

## Série TD 3

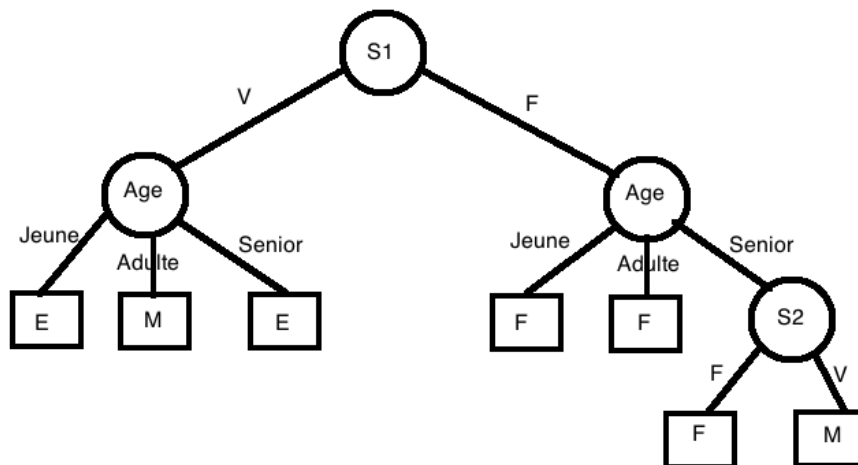
---

### Classification – Classification associative, KNN, et classification bayésienne.

---

#### Exercice 1

Soit l'arbre de décision ci-dessous permettant la prédiction du risque des patients d'avoir une certaine maladie en fonction de leur âge et de deux symptômes booléens (vrai ou faux) appelés S1 et S2. Le risque est évalué selon trois valeurs F (faible), M (moyen) et E (élevé), l'âge est discrétisé selon trois valeurs (jeune, adulte et senior) :



1. Extraire les règles d'association depuis cet arbre.
2. Trouver le risque du patient ayant les attributs (Jeune, V, F) en utilisant la classification bayésienne naïve. (Voir la base de l'exercice 2 – TD 2).

#### Exercice 2

Une banque dispose de la table ci-dessous lui permettant de prédire le risque d'un non-remboursement d'un crédit selon l'âge et le prêt de ses clients :

Age	Prêt	Non-remboursement
25	40,000	N
35	60,000	N
45	80,000	N
20	20,000	N
35	120,000	N
52	18,000	N

23	95,000	Y
40	62,000	Y
60	100,000	Y
48	220,000	Y
33	150,000	Y

1. Trouver le risque du client ayant les attributs (Age=48 et Prêt=142,000) en utilisant l'algorithme des k plus proches voisins.

**NB** : On pose  $K = 3$  et distance = distance euclidienne.

## Correction

### Exercice 1

1 – Les règles :

IF (S1 = V) AND (Age = Jeune) THEN (Risque = E)  
IF (S1 = V) AND (Age = Adulte) THEN (Risque = M)  
IF (S1 = V) AND (Age = Senior) THEN (Risque = E)

IF (S1 = F) AND (Age = Jeune) THEN (Risque = F)  
IF (S1 = F) AND (Age = Adulte) THEN (Risque = F)  
IF (S1 = F) AND (Age = Senior) AND (S2 = F) THEN (Risque = F)  
IF (S1 = F) AND (Age = Senior) AND (S2 = V) THEN (Risque = M)

2 - Le risqué du patient X en utilisant la classification naïve bayésienne :

**X** = (Jeune, V, F)

- $P(\text{Faible} \mid \text{Jeune}, \text{V}, \text{F}) = [P(\text{Jeune} \mid \text{F}) * P(\text{V} \mid \text{F}) * P(\text{F} \mid \text{F}) * P(\text{Faible})] / P(\text{X}) = [2/4 * 0/4 * 3/4 * 4/10] / P(\text{X}) = 0$
- $P(\text{Moyen} \mid \text{Jeune}, \text{V}, \text{F}) = [P(\text{Jeune} \mid \text{M}) * P(\text{V} \mid \text{M}) * P(\text{F} \mid \text{M}) * P(\text{Moyen})] / P(\text{X}) = [0/3 * 2/3 * 1/3 * 3/10] / P(\text{X}) = 0$
- $P(\text{Elevé} \mid \text{Jeune}, \text{V}, \text{F}) = [P(\text{Jeune} \mid \text{E}) * P(\text{V} \mid \text{E}) * P(\text{F} \mid \text{E}) * P(\text{Elevé})] / P(\text{X}) = [1/3 * 3/3 * 1/3 * 3/10] / P(\text{X}) = 0.1/P(\text{X})$

⇒ Risque : **Elevé**

### Exercice 2

1 – Le risque d'un non-remboursement du client : **X** = (Age=48 et Prêt=142,000)

Distance Euclidienne :

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

**K = 3** plus proches voisins.

Age	Prêt	Non-remboursement	Distance
25	40,000	N	$\text{sqrt}((25-48)^2+(40000-142000)^2) = 102\ 000$
35	60,000	N	82 000
45	80,000	N	62 000
20	20,000	N	122 000
35	120,000	N	<b>22 000</b>
52	18,000	N	124 000
23	95,000	Y	47 000
40	62,000	Y	80 000
60	100,000	Y	<b>42 000</b>
48	220,000	Y	78 000
33	150,000	Y	<b>8000</b>

N	<b>22 000</b>	→	<b>Y est majoritaire →</b> <b>Le risque de X est Y</b>
Y	<b>42 000</b>		
Y	<b>8000</b>		