Data

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F
	6	0	0	0	1	1	F
	7	0	0	0	0	0	F
	8	1	0	0	0	0	F
	9	0	0	1	1	0	M
	10	0	1	0	1	0	F
	11	…	…	…	…	…	…
	12	…	…	…	…	…	…

Data row Split

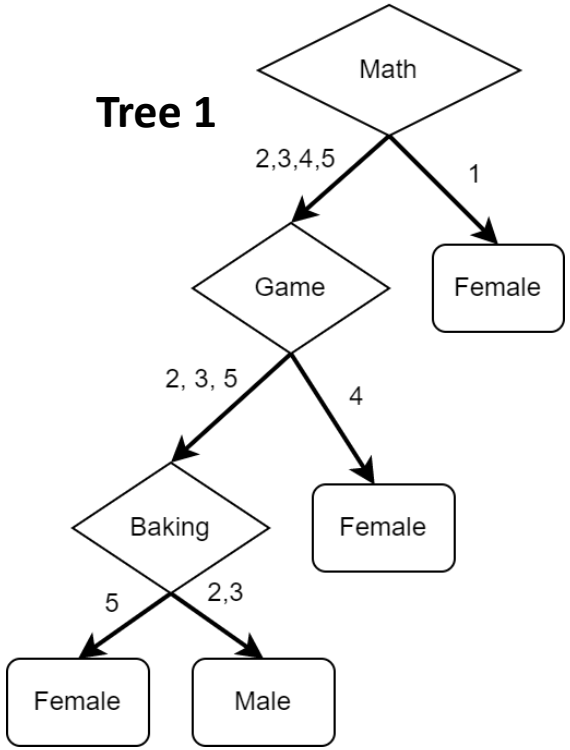
	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training1	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F
Test	6	1	1	1	1	1	?

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training2	6	0	0	0	1	1	F
	7	0	0	0	0	0	F
	8	1	0	0	0	0	F
	9	0	0	1	1	0	M
	10	0	1	0	1	0	F
Test	11	…	…	…	…	…	?

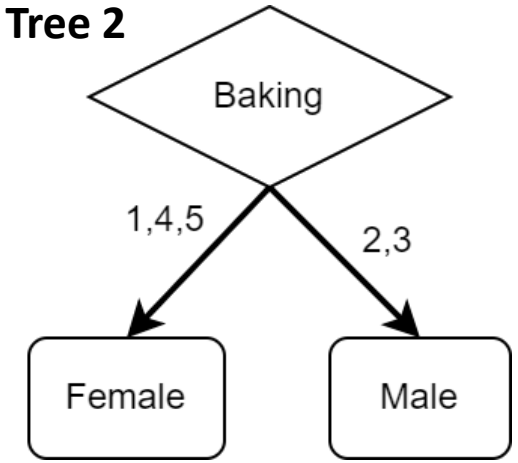
Training Phase Feature Selection

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training1	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F

