

21 oct 2025

1-) Retour sur la régularisation (Reg. v. m. !)

2-) Approches d'ensemble en classification

- Bagging

- Boosting -

3- Pratique: - Regularisation.  
- Random Forest.

4-) TP: - progression -

- finalisation ppt presentation

## Dev Model Prediction:

- 1) Dataset: Absence de volume de Données ✓
- Nombre de predicteurs  
(Characteristics / Descripteurs)
  - Algo choisi est un approprié pour le contexte (Domaine).

## 2) Processus de Dev Model.

objectif: Modèle dont les métriques dépassent le Baseline

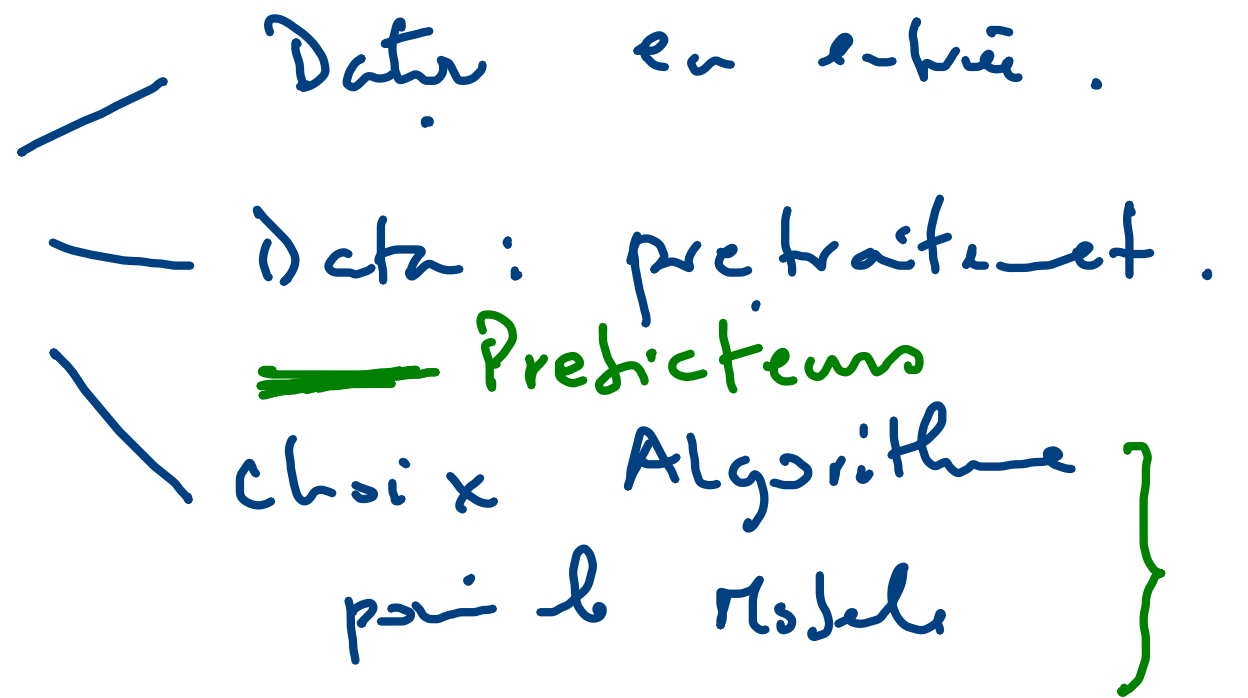
Training - Modèle avec Biais faible

Prod - Modèle doit généraliser en présence de Live Data

underfit / Overfit

Trade off Bias-variance.

