

21 oct 2025

1-) Retour sur la regularisation (Régression)

2-) Approches discutées en classification

- Bagging

- Boosting -

3- Pratique: - Regularisation.

- Random Forest.

4- TP: . progression -

- finalisation ppt presentation

## Dev Model

## Predictions

1) DataSet : - Absence de valeur de Données ✓

- Nombre de prédicteurs  
(Caractéristiques / Descripteurs)

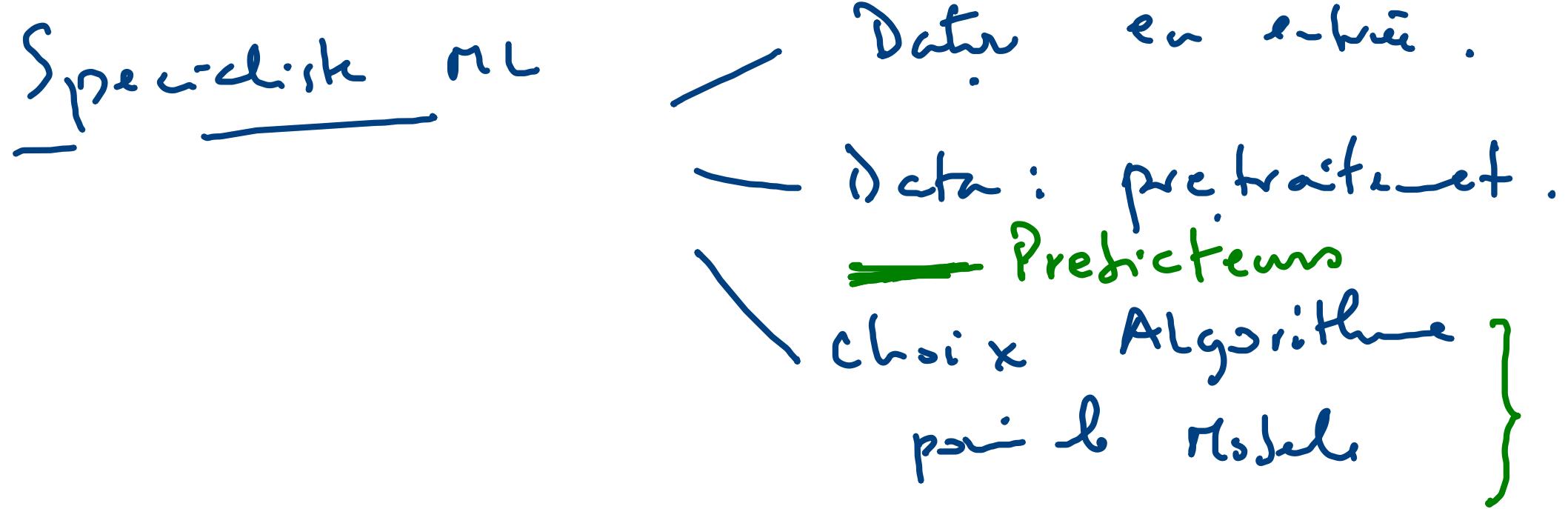
- Algo choisi est un approprié  
pour le contexte (Domaine).

2) Processus de Dev Modèle.

objectif: Modèle dont les métriques dépassent  
la Baseline

Training - Modèle avec Basis taillé

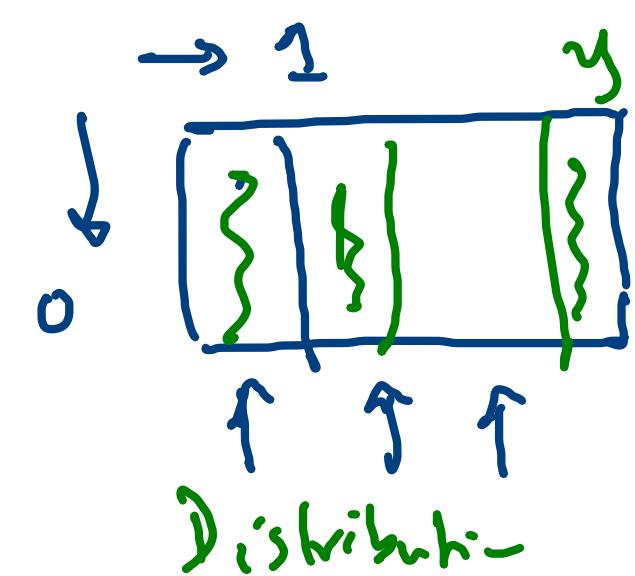
Prod - Modèle doit généraliser la  
précision de Live Data  
Underfit/Overfit Trade off  
Bias-variance.



