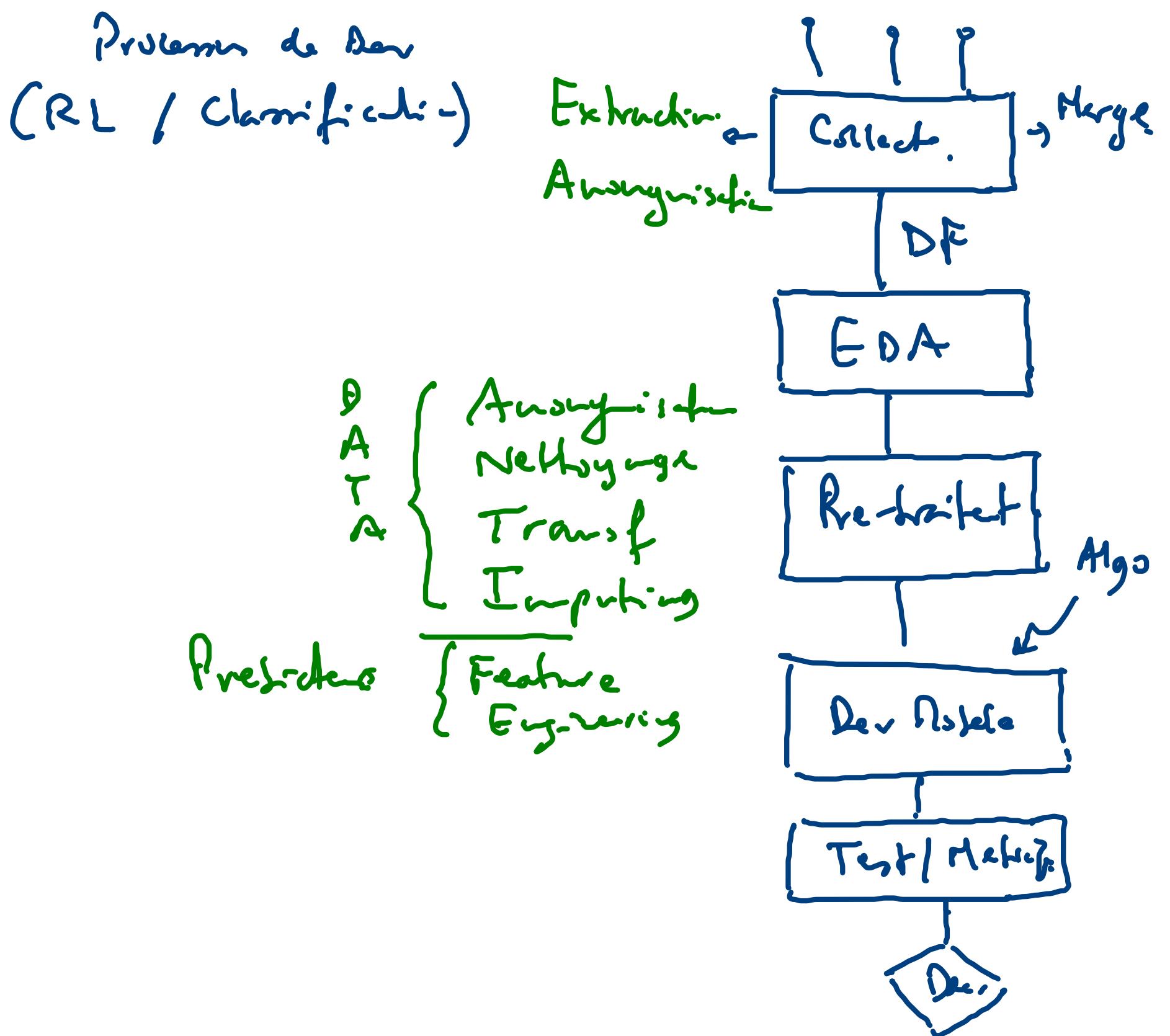


16 Sept 2025

- 1-> Onderwijs et Procesus Der Modelle.
- 2-> Intro classification.
- 3-> Algo kNN - classifier.