Spécialité ML

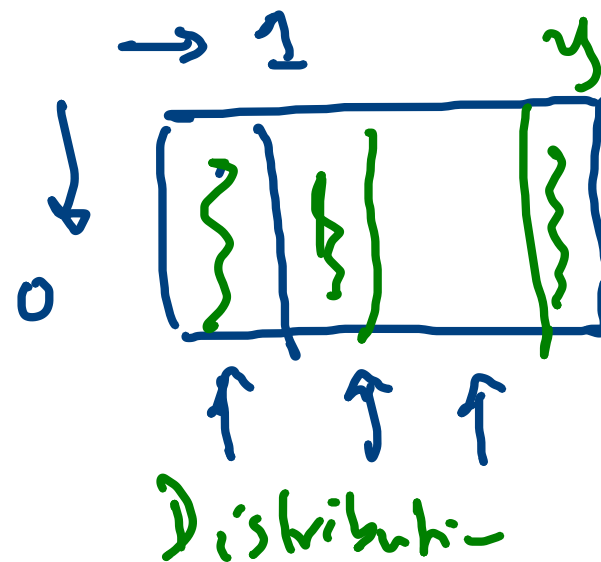


3 éléments  $\Rightarrow$  Qualité du  
Modèle obtenu

TP

1) Data + pretraitement

axis 0: Evaluer puis  
appliquer des opérateurs



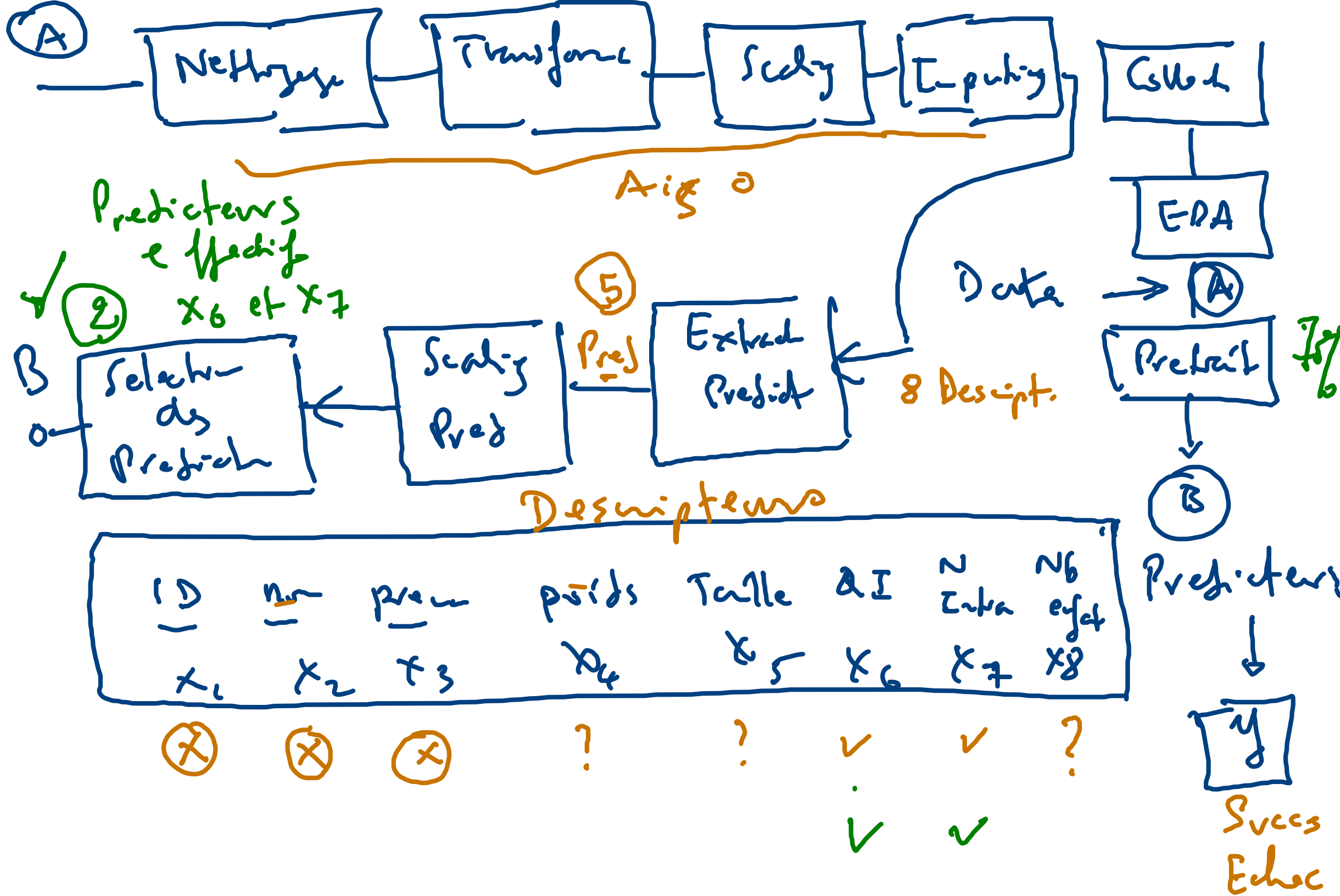
2) Choix de prédiction.

