

16 Sept 2025

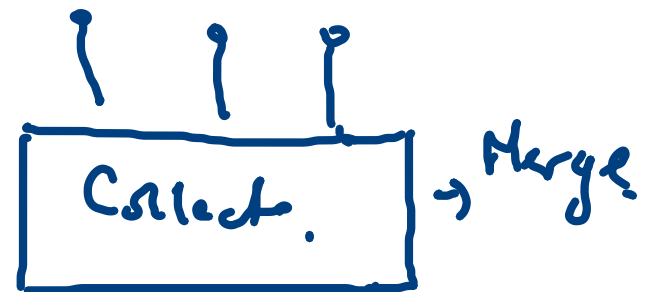
1-> Outline et Processus Des Modele.

2-> Intro classification.

3-> Algo KNN - classification.

Problem de Dev  
(RL / Classification)

Extraction  
Analytische



DF



DATA

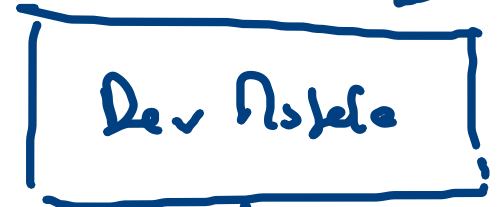
{  
Analytische  
Netzkunde  
Transf  
Imputing

Prozess

{  
Feature  
Engineering



Algo



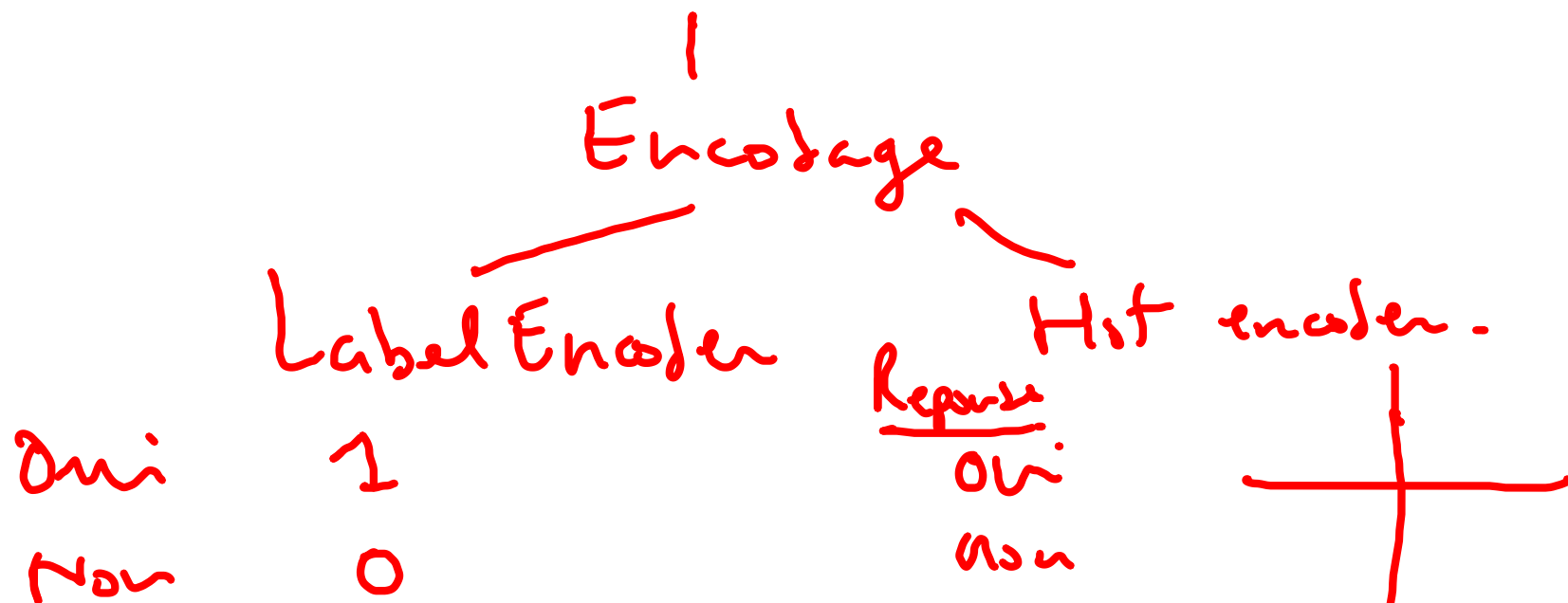
Regression

RMSE  $\rightarrow 0$

R-squared  $\rightarrow 1$

Prob: Données catégorielles -

$X_i$  — Numerique  
      — Catégoriel.



