

Série de TD N°5

Arbres de Décision

Exercice 1 : La table suivante est un ensemble d'apprentissage construit à partir des données relatives aux lecteurs MP3 et la satisfaction des clients ayant utilisés ces lecteurs.

Sur la base de cet ensemble, construire l'arbre de décision en utilisant l'algorithme de base. L'attribut «Customer Satisfaction » est la colonne de classification. Utiliser le Gain en information pour la sélection des attributs.

Rappel : $\log_2(a) = \log(a)/\log(2)$

Memory	Battery Life	Price	Customer Satisfaction
<=4	long	<=150	yes
>4	long	>150	yes
>4	long	<=150	yes
<=4	long	>150	yes
>4	long	>150	yes
>4	low	>150	yes
<=4	low	>150	no
<=4	low	>150	no
>4	low	<=150	yes
<=4	low	<=150	no
<=4	medium	<=150	no
>4	medium	<=150	no
<=4	medium	>150	yes
>4	medium	>150	yes
>4	medium	<=150	no

Exercice 2 : Une banque dispose des informations suivantes sur un ensemble de clients:

Client	Montant	Age	Résidence	Etudes	Internet
1	moyen	moyen	village	oui	oui
2	élevé	moyen	bourg	non	non
3	faible	âgé	bourg	non	non
4	faible	moyen	bourg	oui	oui
5	moyen	jeune	ville	oui	oui
6	élevé	âgé	ville	oui	non
7	moyen	âgé	ville	oui	non
8	faible	moyen	village	non	non

L'attribut *Montant* décrit la moyenne des montants sur le compte client. Le second attribut *Age* donne la tranche d'âge du client. Le troisième attribut *Résidence* décrit la localité de résidence du client. Le dernier attribut *Etudes* a la valeur oui si le client a un niveau d'études supérieures. La classe associée à chacun de ces clients correspond au contenu de la colonne *Internet*. La valeur *oui* correspond à un client qui effectue une consultation de ses comptes bancaires en utilisant Internet.

- 1- On souhaite trouver l'arbre de décision qui soit capable de prédire si un client effectue des consultations de ses comptes par Internet en connaissant les valeurs des attributs montant, âge, résidence et études pour ce client. On se basera sur le critère d'entropie pour déterminer les différents nœuds de l'arbre.
- 2- Dites pourquoi, dans un arbre de décision, les variables doivent être catégoriques (modales ou qualitatives) ? Que faire dans le cas d'une variable numérique continue.
- 3- Etant donné un arbre de décision, vous avez la possibilité de (a) convertir l'arbre de décision en règles ensuite effectuer l'élagage des règles, ou (b) procéder à l'élagage de l'arbre de décision ensuite le convertir en règles. Quel est l'avantage de (a) sur (b)?