axis 1

- Mauvaise extraction
- Mauvaise sélection ← P.C.A

3-) Modèle: Algo

- choix mal fait de l'algo.
- Tuning de l'algo est mal fait
  - Hyperparameters (paramètres qui va dépendent pour de Data).



# Descripteurs

ID	nom	prénom	poids	Taille	AI	N Inter	Nb eff
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$

⊗

⊗

⊗

?

?

✓

✓

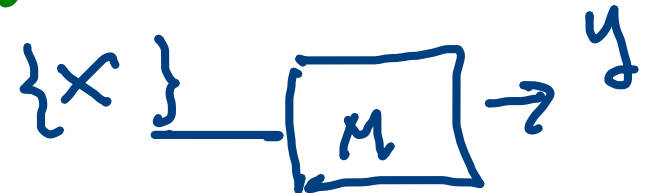
?

$x_i$  doit avoir un effet sur  $y$ .

statistique

✓

✓



## Test hypoth.

$H_0$ :  $x_i$  n'a pas d'effet

$H_1$ :  $x_i$  a un effet

Statistique de Test

→ Comparer à une  $z_{\alpha}$

Si  $H_1$  est retenue

⇒  $x$  a un effet

Z-test

t-test.

Advanced Algo  
Base



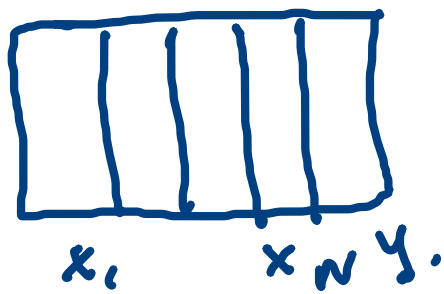
Algo  
Regression - regularisation - Ridge - Lasso

Classification - Ensemble.

Bagging

Boosting.

Prod



Regression

$\begin{cases} x_i \\ \vdots \\ x_N \end{cases}$   
90 82 ....