3 étapes  $\Rightarrow$  Qualité du  
modèle obtenu

TP  
1) Data + prétrait

axis 0 : Evaluer puis appliquer des opérat.



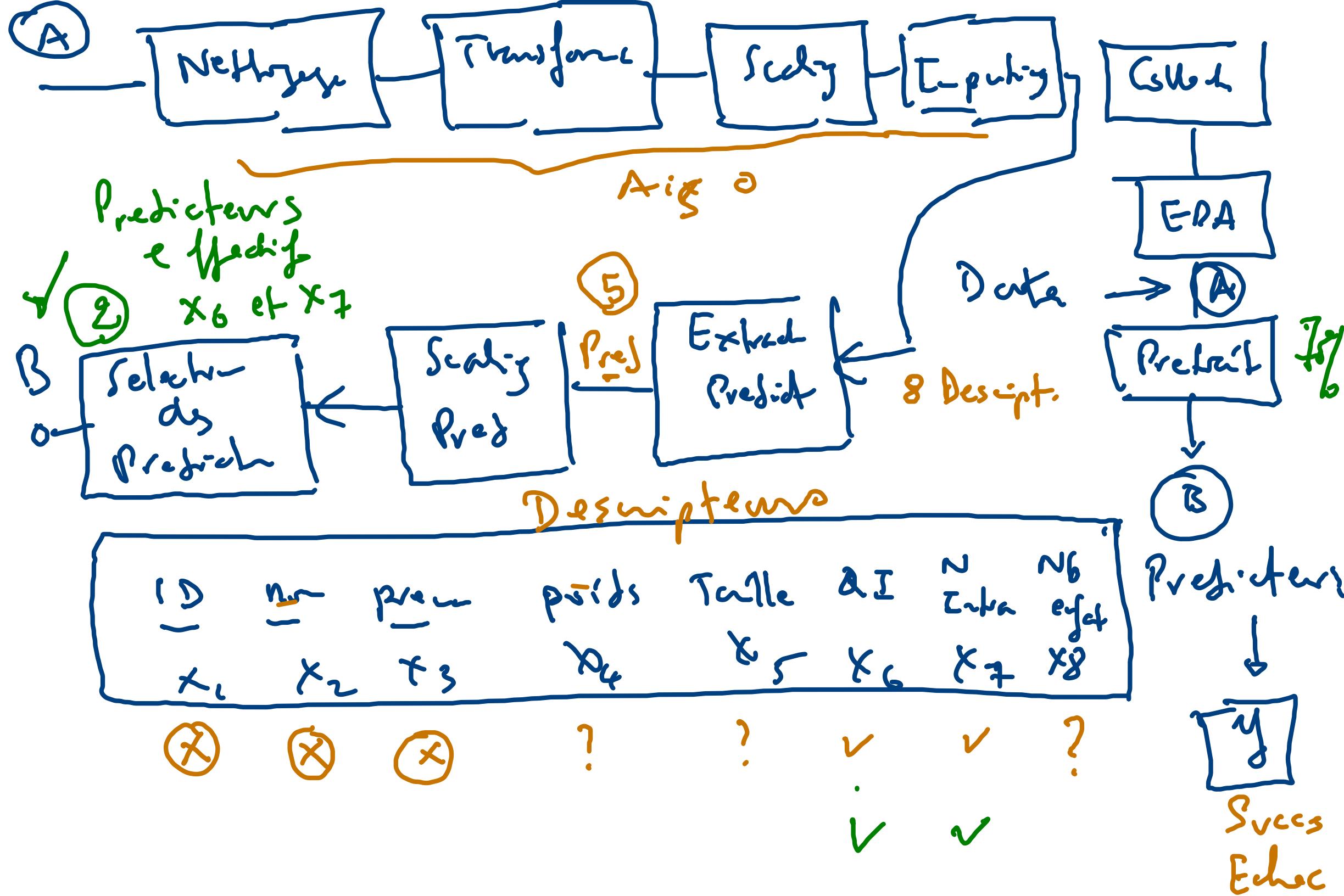
2) Choix des predict.