ID	<sup>X</sup> Reponse	test	Reponse_Oui	Oui non
	Oui		1	O
	non		0	1
	non		0	1
	Oui		1	O
	Oui		1	O

TP 1) Plusieurs Models avec 3 différents Algo.

2) Calcul des Métriques par chacun des Models.

Métrique \ Model	Model 1	Model 2	Model
Rmse			
R-squared			

# Demande TP

Organiser le projet.

1-) Plusieurs notebooks Projet ✓

✓ 2-) Utiliser GitHub pour le travail en Équipe.

3-) Créer une structure de projet standard avec CookieCutter

4-) Effectuer le Build Automatique

⇒ Générer automatiquement RL classification

Classification:

values directly from  $y$

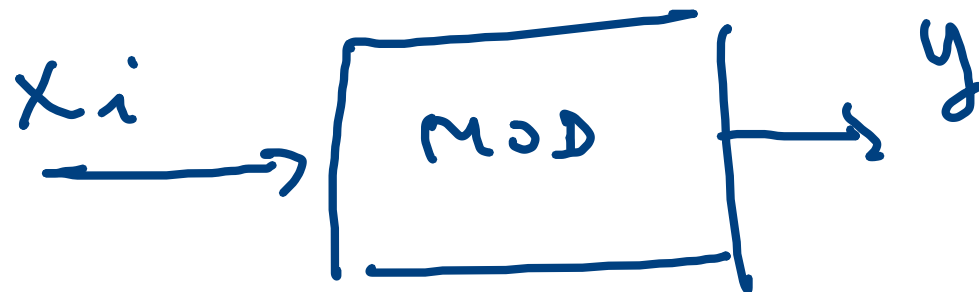
$y \in \{ \text{oui}, \text{non} \} \rightarrow \text{Labels}$

$y \in \{ \text{vrt}, \text{bleu}, \text{range}, \text{orange} \}$

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

$x_1$   $x_n$   $y \rightarrow \text{cible.}$

				oui
				non
				oui
				ou





# Test d'hypothèse Binaire.

$H_0$ :  $y$  est Non.  
nulle

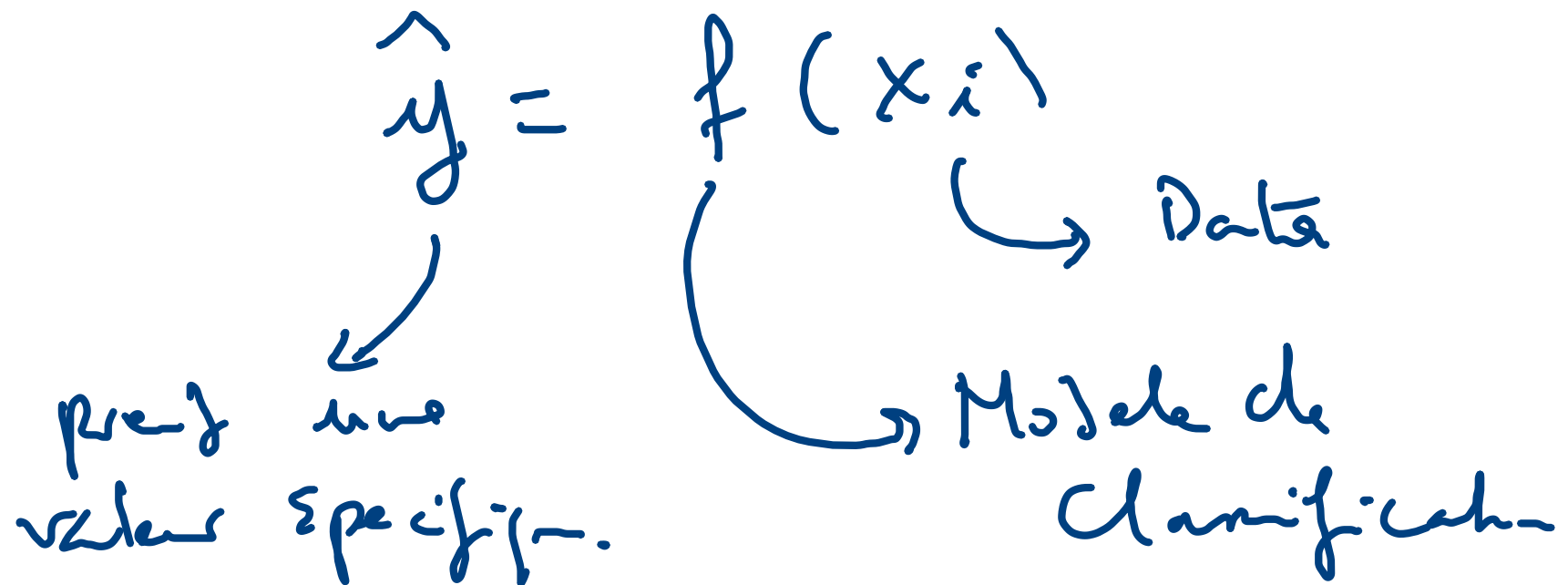
$H_1$ :  $y$  n'a pas  
la valeur NON  
alternative