$\hat{y}$

$y_{\text{PRED.}}$

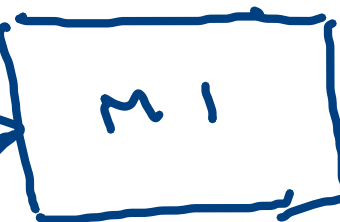
91.0

$\begin{cases} \text{RMSE} \\ R^2 \end{cases}$

Alternative

$x_1 \quad x_4 \quad x_8 \quad x_{10} \quad x_{11}$

$x_1 \quad x_4 \quad x_8$



$\hat{y}_1$

$x_1 \quad x_{10} \quad x_{11}$

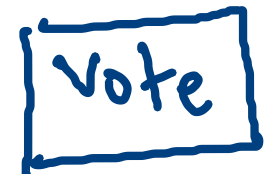


$\hat{y}_2$

$\vdots$



$\hat{y}_m$



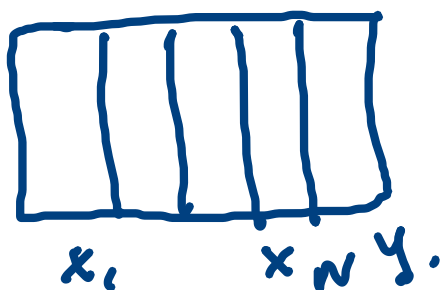
Moy  
Med

Lissage

Parallèle

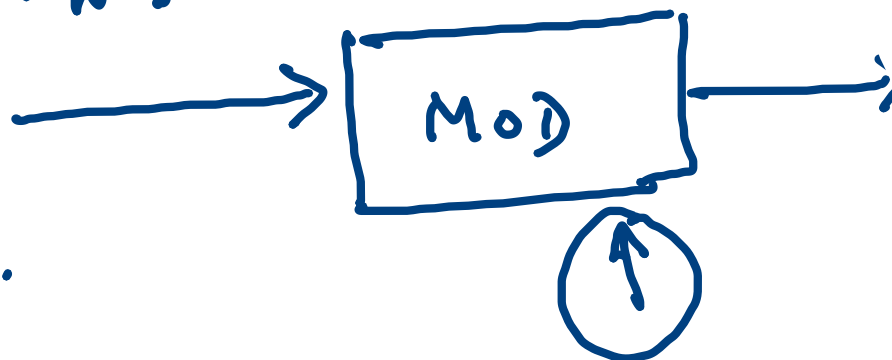


Prod



classification

$\left\{ \begin{matrix} x_i \\ \vdots \\ x_N \end{matrix} \right\}$   
 $\begin{matrix} 90 \\ 82 \\ \dots \end{matrix}$



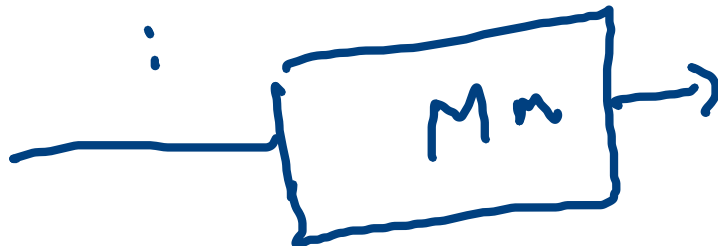
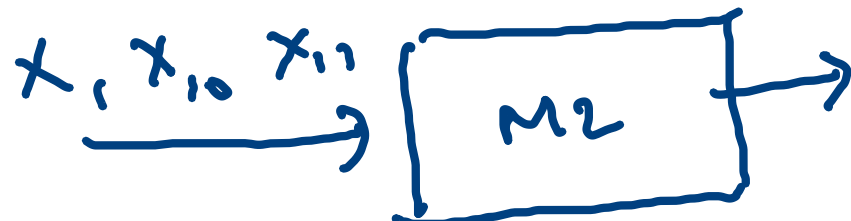
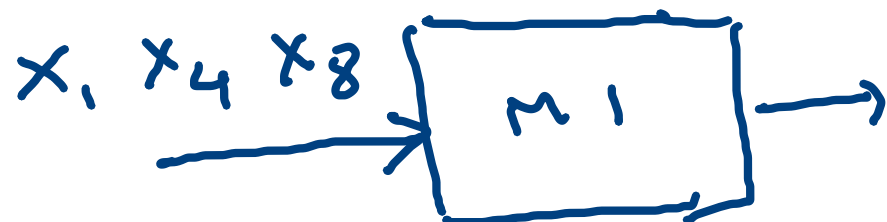
Result: Succes  
Echec.