- axis 1
- Mauvaise extraction
  - Mauvaise sélection  $\leftarrow$  P.C.-A

—

3 -> Modèle : Algo

- choix mal fait de l'algo.
- Tuning de l'algo et mal fait
  - Hyperparameters (paramètres qui ne dépendent pas du Data).



## Descripteurs

ID	nr prem poids	Taille	AI	$\Sigma$	Nb				
$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	Interv	effet

✗

✗

✗

?

?

✓

✓

?

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

statistique

✓

✓

{x}



### Test hypothèse

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

$H_1$ :  $x_i$  a un effet

statistique de Test

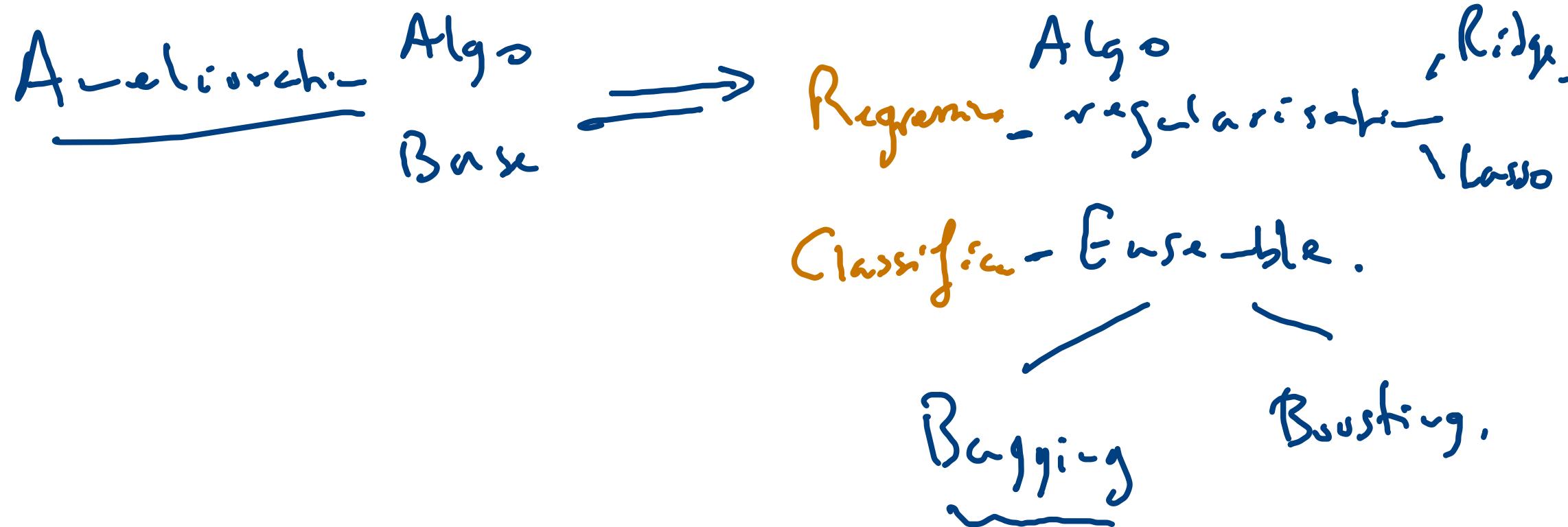
→ Comparer à une Z ou t.

Si  $H_1$  est rejetée

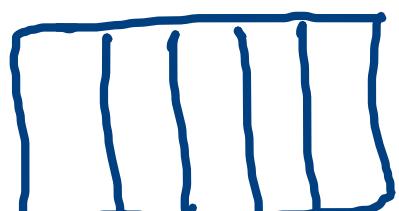
⇒ x a un effet

Z-test

t-test.