Developper une statistique basée sur la  
Data : nous dire si on est dans les  
Conditions de  $H_0$ ,  $H_1$ .

## Classification

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

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

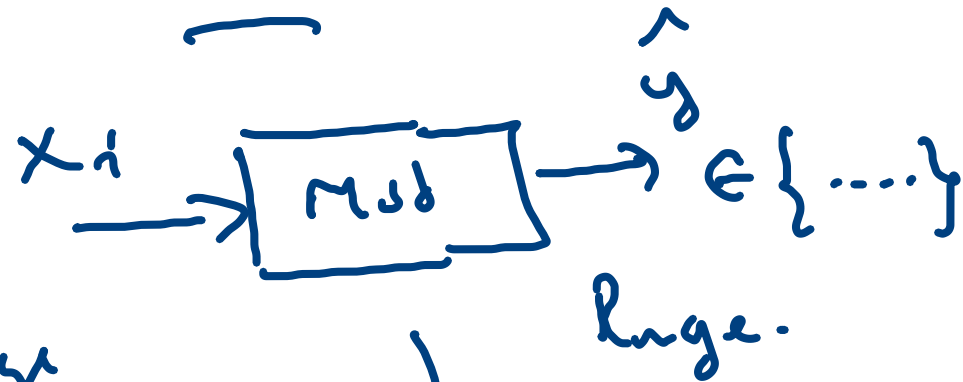


Objectif : Développer un modèle de  
classification (Data Training,  
Data Test)

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

# classification

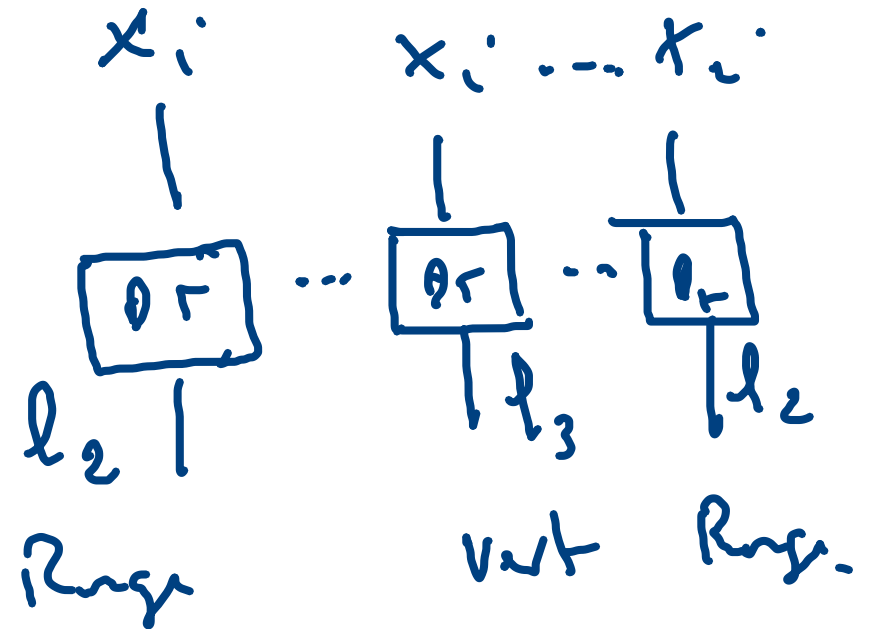
Base.



