

# AdaBoost

1



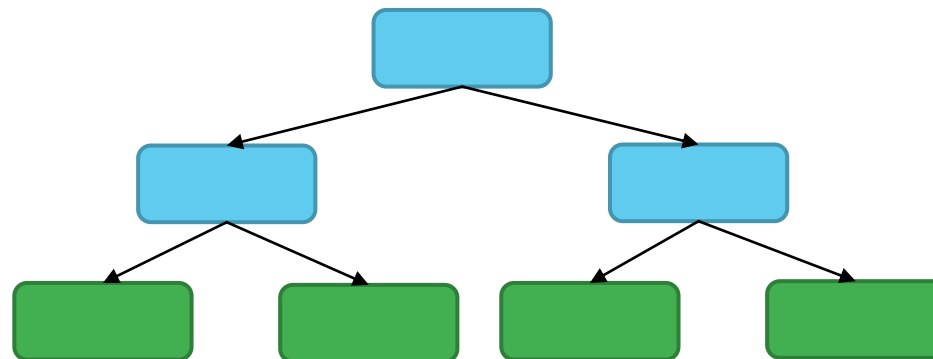
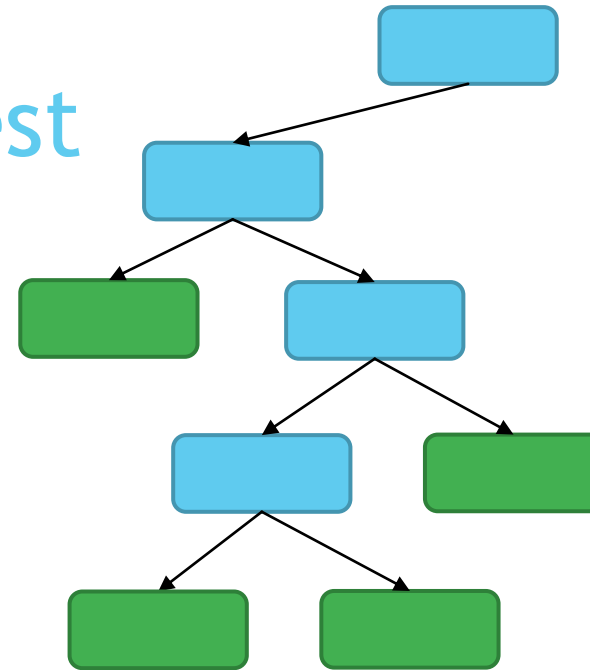
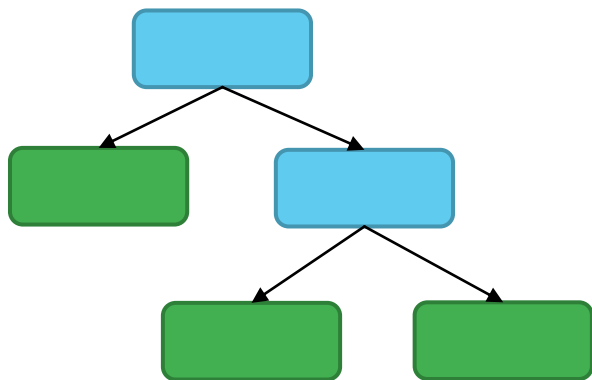
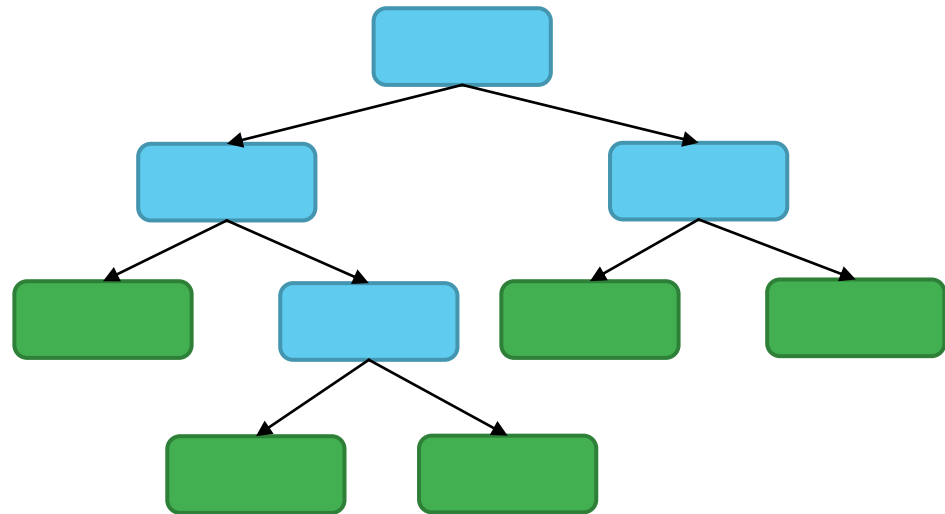


# Prérequis

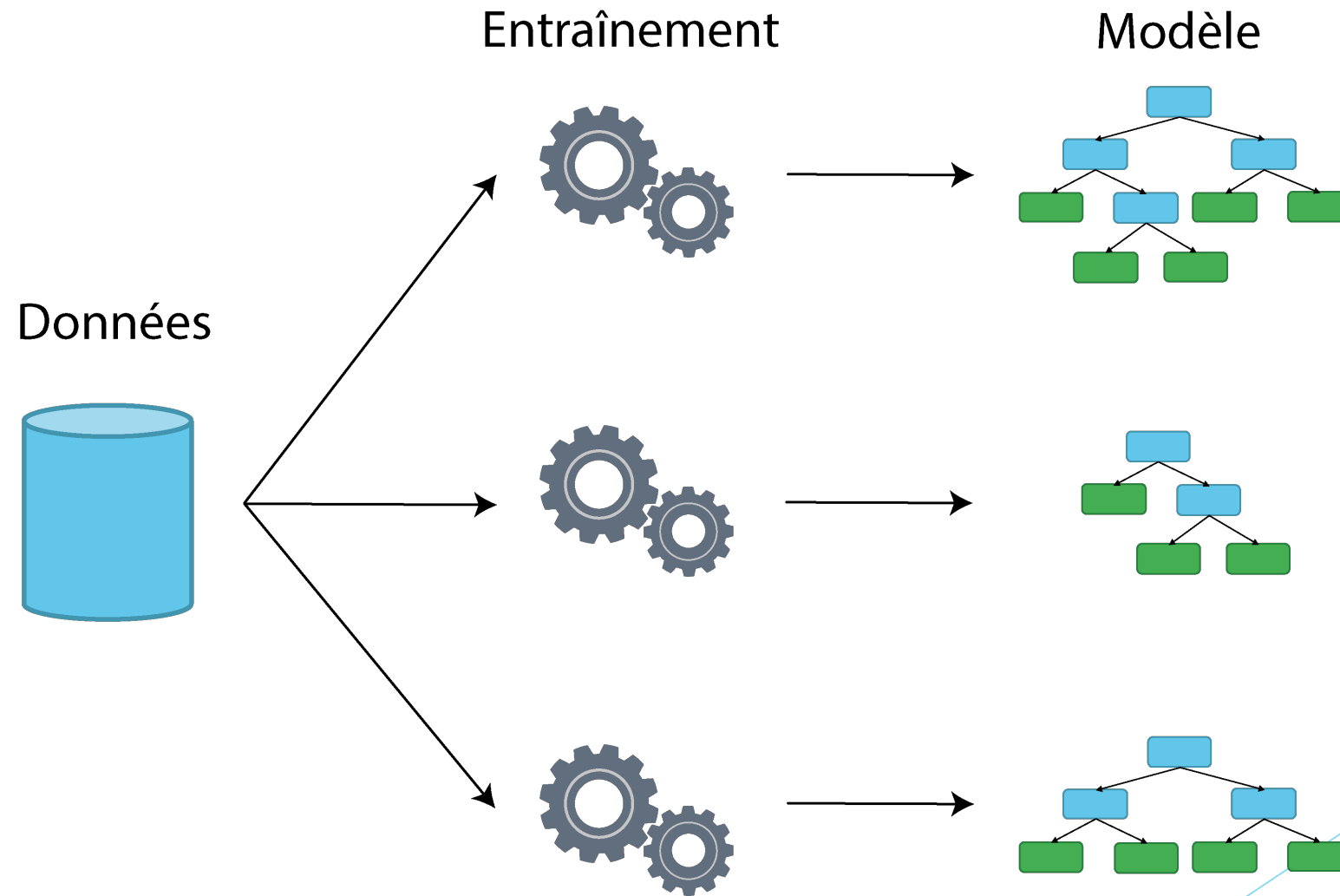
Pour comprendre la théorie d'AdaBoost il faut bien comprendre :