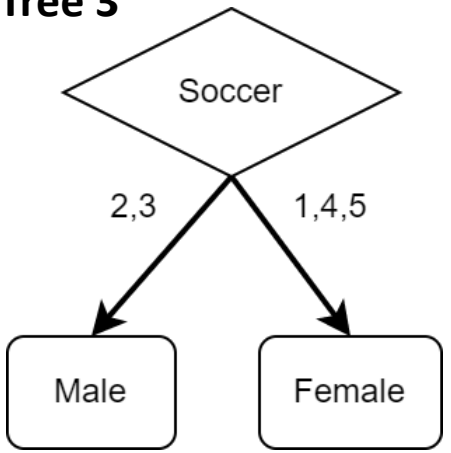
Tree 1



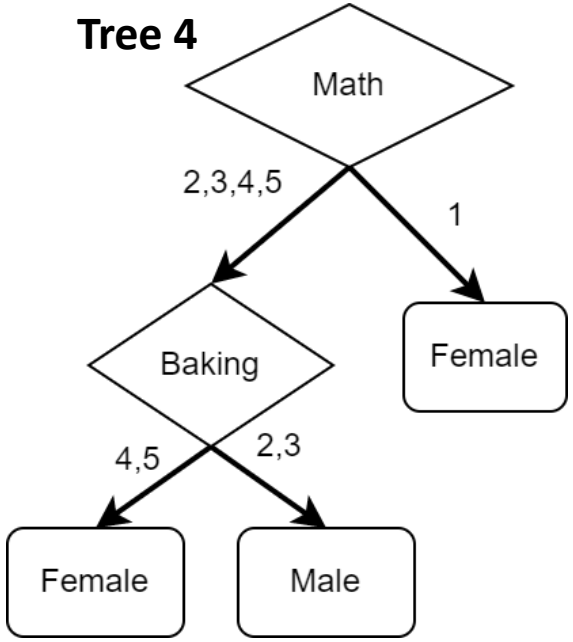
Tree 2



Tree 3



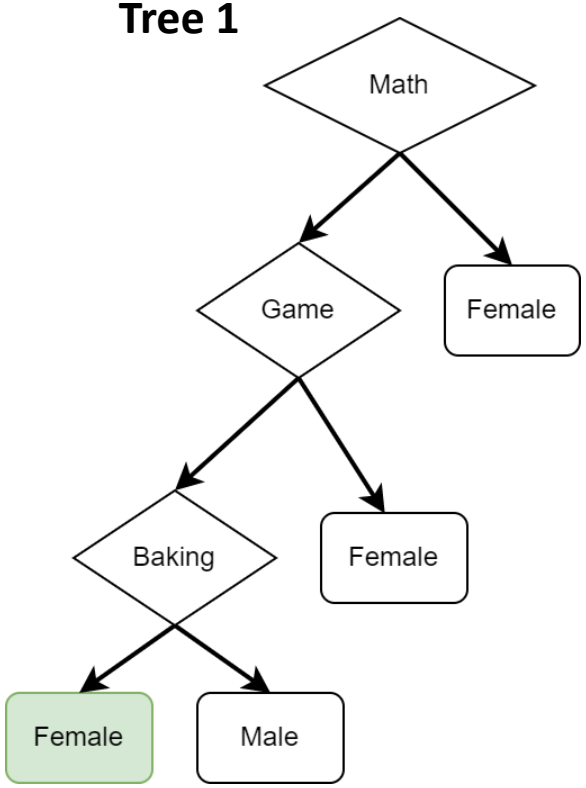
Tree 4



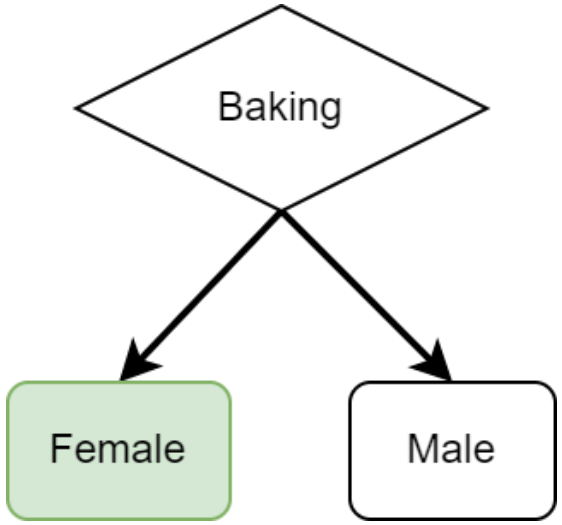
Aggregation

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training1	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F
Test	6	1	1	1	1	1	?