Prod



Regression

$\{x_i\}$   
 $a_I$   
 $q_0$

$n_I$   
 $82$

$\dots$



$\hat{y}$

91.0

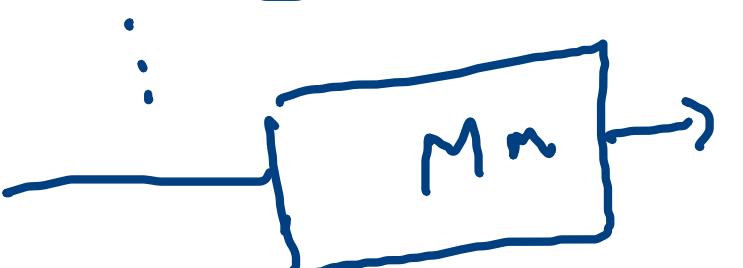
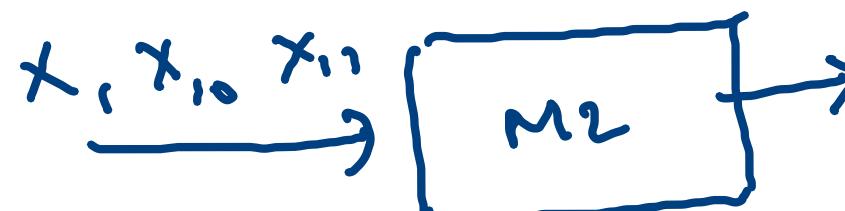
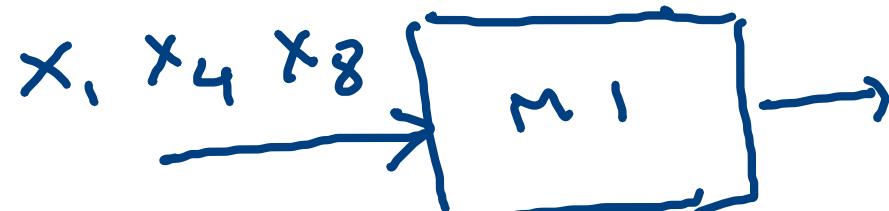
$y_{PRED.}$