- L'ensemble learning
- L'algorithme des arbres de décision
- L'algorithme du random forest

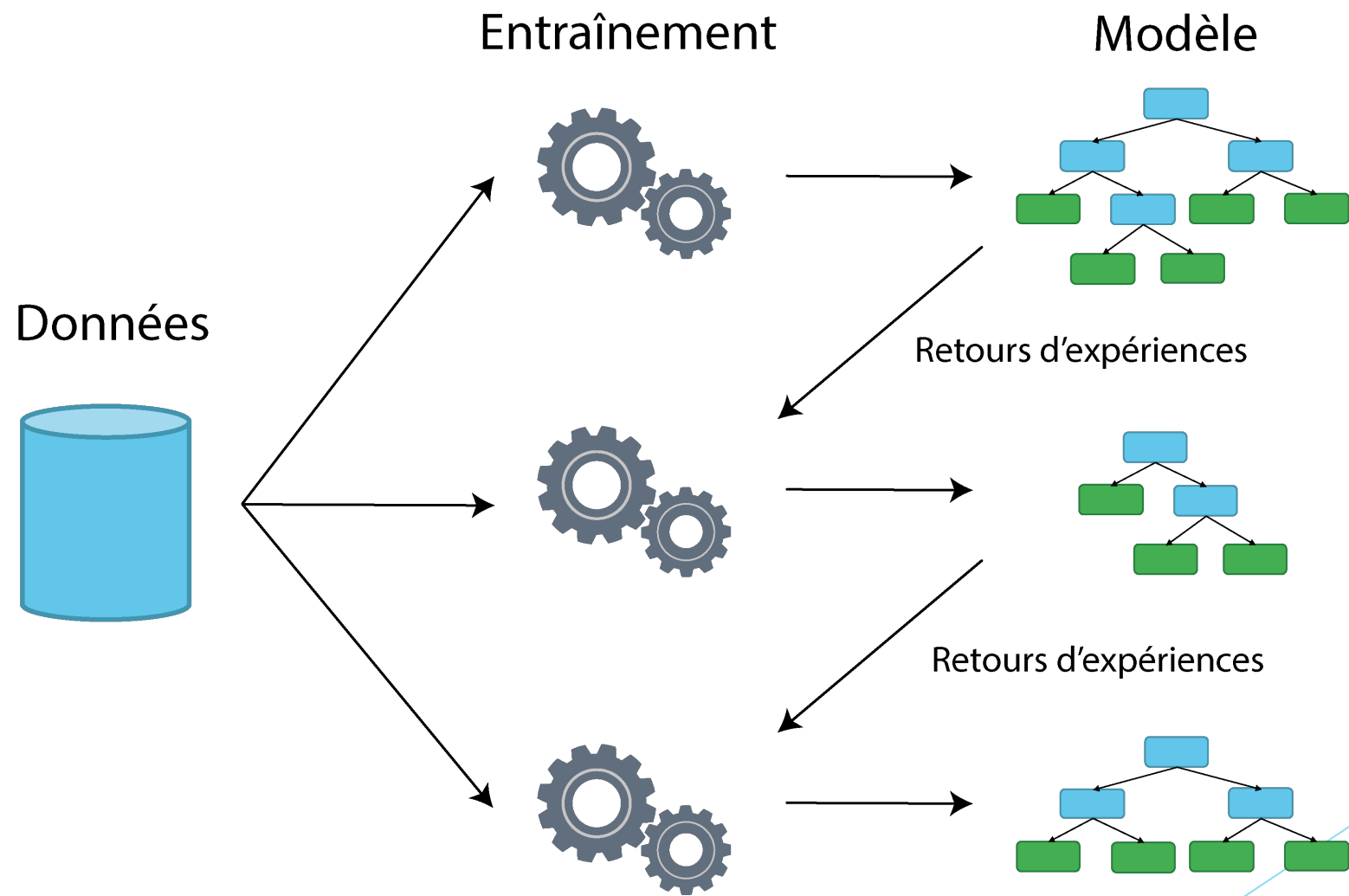
# Le problème du random forest



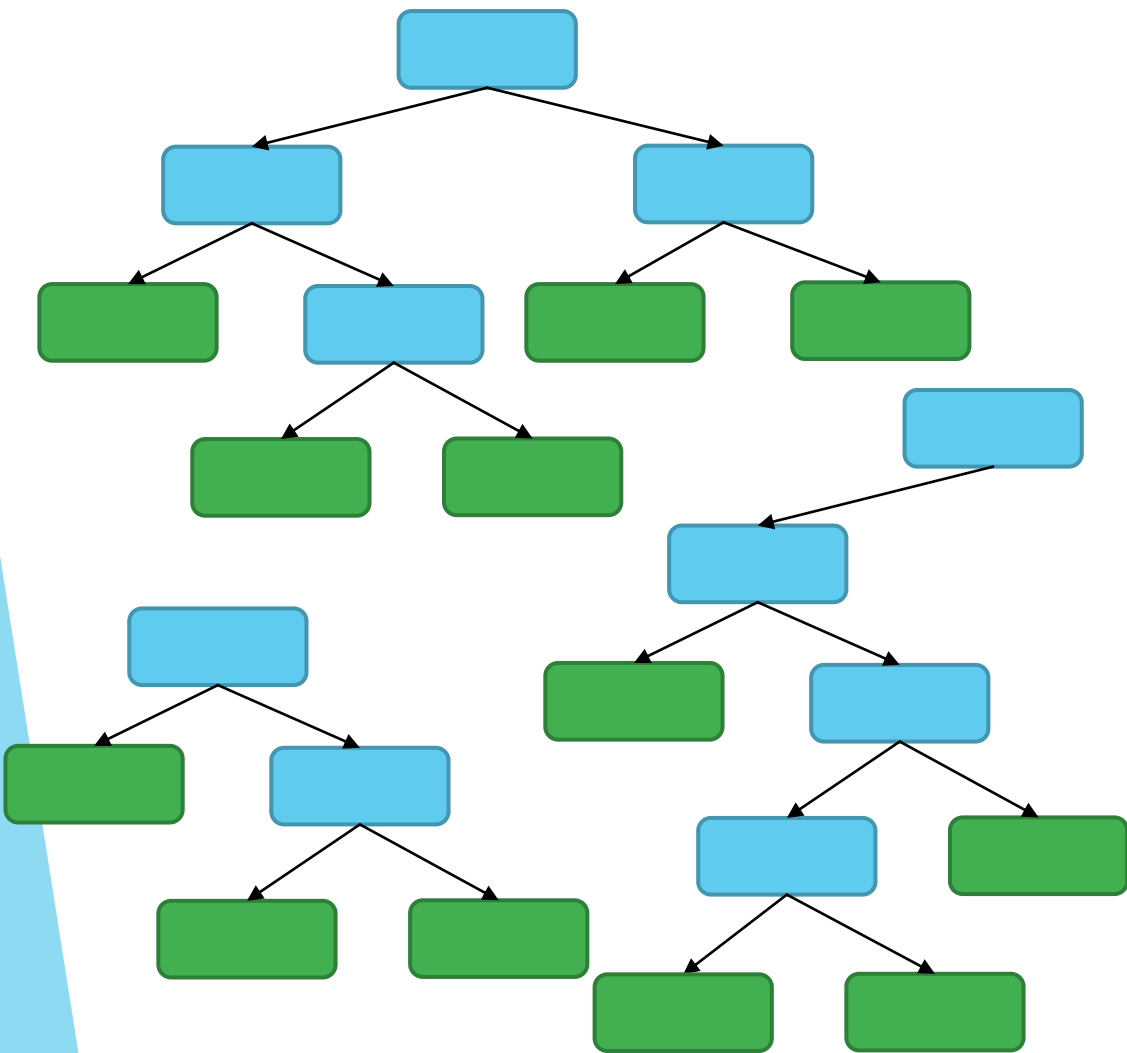
# De l'aléatoire au boosting



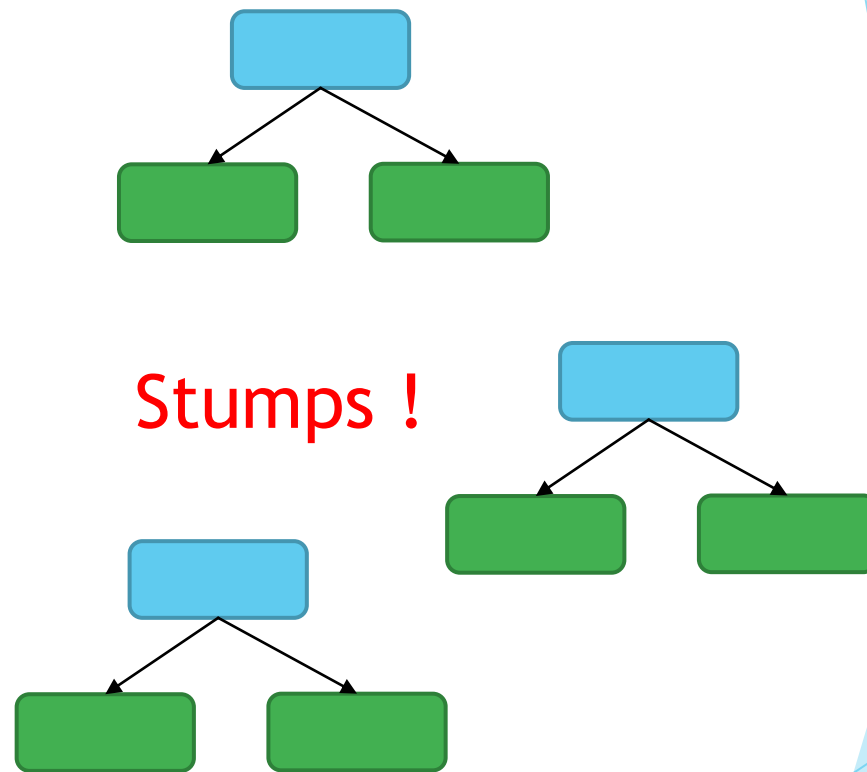
# De l'aléatoire au boosting



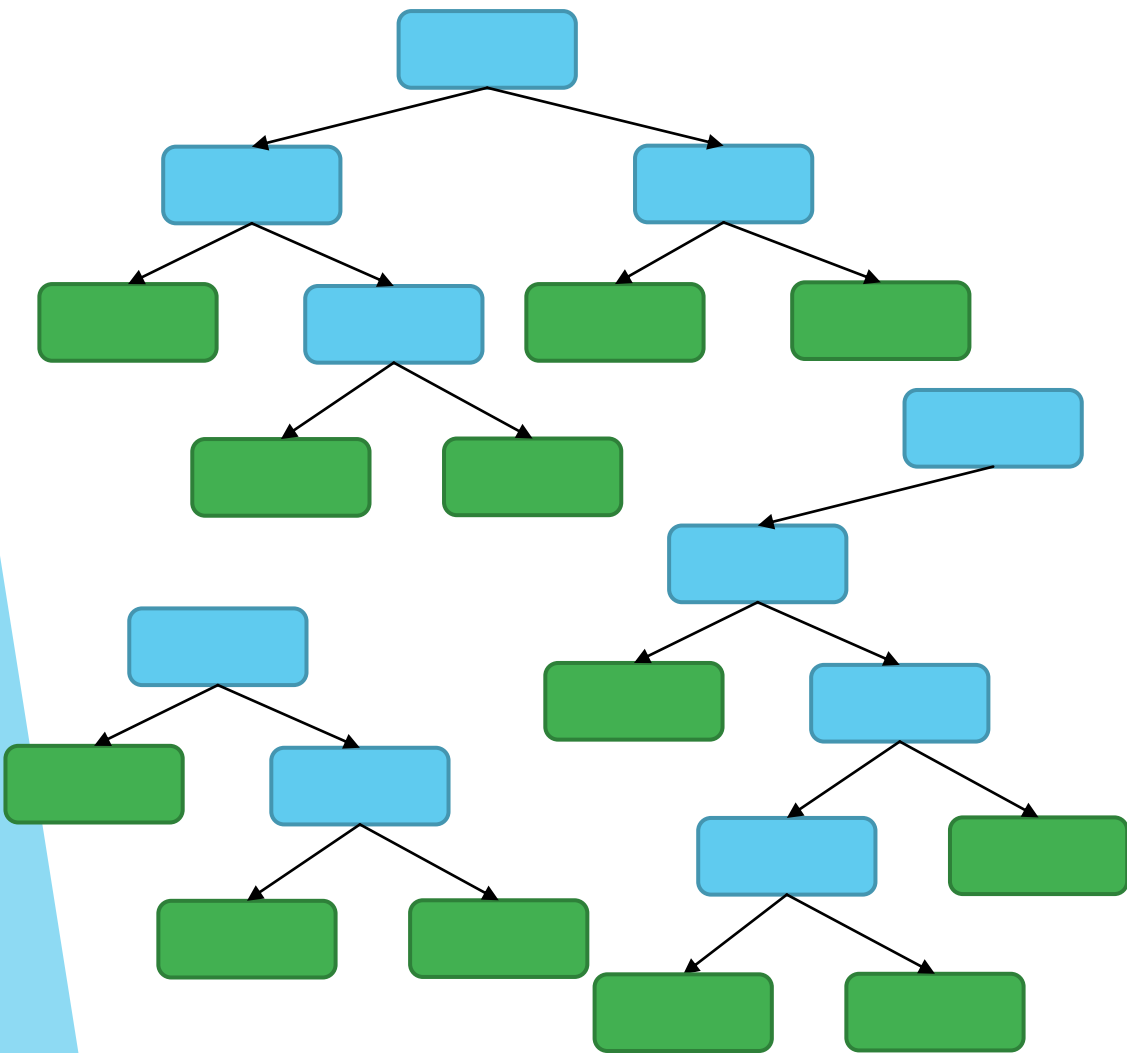
## Random forest



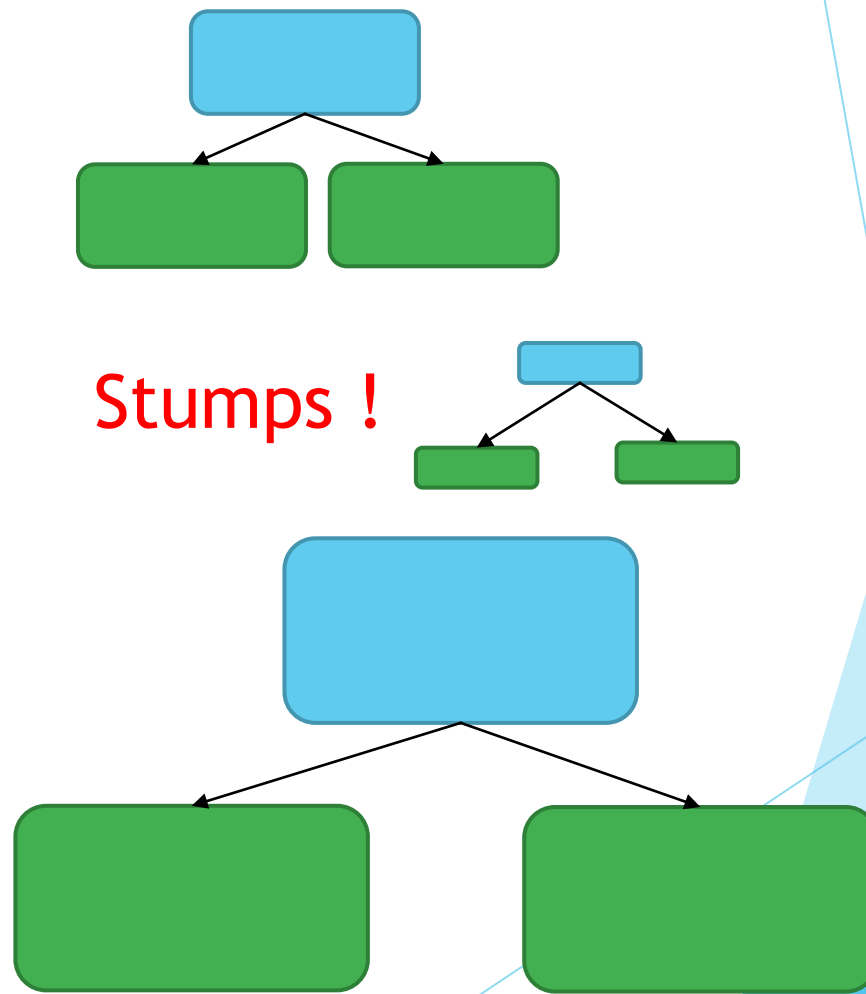
## AdaBoost



## Random forest



## AdaBoost





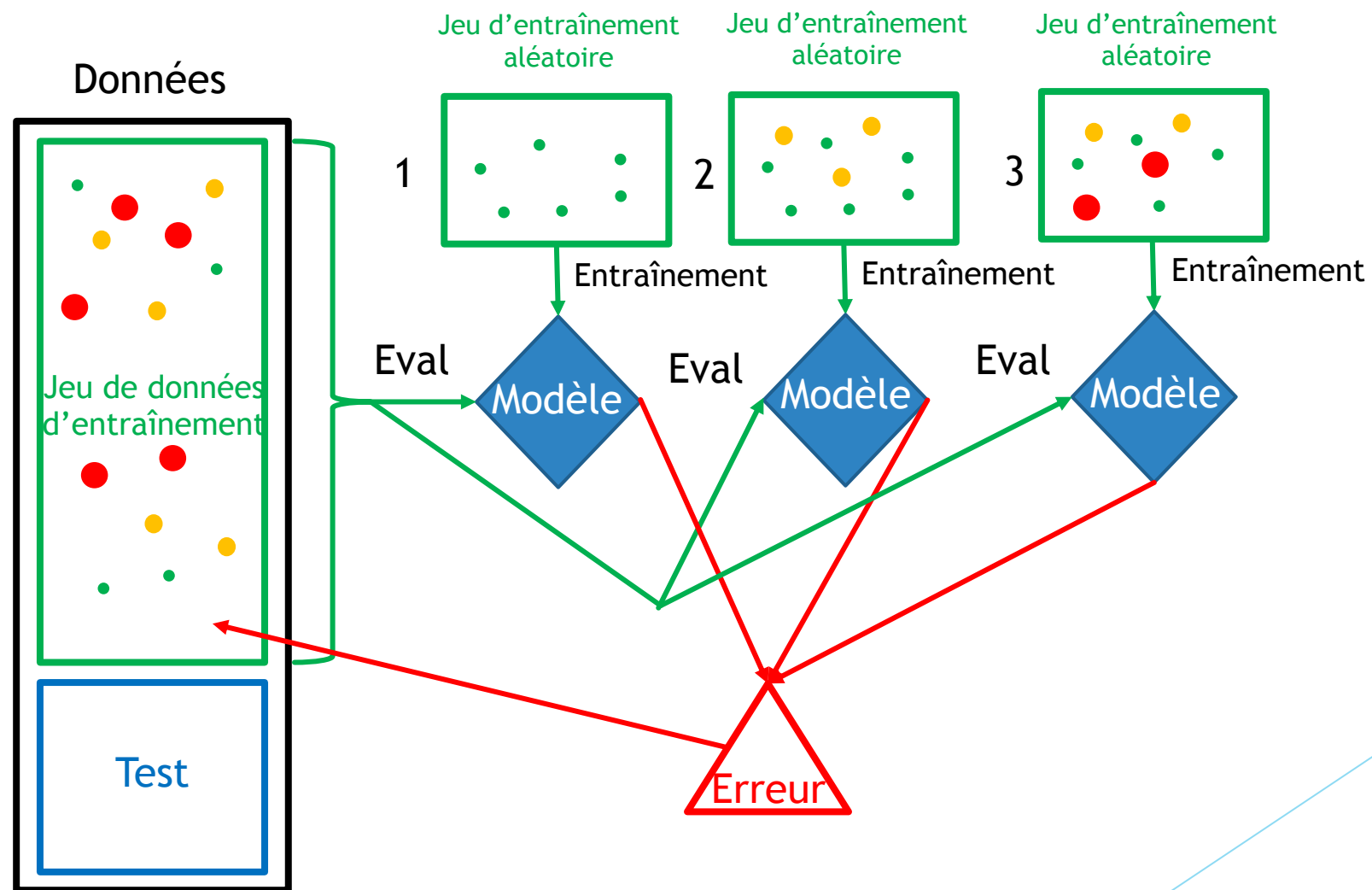
