

# EPISEN – ING3. SI

## Machine Learning



**Abdallah EL HIDALI**

Tech Lead Sita For Aircraft  
abdallah.el-hidali@sit.aero

**EPISEN**

2024/2025

# IV. Les arbres de décision

# Les arbres de décisions: introduction

Gender	Occupation	App
F	Study	Pokemon Go 
F	Work	WhatsApp 
M	Work	Snapchat 
F	Work	WhatsApp 
M	Study	Pokemon Go 
M	Study	Pokemon Go 

Pour une femme qui travaille dans un bureau, quelle application recommandons-nous ?

- ☐ Pokemon Go
- ☐ Whatsapp
- ☐ Snapchat

Pour un homme qui travaille dans une usine, quelle application recommandons-nous ?

- ☐ Pokemon Go
- ☐ Whatsapp
- ☐ Snapchat

Pour une fille qui va au lycée, quelle application recommandons-nous ?

- ☐ Pokemon Go
- ☐ Whatsapp
- ☐ Snapchat

# Les arbres de décisions: introduction

Gender	Occupation	App
F	Study	Pokemon Go 
F	Work	WhatsApp 
M	Work	Snapchat 
F	Work	WhatsApp 
M	Study	Pokemon Go 
M	Study	Pokemon Go 

Pour une femme qui travaille dans un bureau, quelle application recommandons-nous ?

- ☐ Pokemon Go
- ☒ **Whatsapp**
- ☐ Snapchat




Pour un homme qui travaille dans une usine, quelle application recommandons-nous ?

- ☐ Pokemon Go
- ☐ Whatsapp
- ☒ **Snapchat**

Pour une fille qui va au lycée, quelle application recommandons-nous ?

- ☒ **Pokemon Go**
- ☐ Whatsapp
- ☐ Snapchat




# Les arbres de décisions: introduction

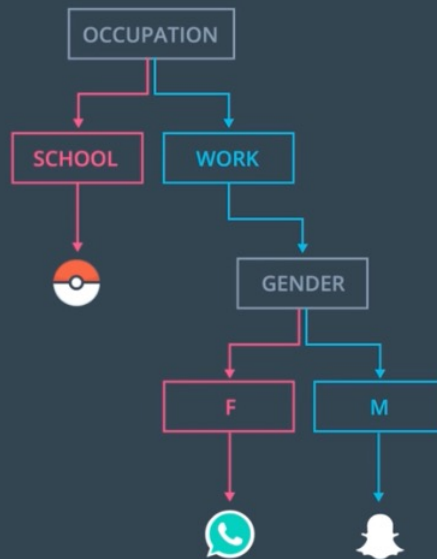
Gender	Occupation	App
F	Study	Pokemon Go 
F	Work	WhatsApp 
M	Work	Snapchat 
F	Work	WhatsApp 
M	Study	Pokemon Go 
M	Study	Pokemon Go 

**Essayez de créer un arbre de décision en vous basant sur le tableau ci-dessus**

# Les arbres de décisions: introduction

## Recommending Apps

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	



Dans la suite du cours, nous examinerons comment déterminer que la colonne "occupation" est la plus pertinente pour effectuer la première division de notre arbre de décision

# L'entropie dans les arbres de décision:

## Définition :

- L'entropie mesure l'impureté ou le désordre dans un ensemble de données.

## Formule:

$$H(S) = - \sum_{i=1}^c p_i \log_2(p_i)$$

c: nombre de classes

pi: la proportion d'éléments de la classe

## Propriétés:

- Varie entre 0 (homogénéité parfaite) et 1 (impureté maximale)
- Entropie faible = meilleure séparation des classes

