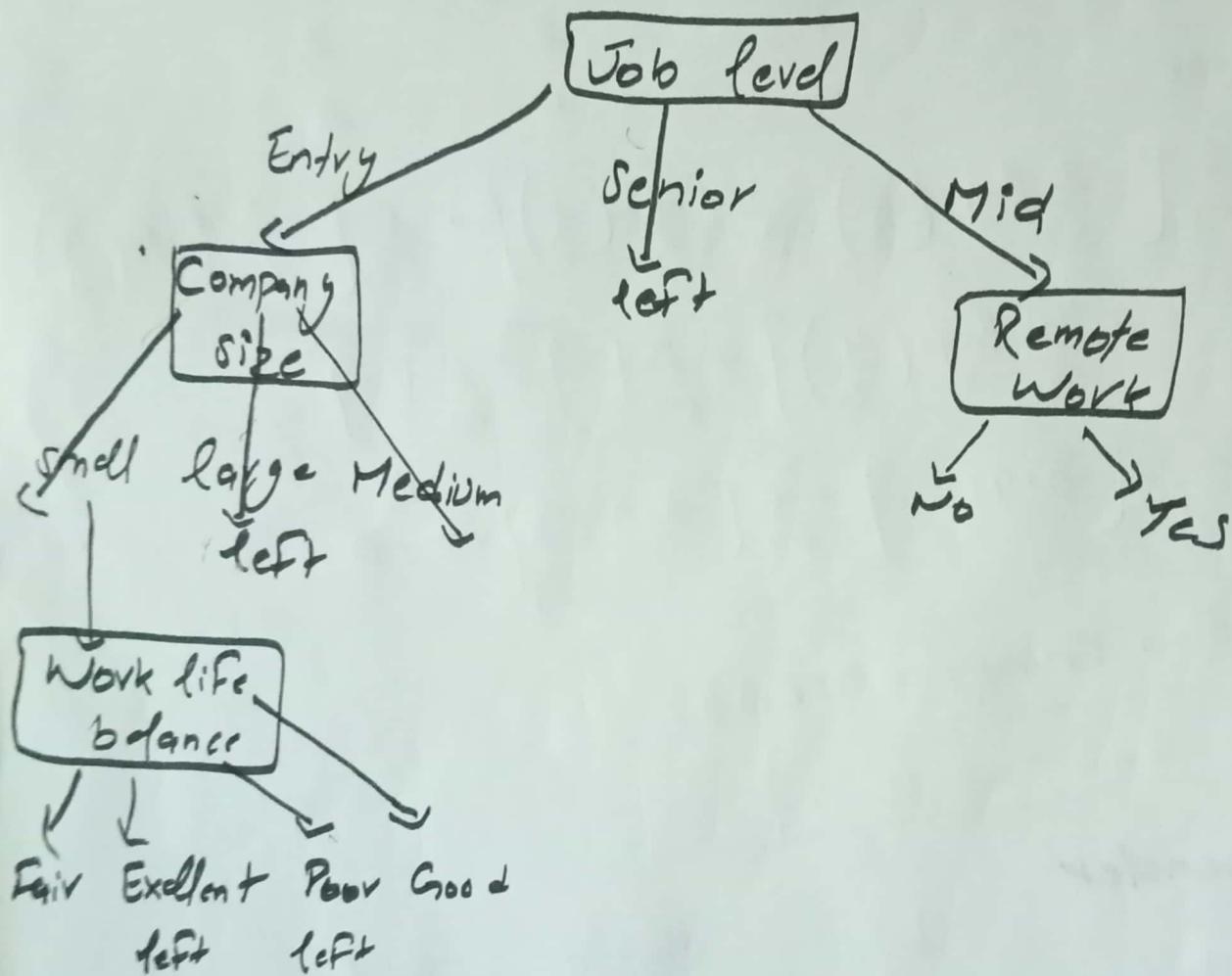
$$\begin{aligned}
 E &= \frac{9}{51} \left[-\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) \right] + \\
 &\quad \frac{4}{51} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right]
 \end{aligned}$$

$$E = \frac{9}{51} [0.9183] + \frac{4}{51} [0.8113]$$

Weighted Entropy = 0.1985

$$\text{Info-Gain} = 0.1104$$

⇒ Work life balance has high information gain, so it is root node now.



⑤ Remote work No split:

- ~~Company size~~

	Left	Stayed	Total
Medium	15	9	24
Small	6	3	9
Large	5	0	5

$$E = \frac{24}{57} \left[-\frac{15}{24} \log_2 \left(\frac{15}{24} \right) - \frac{9}{24} \log_2 \left(\frac{9}{24} \right) \right] + \frac{9}{57} \left[-\frac{6}{9} \log_2 \left(\frac{6}{9} \right) - \frac{3}{9} \log_2 \left(\frac{3}{9} \right) \right] + \frac{5}{57} [0]$$

Weighted Entropy = 0.8203

$$\boxed{\text{Info. Gain} = 0.0795}$$

• Company Size:

	Left	Stayed	Total
Medium	15	9	24
Small	6	3	9
Large	5	0	5

$$E = \frac{24}{57} \left[-\frac{15}{24} \log_2 \left(\frac{15}{24} \right) - \frac{9}{24} \log_2 \left(\frac{9}{24} \right) \right] + \frac{9}{57} \left[-\frac{6}{9} \log_2 \left(\frac{6}{9} \right) - \frac{3}{9} \log_2 \left(\frac{3}{9} \right) \right] + \frac{5}{57} [0]$$

$$= \frac{24}{57} [0.9544] + \frac{9}{57} [0.9183] + 0$$

Weighted Entropy = 0.8203

$$\boxed{\text{Info. Gain} = 0.0795}$$

• Work-life balance:

	Left	Stayed	Total
Poor	6	8	8
Fair	8	3	11
Excellent	2	2	4
Good	10	5	15
$E =$	$\frac{8}{51} \left[-\frac{6}{8} \log_2 \left(\frac{6}{8} \right) - \frac{2}{8} \log_2 \left(\frac{2}{8} \right) \right] +$		
	$\frac{11}{51} \left[-\frac{8}{11} \log_2 \left(\frac{8}{11} \right) - \frac{3}{11} \log_2 \left(\frac{3}{11} \right) \right] +$		
	$\frac{4}{51} [1] + \frac{15}{51} \left[-\frac{10}{15} \log_2 \left(\frac{10}{15} \right) - \frac{5}{15} \log_2 \left(\frac{5}{15} \right) \right]$		
$E =$	$\frac{8}{51} [0.8113] + \frac{11}{51} [0.8454] + \frac{4}{51} + \frac{15}{51} [0.9183]$		

Weighted entropy = 0.8833

$$\boxed{\text{Info-Gain} = 0.0165}$$

• Gender:

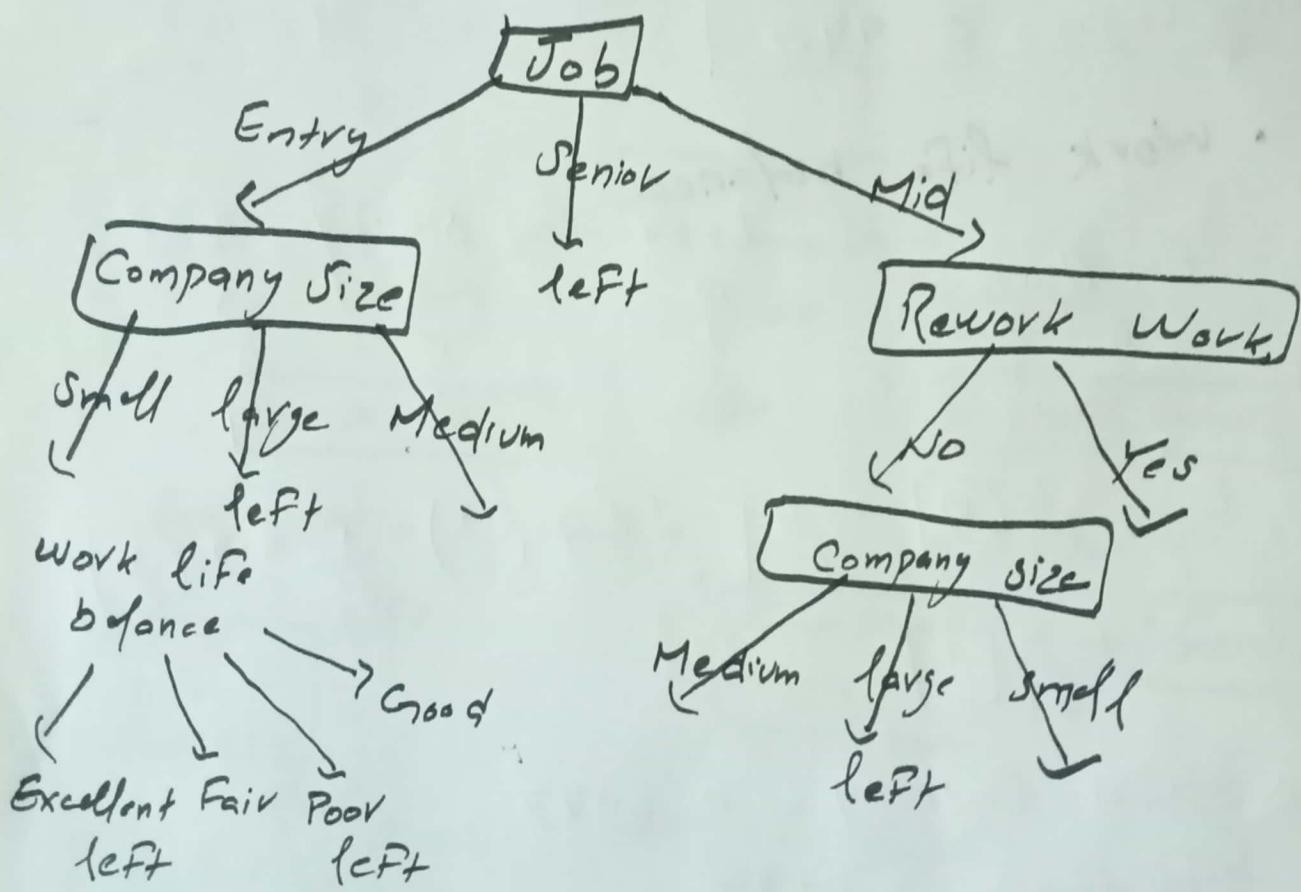
	Left	Stayed	Total
Male	12	8	20
Female	14	4	18
$E =$	$\frac{20}{51} \left[-\frac{12}{20} \log_2 \left(\frac{12}{20} \right) - \frac{8}{20} \log_2 \left(\frac{8}{20} \right) \right]$		
	$+ \frac{18}{51} \left[-\frac{14}{18} \log_2 \left(\frac{14}{18} \right) - \frac{4}{18} \log_2 \left(\frac{4}{18} \right) \right]$		

$$= \frac{18}{51} [0.9710] + \frac{18}{51} [0.7642]$$

Weighted entropy = 0.8730

$$\text{Info. Gain} = 0.0267$$

\Rightarrow Company size has high Info. Gain, so it is root



⑥ Remote Work Yes Split;

- Company size.

	Left	Stayed	Total
Medium	2	2	4
Large	1	4	5
Small	0	4	4

$$E = \frac{4}{57} [1] + \frac{4}{57} [0] + \frac{5}{57} \left[-\frac{4}{5} \log_2 \left(\frac{4}{5} \right) - \frac{1}{5} \log_2 \left(\frac{1}{5} \right) \right]$$

$$= \frac{4}{57} + 0 + \frac{6}{57} [0' \cancel{0.5859}]$$

Weighted Entropy = 0.5859

Info. Gain = 0.1990

• Work-life balance

	Left	Stayed	Total
Excellent	0	3	3
Good	1	5	6
Fair	2	2	4

$$E = \frac{3}{57} [0] + \frac{6}{57} \left[-\frac{1}{6} \log_2 \left(\frac{1}{6} \right) - \frac{5}{6} \log_2 \left(\frac{5}{6} \right) \right]$$

