

Représentations des Arbres de Décision

Nicolas PASQUIER
Université Côte d'Azur
Département Informatique
Laboratoire I3S (UMR-7271 UCA/CNRS)
<http://www.i3s.unice.fr/~pasquier>



Rappel : Exemple de Prédiction d'Appétence

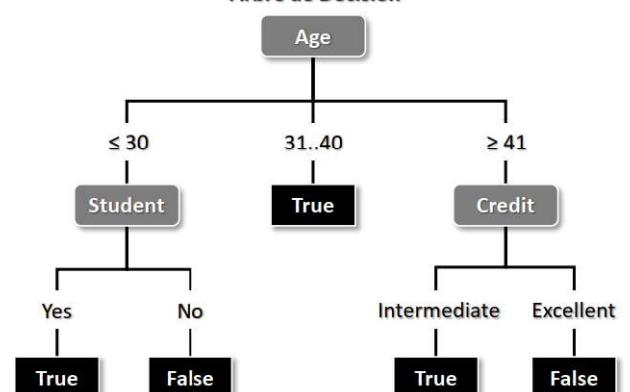
Ensemble d'Apprentissage

ID	Age	Income	Student	Credit	Buyer
1	28	High	No	Intermediate	False
2	24	High	No	Excellent	False
3	39	High	No	Intermediate	True
4	47	Medium	No	Intermediate	True
5	41	Low	Yes	Intermediate	True
6	52	Low	Yes	Excellent	False
7	35	Low	Yes	Excellent	True
8	19	Medium	No	Intermediate	False
9	22	Low	Yes	Intermediate	True
10	54	Medium	Yes	Intermediate	True
11	23	Medium	Yes	Excellent	True
12	34	Medium	No	Excellent	True
13	37	High	Yes	Intermediate	True
14	44	High	No	Intermediate	True

Apprentissage de l'arbre de décision

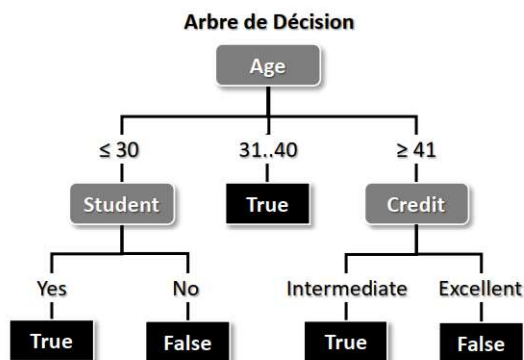
- Variable cible : Buyer
- Variables prédictives : Age, Income, Student, Credit

Arbre de Décision



Interprétation des Arbres de Décision

- Un arbre de décision fournit de par sa structure des informations sur l'espace des données défini par l'ensemble d'apprentissage



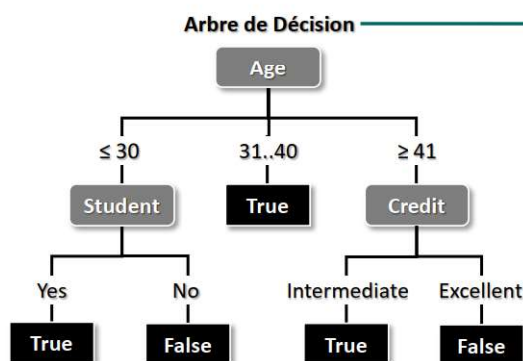
➤ Considérées indépendamment (assumption d'indépendance) :

- La variable Age est la plus utile pour prédire la classe
- Les variables Student et Credit sont également utiles, mais moins que Age
- La variable Income est la moins utile (inutilisée dans cet arbre)

- Cet arbre de décision distingue cinq cas (nœuds feuilles) de combinaisons de valeurs des variables Age, Student et Credit : trois cas de prédiction de la classe Buyer = True et deux cas de prédiction de la classe Buyer = False

Traduction en Règles de Classification

- Branche de l'arbre de décision : chemin du nœud racine jusqu'à un nœud feuille
- Chaque branche peut être traduite en une règle de classification de la forme : Si conditions alors classe prédite
 - Les nœuds internes (variables) et leur arc sortant (valeur) dans la branche constituent la conjonction de *conditions*
 - Le nœud feuille de la branche constitue la *classe prédite*