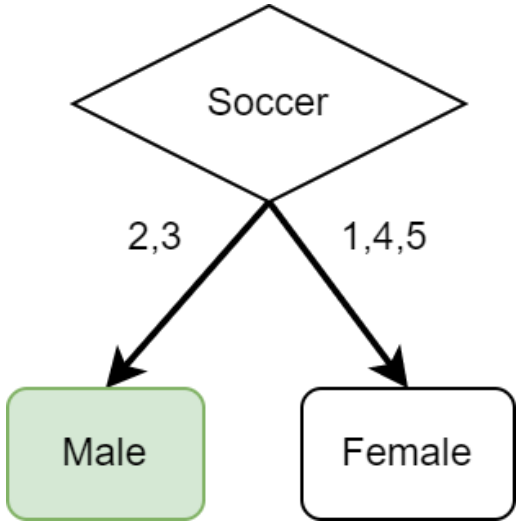
Tree 1



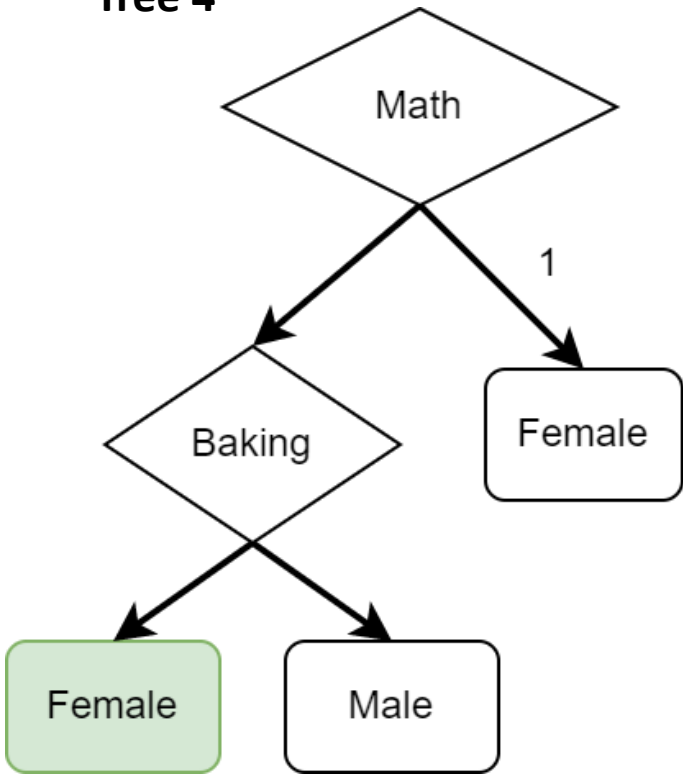
Tree 2



Tree 3



Tree 4



Female: 3
Male: 1

Winner = Female

Testing Phase

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training1	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F
Test	6	1	1	1	1	1	**F**

Aggregation

Female: 3
Male: 1

Winner = Female

$Enthropy = 1000$

$Enthropy = 1000000!$

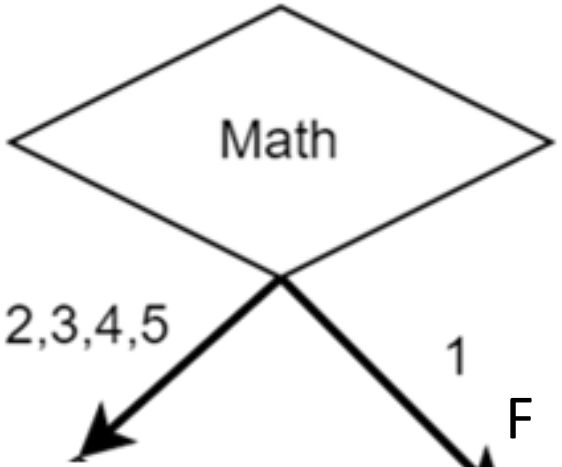
$$Enthropy = 0$$

- **Rate of disorder,**
- **Rate of randomness**

$$Enthropy = \sum_{i=1}^{N} -p_i \log_2 \left(p_i \right)$$

N = Number of class

p_i = Probaility of class i



$$E = -\frac{1}{2} \times \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \times \log_2 \left(\frac{1}{2} \right)$$