# Du random forest à AdaBoost

- ▶ Les arbres sont créés séquentiellement l'un après l'autre afin de minimiser l'erreur des arbres précédemment créés.
- ▶ Les arbres sont des stumps c'est-à-dire des arbres de profondeur une avec seulement deux feuilles.
- ▶ Les arbres peuvent avoir des poids différents en fonction de leur importance dans la forêt.





# AdaBoost



# Poids des observations

Sexe	Age	Classe	Survie	Poids
1	30	3	0	1/n
2	24	3	1	1/n
1	10	1	1	1/n
2	16	2	1	1/n
1	67	1	0	1/n

$$\frac{1}{\text{nombre d'exemples}}$$



# Poids des observations

Sexe	Age	Classe	Survie	Poids
1	30	3	0	1/5
2	24	3	1	1/5
1	10	1	1	1/5
2	16	2	1	1/5
1	67	1	0	1/5

$$\frac{1}{\text{nombre d'exemples}}$$

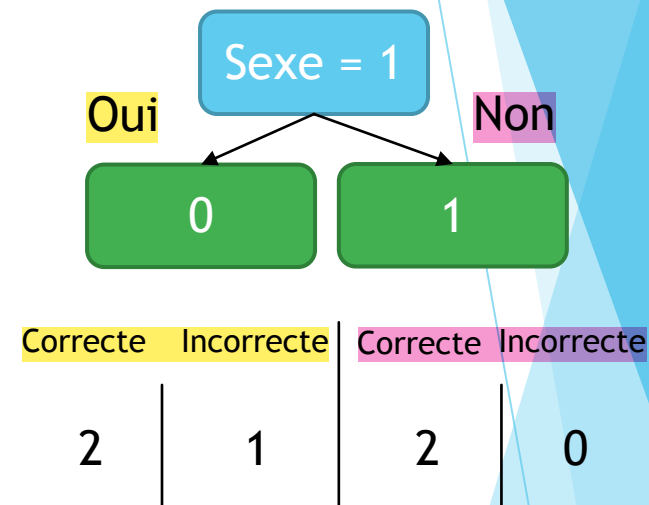


# Création d'un stump

étonnant que Oui soit à gauche avec valeur 0....à vérifier

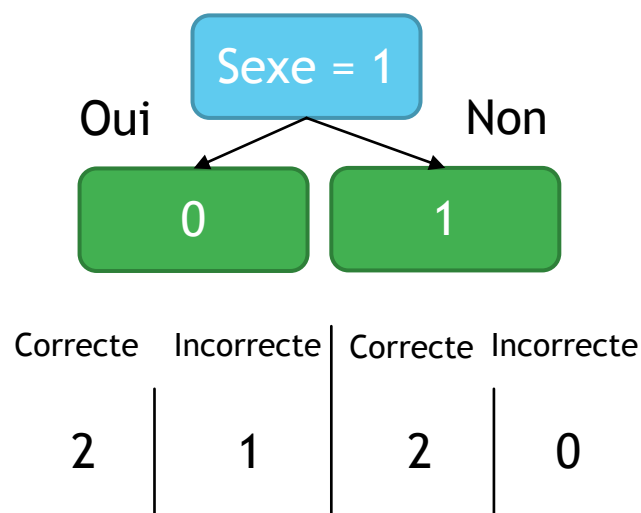
Sexe	Age	Classe	Survie (y)
1	30	3	0
2	24	3	1
1	10	1	1
2	16	2	1
1	67	1	0

Poids (p)
1/5
1/5
1/5
1/5
1/5





# Poids du votes de l'arbre



$$Erreur\ Total = \sum_{i=1}^n (y_i \neq \hat{y}_i) * p_i$$