$y_{pred}$

Algorithme

$x_1 \quad x_4 \quad x_8 \quad x_{10} \quad x_{11}$

Parallèle



Echec

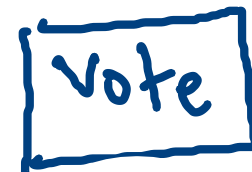
$\hat{y}_1$

$\hat{y}_2$

$\hat{y}_m$

Succes

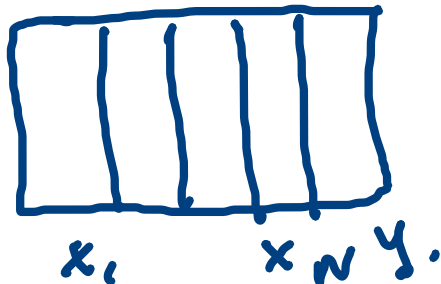
Echec



Mode.

(voter frequency)

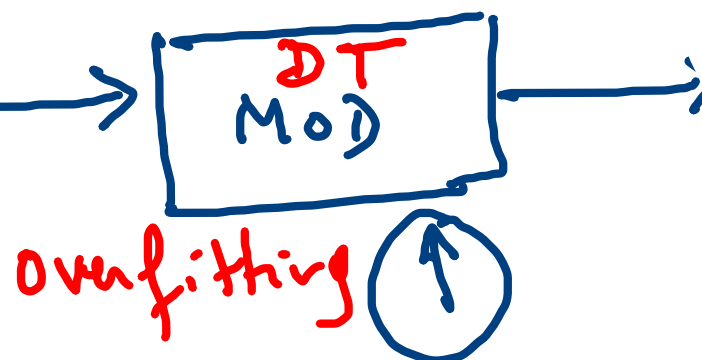
Prod



classification - Decision Tree

Result: Success  
Echec.

$\begin{cases} x_i \\ \vdots \\ x_n \end{cases}$   
90 82 ....



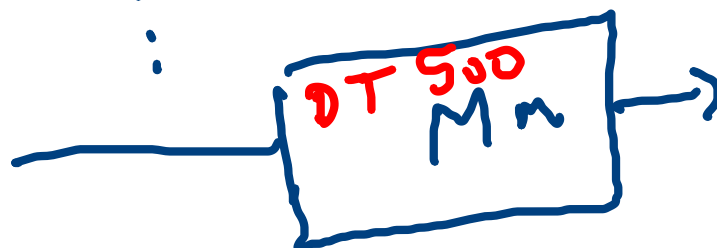
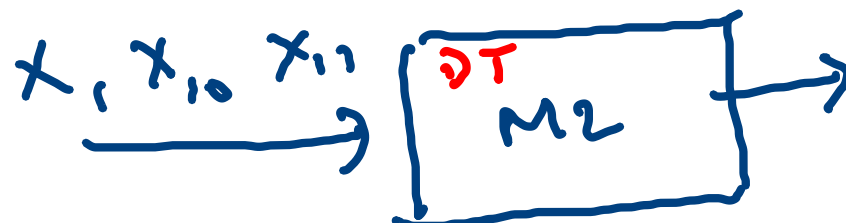
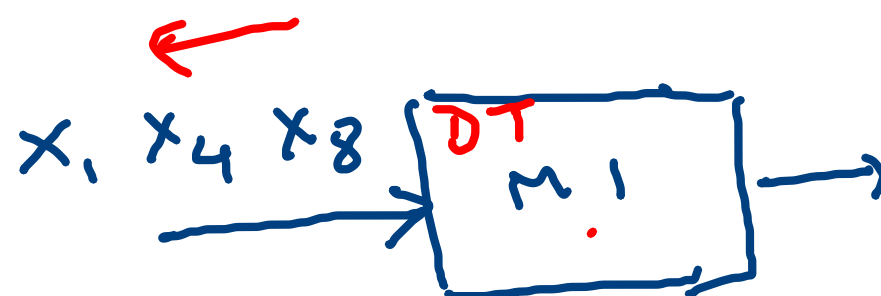
$y_{\text{pred.}}$

$\in \{ \underline{l_1}, \underline{l_2}, \dots, \underline{l_n} \}$

Algorithme.