$$E = 1$$

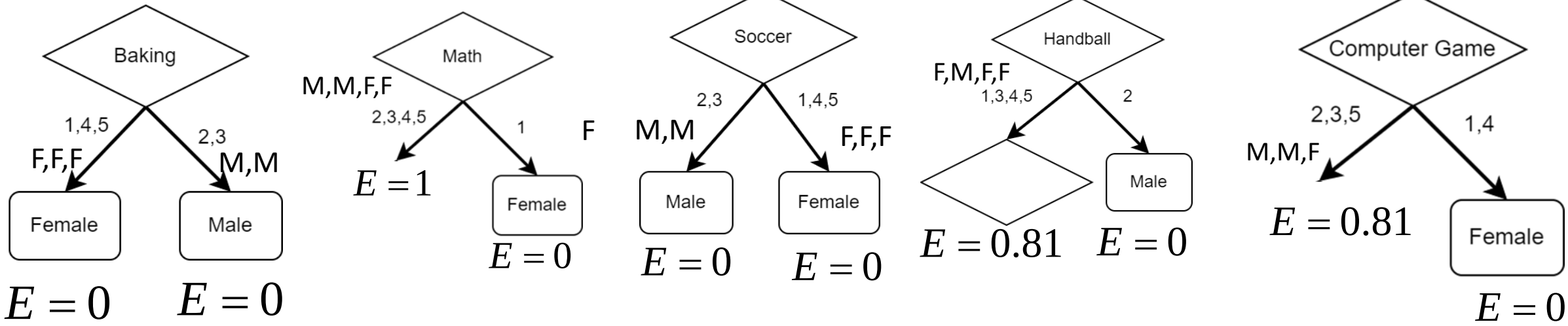
$$E = 0$$

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F

$$Enthropy = \sum_{i=1}^{N} -p_i \log_2 \left(p_i \right)$$

N = Number of class

p_i = Probaility of class i



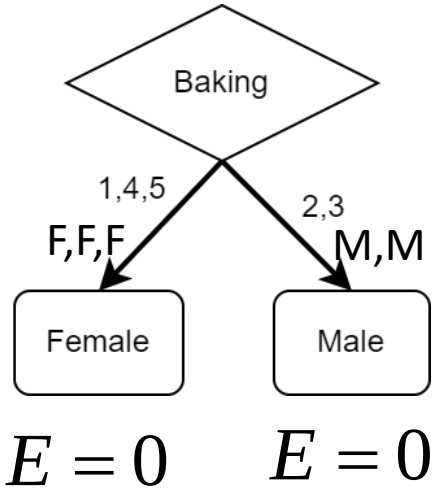
	Young people 10-25						
	ID	Baking	Math	Soccer	Handball	Cptr Game	Genre
Training	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F

INFORMATION GAIN

Find the code here

https://github.com/tchamna/information_gain_and_entropy_decision_tree

$$IG = E\left(Parent\right) - \sum_{i=1}^{N_{children}} \omega_i E\left(child_i\right)$$



ω_i = weight of child i

Parent: F,M,M,F,F

$$E\left(parent\right) = -\tfrac{3}{5}\log_2\left(\tfrac{3}{5}\right) - \tfrac{2}{5}\log_2\left(\tfrac{2}{5}\right)$$

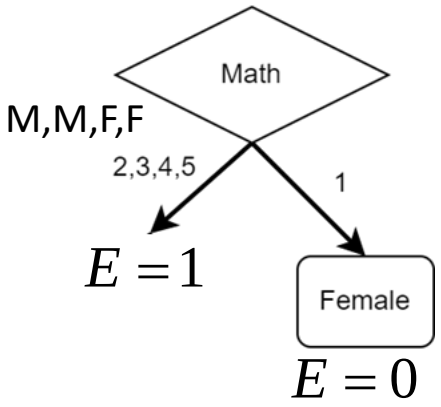
E(parent) = 0.97

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F

$$\sum_{i=1}^{N_{children}} \omega_i E\left(child_i\right) = \frac{3}{5}\times 0 + \frac{2}{5}\times 0 = 0$$

$$IG = E\left(Parent\right) = 0.97$$

$$IG = E(Parent) - \sum_{i=1}^{N_{children}} \omega_i E(child_i)$$



ω_i = weight of child i

Parent: F,M,M,F,F

$$E(parent) = -\tfrac{1}{5}\log_2\left(\tfrac{1}{5}\right) - \tfrac{4}{5}\log_2\left(\tfrac{4}{5}\right)$$

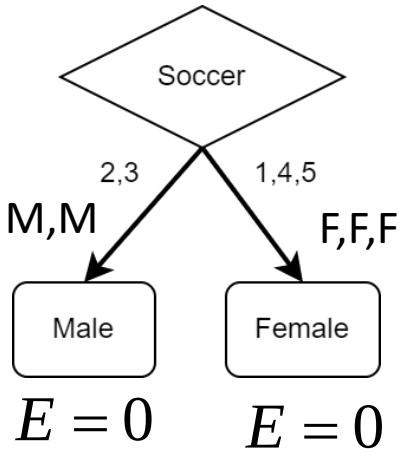
E(parent) = 0.72

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F

$$\sum_{i=1}^{N_{children}} \omega_i E(child_i) = \frac{4}{5}\times 1 + \frac{1}{5}\times 0 = 0.8$$

$$IG = E(Parent) = -0.08$$

$$IG = E(Parent) - \sum_{i=1}^{N_{children}} \omega_i E(child_i)$$



ω_i = weight of child i

Parent: F,M,M,F,F

$$E(parent) = -\tfrac{2}{5}\log_2\left(\tfrac{3}{5}\right) - \tfrac{3}{5}\log_2\left(\tfrac{2}{5}\right)$$