$$+ \frac{4}{57} [1]$$

Weighted Entropy = 0.6077

Info. Gain = 0.1716

• Gender

	Left	Stayed	Total
Male	7	0	7
Female	3	3	6

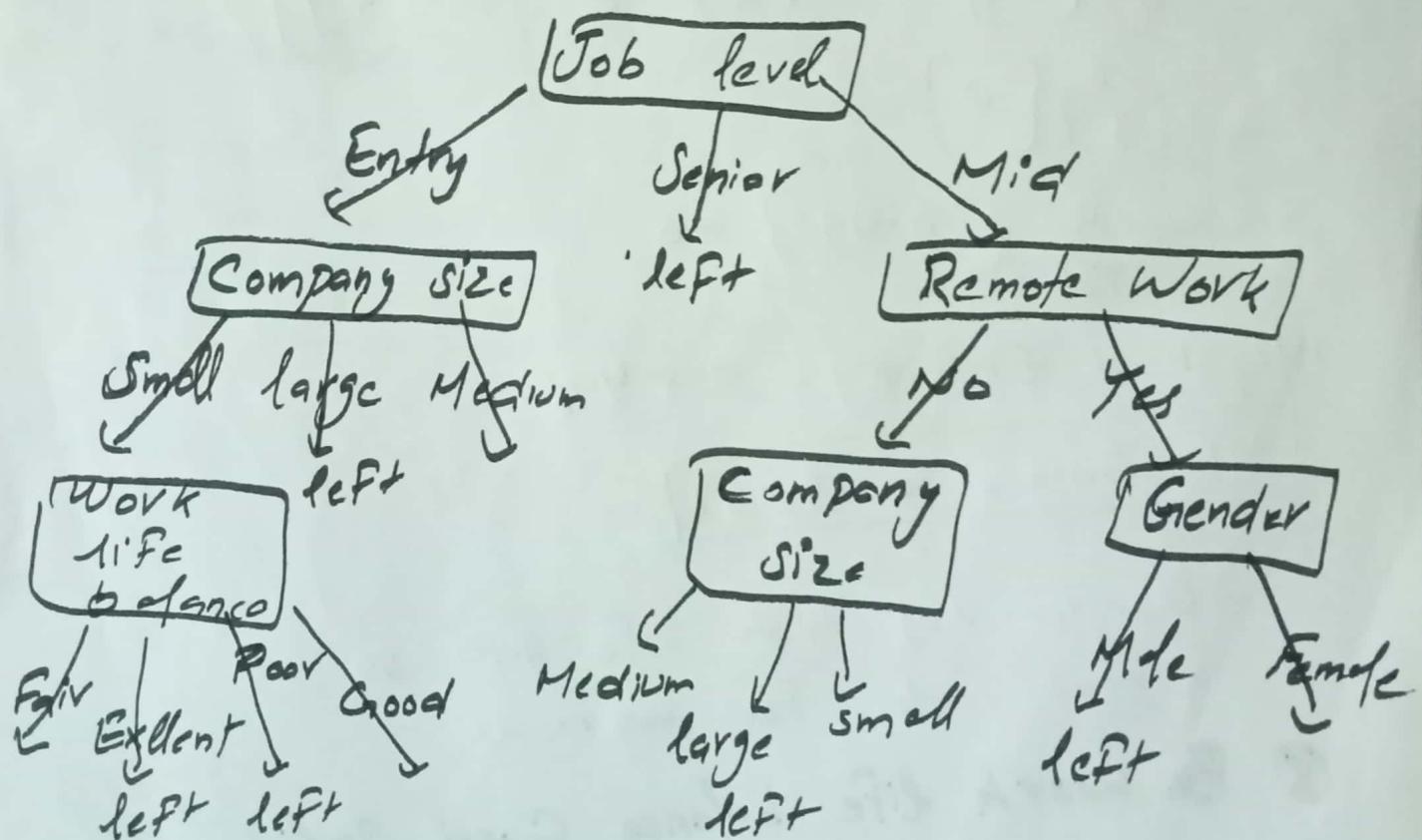
$$E = \frac{7}{51} [0] + \frac{6}{51} \left[-\frac{3}{6} \log_2 \left(\frac{3}{6} \right) - \frac{3}{6} \log_2 \left(\frac{3}{6} \right) \right]$$

$$wE = 0 + \frac{6}{51} [1]$$

Weighted Entropy = 0.4615

Information Gain = 0.3178

→ Gender has high, so it is root node for now.



- ④ Work life balance Fair Split,
• Remote Work

	Left	Stayed	Total
Yes	2	2	4
No	8	3	11

$$E = \frac{4}{51} [1] + \frac{47}{51} \left[-\frac{8}{11} \log_2 \left(\frac{8}{11} \right) - \frac{3}{11} \log_2 \left(\frac{3}{11} \right) \right]$$

Weighted Entropy = 0.8866

Info-Gain = 0.0317

• Gender

	Left	Stayed	Total
Male	4	5	9
Female	6	0	6

$$E = \frac{9}{51} \left[-\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) \right] +$$

$$\frac{6}{51} [0]$$

$$= \frac{9}{51} [0.991] + 0$$

Weighted Entropy = 0.5946

Info-Gain = 0.3237

⇒ Gender has high info-gain so it is root node for now.

② Work life balance Good Split,

• Remote work

	Left	Stayed	Total
Yes	1	8	9
No	10	5	15

$$E = \frac{6}{51} \left[-\frac{1}{6} \log_2 \left(\frac{1}{6} \right) - \frac{5}{6} \log_2 \left(\frac{5}{6} \right) \right] + \frac{15}{51} \left[-\frac{10}{15} \log_2 \left(\frac{10}{15} \right) - \frac{5}{15} \log_2 \left(\frac{5}{15} \right) \right]$$

$$= \frac{6}{51} [0.6500] + \frac{15}{51} [0.9183]$$

Weighted Entropy = 0.8416

Info. Gain = 0.1567

• Gender:

	Left	Stayed	Total
Female	8	4	12
Male	3	6	9

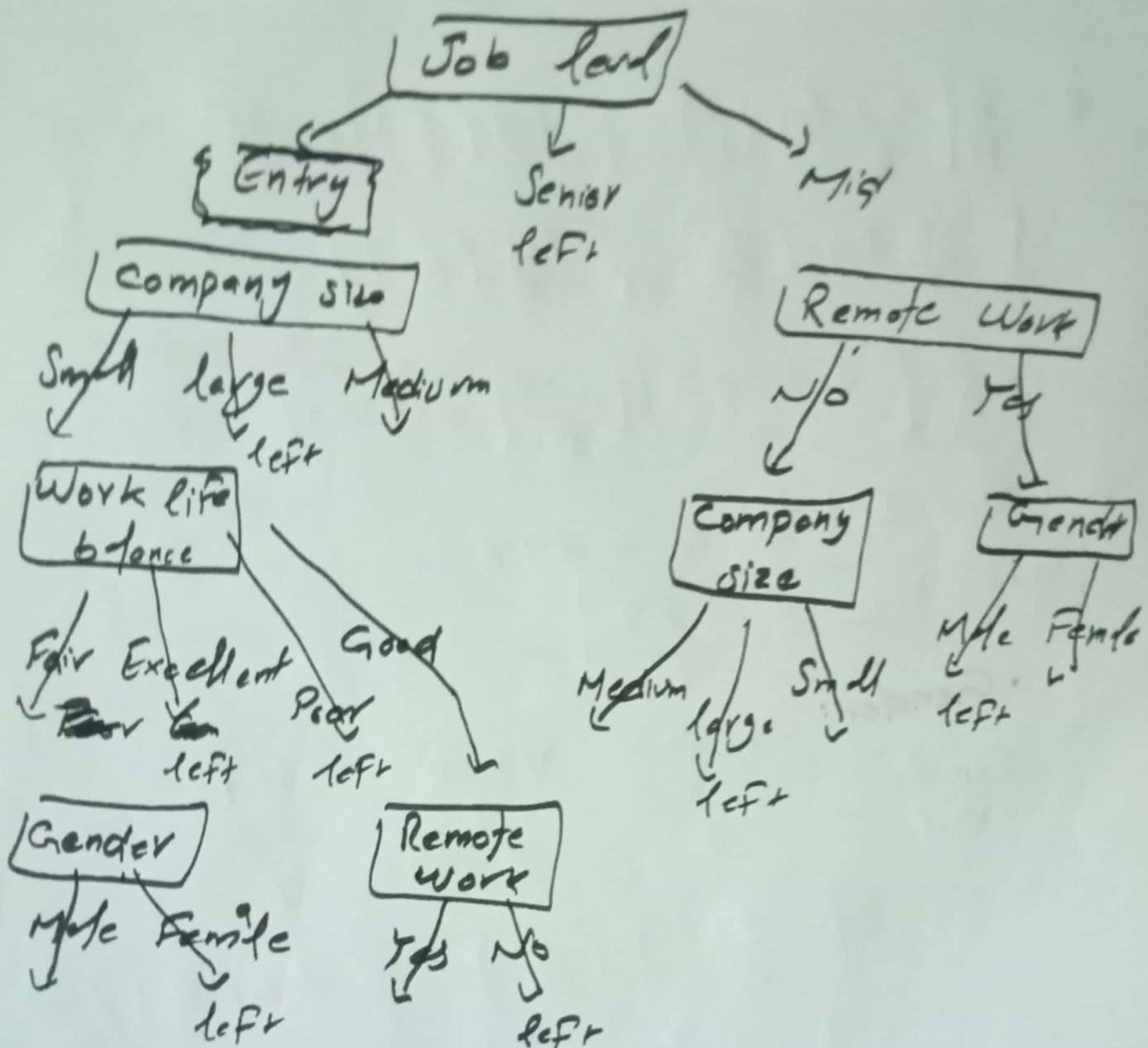
$$E = \frac{12}{51} \left[-\frac{8}{12} \log_2 \left(\frac{8}{12} \right) - \frac{4}{12} \log_2 \left(\frac{4}{12} \right) \right] + \frac{9}{51} \left[-\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) \right]$$