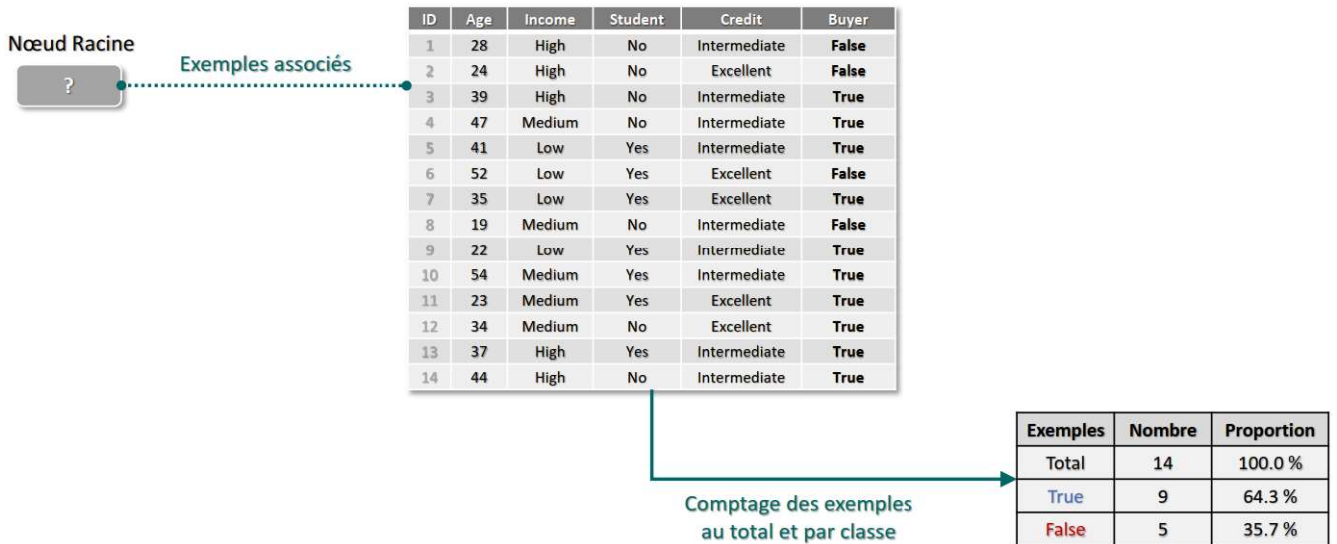
→ Création d'une règle pour chaque branche

Règles de Classification

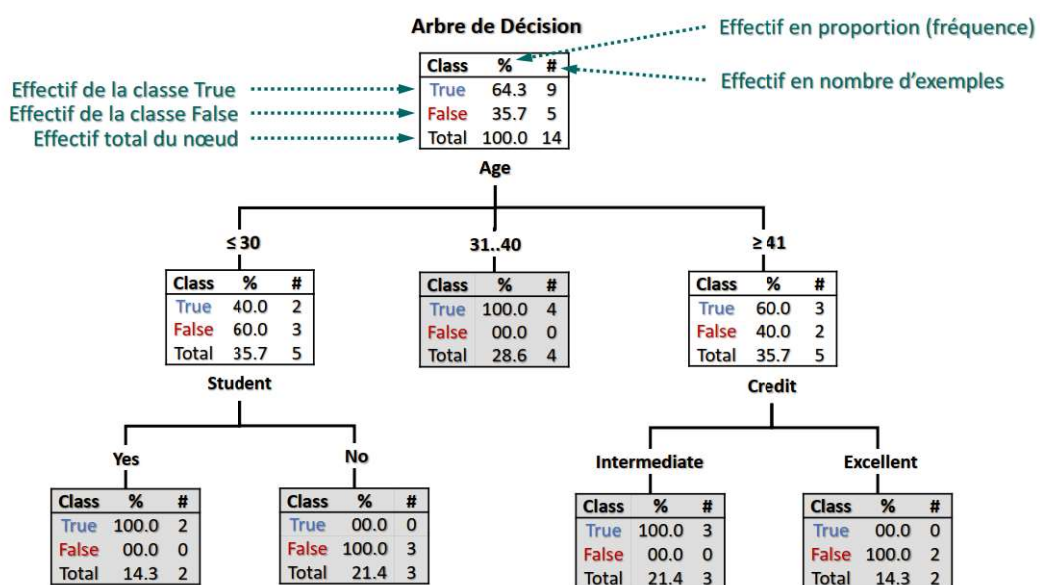
- Si Age ≤ 30 et Student = Yes alors Buyer = True
- Si Age ≤ 30 et Student = No alors Buyer = False
- Si Age $\in [31..40]$ alors Buyer = True
- Si Age ≥ 41 et Credit = Intermediate alors Buyer = True
- Si Age ≥ 41 et Credit = Excellent alors Buyer = False

Représentation des Effectifs des Nœuds

- Pour chaque nœud peuvent être représentés les effectifs (nombre d'exemples de l'ensemble d'apprentissage) correspondant au nœud

