

Classement

Y à $K \geq 2$ modalités !

Base initiale, données étiquetées

Variables prédictives (X_1, \dots, X_p)	Cible (Y)
(0)	

≈ 70%

≈ 30%

Base d'apprentissage (TRAIN SET)

Variables prédictives (X_1, \dots, X_p)	Cible (Y)
(1)	

Base de test (TEST SET)

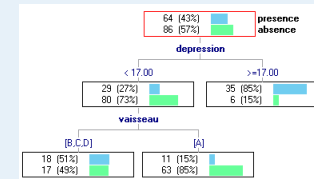
Variables prédictives (X_1, \dots, X_p)	Cible (Y)
(3)	

Subdivision des données (0) en échantillons d'apprentissage (1) et de test (3).

Modélisation (2)

(2)

Modélisation : Base de règles ou Classifieur linéaire



$$\text{logit}(y_k, X) = a_{0,k} + a_{1,k}X_1 + a_{2,k}X_2 + \dots$$

Prédiction

Application des règles prédictives sur l'échantillon test. Construction de la prédiction \hat{Y}

(4)

Cible (Y)	\hat{Y}

Mesure des performances

Confrontation entre valeurs observées de Y et prédictions

(5)

Matrice de confusion

		\hat{Y}	
		$+\hat{}$	$-\hat{}$
Y	+	a	b
	-	c	d

Indicateurs

- Taux d'erreur
- Rappel (sensibilité)
- Précision

Evaluation du modèle prédictif (4 et 5)