$$Poids\ de\ l'arbre = \frac{1}{2} \log \left( \frac{1 - Erreur\ Total}{Erreur\ Total + \epsilon} + \epsilon \right)$$

epsilon ajouté pour éviter une division par 0 ou un log de x négatif

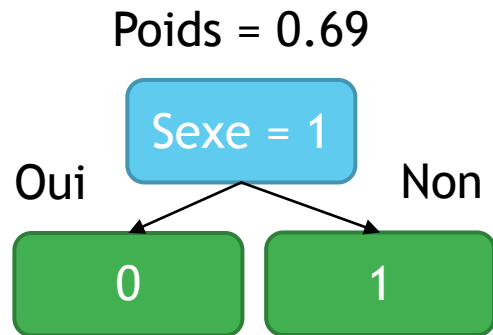


# Poids du votes de l'arbre

Sexe	Age	Classe	Survie (y)	Poids (p)
1	30	3	0	1/5
2	24	3	1	1/5
1	10	1	1	1/5
2	16	2	1	1/5
1	67	1	0	1/5

$$\begin{aligned} \text{Poids de l'arbre} &= \frac{1}{2} \log \left( \frac{1 - \text{Erreur Total}}{\text{Erreur Total}} \right) \\ &= \frac{1}{2} \log \left( \frac{1 - \frac{1}{5}}{\frac{1}{5}} \right) \\ &= 0.69 \end{aligned}$$