## Utilisation:

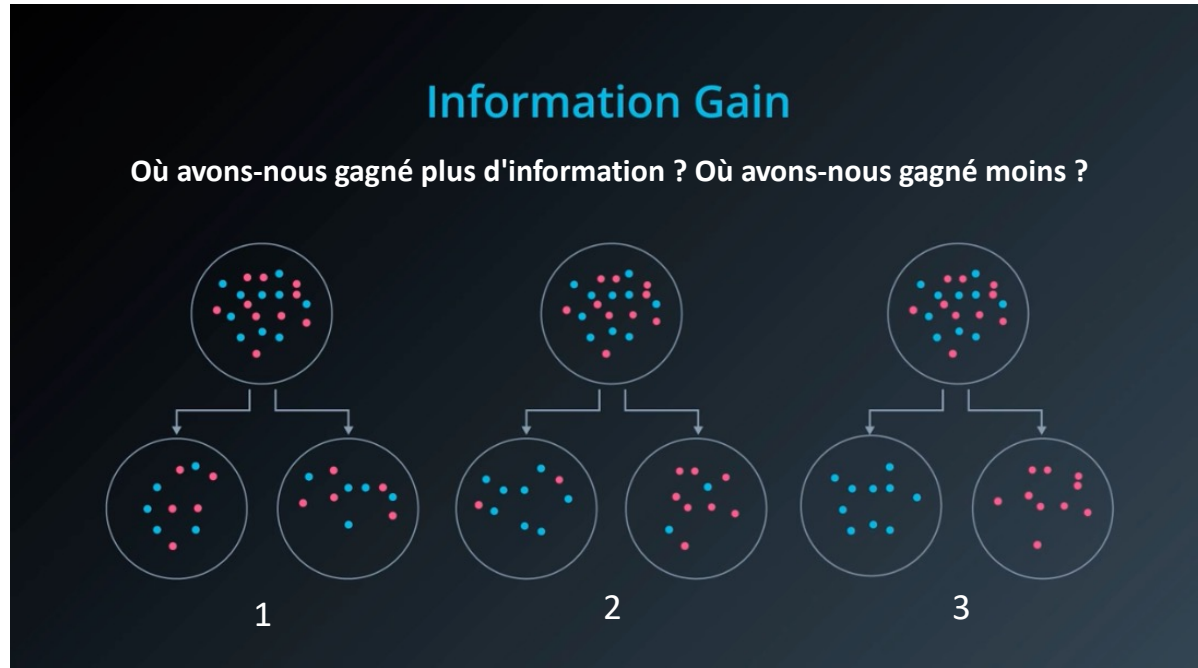
- Sélectionner le meilleur attribut pour diviser un nœud
- Calculer le gain d'information