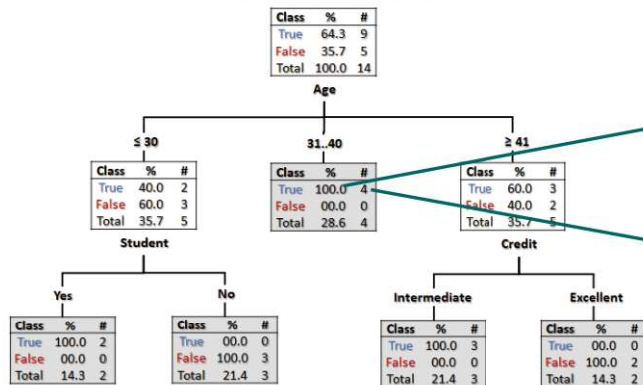


Représentation des Effectifs des Nœuds



Interprétation des Effectifs des Nœuds

Arbre de Décision



Règles de Classification

Antécédant	Classe	Précision	Exemples
Si Age ≤ 30 et Student = Yes	True	100 %	2
Si Age ≤ 30 et Student = No	False	100 %	3
Si Age ∈ [31..40]	True	100 %	4
Si Age ≥ 41 et Credit = Intermediate	True	100 %	3
Si Age ≥ 41 et Credit = Excellent	False	100 %	2

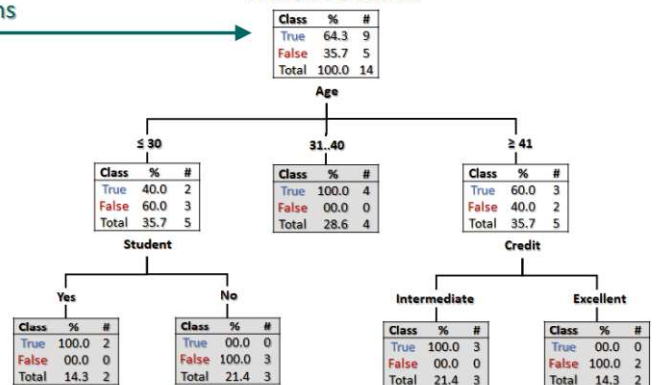
Classe Prédite et Précision des Prédictions

Ensemble de Test

ID	Age	Income	Student	Credit	Buyer	Prediction
15	19	Low	No	Intermediate	False	?
16	32	Medium	No	Excellent	True	?
17	47	Medium	No	Excellent	True	?
18	23	Low	Yes	Intermediate	True	?
19	40	Medium	No	Excellent	False	?
20	25	High	Yes	Excellent	True	?

Prédictions

Arbre de Décision



Ensemble de Test

ID	Age	Income	Student	Credit	Buyer	Prediction	P(True)	P(False)
15	19	Low	No	Intermediate	False	False	0.00	1.00
16	32	Medium	No	Excellent	True	True	1.00	0.00
17	47	Medium	No	Excellent	True	False	0.00	1.00
18	23	Low	Yes	Intermediate	True	True	1.00	0.00
19	40	Medium	No	Excellent	False	True	1.00	0.00
20	25	High	Yes	Excellent	True	True	1.00	0.00

Résultat pour chaque exemple :

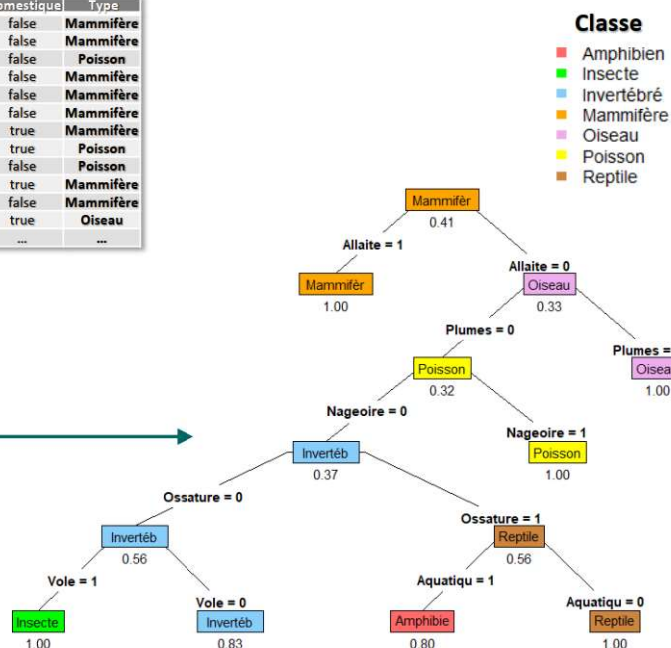
- Classe prédite
- Probabilité de la classe True
- Probabilité de la classe False

Exemple de Prédiction du Type d'Animaux

[illegible]

