

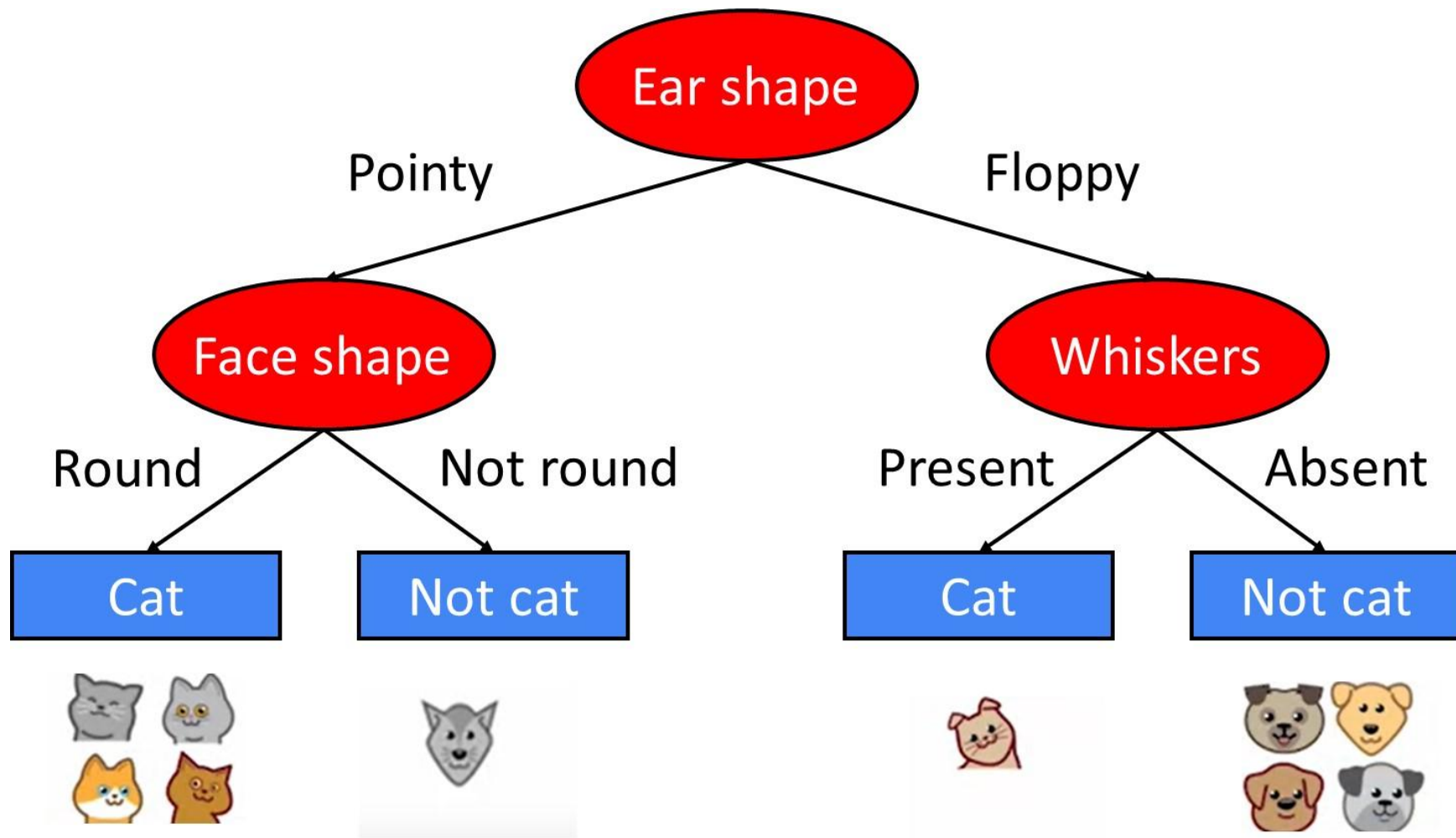
Intelligence Artificielle

Chapitre 2 – Classification avec apprentissage supervisé

Les arbres de décision

Bezzaz Soumia

2025/2026 – Semestre 2



Qu'est-ce qu'un arbre de décision ?

Imaginez un jeu de devinettes : « Je pense à un animal... devine lequel ! »

Le jeu des 20 questions

Question 1 : Est-ce que ça vit dans l'eau ?