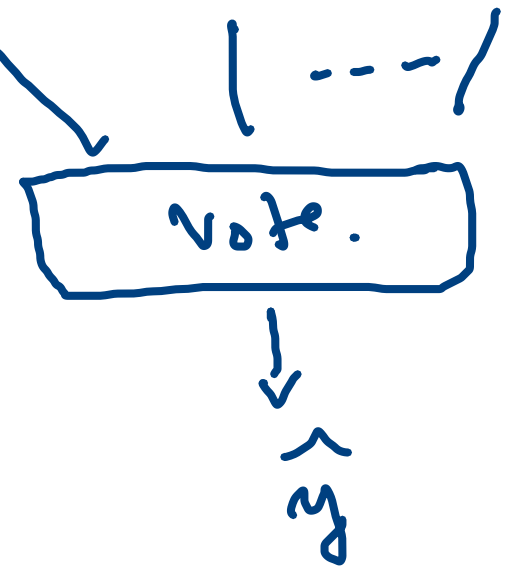
Rule  
vnt  
Blanc

Decision  
Tree.

Ensemble.



Random  
Forest



# Algo de KNN

Qui se ressemble  
s'assemble

K-nearest neighbor

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

$e_c \rightarrow$   
 $e_1$   
 $e_2$   
 $e_3$   
 $\vdots$

	$x_1$	$x_2$	$\dots$	$x_n$	$y$
$e_1$	:	.			R
$e_2$					R
$\vdots$					V
$e_m$					B

$e_c = \{x_1=?, x_2=?, \dots, x_n=?\}$   
 $\hookrightarrow$  marque  $\hat{y} \in \begin{matrix} R \\ V \\ B \end{matrix}$

## Objectif

- 1-) Définir une Distance basée sur les  $x_i$
- 2-) Pour 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).

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

Trier les distances par  
ordre croissant

$k_1 \rightarrow d_{25} \rightarrow e_{25} \rightarrow V$   
 $d_4 \rightarrow e_4 \rightarrow R$   
 $d_{11} \rightarrow e_{11} \rightarrow V$   
 $\vdots$

$K$   
 $i_{\text{input}} \rightarrow$   
 $\underline{K=1}$   
 $K=2$   
 $\underline{K=3}$