$$E = \frac{12}{51} [0.9183] + \frac{9}{51} [0.9183]$$

Weighted Entropy = 0.9183

Info. Gain = 0.0801

⇒ Remote work high IG, so it is root node for now



① Company size Medium Split:

• Overline Remote work:

	Left	Stayed	Total
No	15	9	24
Yes	2	2	4

$$E = \frac{24}{51} \left[-\frac{15}{24} \log_2 \left(\frac{15}{24} \right) - \frac{9}{24} \log_2 \left(\frac{9}{24} \right) \right] + \frac{4}{51} \left[-\frac{2}{4} \right]$$

Weighted Entropy = 0.9609

$$\boxed{\text{Info - Grain} = 0.0057}$$

• Work life balance:

	Left	Stayed	Total
Poor	2	2	4
Excellent	1	2	3
Fair	6	4	10
Good	8	3	11

$$E = \frac{4}{51} \left[\#1 \right] + \frac{3}{51} \left[\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right]$$

$$+ \frac{10}{51} \left[-\frac{6}{10} \log_2 \left(\frac{6}{10} \right) - \frac{4}{10} \log_2 \left(\frac{4}{10} \right) \right] +$$

$$\frac{11}{51} \left[-\frac{8}{11} \log_2 \left(\frac{8}{11} \right) - \frac{3}{11} \log_2 \left(\frac{3}{11} \right) \right]$$

$$E = \frac{4}{51} + \frac{3}{51} \left[0.9183 \right] + \frac{10}{51} \left[0.9710 \right] +$$

$$\frac{11}{51} \left[0.8454 \right]$$

Weighted Entropy = 0.9201

$$\boxed{\text{Info - Grain} = 0.0465}$$

• Gender:

	Left	Stayed	Total
Male	7	8	15
Female	10	3	13

$$E = \frac{15}{51} \left[-\frac{7}{15} \log_2 \left(\frac{7}{15} \right) - \frac{8}{15} \log_2 \left(\frac{8}{15} \right) \right] + \frac{13}{51} \left[-\frac{10}{13} \log_2 \left(\frac{10}{13} \right) - \frac{3}{13} \log_2 \left(\frac{3}{13} \right) \right]$$

Weighted Entropy = 0.8958

Info. Gain = 0.0708

\Rightarrow Gender has high IG.

⑩ Company size ~~14~~ ^{Small} split,

• Work life balance:

	Left	Stayed	Total
Poor	2	10	12
Excellent	0	2	2
Fair	2	1	3
Good	2	4	6

$$E = \frac{2}{12} [0] + \frac{2}{12} \left[\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right]$$

$$+ \frac{3}{12} \left[-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right] +$$

$$\frac{6}{12} \left[-\frac{2}{6} \log_2 \left(\frac{2}{6} \right) - \frac{4}{6} \log_2 \left(\frac{4}{6} \right) \right]$$

$$E = 0.9183 + 0.9183 + \frac{2}{51} [0.9183] + \frac{6}{51} [0.9183]$$

Weighted Entropy = 0.6357

$$\text{Info-Gain} = 0.466$$

• Gender

	Left	Stayed	Total
Male	73	86	159
Female	13	13	26

$$E = \frac{9}{159} \left[-\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{86}{98} \log_2 \left(\frac{86}{98} \right) \right],$$

$$\frac{4}{26} \left[-\frac{13}{40} \log_2 \left(\frac{13}{40} \right) - \frac{13}{13} \log_2 \left(\frac{13}{13} \right) \right]$$

Weighted Entropy = ~~0.8988~~ 0.8854

$$\text{Info-Gain} = 0.466$$

\Rightarrow Work life balance
root node has high I.G., so it is now.

② Company size ~~Medium~~ split

• Gender

	Left	Stayed	Total
Male	7	8	15
Female	10	3	13

$$E = \frac{15}{51} \left[-\frac{7}{15} \log_2 \left(\frac{7}{15} \right) - \frac{8}{15} \log_2 \left(\frac{8}{15} \right) \right],$$

$$\frac{13}{51} \left[-\frac{10}{13} \log_2 \left(\frac{10}{13} \right) - \frac{3}{13} \log_2 \left(\frac{3}{13} \right) \right]$$

Weighted Entropy = 0.8958

Info-Gain = 0.0708

• Work life balance

	Left	Stayed	Total
Poor	2	2	4
Excellent	1	2	3
Fair	6	4	10
Good	8	3	11

$$E = \frac{4}{51} \left[-1 \right] + \frac{3}{51} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right] + \\ + \frac{10}{51} \left[-\frac{6}{10} \log_2 \left(\frac{6}{10} \right) - \frac{4}{10} \log_2 \left(\frac{4}{10} \right) \right] + \\ + \frac{11}{51} \left[-\frac{8}{11} \log_2 \left(\frac{8}{11} \right) - \frac{3}{11} \log_2 \left(\frac{3}{11} \right) \right]$$

Weighted Entropy = 0.9201

Info-Gain = 0.04605

\Rightarrow Gender has high I.G., so it's root node now.