Oui → Poisson, Baleine... / Non → Continue

Question 2 : Est-ce que ça a 4 pattes ?

Oui → Chien, Chat... / Non → Oiseau ?

Question 3 : Est-ce que ça vole ?

Oui → Aigle, Pigeon... / Non → Serpent ?

Chaque question divise les possibilités en groupes plus petits et plus homogènes.

Un arbre de décision, c'est pareil !

Définition

Un algorithme de classification qui pose une série de questions sur les données pour les ranger dans des catégories.

Principe

Chaque nœud de l'arbre = une question. Chaque branche = une réponse possible. Chaque feuille = une décision finale (classe).

Objectif

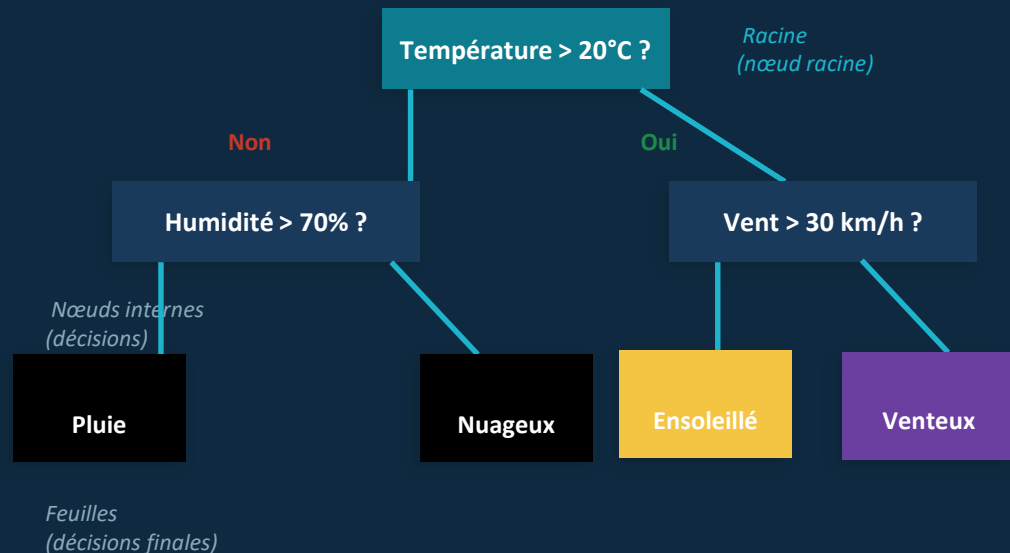
Trouver les questions qui séparent le mieux les données. Plus les groupes sont purs (homogènes), mieux c'est.

Pourquoi « arbre » ?

Visuellement, ça ressemble à un arbre renversé : racine en haut, branches au milieu, feuilles (décisions) en bas.

Anatomie d'un arbre de décision

Exemple : Classifier le temps qu'il fait



Les 3 composants d'un arbre

Nœud racine

La première question posée. C'est la question la plus importante, celle qui divise le mieux les données.

Nœuds internes

Les questions intermédiaires. Chaque nœud teste une condition sur une variable (ex: température, humidité...).

Feuilles (nœuds terminaux)

Les décisions finales. Chaque feuille contient une classe : Pluie, Soleil, Neige, etc.

Branches

Les chemins entre les nœuds. Représentent les réponses possibles (Oui/Non, ou plusieurs catégories).

Profondeur de l'arbre

Le nombre de niveaux de questions. Un arbre trop profond peut mémoriser les données (surapprentissage).

Construire un arbre pas à pas

Objectif : classer des terrains en « Constructible » ou « Non constructible »

Terrain	Altitude	Pente	Sol	Classe
T1	Basse	Faible	Stable	✔ Constructible
T2	Basse	Forte	Instable	✗ Non
T3	Haute	Faible	Stable	✔ Constructible
T4	Haute	Forte	Stable	✗ Non
T5	Basse	Faible	Instable	✗ Non
T6	Basse	Faible	Stable	✔ Constructible

Étapes de construction

1- Choisir la première question

Quelle variable sépare le mieux ? → Stabilité du Sol !
(divise parfaitement Stable vs Instable)

2- Diviser les données

Groupe 1 (Sol Stable) → T1, T3, T4, T6
Groupe 2 (Sol Instable) → T2, T5 (tous non constructibles)

3- Continuer pour les groupes impurs

Groupe 1 encore mélangé → nouvelle question : Pente > Forte ? Si Oui → Non constructible. Si Non → Constructible.