RMSE  
R-2

Aleatoric

$x_1, x_4, x_8, x_{10}, x_{11}$

Parallel



$\hat{y}_1$

$\hat{y}_2$

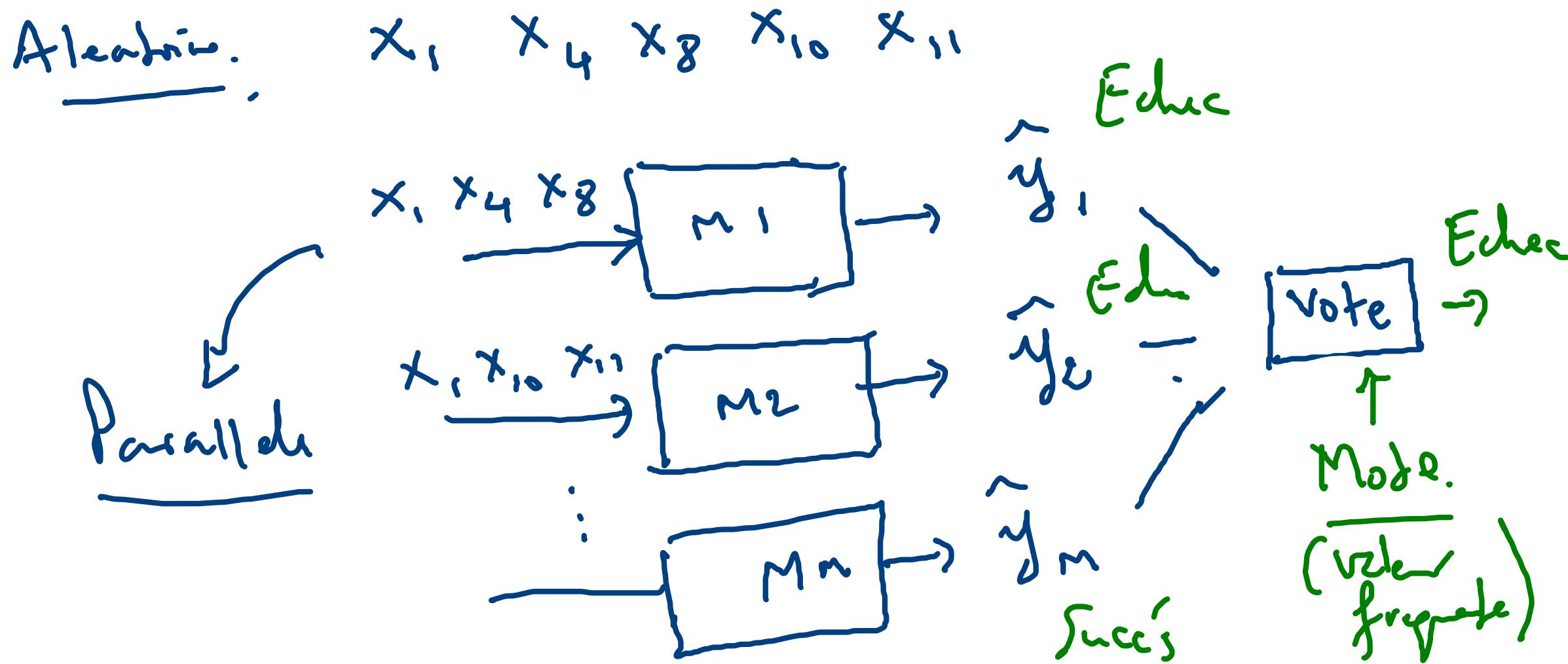
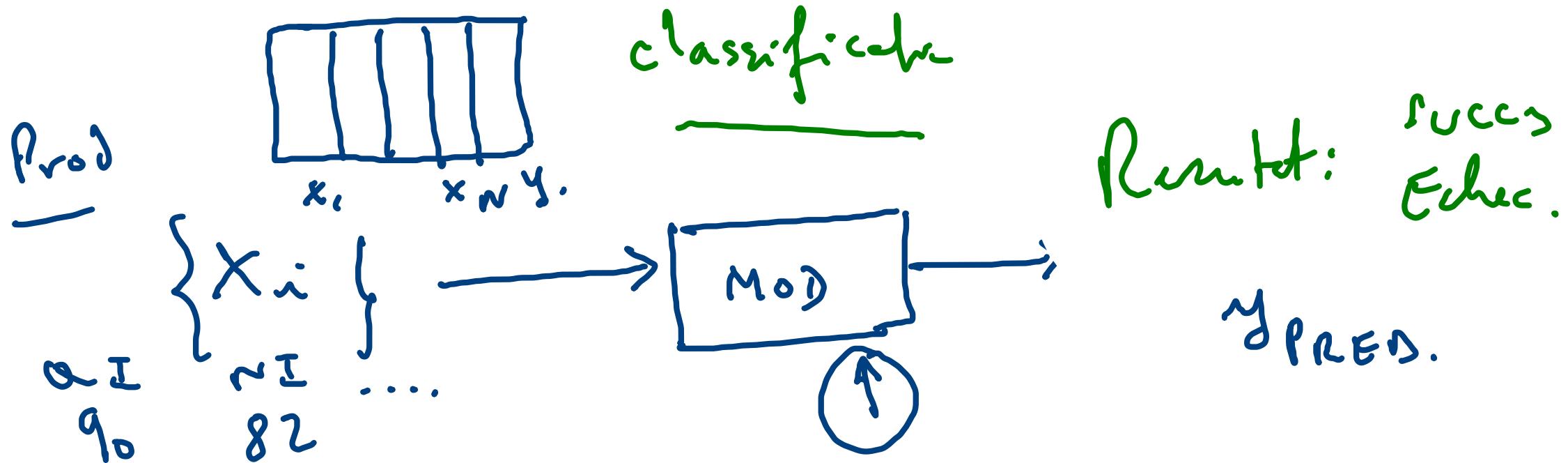
$\hat{y}_m$