⑫ Gender Female Split,

• Company size,

	Left	Stayed	Total
Large	4	3	7
Medium	10	3	13
Small	8	1	9

$$E = \frac{2}{51} \left[-\frac{4}{7} \log_2 \left(\frac{4}{7} \right) - \frac{3}{7} \log_2 \left(\frac{3}{7} \right) \right] + \frac{13}{51} \left[-\frac{10}{13} \log_2 \left(\frac{10}{13} \right) - \frac{3}{13} \log_2 \left(\frac{3}{13} \right) \right] + \frac{4}{51} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right]$$

Weighted Entropy = 0.8447

[Info-Gain = 0.0261]

• Work life balance:

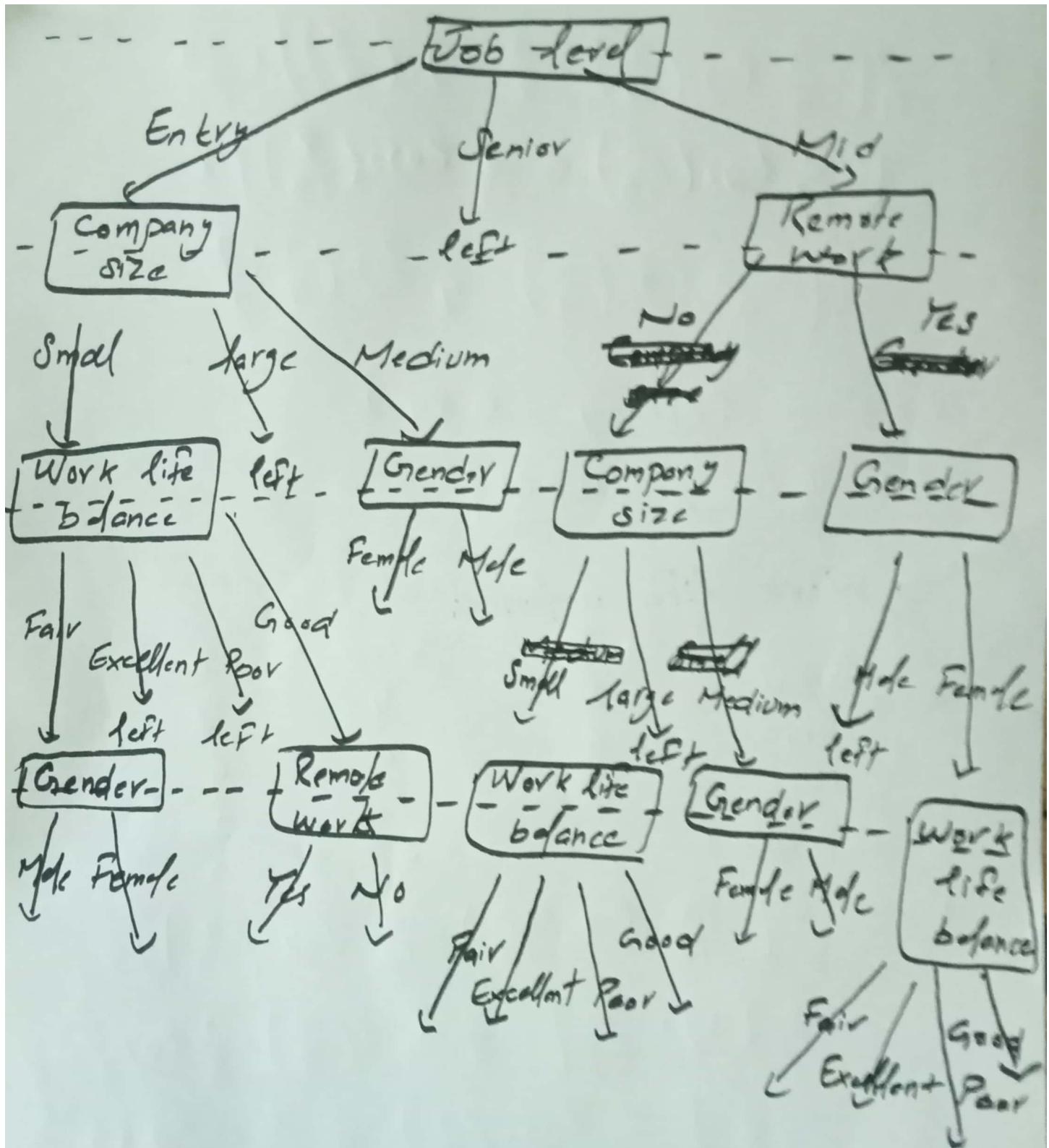
	Left	Stayed	Total
Good	8	4	12
Fair	6	0	6
Poor	2	1	3
Excellent	1	2	3

$$E = \frac{12}{51} \left[-\frac{8}{12} \log_2 \left(\frac{8}{12} \right) - \frac{4}{12} \log_2 \left(\frac{4}{12} \right) \right] + \frac{6}{51} [0] + \frac{3}{51} \left[-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right] + \frac{3}{51} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right]$$

Weighted Entropy = 0.6887

[Info. Gain = 0.1821]

=> Work life balance high so, so root node now.



• SECOND DECISION TREE:

→ Now I select second highest information gain attribute (Remote work)

Remote Work

① Split on Remote work No:

• Job level:

$$E = \frac{14}{51} \left[-\frac{4}{14} \log_2 \left(\frac{4}{14} \right) - \frac{10}{14} \log_2 \left(\frac{10}{14} \right) \right] + \frac{19}{51} \left[-\frac{3}{19} \log_2 \left(\frac{3}{19} \right) - \frac{16}{19} \log_2 \left(\frac{16}{19} \right) \right] + 0$$

Weighted Entropy = 0.6326

[Info-Grain = 0.267]

• Company size:

$$E = \frac{24}{51} \left[-\frac{9}{24} \log_2 \left(\frac{9}{24} \right) - \frac{15}{24} \log_2 \left(\frac{15}{24} \right) \right] + \frac{9}{51} \left[-\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) \right] + 0$$

Weighted Entropy = 0.8203

[Info-Grain = 0.0795]

• Gender:

$$E = \frac{20}{51} \left[-\frac{8}{20} \log_2 \left(\frac{8}{20} \right) - \frac{12}{20} \log_2 \left(\frac{12}{20} \right) \right] + \frac{18}{51} \left[-\frac{4}{18} \log_2 \left(\frac{4}{18} \right) - \frac{14}{18} \log_2 \left(\frac{14}{18} \right) \right]$$

Weighted Entropy = 0.8730

[Info-Grain = 0.0267]

\Rightarrow So now the root node is the Job level.