4- S'arrêter quand c'est pur

Quand tous les exemples d'un groupe ont la même classe, on crée une feuille. Fin de la branche.

Comment choisir la meilleure question ?

On cherche la question qui crée les groupes les plus purs (homogènes) possible.

Vous avez un sac mélangé de bonbons rouges et bleus. Vous voulez les séparer en deux pots.

Mauvaise question : « Est-ce que c'est un bonbon ? » → Oui pour tous ! Ça ne sépare rien.

Bonne question : « Est-ce que c'est rouge ? » → Oui = tous rouges dans un pot, Non = tous bleus dans l'autre. Parfaitement séparé !

Un arbre de décision :

- Pose des questions simples
- Avance étape par étape
- Arrive à une décision finale
- Est utilisé en intelligence artificielle
- Fait partie de l'apprentissage supervisé

Avantages et limites des arbres de décision

Ce que l'arbre fait très bien

Facile à comprendre

C'est l'algorithme le plus intuitif ! On peut suivre chaque décision et l'expliquer clairement.

Visualisable

On peut dessiner l'arbre et voir exactement comment il prend ses décisions. Parfait pour présenter aux non-experts.

Rapide

Classifier un nouveau cas = suivre un chemin dans l'arbre. C'est très rapide même avec beaucoup de données.

Données mixtes

Fonctionne avec des nombres (température) ET des catégories (couleur, type de sol) sans transformation.

Ses points faibles

Surapprentissage facile

L'arbre peut devenir trop profond et mémoriser les données d'entraînement au lieu d'apprendre des règles générales.

Instable

Changer un seul exemple peut complètement modifier l'arbre. L'algorithme est très sensible aux données.

Biais sur classes déséquilibrées

Si une classe est beaucoup plus fréquente, l'arbre va favoriser cette classe et ignorer les autres.

Les arbres de décision

Les arbres sont particulièrement utiles quand on doit expliquer nos décisions aux décideurs et au public.

Aptitude à la construction

Classier des parcelles en Constructible / Non constructible selon pente, sol, risque sismique, accès eau.

Avantage : Les urbanistes peuvent suivre le raisonnement et le justifier aux citoyens.

Cartographie des risques d'inondation

Prédire le risque (Faible / Moyen / Fort) selon altitude, proximité rivière, historique, type de sol.

Exemple : « Si altitude < 50m ET près rivière → Risque Fort »

Classification de l'occupation des sols

À partir d'images satellites, classer chaque pixel : Forêt, Culture, Eau, Urbain.

Données : bandes spectrales Sentinel-2 (rouge, vert, proche infrarouge...)

Analyse de la stabilité des pentes

Prédire les glissements de terrain selon inclinaison, type de roche, végétation, précipitations.

Utilisé pour les routes de montagne et les zones à risque.

Exercice 1 – Construire un arbre simple

Contexte :

Un géographe étudie des zones pour identifier celles à risque d'incendie de forêt. Il dispose de données sur 8 zones déjà connues. Votre mission : construire un arbre de décision pour prédire le risque des nouvelles zones.

Zone	Température	Végétation	Humidité	Risque incendie
Z1	Haute	Dense	Faible	Élevé
Z2	Basse	Dense	Élevée	Faible
Z3	Haute	Clairsemée	Faible	Élevé
Z4	Basse	Clairsemée	Élevée	Faible
Z5	Haute	Dense	Élevée	Modéré
Z6	Haute	Clairsemée	Élevée	Faible
Z7	Basse	Dense	Faible	Modéré
Z8	Haute	Dense	Faible	Élevé

Questions

Q1. Quelle est la première question à poser ? (Quelle variable sépare le mieux les données ?)

Q2. Dessinez l'arbre complet avec tous les nœuds et toutes les feuilles.

Q3. Quelle serait la prédiction pour une nouvelle zone : Température Haute, Végétation Dense, Humidité Faible ?

Astuce : commencez par la variable qui sépare le mieux les 3 classes (Élevé / Modéré / Faible).

Solution – Exercice 1

Correction détaillée avec arbre complet

Q1 – Première question

Réponse : Température > Haute ?

