

Data Mining : Arbres de décision

TD N : 1

Exercice 1

Supposons qu'on veut utiliser l'algorithme CART pour construire un arbre de décision permettant de décider si un patient est malade ou pas. Le tableau suivant contient des données d'un ensemble de patients. Les attributs sont : "Fièvre", "Douleur" et "Toux". Les patients sont répartis selon s'ils sont malade ou pas, alors trois classes sont identifiées : Appendicite, rhume et mal de gorge.

Travail à faire :

Utiliser les données du tableau pour construire l'arbre de décision. Montrez toutes les étapes de calcul. Dessinez l'arbre final.

	Fièvre	Douleur	Toux	Maladie
1	oui	Abdomen	non	Appendicite
2	non	Abdomen	oui	Appendicite
3	oui	gorge	non	rhume
4	oui	gorge	oui	rhume
5	non	gorge	oui	mal de gorge

Exercice 2

Le tableau suivant contient des données sur les résultats obtenus par des étudiants de Tronc Commun (première année à l'Université). Chaque étudiant est décrit par 3 attributs : Est-il doublant ou non, la série du Baccalauréat obtenu et la mention. Les étudiants sont répartis en deux classes : Admis et Non Admis.

	Doublant	Série	Mention	Classe
1	Non	Maths	ABien	Admis
2	Non	Techniques	ABien	Admis
3	Oui	Science	ABien	Non Admis
4	Oui	Science	Bien	Admis
5	Non	Maths	Bien	Admis
6	Non	Techniques	Bien	Admis
7	Oui	Science	Passable	Non Admis
8	Oui	Maths	Passable	Non Admis
9	Oui	Techniques	Passable	Non Admis
10	Oui	Maths	TBien	Admis
11	Oui	Techniques	TBien	Admis
12	Non	Science	TBien	Admis
13	Oui	Maths	Bien	Admis
14	Non	Science	ABien	Non Admis
15	Non	Maths	TBien	Admis
16	Non	Maths	Passable	Non Admis

On veut construire un arbre de décision à partir des données du tableau, pour rendre compte des éléments qui influent sur les résultats des étudiants en Tronc Commun. Les lignes de 1 à 12 sont utilisées comme données d'apprentissage. Les lignes restantes (de 13 à 16) sont utilisées comme données de tests.

Travail à faire :

1. Utiliser les données des lignes de 1 à 12 pour construire l'arbre en utilisant l'algorithme C4.5. Montrez toutes les étapes de calcul. Dessinez l'arbre final.
2. Quels sont les résultats de test de l'arbre obtenu sur les données des lignes de 13 à 16 ?