

Arbres de décision - Exercice

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.

On veut construire un arbre de décision à partir des données ci dessous pour rendre compte des éléments qui influent sur les résultats des étudiants en Tronc Commun.

Données d'apprentissage

Doublant	Série	Mention	Classe
Non	Maths	ABien	Admis
Non	Techniques	ABien	Admis
Oui	Sciences	ABien	Non Admis
Oui	Sciences	Bien	Admis
Non	Maths	Bien	Admis
Non	Techniques	Bien	Admis
Oui	Sciences	Passable	Non Admis
Oui	Maths	passable	Non Admis
Oui	techniques	Passable	Non Admis
Oui	Maths	TBien	Admis
Oui	Techniques	TBien	Admis
Oui	Sciences	TBien	Admis
Oui	Maths	Bien	Admis
Non	Sciences	ABien	Non Admis
Non	Maths	TBien	Admis
Non	Maths	Passable	Non Admis

Données de test et de validation

Doublant	Série	Mention	Classe
Oui	Maths	Bien	Admis
Non	Sciences	ABien	Non Admis
Non	Maths	TBien	Admis
Non	Maths	Passable	Non Admis

1 - Utiliser les données d'apprentissage pour construire l'arbre de décision en utilisant l'algorithme ID3. Montrez toutes les étapes de calcul. Dessinez l'arbre final.

2 Quels sont les résultats de performance de l'arbre obtenus sur les données de test.

3 – Construire un arbre élagué de l'arbre déduit des questions précédentes en utilisant la méthode CART. Montrez toutes étapes de calcul.