Pourquoi ? Toutes les zones à risque Élevé ont une température Haute (Z1, Z3, Z8). Cette variable sépare déjà bien les cas les plus dangereux. C'est donc la meilleure première question.

Q3 – Prédiction pour une nouvelle zone

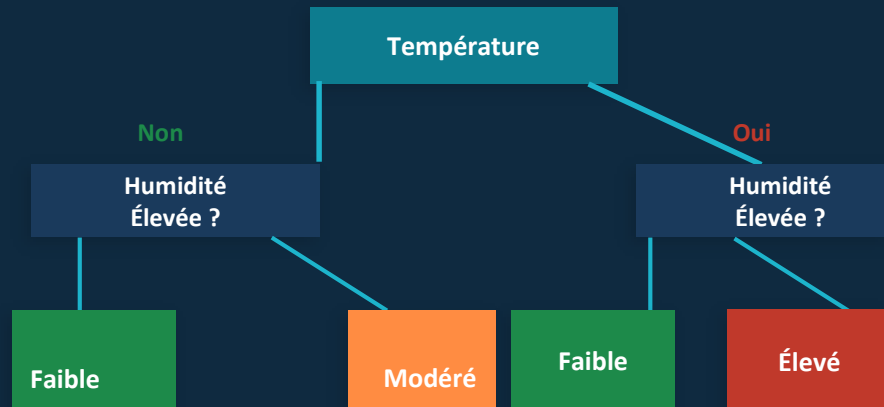
Données : Température Haute, Végétation Dense, Humidité Faible

Chemin dans l'arbre :

1. Température Haute ? → OUI → branche droite
2. Humidité Élevée ? → NON → branche droite
3. Arrivée à la feuille : Risque Élevé

Prédiction finale : Risque Élevé d'incendie

Q2 – Arbre de décision



Exercice 2

Testez votre compréhension des concepts — répondez aux questions suivantes

Partie A – Vrai ou Faux ?

1. Un arbre de décision peut seulement classer en 2 catégories.
2. L'arbre de décision est l'algorithme le plus facile à interpréter et à expliquer.
3. Un arbre trop profond risque de mémoriser les données (surapprentissage).
4. Les arbres de décision sont des algorithmes d'apprentissage non supervisé.
5. Un petit changement dans les données peut complètement modifier l'arbre.

V

F

V

F

V

F

V

F

V

F

Partie B – Questions ouvertes

- Q1. Donnez un exemple concret où un arbre de décision serait préférable à un SVM. Pourquoi ?
- Q2. Pourquoi dit-on qu'un arbre trop profond « mémorise » au lieu « d'apprendre » ?

Solution – Exercice 2

Partie A – Réponses Vrai / Faux

FAUX

1. Un arbre peut classer en autant de catégories que nécessaire : 2, 3, 10, 100... Il n'y a pas de limite.

VRAI

2. C'est le point fort principal des arbres ! On peut suivre chaque décision et l'expliquer à n'importe qui.

VRAI

3. Un arbre avec trop de niveaux s'adapte parfaitement aux données d'entraînement mais échoue sur les nouvelles données.

FAUX

4. Les arbres de décision sont généralement des algorithmes d'apprentissage supervisé, pas non supervisé..

VRAI

5. C'est une des faiblesses des arbres : ils sont instables et sensibles aux variations des données.

Partie B – Éléments de réponse attendus

Q1 : Exemple : Classification des zones à risque d'inondation pour un rapport public. Avantage de l'arbre : les décideurs peuvent suivre le raisonnement (« Si altitude < 50m ET près rivière → Risque élevé »).

Q2 : Un arbre trop profond crée des règles ultra-spécifiques qui collent parfaitement aux données d'entraînement, mais ne généralisent pas. C'est comme apprendre par cœur sans comprendre : ça marche sur les exemples connus, mais échoue sur les nouveaux cas.

1

Chaque nœud pose une question qui divise les données. Les feuilles donnent la réponse finale.

2

On choisit la meilleure question (celle qui sépare le mieux), puis on recommence sur chaque groupe.

3

L'algorithme le plus facile à expliquer. On peut dessiner l'arbre et suivre chaque décision.

4

Un arbre trop profond mémorise au lieu d'apprendre. Il faut limiter la profondeur.

5

Classification des sols, risques naturels, aptitude à la construction — tout est explicable !

Attention au surapprentissage

Parfait pour la géographie