## Objectif:

- Minimiser l'entropie à chaque division pour construire un arbre optimal

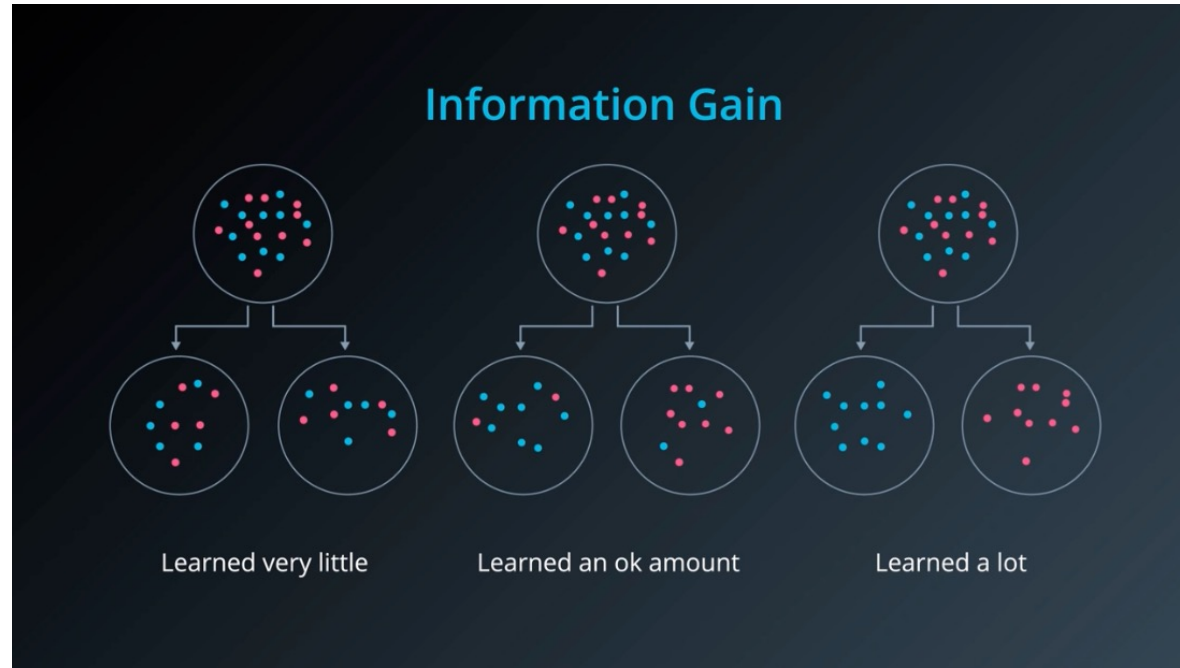


# Le gain d'information



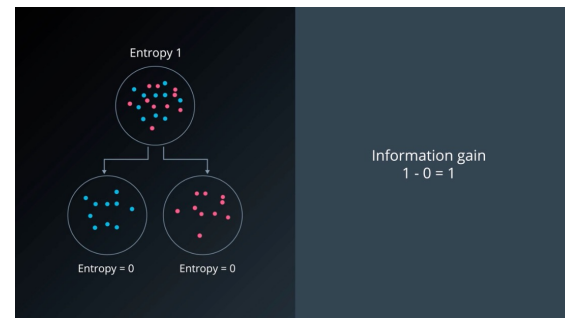
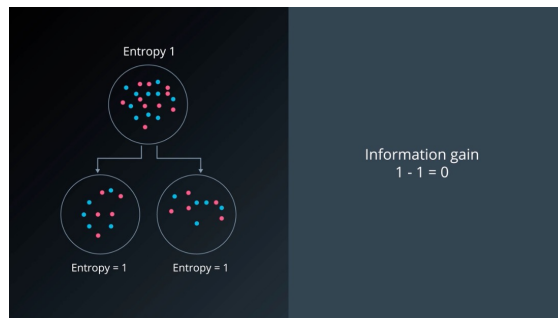
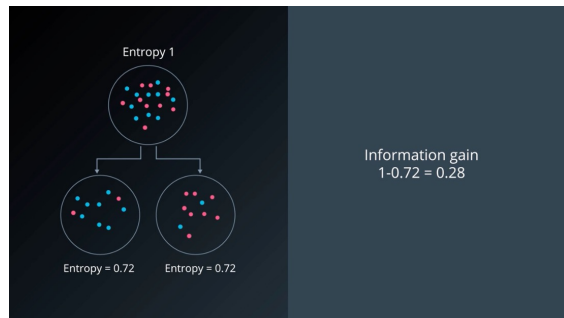


# Le gain d'information



$$\text{InformationGain} = \text{Entropy}(\text{Parent}) - \left( \frac{m}{m+n} \text{Entropy}(\text{Child1}) + \frac{n}{m+n} \text{Entropy}(\text{Child2}) \right)$$

# Le gain d'information



Le découpage optimal est celui qui maximise le gain d'information.

# Construction d'un arbre de décision

- ✓ Examiner tous les points de division possibles pour chaque attribut (colonne/feature) du jeu de données.
- ✓ Pour chaque point de division potentiel :
  - Calculer le gain d'information résultant de cette division.
- ✓ Sélectionner l'attribut et le point de division qui maximisent le gain d'information.
- ✓ Effectuer la division du jeu de données selon l'attribut et le point choisis.
- ✓ Répéter récursivement ce processus pour chaque sous-ensemble résultant de la division, jusqu'à ce qu'un critère d'arrêt soit atteint (par exemple, profondeur maximale de l'arbre ou gain d'information minimal).

Gender	Occupation	App
F	Study	
F	Work	
M	Work	
F	Work	
M	Study	
M	Study	

# Construction d'un arbre de décision

## Recommending Apps

Gender	Occupation	App
F	Study	Pokemon
F	Work	WhatsApp
M	Work	WhatsApp
F	Work	WhatsApp
M	Study	Pokemon
M	Study	Pokemon

$$\text{Entropy} = -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right) - \frac{1}{6}\log_2\left(\frac{1}{6}\right) = 1.46$$

## Recommending Apps

Gender	Occupation	App
F	Study	Pokemon
F	Work	WhatsApp
M	Work	WhatsApp
F	Work	WhatsApp
M	Study	Pokemon
M	Study	Pokemon

Gender

F M

0.92 0.92

Entropy

Information gain = 1.46 - 0.92 = 0.54

## Recommending Apps

Gender	Occupation	App
F	Study	Pokemon
F	Work	WhatsApp
M	Work	WhatsApp
F	Work	WhatsApp
M	Study	Pokemon
M	Study	Pokemon