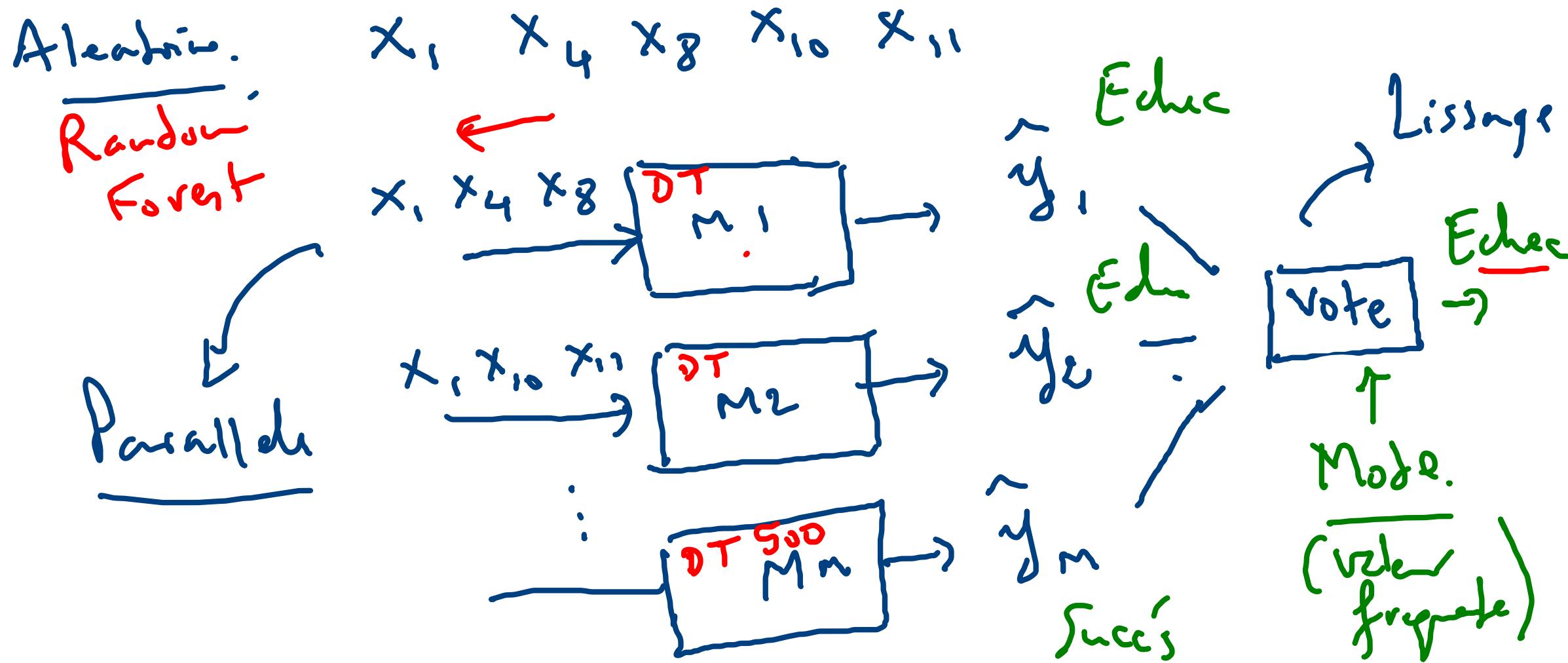
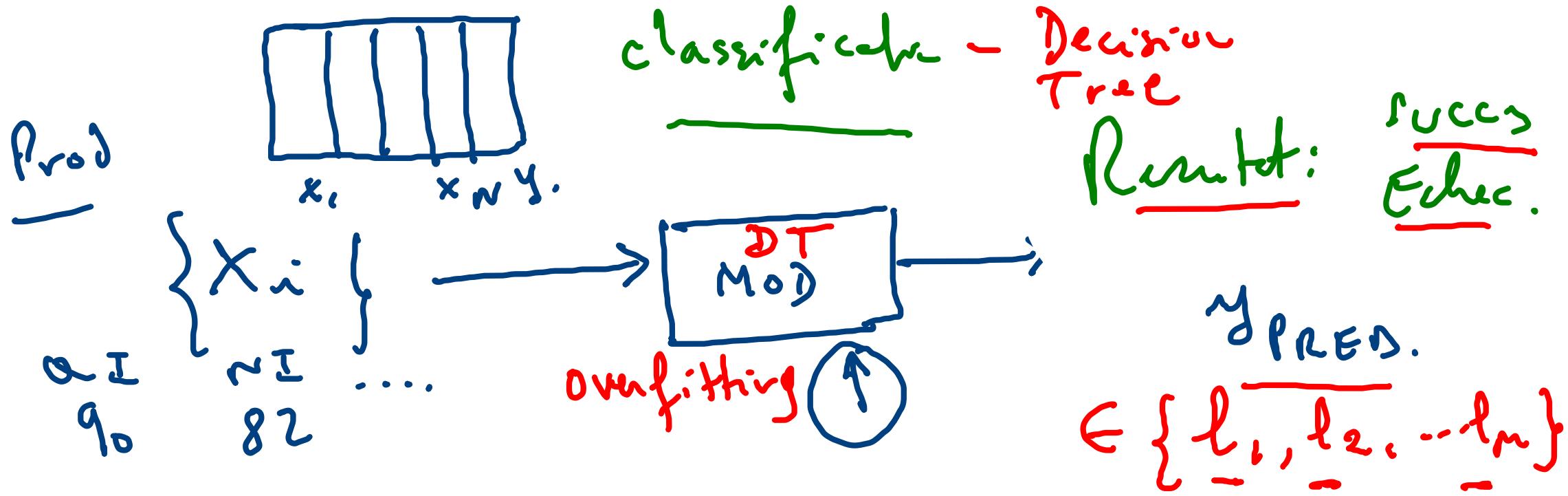
Vote