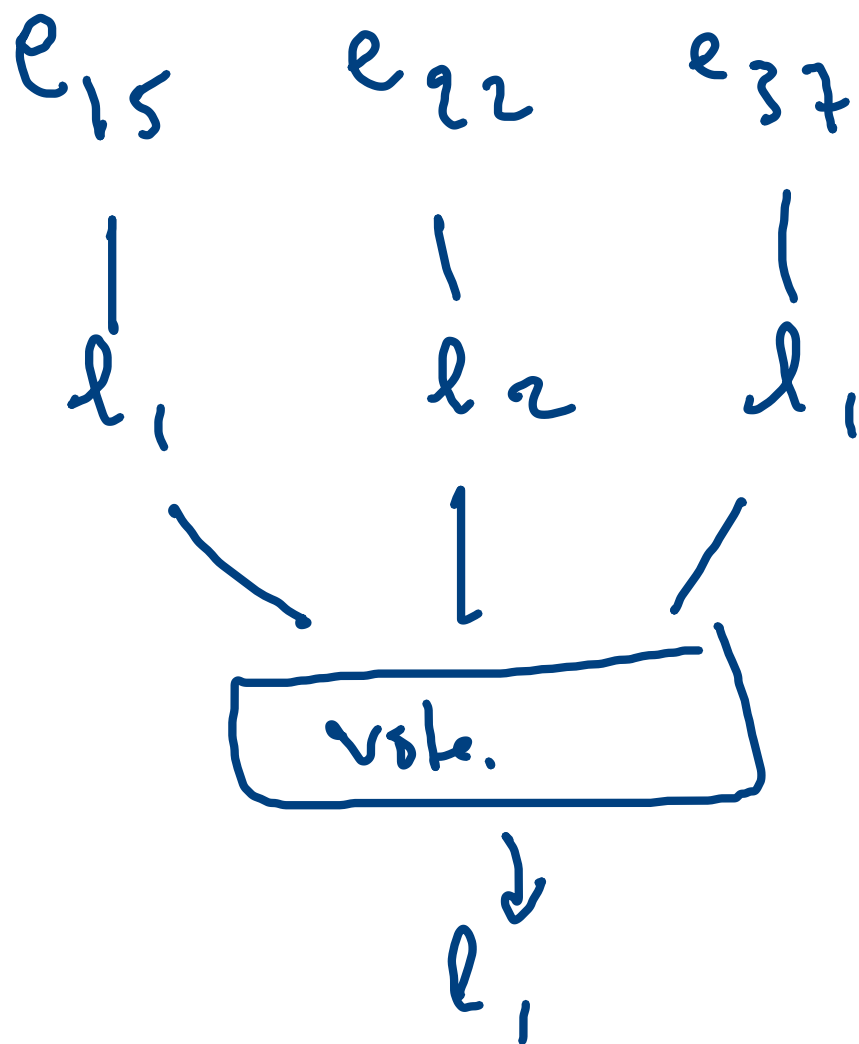
$y_c \rightarrow V$   
 $y_c ? V / R$   
 $y_c \rightarrow V$

$KNN$

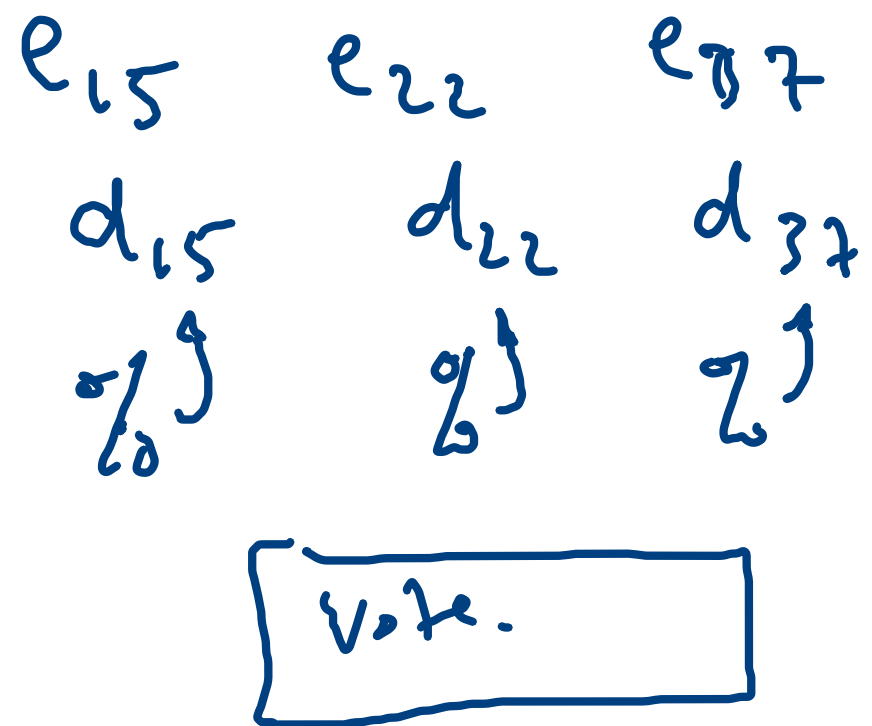
Vote sur  
 $y / K.$   
 $\Rightarrow$  label  
 majoritaire

Vote pour obtenir le label correspondant.

majoritaire  
sur  $K$  points.



Ponderé: le vote sera pondéré par la distance obtenue.





KNN n'a pas de Training.

Choisir la valeur de  $K$ .

erreur  
(Métrique)