Random Forest

$x_1 \quad x_4 \quad x_8 \quad x_{10} \quad x_{11}$



Echec

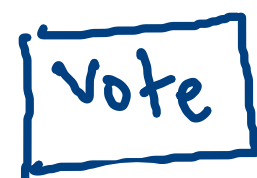
$\hat{y}_1$

$\hat{y}_2$

$\hat{y}_n$

Success

Liège



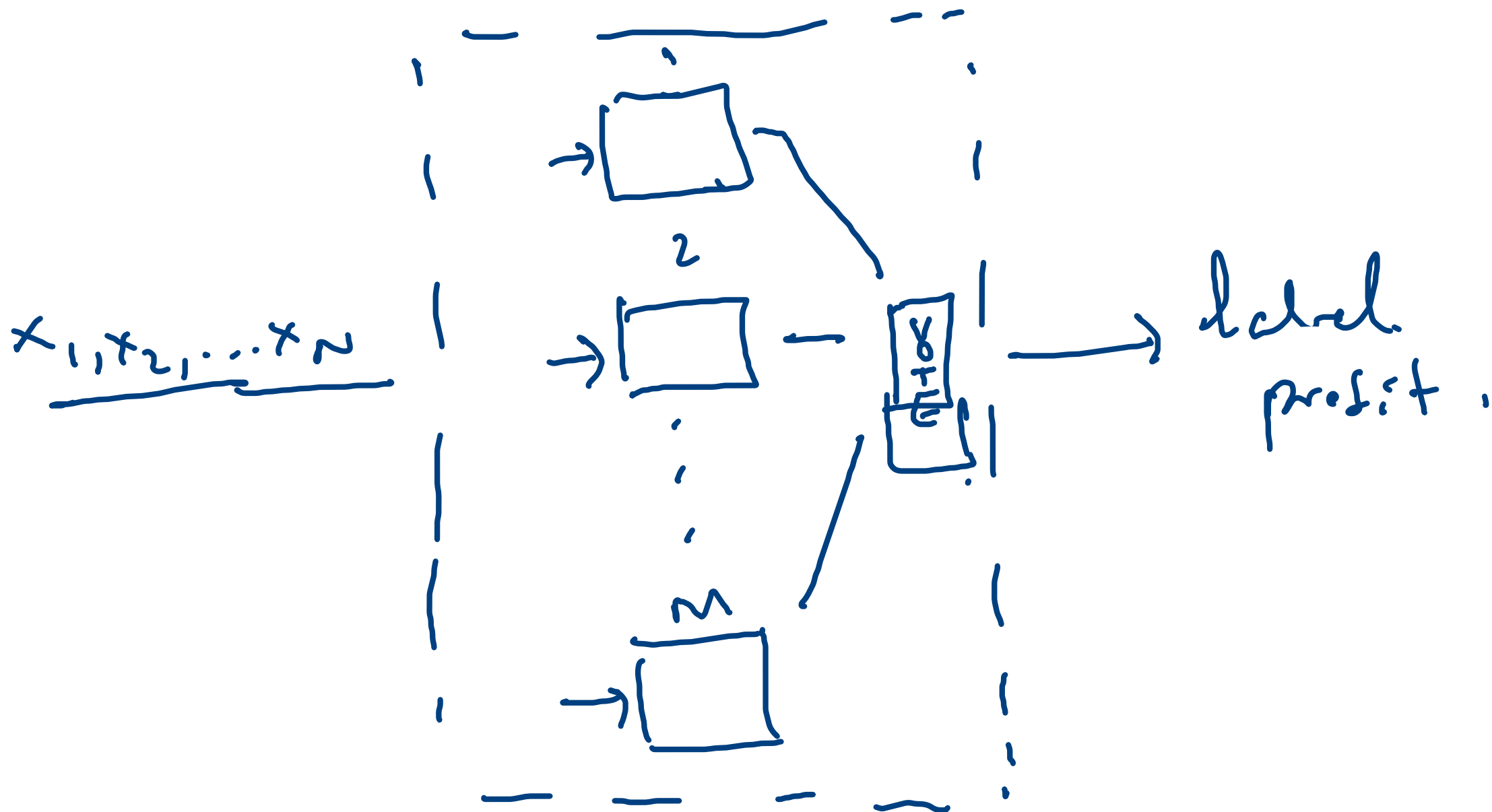
Echec

Mode.  
(voter frequency)