Moy

Maj

Lissage





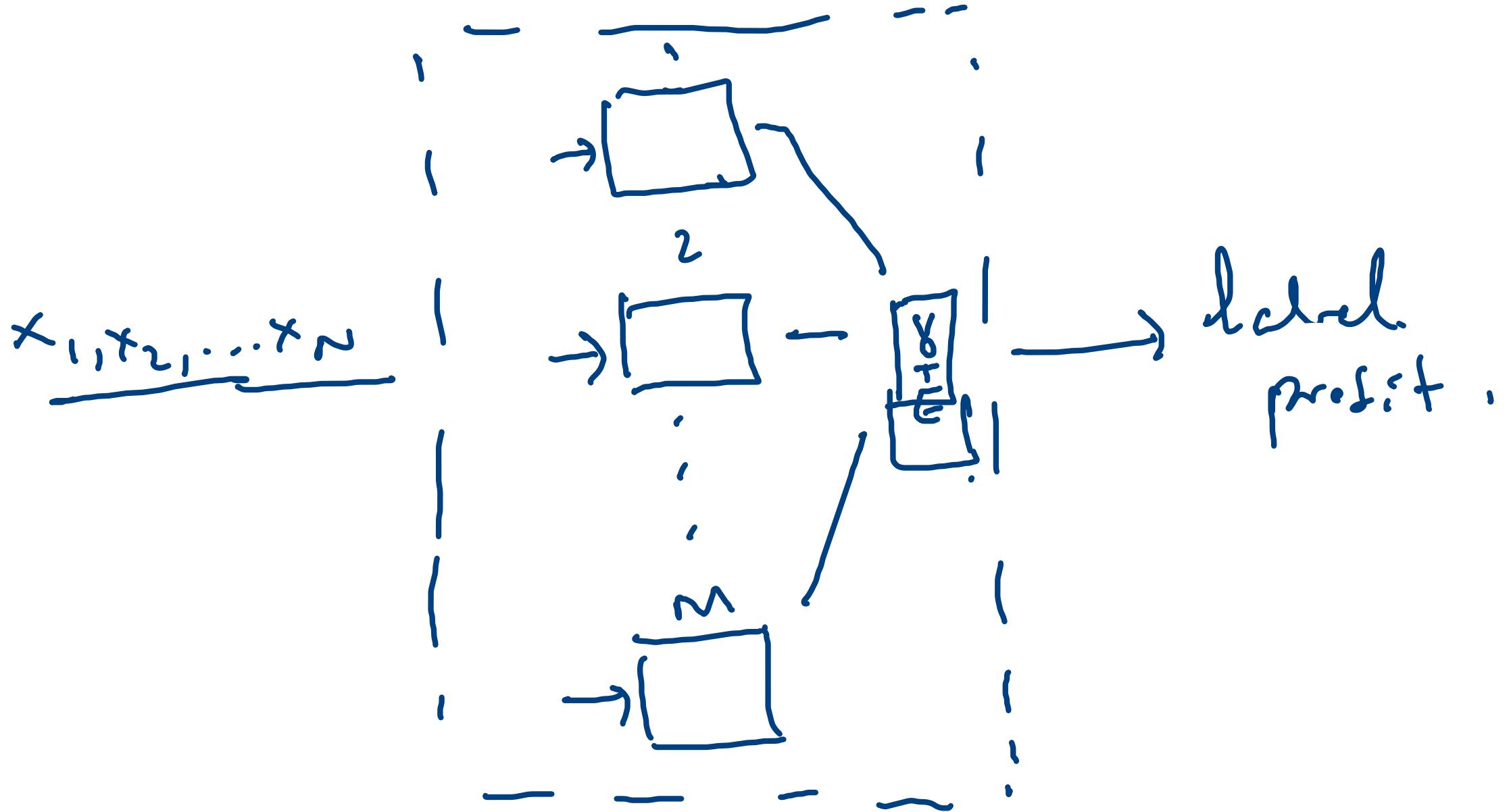
D.T. sujet à l'over fitting

Bagging

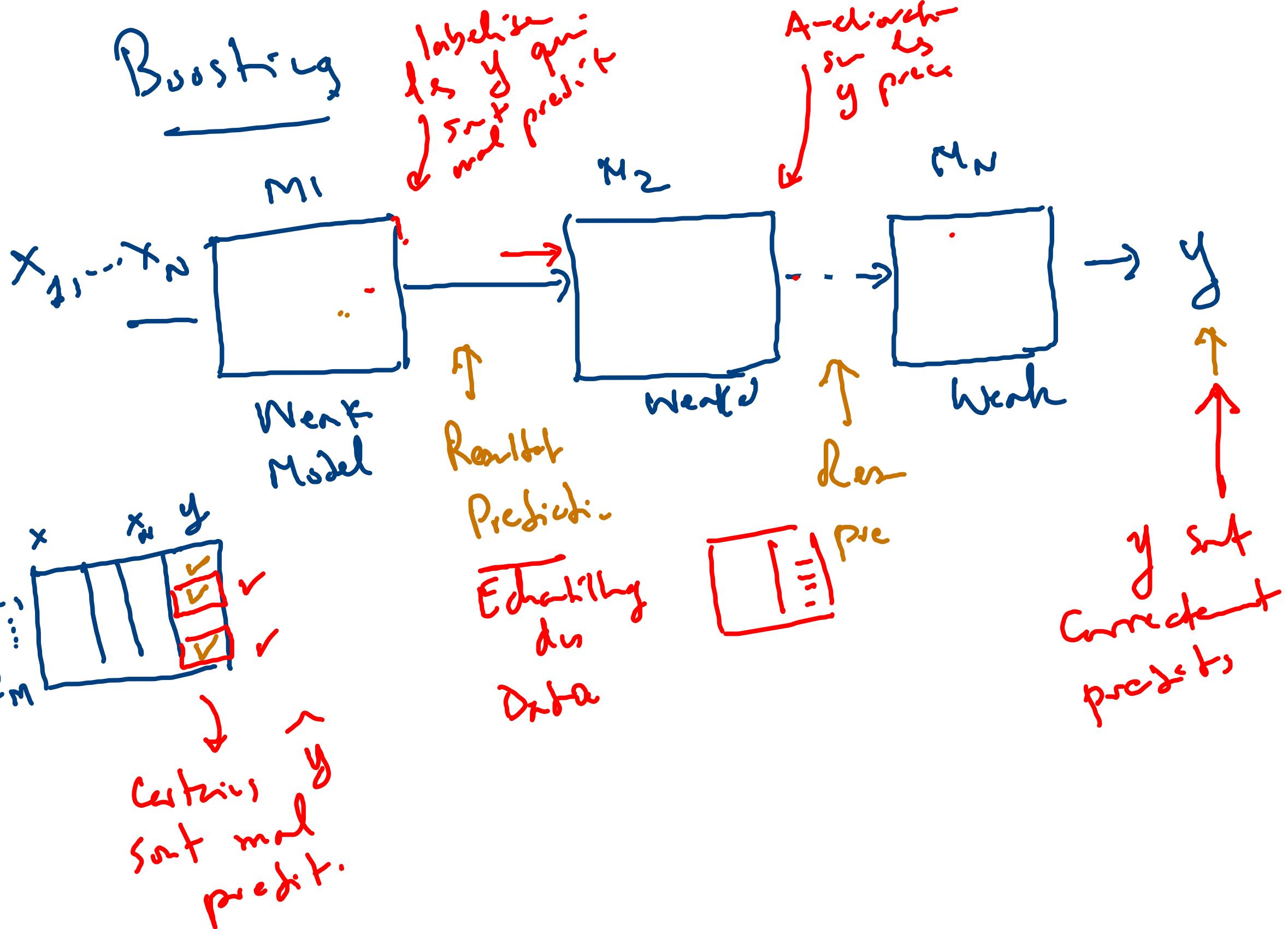
Modèle

- Crée un ensemble de Modèle en parallèle  $\Rightarrow$  Réduire le overfit
- Archiver la sélection du meilleur label.

Approche parallèle.



# Boosting



# Modèle à Développer

---

(Robuste)  overfit !

Algo de Base.

- Ajuster le  
paramètre -

- Collecter des  
données

graphique

- Hyperparamètre

- Cross validation.

Regularisation Approach  
Lasso Ridge,   
d'ensemble.  
! { Bagging Boosting