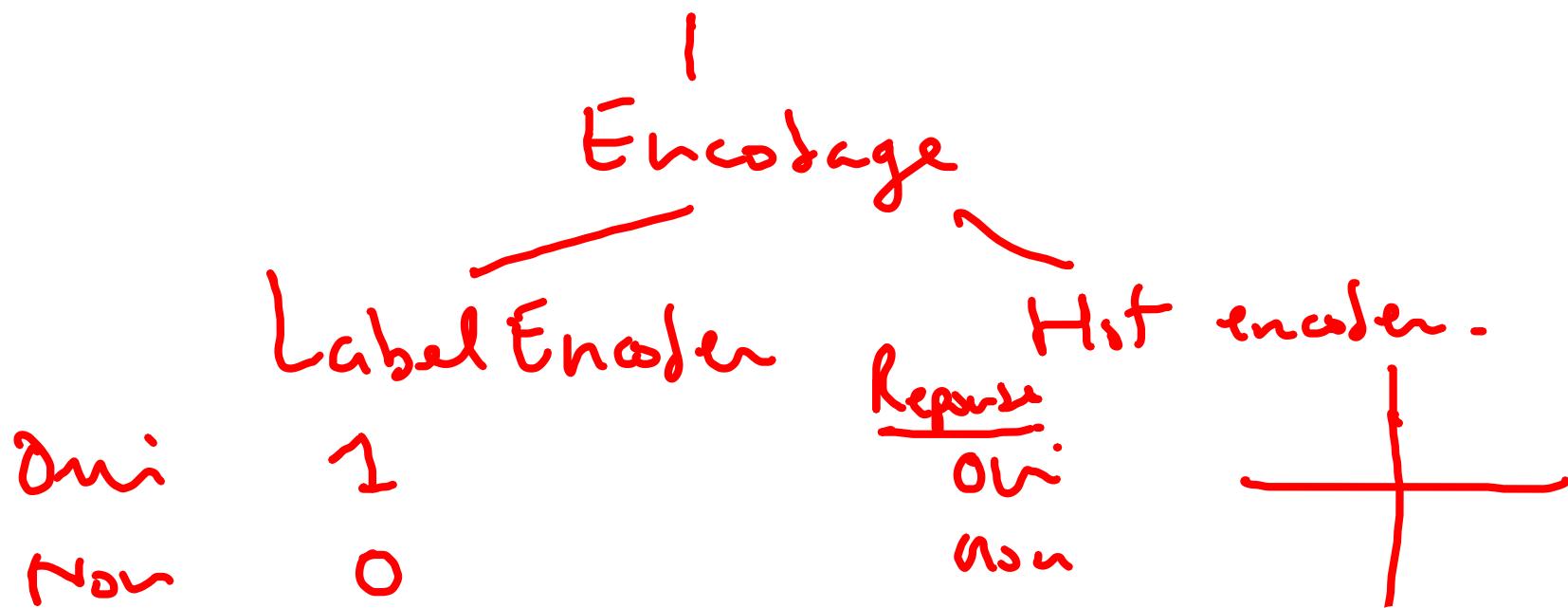
Regression

RMS \bar{E} $\rightarrow 0$

R-squared $\rightarrow 1$

Prob: Données catégorielles -

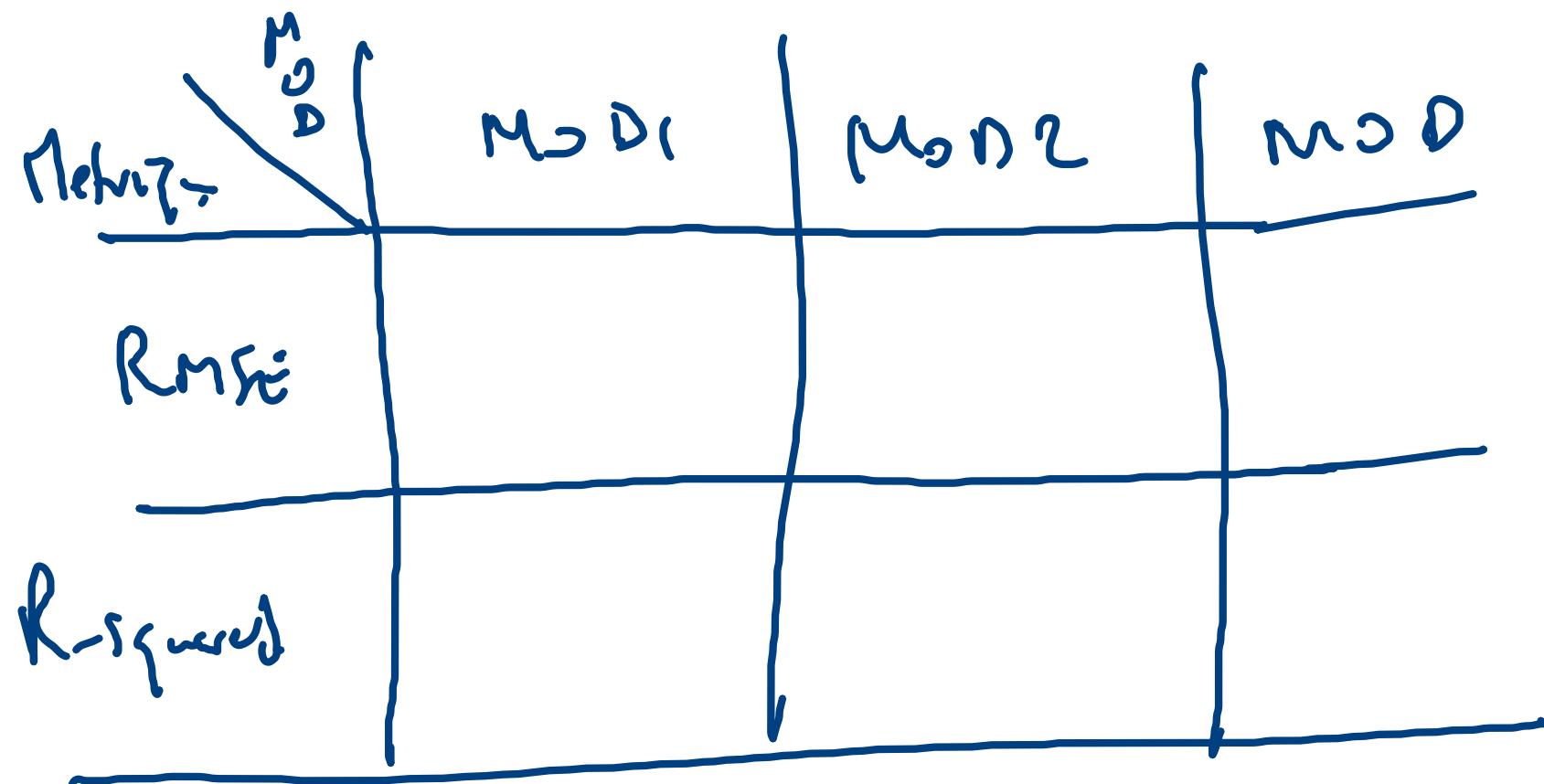
x_i $\xrightarrow{\text{Natur}}$
 $\xrightarrow{\text{Catégoriel.}}$



X	ID	Report	TSL	Report_oui	Rep_oui	Rep_non
	1	oui		1	0	0
	2	non		0	1	1
	3	non		0	0	1
	4	oui		1	1	0
	5	oui		1	0	0

TP 1) Plusieurs modèles avec 3 différents algo.

e) Calcul des Métriques pour chaque des modèles.



Demande TF

Organise le projet.

- 1-) Plusieurs notebooks Projet ✓
- ✓ 2-) Utiliser GitHub pour le travail en Equipe.
- { 3-) Creer une structure de projet standard avec CookieCutter
-
- 4-) Effectuer le Build Automatique
⇒ Générer automatiquement RL
classification

Classification:

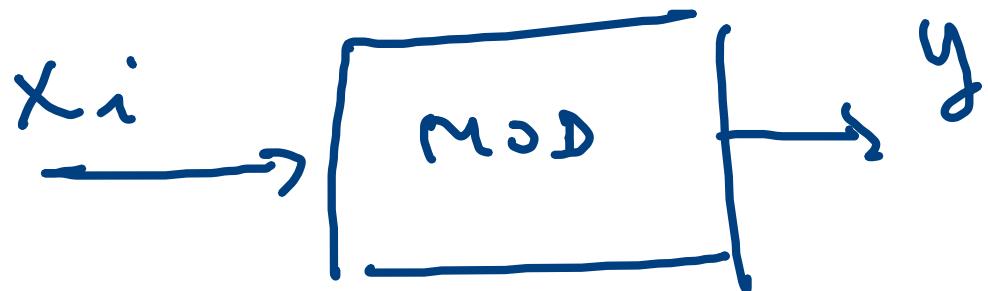
values directly from y

$$y \in \{ \text{uni, no} \} \rightarrow \text{Labels}$$

$$y \in \{ \text{vvt, blen, range, orig} \}$$

$$y \in \{ 1, 2, 5 \}$$

	x_1	x_n	$y \rightarrow \text{cible}$
			oui
			non
			oui
			oui



Test d'hypothèse

binaire.

H_0 : y est Nor.
nulle

H_1 : y n'a pas
alternativement la valeur NON

Développer une statistique basée sur le
Datal : nous dire si on est dans les
conditions de H_0 , H_1 .

Classification

Soient les valeurs des $\{x_i\}$, trouve le label (valeur) de la cible y correspondante.