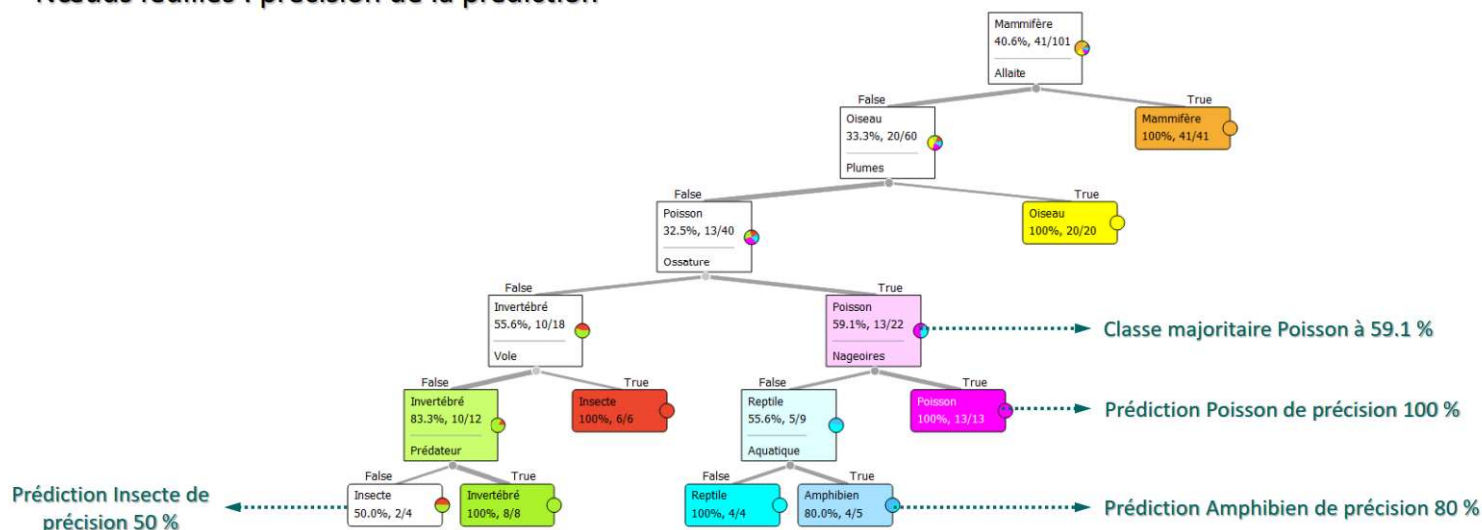
- Variable cible : Type (7 classes)

- Variable cible : Type (7 classes)
- Variables prédictives : Poils, Plumes, Œufs, Allaité, Vole, Aquatique, Prédateur, Dents, Ossature, Respire, Venimeux, Nageoires, Pattes, Queue, Domestique



Représentation de la Précision des Prédictions

- Les couleurs représentent la classe majoritaire dans le nœud et leur intensité la proportion de la classe
- Nœuds feuilles : précision de la prédiction



Représentation Textuelle Hiérarchique

- Permet la lecture et l'exploration des arbres de décision de très grande taille
- Indentation (décalage) et numéros représentent les liens de parenté

Règles de Classification

	#	Nœud	Exemples	Feuille	Classe	Proportion	Effectif
Nœuds enfants	1	Root	101		Mammifère	40.6 %	41
	2> Allaite=T	41	*	Mammifère	100.0 %	41
	3> Allaite=F	60		Oiseau	33.2 %	20
Nœuds frères	4	Plumes = T	20	*	Oiseau	100.0 %	20
	5	Plumes = F	40		Poisson	32.5 %	13
	6> Ossature = T	22		Poisson	59.1 %	13
	8	Nageoires = T	13	*	Poisson	100.0 %	13
	9	Nageoires = F	9		Reptile	55.6 %	5
	10	Aquatique = T	5	*	Amibien	80.0 %	4
	11	Aquatique = F	4	*	Reptile	100.0 %	4
	7> Ossature = F	18		Invertébré	55.6 %	10
	12	Vole = T	6	*	Insecte	100.0 %	6
	13	Vole = F	12		Invertébré	83.3 %	10
	14	Prédateur = T	8	*	Invertébré	100.0 %	8
	15	Prédateur = F	4	*	Insecte	50.0 %	2

→ Branche : Si Allaite = F et Plumes = T Alors Type = Oiseau [Précision = 100 %, Support = 20/101 (# exemples)]

Références et Bibliographie

- Principales Librairies R
 - [rpart.plot](#) : représentations graphiques avancées des arbres de rpart
 - [tree](#) : fonctions plot.tree() et text.tree() de représentation des arbres tree
 - [partykit](#) : représentations graphiques avancées des arbres party (fonction ctree()), C50, rpart et RWeka
 - [C50](#) : fonction plot.C5.0() de représentation des arbres de décision C50 et fonction summary.C5.0() de représentation des règles de classification C50
- Bibliographie
 - [Plotting rpart trees with the rpart.plot package](#). Stephen Milborrow
 - [Constant Partying: Growing and Handling Trees with Constant Fits](#). Torsten Hothorn & Achim Zeileis