Occupation

S M

0 0.92

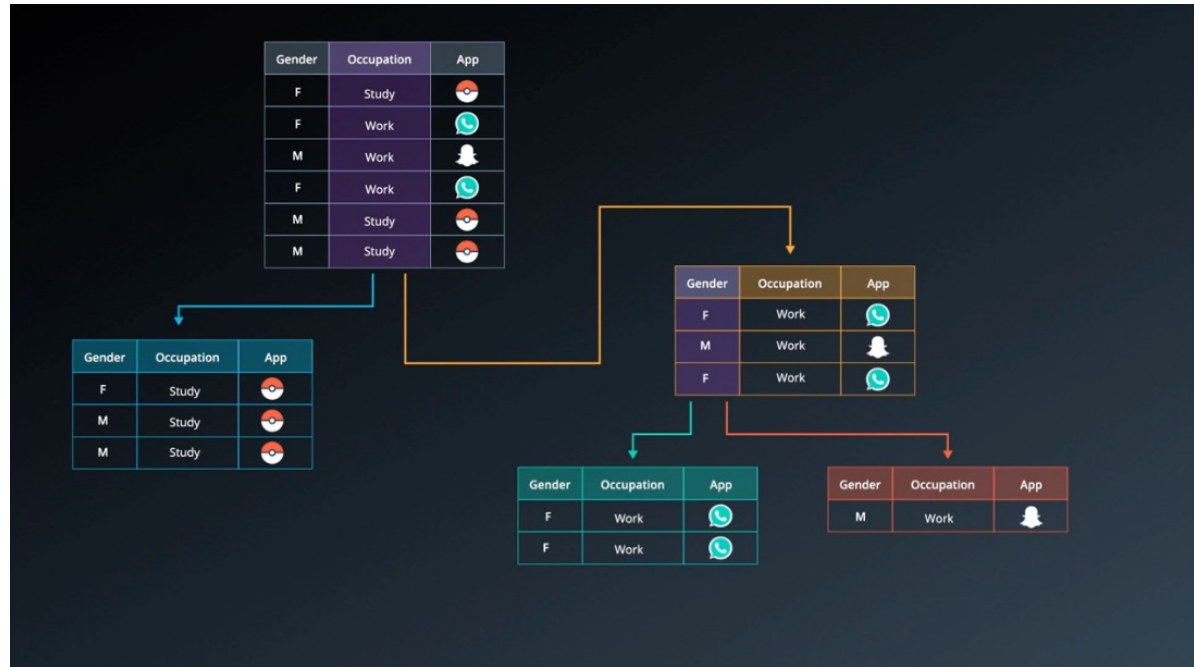
Entropy

Information gain = 1.46 - 0.46 = 1

## Recommending Apps

Gender	Occupation	App
F	Study	Pokemon
F	Work	WhatsApp
M	Work	WhatsApp
F	Work	WhatsApp
M	Study	Pokemon
M	Study	Pokemon

# Construction d'un arbre de décision



# Arbres de décision : cas des variables continues



Pour les variables continues, il faut évaluer tous les points de division potentiels afin de déterminer celui qui maximise le gain d'information.

# Les hyperparamètres

Afin de créer des arbres de décision qui se généraliseront bien à de nouveaux problèmes, nous pouvons ajuster plusieurs aspects différents de ces arbres.

## Profondeur maximale

La profondeur maximale d'un arbre de décision est simplement la plus grande longueur entre la racine et une feuille. Un arbre de longueur maximale  $k$  peut avoir au maximum  $2^k$  feuilles.



Depth = 1



Depth = 2



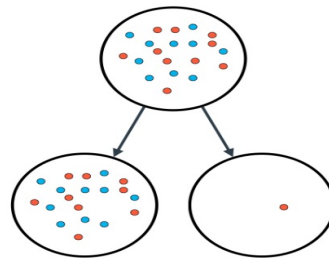
Depth = 3



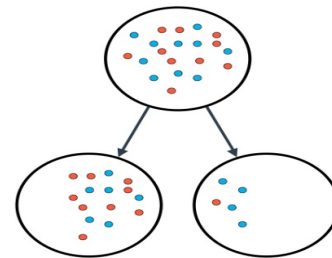
Depth = 4

## Nombre minimum d'échantillons par feuille

Pour éviter des divisions déséquilibrées (par exemple, 99 échantillons d'un côté et 1 de l'autre), on peut fixer un nombre minimum d'échantillons par feuille. Cela optimise le processus et évite le surapprentissage.



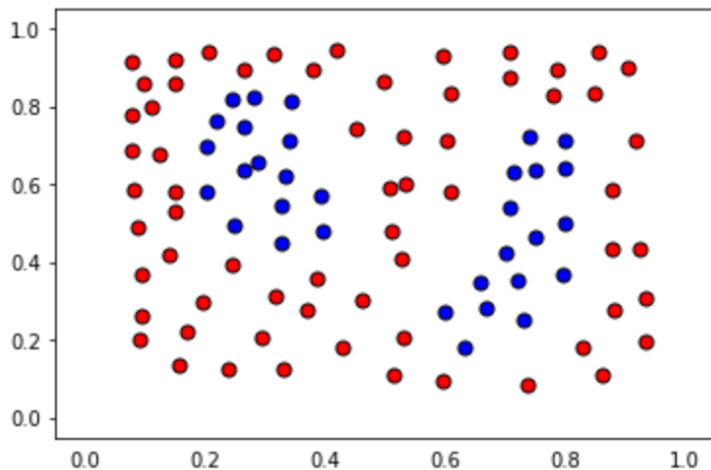
Minimum samples per leaf = 1



Minimum samples per leaf = 5

# Exercice

Dans cet exercice, on vous fournira l'ensemble de données échantillon suivant, et votre objectif sera de définir un modèle qui atteigne une précision de 100% sur celui-ci.



<https://github.com/elhidali/EPISEN-2024>