② Split on Remote work Yes:

• Job level:

$$E = \frac{6}{51} [0] + \frac{5}{51} \left[-\frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} \log_2 \left(\frac{3}{5} \right) \right] + \frac{2}{51} [0]$$

Weighted Entropy = 0.3734

Info-Gain = 0.4059

• Company size:

$$E = \frac{4}{51} [1] + \frac{5}{51} \left[-\frac{1}{5} \log_2 \left(\frac{1}{5} \right) - \frac{4}{5} \log_2 \left(\frac{4}{5} \right) \right] + \frac{4}{51} [0]$$

Weighted Entropy = 0.5854

Entropy

Info-Gain = 0.1940

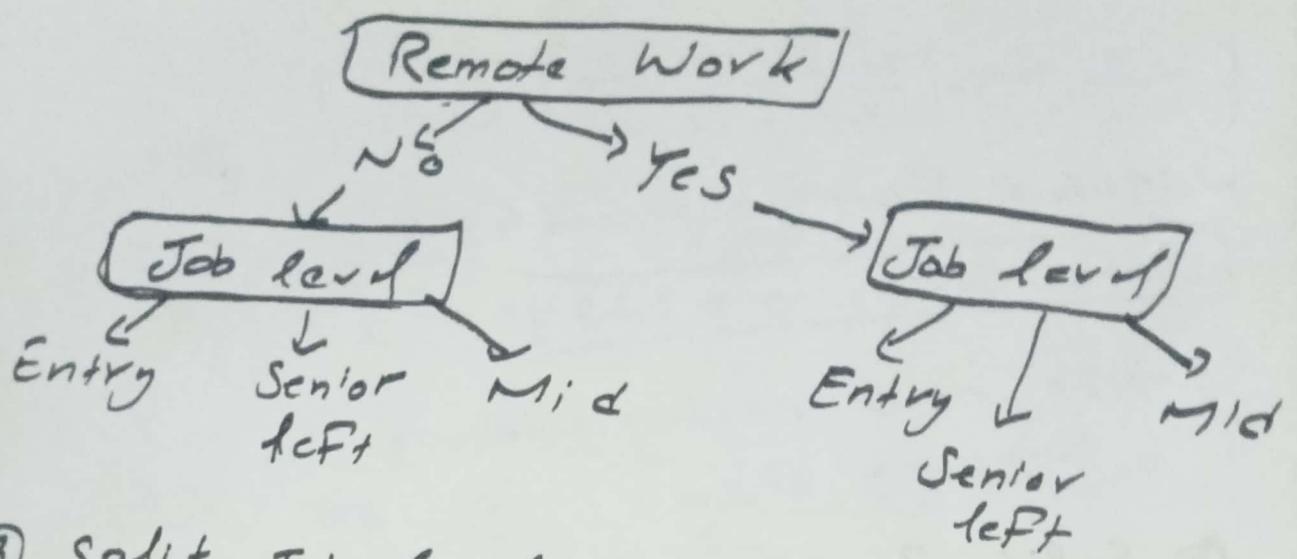
• Gender:

$$E = \frac{7}{51} [0] + \frac{6}{51} [1]$$

Weighted Entropy = 0.4615

Info-Gain = 0.3178

\Rightarrow So now root node is the Job level.



③ Split Job level on Entry:

- Company size;

$$E = \frac{15}{51} \left[-\frac{2}{15} \log_2 \left(\frac{2}{15} \right) - \frac{13}{15} \log_2 \left(\frac{13}{15} \right) \right] + \frac{7}{51} [0] + \frac{8}{51} \left[-\frac{3}{8} \log_2 \left(\frac{3}{8} \right) - \frac{2}{8} \log_2 \left(\frac{2}{8} \right) \right]$$

Weighted Entropy = 0.5583
Info-Gain = 0.1819

- Work life balance;

$$E = \frac{11}{51} \left[-\frac{8}{11} \log_2 \left(\frac{8}{11} \right) - \frac{3}{11} \log_2 \left(\frac{3}{11} \right) \right] + \frac{8}{51} \left[-\frac{1}{8} \log_2 \left(\frac{1}{8} \right) - \frac{7}{8} \log_2 \left(\frac{7}{8} \right) \right] + \frac{2}{51} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right]$$

Weighted Entropy = 0.6834
Info-Gain = 0.0549

- Gender:

$$E = \frac{13}{51} \left[-\frac{1}{13} \log_2 \left(\frac{1}{13} \right) - \frac{12}{13} \log_2 \left(\frac{12}{13} \right) \right] + \frac{11}{51}$$

$$\left[-\frac{4}{11} \log_2\left(\frac{4}{11}\right) - \frac{7}{11} \log_2\left(\frac{7}{11}\right) \right]$$

Weighted Entropy = 0.6454

$$\text{Info-Gain} = 0.0929$$

\Rightarrow So now root node is company size.

④ Split Remote work on Mid:

- Company size:

$$E = \frac{16}{51} \left[-\frac{4}{10} \log_2\left(\frac{4}{10}\right) - \frac{6}{10} \log_2\left(\frac{6}{10}\right) \right] + \frac{5}{51} [0] + \frac{6}{51} \left[-\frac{2}{8} \log_2\left(\frac{2}{8}\right) - \frac{4}{8} \log_2\left(\frac{4}{8}\right) \right]$$

Weighted Entropy = 0.9610

$$\text{Info-Gain} = 0.0390$$

- Work life balance,

$$E = \frac{4}{51} [0] + \frac{2}{51} [0] + \frac{7}{51} \left[-\frac{4}{7} \log_2\left(\frac{4}{7}\right) - \frac{3}{7} \log_2\left(\frac{3}{7}\right) \right]$$

Weighted Entropy = 0.6897

$$\text{Info-Gain} = 0.3103$$

- Gender:

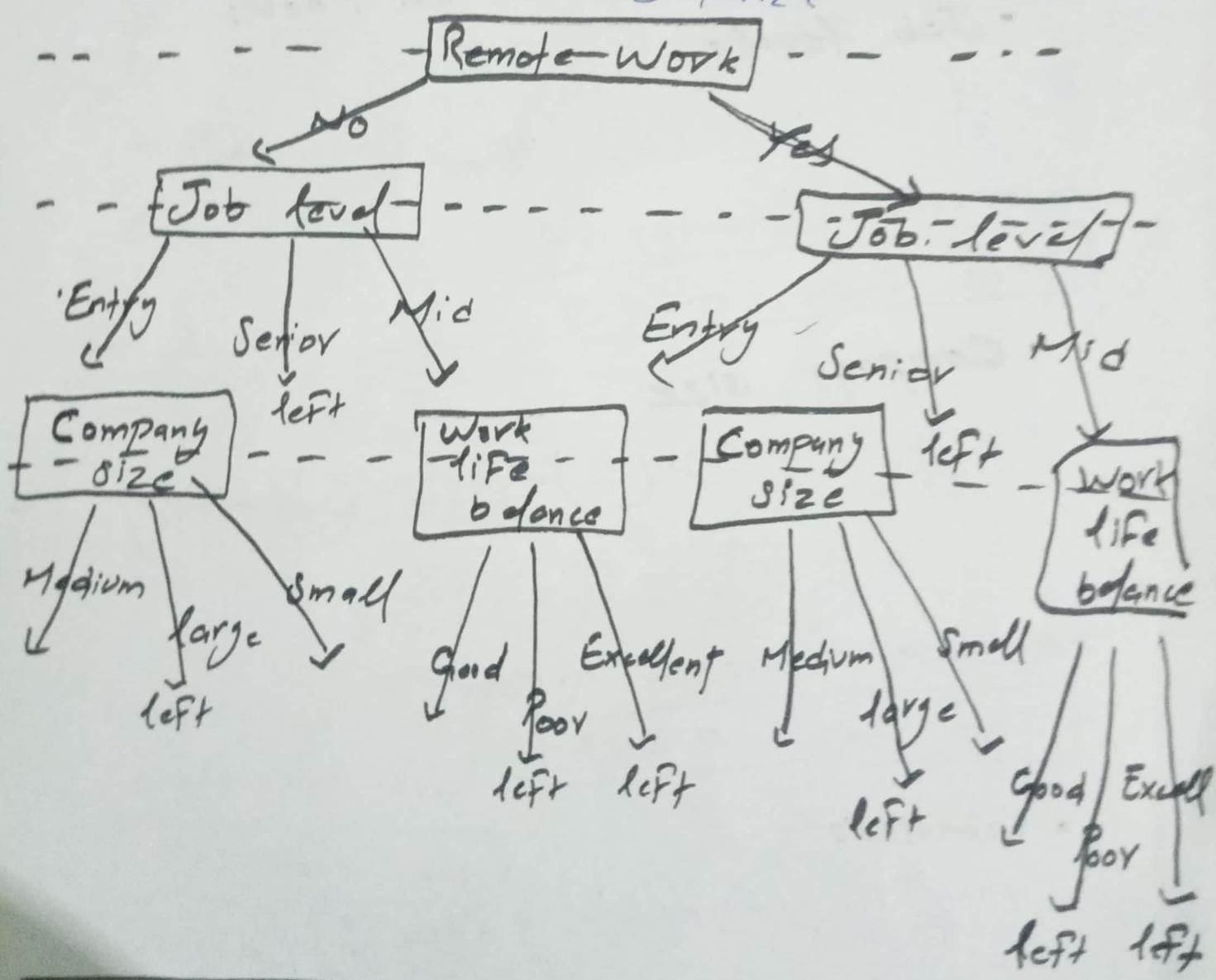
$$E = \frac{13}{51} \left[-\frac{8}{13} \log_2\left(\frac{8}{13}\right) - \frac{5}{13} \log_2\left(\frac{5}{13}\right) \right] + \frac{7}{51}$$

$$\left[-\frac{2}{7} \log_2\left(\frac{2}{7}\right) - \frac{5}{7} \log_2\left(\frac{5}{7}\right) \right]$$

Weighted Entropy = 0.9269

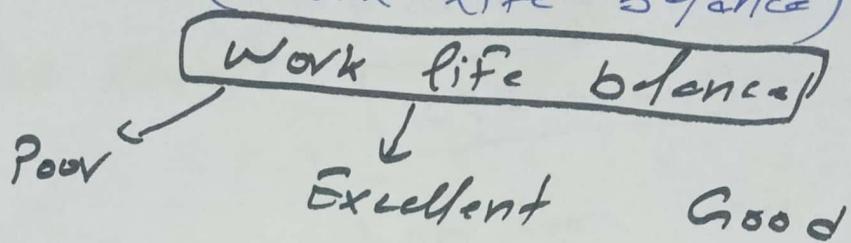
$$\text{Info-Gain} = 0.0731$$

So, now root node is Company size. work life balance



• THIRD DECISION TREE:

Now select third highest information gain element (work life balance)



- ① Split work life balance on Poor;

- Job level;

$$E = \frac{4}{51} [0] + \frac{2}{51} [0] + \frac{2}{51} [0]$$

Weighted Entropy = 0

Info - Gain = 0.8113

- Company size

$$E = \frac{4}{51} \left[-\frac{2}{7} \log_2 \left(\frac{2}{7} \right) - \frac{2}{7} \log_2 \left(\frac{2}{5} \right) \right] \frac{7}{51} [0] + \frac{2}{51} [0]$$

Weighted Entropy = 0.58008

Info - Gain = 0.8113

- Gender;

$$E = \frac{5}{51} \left[-\frac{1}{5} \log_2 \left(\frac{1}{5} \right) - \frac{4}{5} \log_2 \left(\frac{4}{5} \right) \right] + \frac{3}{51} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right]$$

Weighted Ent = 0.7956

Info - Gain = 0.157

\Rightarrow Now Job level is the root node.

② Work life balance split on Excellent;

• Job level:

$$E = \frac{2}{5} [0] + \frac{2}{5} [0] + \frac{3}{5} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right]$$

Weighted Entropy = 0.3936

$$\boxed{\text{Info-Gain} = 0.4096}$$

• Remote work:

$$E = \frac{3}{5} [0] + \frac{2}{5} [1]$$

Weighted Ent = 0.5714

$$\boxed{\text{Info-Cgain} = 0.2917}$$

• Company size:

$$E = \frac{3}{5} \left[-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right] + \frac{2}{5} [1] + \frac{2}{5} [0]$$

Weighted Ent = 0.6793

$$\boxed{\text{Info-Gain} = 0.1839}$$

\Rightarrow Now Job level is root node.

③ Work life balance spread on Good,

• Job level:

$$E = \frac{7}{51} \left[-\frac{5}{7} \log_2 \left(\frac{5}{7} \right) - \frac{3}{7} \log_2 \left(\frac{3}{7} \right) + \frac{11}{51} \left(-\frac{3}{11} \log_2 \left(\frac{3}{11} \right) - \frac{8}{11} \log_2 \left(\frac{8}{11} \right) \right) \right] + \frac{3}{51} [0]$$

Weighted Ent = 0.7712
Info Gain = 0.2272

• Remote Work:

$$E = \frac{6}{51} \left[-\frac{5}{6} \log_2 \left(\frac{5}{6} \right) - \frac{1}{6} \log_2 \left(\frac{1}{6} \right) \right] + \frac{15}{51} \left[-\frac{5}{15} \log_2 \left(\frac{5}{15} \right) - \frac{10}{15} \log_2 \left(\frac{10}{15} \right) \right]$$