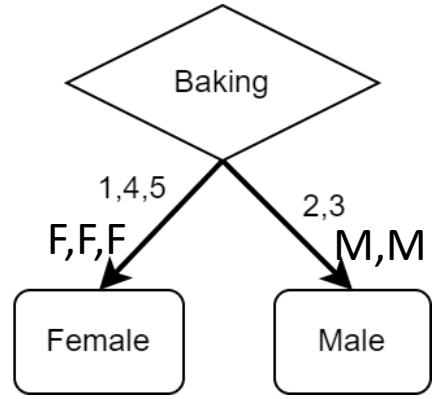
$$E(parent) = 0.97$$

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F

$$\sum_{i=1}^{N_{children}} \omega_i E(child_i) = 0$$

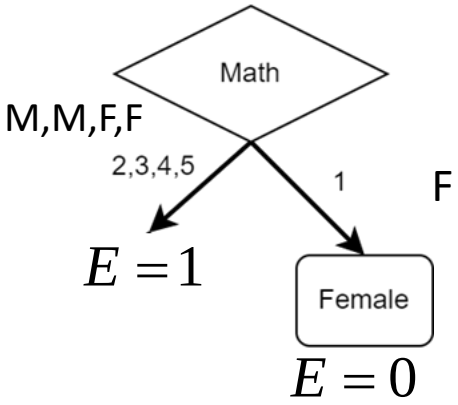
$$IG = E(Parent) = 0.97$$

$$IG = E(Parent) - \sum_{i=1}^{N_{children}} \omega_i E(child_i)$$



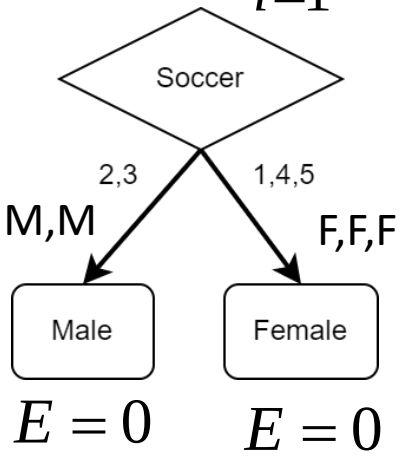
$E = 0 \quad E = 0$

$IG = 0.97$



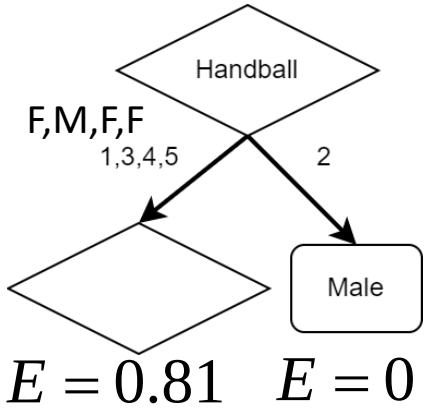
$E = 1 \quad E = 0$

$IG = -0.08$



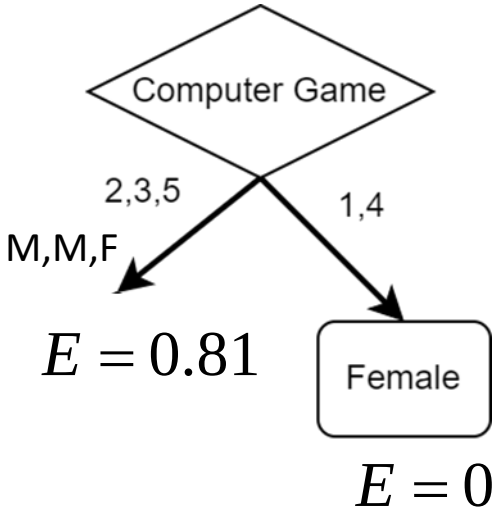
$E = 0 \quad E = 0$

$IG = 0.97$



$E = 0.81 \quad E = 0$

$IG = 0.072$



$E = 0.81 \quad E = 0$

$IG = 0.072$

	Young people 10-25						
	ID	**Baking**	**Math**	**Soccer**	**Handball**	**Cptr Game**	**Genre**
Training	1	1	0	0	1	0	F
	2	0	1	1	0	1	M
	3	0	1	1	1	1	M
	4	1	1	0	1	0	F
	5	1	1	0	1	1	F