



LYCÉE LECONTE DE LISLE

Algorithme ID3 : exemple

Vincent Picard

Nom	Fille ?	Courage ?	Intelligence ?	Malice ?	Maison
Harry	non	oui	non	oui	Gryffondor
Ron	non	oui	non	non	Gryffondor
Hermione	oui	oui	oui	non	Gryffondor
Neville	non	oui	non	non	Gryffondor
Drago	non	non	oui	oui	Serpentard
Goyle	non	non	non	oui	Serpentard
Cédric	non	oui	oui	non	Poufsouffle
Luna	oui	non	oui	non	Serdaigle

- Les attributs d'apprentissage sont : Fille, Courage, Intelligence et Malice.
- L'attribut cible (classe) est Maison.
- On souhaite construire un arbre binaire de décision pour décider la maison d'un nouvel élève → on utilise l'algorithme ID3.

- Le calcul de l'entropie initiale de l'ensemble d'exemples S donne :

$$H(S) = -\frac{4}{8}\log_2\frac{4}{8} - \frac{2}{8}\log_2\frac{2}{8} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{8}\log_2\frac{1}{8} = 1,75$$

- Le calcul des gains d'information pour chaque attribut donne :

- ▶ $G(S, \text{Fille}) = 0,41$

- ▶ $G(S, \text{Courage}) = 0,95$

- ▶ $G(S, \text{Intelligence}) = 0,34$

- ▶ $G(S, \text{Malice}) = 0,55$

- On en déduit que l'attribut le plus discriminant est Courage

- On crée un nœud étiqueté Courage et on recommence pour chaque fils

Courage ?

Cas : Courage = non

Nom	Fille ?	Courage ?	Intelligence ?	Malice ?	Maison
Drago	non	non	oui	oui	Serpentard
Goyle	non	non	non	oui	Serpentard
Luna	oui	non	oui	non	Serdaigle

- Le calcul de l'entropie initiale de l'ensemble d'exemples S donne :

$$H(S) = 0.92$$

- Le calcul des gains d'information pour chaque attribut donne :

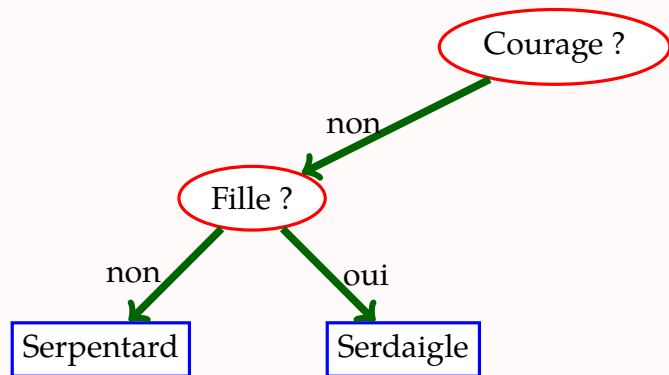
- ▶ $G(S, Fille) = 0,92$

- ▶ $G(S, Intelligence) = 0,25$

- ▶ $G(S, Malice) = 0,92$

- On en déduit que l'attribut le plus discriminant est soit Fille, soit Malice (et par ailleurs le gain d'information est maximal). On choisit Fille (par exemple).
- Dans les deux sous-cas obtenus les exemples sont dans la même classe : on obtient deux feuilles.

- On complète le sous-arbre gauche avec un nœud **Fille** ayant pour fils gauche la feuille **Serpentard** et pour fils droit la feuille **Serdaigle**.



Cas : Courage = oui

Nom	Fille ?	Courage ?	Intelligence ?	Malice ?	Maison
Harry	non	oui	non	oui	Gryffondor
Ron	non	oui	non	non	Gryffondor
Hermione	oui	oui	oui	non	Gryffondor
Neville	non	oui	non	non	Gryffondor
Cédric	non	oui	oui	non	Poufsouffle

- Le calcul de l'entropie initiale de l'ensemble d'exemples S donne :

$$H(S) = 0.72$$

- Le calcul des gains d'information pour chaque attribut donne :

- ▶ $G(S, Fille) = 0,07$

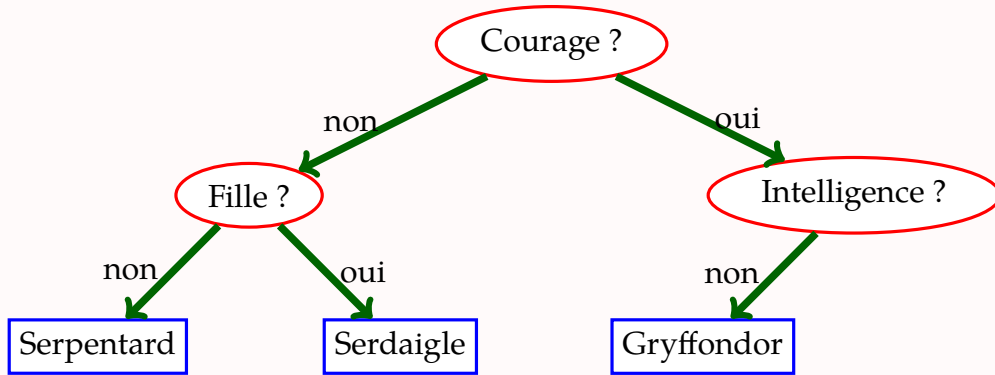
- ▶ $G(S, Intelligence) = 0,32$

- ▶ $G(S, Malice) = 0,07$

- On en déduit que l'attribut le plus discriminant est Intelligence

- Dans le cas Intelligence = non tous les exemples sont étiquetés par Gryffondor le fils gauche est donc une feuille Gryffondor

- On complète le sous-arbre droit avec un nœud Intelligence ayant pour fils gauche la feuille Gryffondor et un fils droit à déterminer.



Cas : Courage = oui, Intelligence = oui

Nom	File ?	Courage ?	Intelligence ?	Malice ?	Maison
Hermione	oui	oui	oui	non	Gryffondor
Cédric	non	oui	oui	non	Poufsouffle

- Le calcul de l'entropie initiale de l'ensemble d'exemples S donne :

$$H(S) = 1$$

- Le calcul des gains d'information pour chaque attribut donne :

- ▶ $G(S, Fille) = 1$

- ▶ $G(S, Malice) = 0$

- On en déduit que l'attribut le plus discriminant est Fille

- Dans chaque cas Fille = oui et Fille = non, les exemples restants ont la même classe : on construit donc deux feuilles.

On obtient l'arbre final suivant :