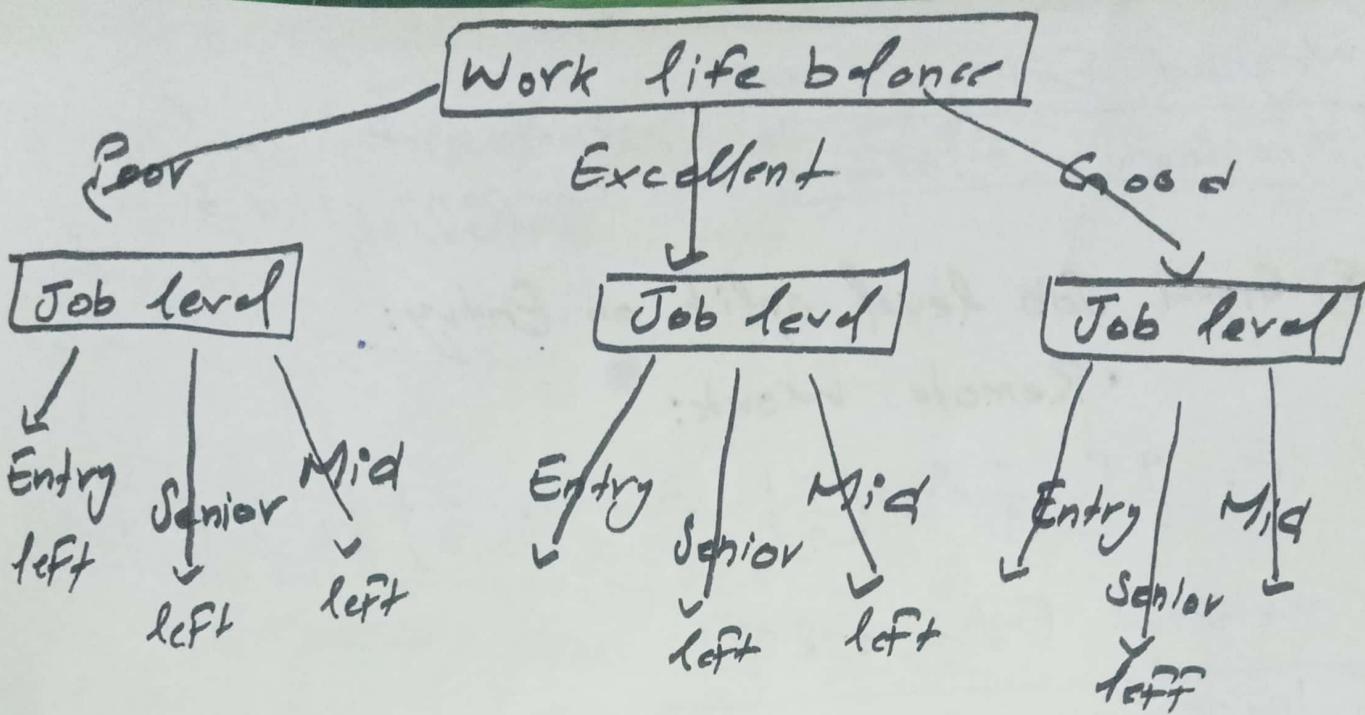
Weighted Ent = 0.8416
Info - Gain = 0.1567

• Company size,

$$E = \frac{4}{51} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right] + \frac{11}{51} \left[-\frac{3}{11} \log_2 \left(\frac{3}{11} \right) - \frac{8}{11} \log_2 \left(\frac{8}{11} \right) \right] + \frac{6}{51} \left[-\frac{4}{6} \log_2 \left(\frac{4}{6} \right) - \frac{2}{6} \log_2 \left(\frac{2}{6} \right) \right]$$

Weighted Ent = 0.8597
Info - Gain = 0.1387

∴ Now Job level is the root.



④ Job level split on entry:

- Remote work:

$$E = \frac{1}{5} [0] + \frac{2}{5} \left[\left(\frac{2}{7} \log_2 \left(\frac{2}{7} \right) \right) \left(\frac{5}{7} \log_2 \left(\frac{5}{7} \right) \right) \right]$$

Weighted Ent = 0. ~~0.871000~~

Info - Gain = 0. ~~0.8839183~~

- Company size

$$E = \frac{2}{5} \left[\left(\frac{1}{7} \log_2 \left(\frac{1}{7} \right) \right) - \frac{6}{7} \log_2 \left(\frac{6}{7} \right) \right] - \frac{1}{5} [0] \\ + \frac{3}{5} \left[\left(\frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right]$$

Weighted Ent = 0. ~~0.888000~~

Info - Gain = 0. ~~0.889183~~

- Gender:

$$E = \frac{2}{5} \left[-\frac{1}{7} \log_2 \left(\frac{1}{7} \right) - \frac{6}{7} \log_2 \left(\frac{6}{7} \right) \right] + \frac{3}{5} [1]$$

$$\begin{array}{l} \text{Weighted Ent} = 0.7402 \\ \boxed{\text{Info-Gain} = 0.1052} \end{array} \Rightarrow \text{Root node is company size}$$

⑤ Good Job level split on Entry:-

- Remote Work:

$$E = \frac{2}{51} \left[-\frac{7}{9} \log_2 \left(\frac{7}{9} \right) - \frac{2}{9} \log_2 \left(\frac{2}{9} \right) \right] + \frac{2}{51} [1]$$

$$\text{Weighted Ent} = 0.8071$$

$$\boxed{\text{Info-Gain} = 0.0383}$$

- Company Size:

$$E = \frac{7}{51} \left[-\frac{1}{7} \log_2 \left(\frac{1}{7} \right) - \frac{6}{7} \log_2 \left(\frac{6}{7} \right) \right] + \frac{1}{51} [0] \\ + \frac{3}{51} \left[-\frac{2}{3} \log_2 \left(\frac{2}{3} \right) - \frac{1}{3} \log_2 \left(\frac{1}{3} \right) \right]$$

$$\text{Weighted Ent} = 0.6270$$

$$\boxed{\text{Info-Gain} = 0.2184}$$

- Gender:

$$E = \frac{7}{51} \left[-\frac{1}{7} \log_2 \left(\frac{1}{7} \right) - \frac{6}{7} \log_2 \left(\frac{6}{7} \right) \right] + \frac{7}{51} [1]$$

$$\boxed{\text{Weighted Ent} = 0.7402}$$

$$\boxed{\text{Info-Gain} = 0.1052}$$

\Rightarrow Now company size is the root node.

⑥ Good Job level split on Mid:

• Remote work:

$$E = \frac{2}{51} [0] + \frac{5}{51} \left[-\frac{2}{5} \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} \log_2 \left(\frac{3}{5} \right) \right]$$

Weighted Ent = 0.6935

Info-Gain = 0.2917

• Company size:

$$E = \frac{1}{51} [0] + \frac{4}{51} [1] + \frac{2}{51} [1]$$

Weighted Ent = 0.8571

Info-Gain = 0.1281

• Gender:

$$E = \frac{8}{51} \left[-\frac{1}{3} \log_2 \left(\frac{1}{3} \right) - \frac{2}{3} \log_2 \left(\frac{2}{3} \right) \right] + \frac{4}{51} \left[-\frac{3}{4} \log_2 \left(\frac{3}{4} \right) - \frac{1}{4} \log_2 \left(\frac{1}{4} \right) \right]$$

Weighted Ent = 0.8571

Info-Gain = 0.1281

⇒ Root node is Remote work.