# Notre forêt





# Proportion de votes d'un arbre

Sexe	Age	Classe	Survie (y)	Poids (p)	Nouveau Poids (p)
1	30	3	0	1/5	
2	24	3	1	1/5	
1	10	1	1	1/5	0.4
2	16	2	1	1/5	
1	67	1	0	1/5	

$$\text{Nouveau poids}_{y \neq \hat{y}} = \text{poids} \cdot e^{\text{Poids de l'arbre}}$$

$$\text{Nouveau poids}_{y \neq \hat{y}} = \frac{1}{5} \cdot e^{0.69} = 0.40$$

$$\frac{1}{5} = 0.2 < 0.4$$



# Proportion de votes d'un arbre

Sexe	Age	Classe	Survie (y)	Poids (p)	Nouveau Poids (p)
1	30	3	0	1/5	0.1
2	24	3	1	1/5	0.1
1	10	1	1	1/5	0.4
2	16	2	1	1/5	0.1
1	67	1	0	1/5	0.1

$Nouveau poids_{y=\hat{y}} = poids \cdot e^{-Poids \text{ de l'arbre}}$

$$Nouveau poids_{y=\hat{y}} = \frac{1}{5} \cdot e^{-0.69} = 0.10$$

$$\frac{1}{5} = 0.2 > 0.1$$

# Normalisation des poids

Sexe	Age	Classe	Survie (y)	Nouveau Poids (p)	Poids norm. (p)
1	30	3	0	0.1	0.1/0.8
2	24	3	1	0.1	0.1/0.8
1	10	1	1	0.4	0.4/0.8
2	16	2	1	0.1	0.1/0.8
1	67	1	0	0.1	0.1/0.8

} = 0.8

# Normalisation des poids

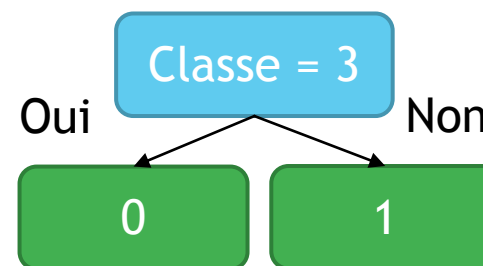
Sexe	Age	Classe	Survie (y)	Nouveau Poids (p)	Poids norm. (p)
1	30	3	0	0.1	0.125
2	24	3	1	0.1	0.125
1	10	1	1	0.4	0.5
2	16	2	1	0.1	0.125
1	67	1	0	0.1	0.125

$\left. \begin{array}{c} \text{Nouveau Poids (p)} \\ \text{Poids norm. (p)} \end{array} \right\} = 0.8$ 
 $\left. \begin{array}{c} \text{Poids norm. (p)} \end{array} \right\} = 1$



# Création d'un nouvel arbre

Sexe	Age	Classe	Survie (y)	Poids (p)
1	30	3	0	0.125
2	24	3	1	0.125
1	10	1	1	0.5
2	16	2	1	0.125
1	67	1	0	0.125

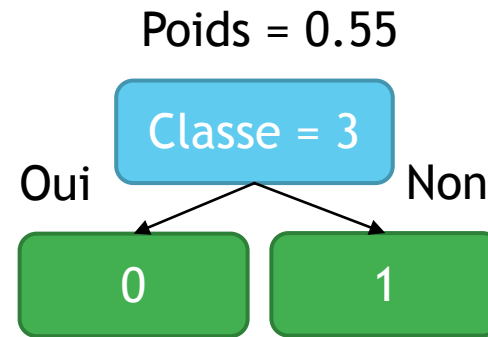
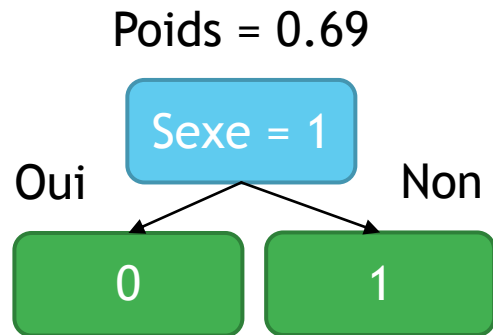


Correcte	Incorrecte	Correcte	Incorrecte
1	1	2	1

$$Erreur\ Total = \sum_{i=1}^n (y_i \neq \hat{y}_i) * p_i = 0.25$$

$$Poids\ de\ l'arbre = \frac{1}{2} \log \left( \frac{1 - Erreur\ Total}{Erreur\ Total + \varepsilon} + \varepsilon \right) = 0.55$$

# Notre forêt





## Calcul des nouveaux poids

$$\text{Nouveau poids}_{y \neq \hat{y}} = \text{poids} \cdot e^{\text{Poids de l'arbre}}$$

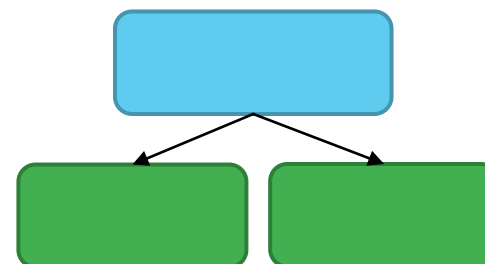
$$\text{Nouveau poids}_{y = \hat{y}} = \text{poids} \cdot e^{-\text{Poids de l'arbre}}$$

Sexe	Age	Classe	Survie (y)	Poids (p)	Nouveau poids (p)	Poids norm. (p)
1	30	3	0	0.125	0.07	0.08
2	24	3	1	0.125	0.22	0.25
1	10	1	1	0.5	0.29	0.34
2	16	2	1	0.125	0.07	0.08
1	67	1	0	0.125	0.22	0.25