Cette valeur de y (label) appartient à un ensemble de valeurs pré-définies

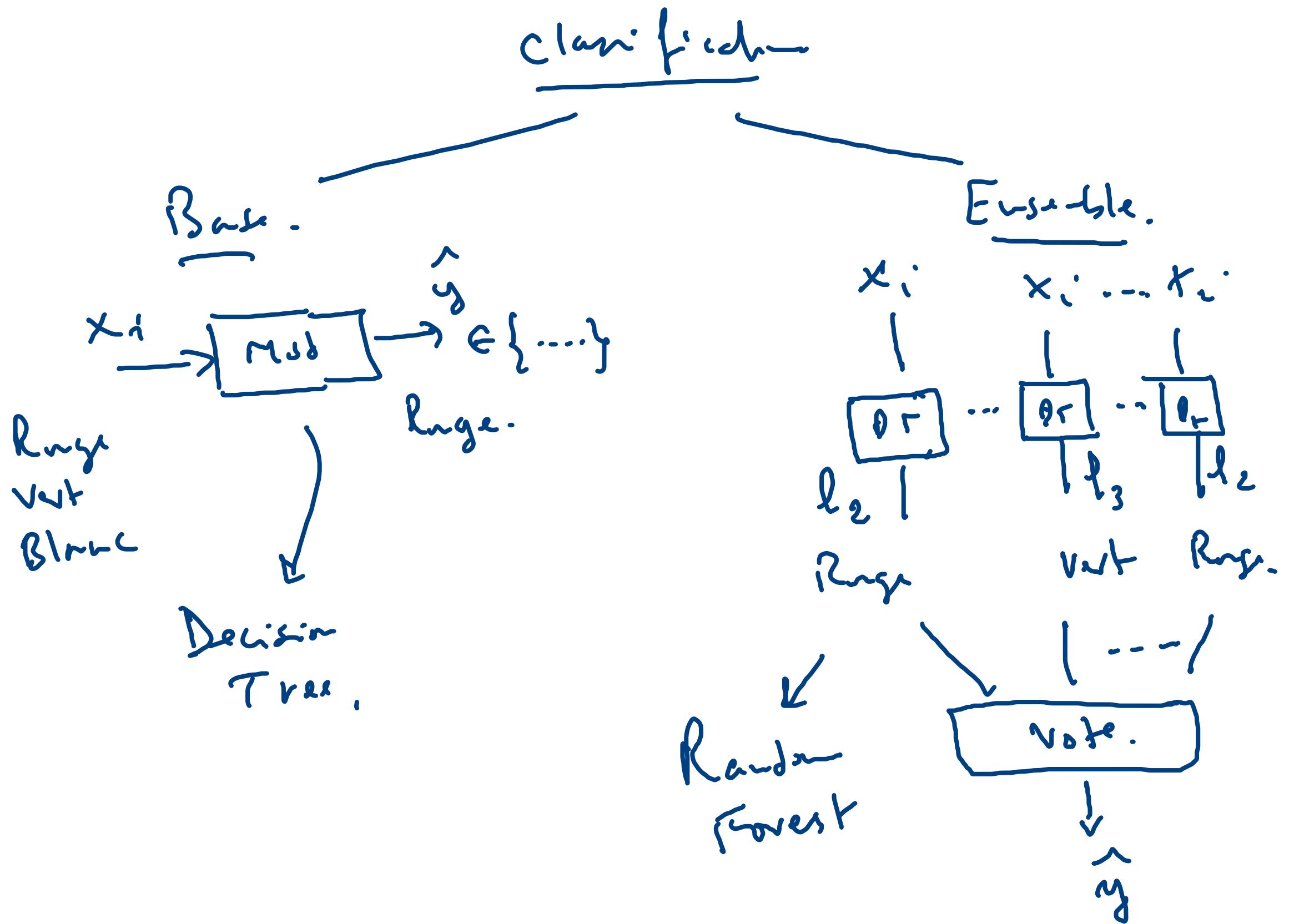
$$\hat{y} = f(x_i)$$

Prej une
valeur spécifiq.

→ Données
→ Modèle de
Classification

Objectif : Développer un modèle de
classification (Data Training,
Data Test)

- 1-> Comprendre les principaux algorithmes de classification
- 2-> Comprendre les métriques à utiliser.
- 3-> Intégrer les clés dans la dev de modèle.



Algs de KNN

Qui se ressemble
s'assimile
K-nearest neighbor

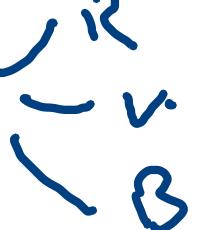
$y \in \{\text{Rouge, vert, Blanc}\}$

$e_c \rightarrow e_1$
 e_2
 e_3
 \vdots

e_c 
 e_1
 e_2
 \vdots
 e_m

x_1	x_2	\dots	x_N	y
:	:			R
				R
				V
				B

$e_c = \{x_1=? , x_2=? , \dots , x_n=?\}$

 marque \hat{y} 

Objectif

- 1-) Définir une [Distance] basée sur les x_i
- 2-) Puis tous les enregistrements, calculer la distance $e_i \rightarrow e_c$
- 3-) La ressemblance est pour le e_i qui a la distance la plus faible (la plus petite).

x_i
 $e_1 \leftrightarrow e_c$ d_1
 $e_2 \leftrightarrow e_c$ d_2
 $e_3 \leftrightarrow e_c$ d_3
 \vdots
 $e_n \leftrightarrow e_c$ d_n