D.T Sujet à l'overfitting

Bagging  
Approach  
parallel.

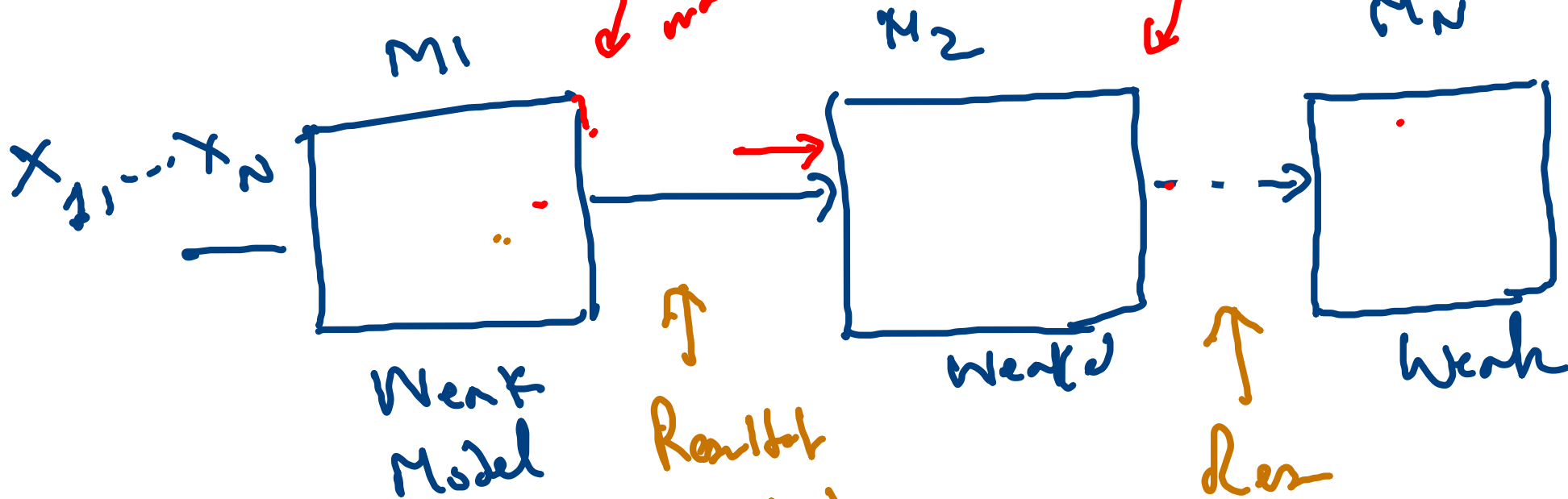
Model : - Créer un ensemble  
de Model en  
parallel  $\Rightarrow$  Réduire le  
overfit  
- Achiever la sélection du  
meilleur label.



# Boosting

Labelize  
 $x$  &  $y$  qui  
 sont mal predict

A-choix  
 sur  
 $y$  price



	$x$	$x_n$	$y$
$e_1$			✓
$\vdots$			
$e_m$			✓

certains  
 sont mal  
 predict.



$y$  sont  
 Correctement  
 predicts

# Modelle zu Entwickeln (Robuste) $\rightarrow$ overfit!

- Algo der Base.
- Anpassen der  
pre-train.
- Collection der Daten  
suchen.
- Hyperparameter.
- Cross validation.