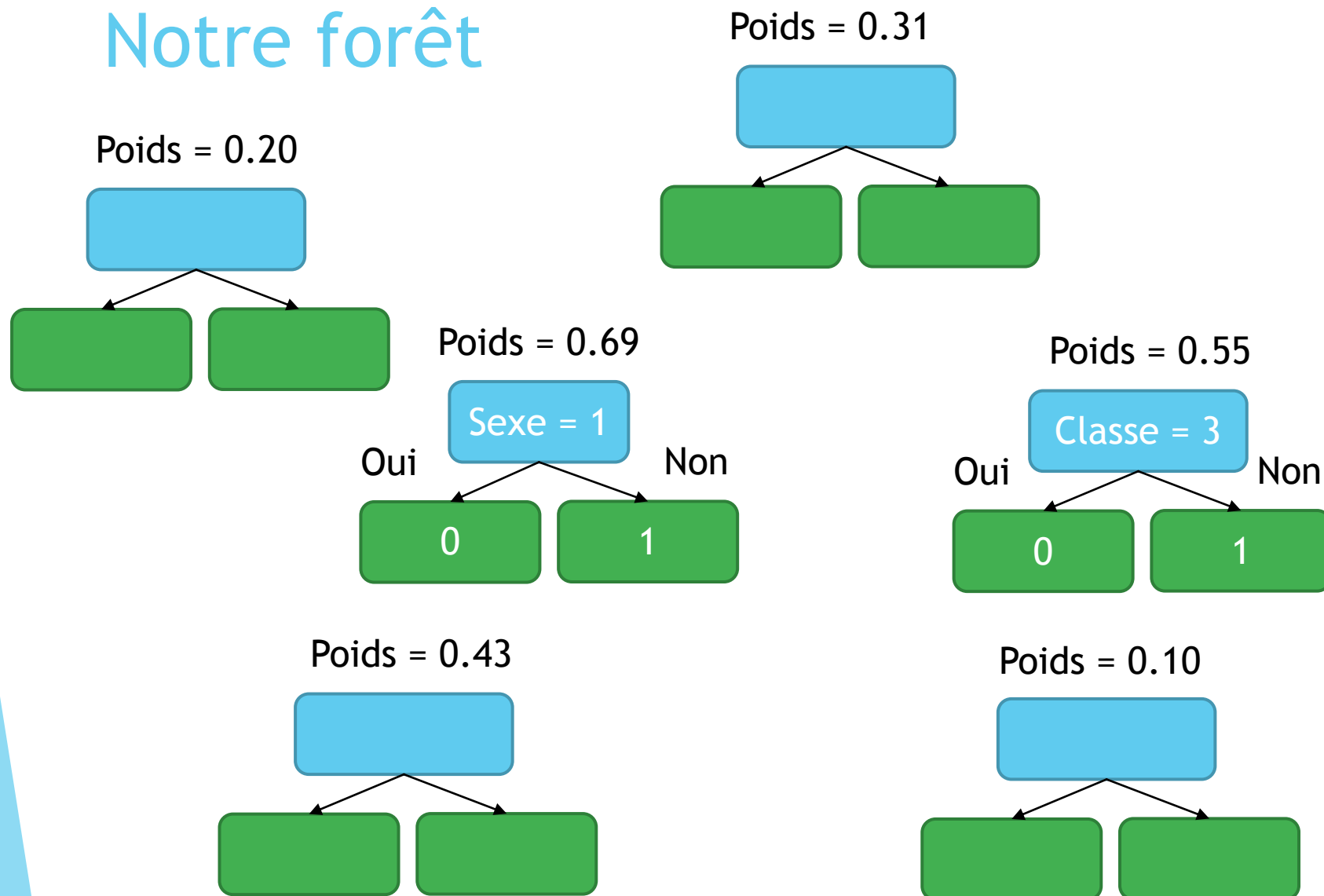
# Création d'un nouvel arbre

Sexe	Age	Classe	Survie (y)	Poids (p)
1	30	3	0	0.08
2	24	3	1	0.25
1	10	1	1	0.34
2	16	2	1	0.08
1	67	1	0	0.25