K
 Input & $\underbrace{K=1}$
 $\underbrace{K=2}$
 $\underbrace{K=3}$

Triez les distances par
ordre croissant

$d_{15} \rightarrow e_{25} \rightarrow V$

$d_{45} \rightarrow e_4 \rightarrow R$

$d_{11} \rightarrow e_{11} \rightarrow V$

\vdots
 \vdots

$y_c \rightarrow V$
 $y_c ? V | R$
 $y_c \rightarrow V$

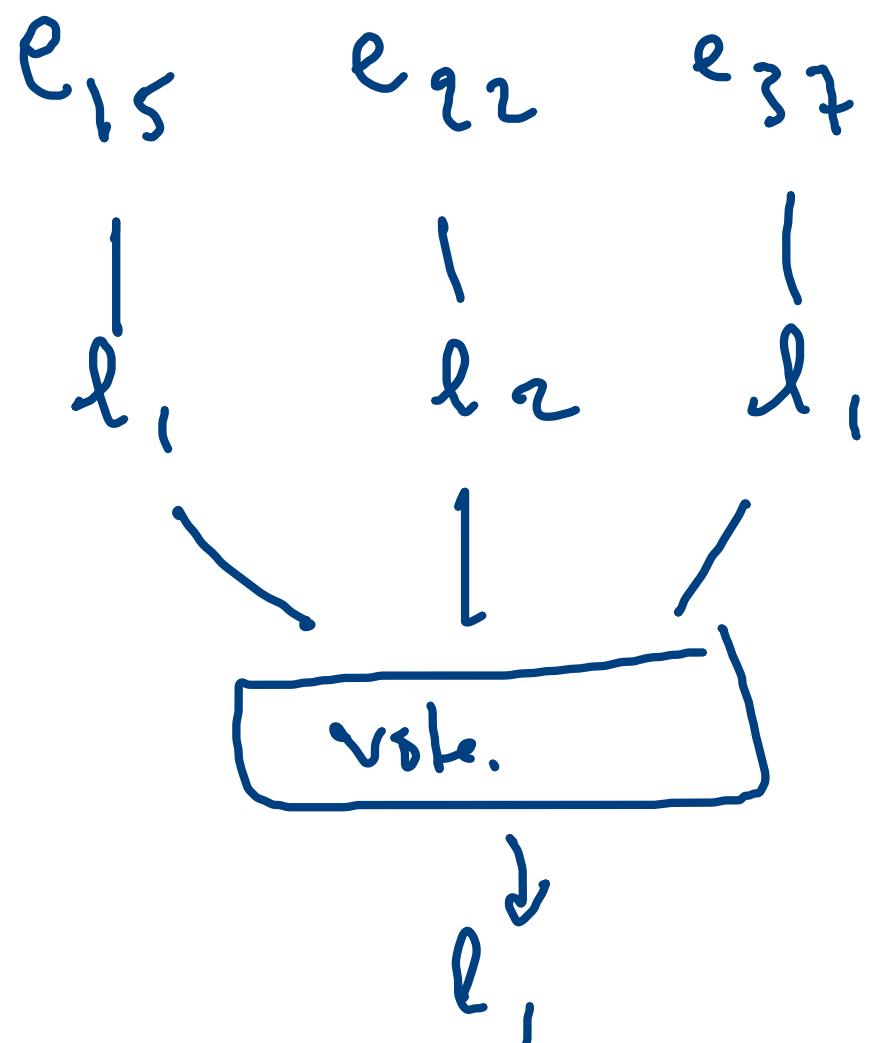
Vote Sur
 $y / K.$
 \Rightarrow label
 majoritaire

$K \approx N$

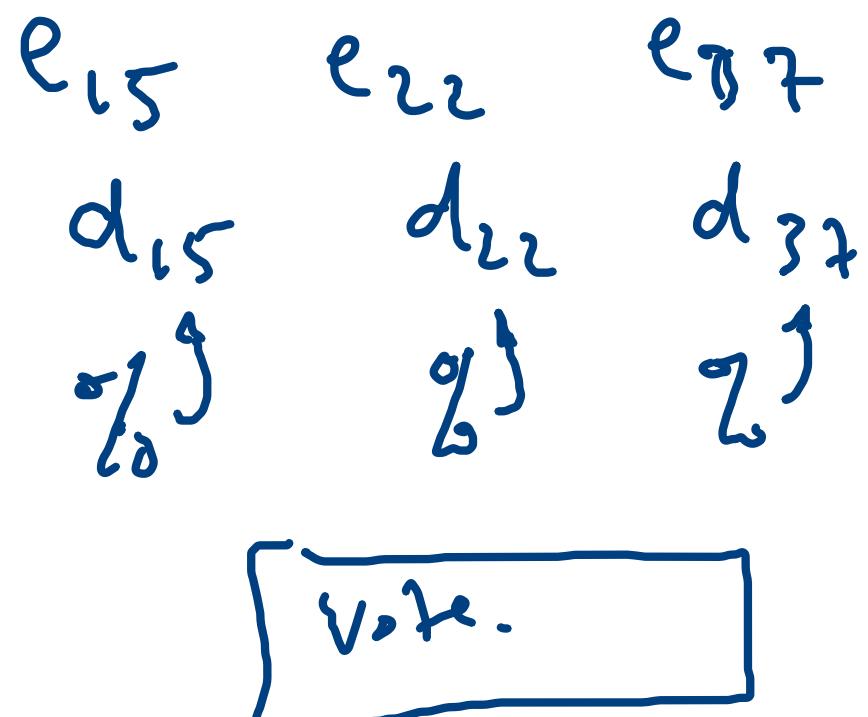
Vote pour obtenir le label correspondant.

majoritaire

sur K points.



Pouderé: le vote sera pondéré par la distance oblique.



kNN n'a pas de Training.

Choisir la valeur de K .

valeur K
postuelle