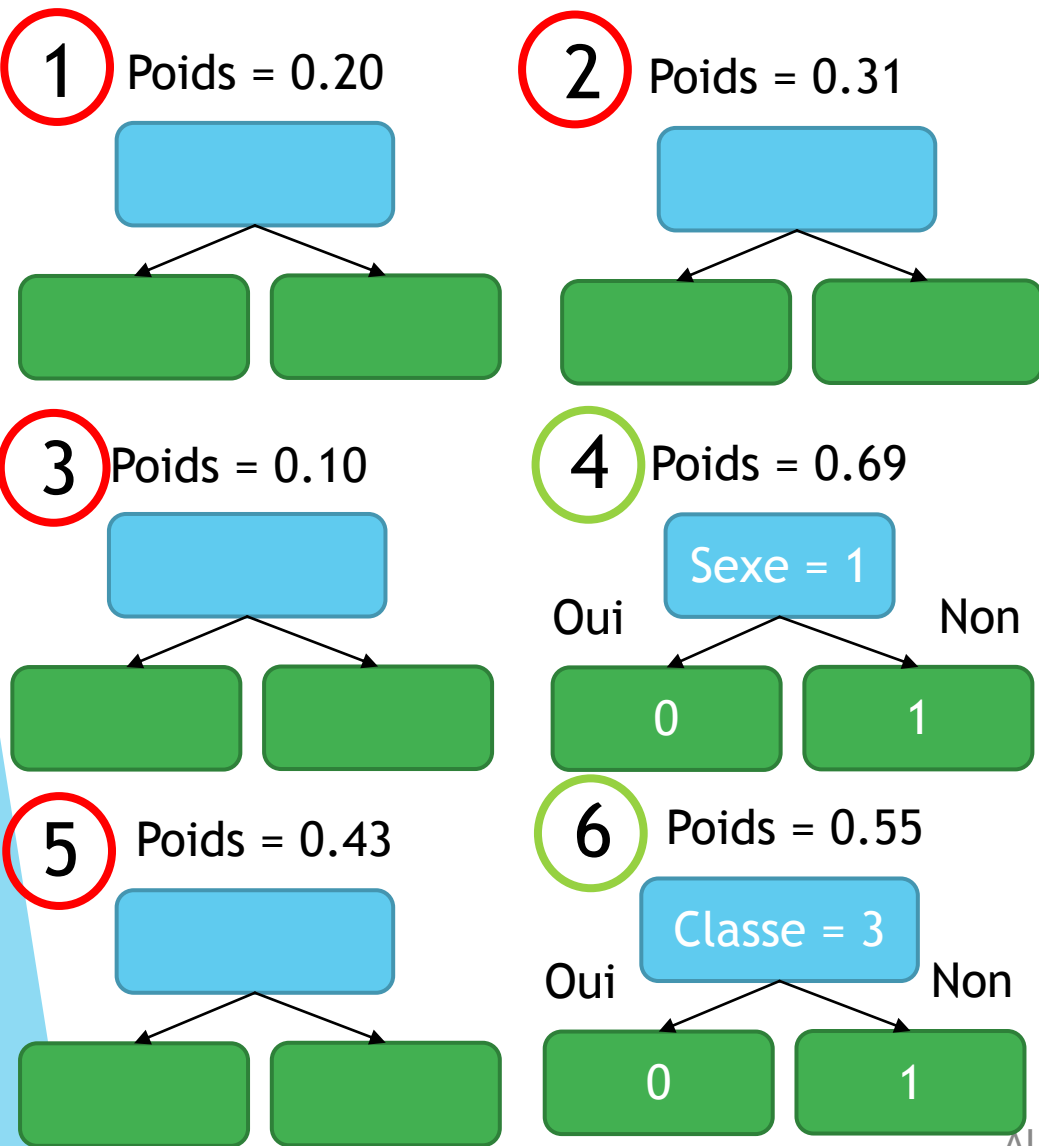


# Notre forêt



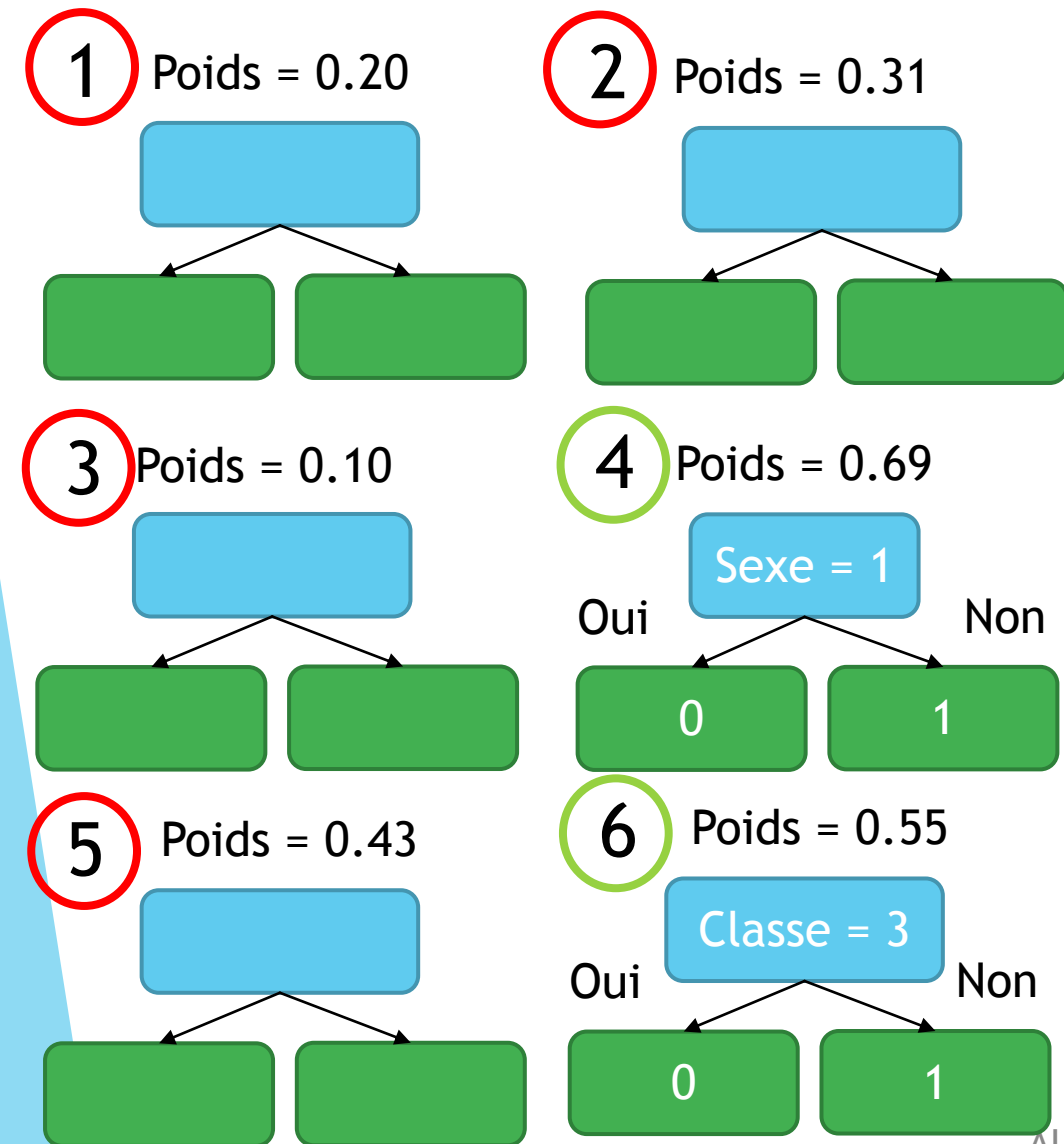


# Prédiction de la forêt



Arbre	Décision
1	0
2	0
3	0
4	1
5	0
6	1

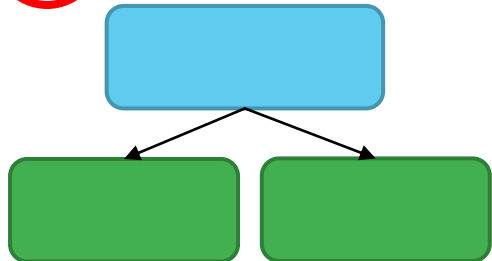
# Prédiction de la forêt



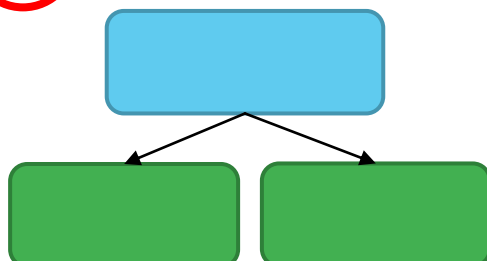
Arbre	Décision	Poids
1	0	0.2
2	0	0.31
3	0	0.10
4	1	0.69
5	0	0.43
6	1	0.55

# Prédiction de la forêt

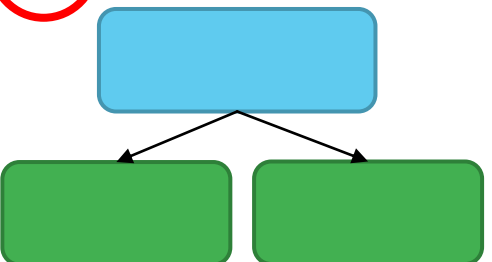
1 Poids = 0.20



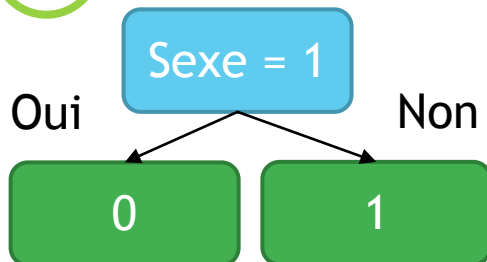
2 Poids = 0.31



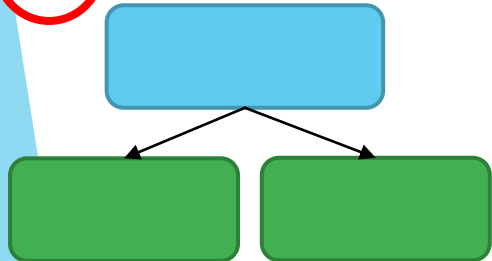
3 Poids = 0.10



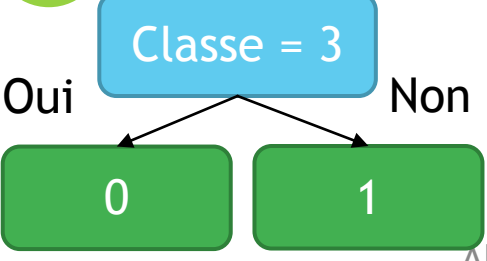
4 Poids = 0.69



5 Poids = 0.43



6 Poids = 0.55



Arbre	Décision	Poids	%
1	0	0.2	9%
2	0	0.31	14%
3	0	0.10	4%
4	1	0.69	30%
5	0	0.43	19%
6	1	0.55	24%

1 - 54%

0 - 46%