

12 Sept 2025

Regression Linéaire.

1-> Retour sur la R.L

2-> Déroulement d'un processus de modèle RL.

3-> TP (suivi): - Formation des Equipes
- Domaine/ Thème .

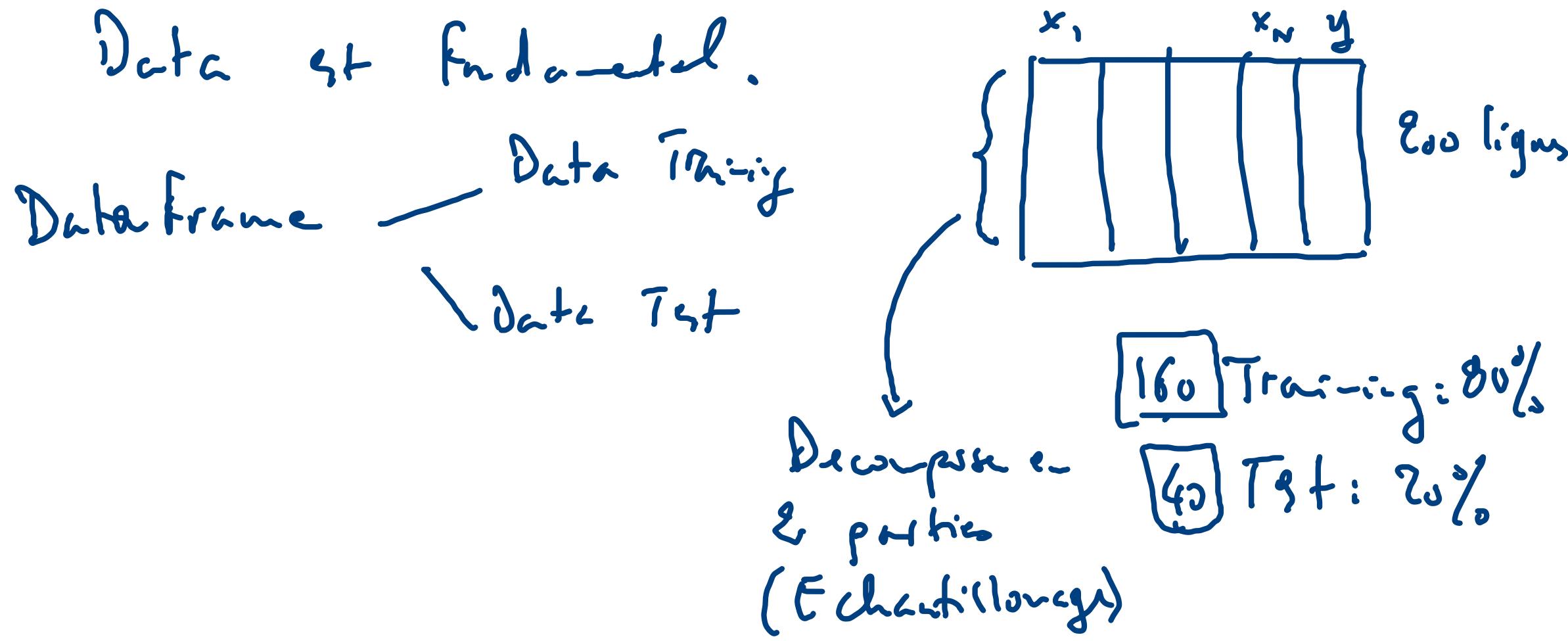
Apprentissage Supervisé

1-) Phase de training: Dev du modèle (sur un processus) selon la data disponible.

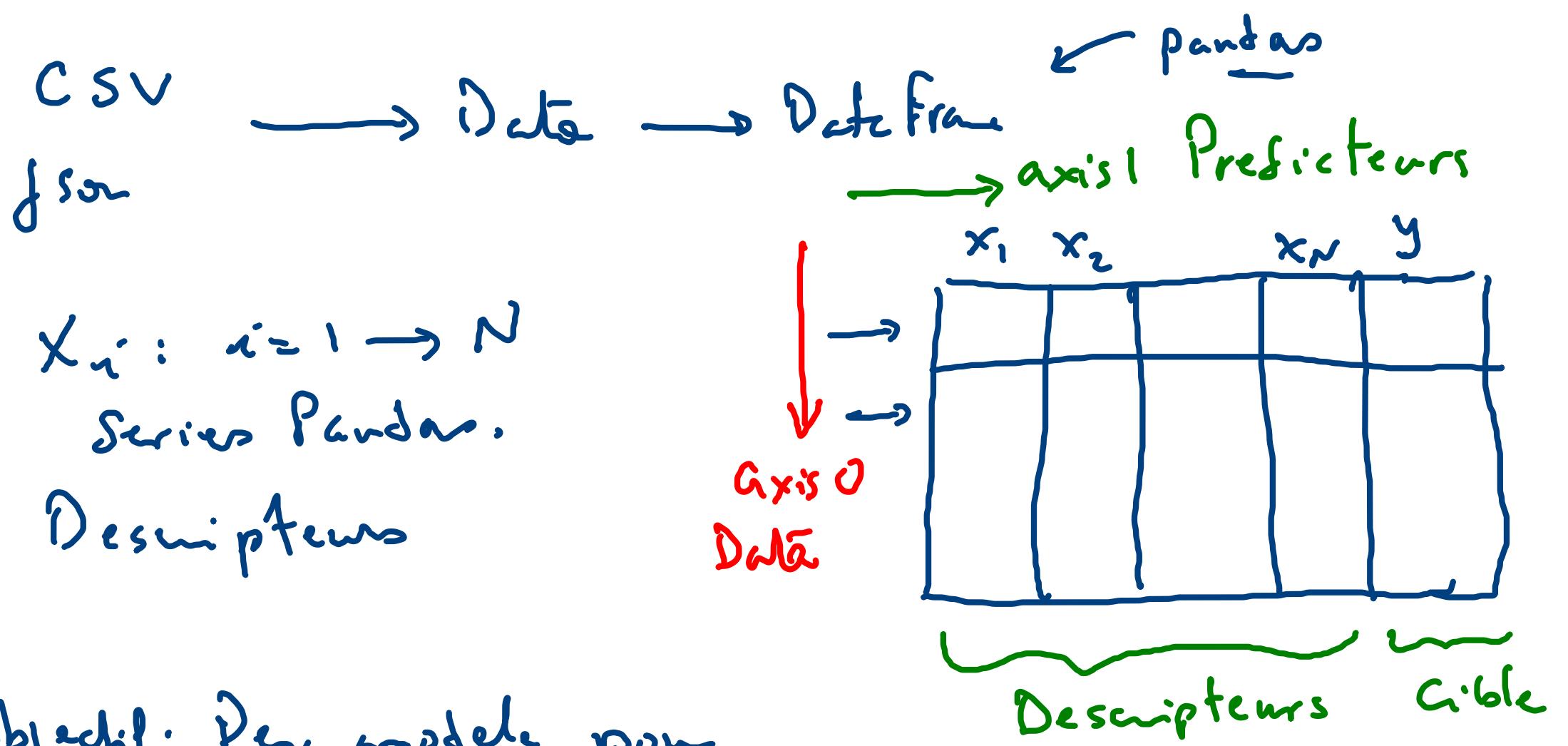
⇒ Data de Training (train)

2-) Phase de Test & Evaluation: Test du modèle candidat ⇒ Utilise du Data de Test

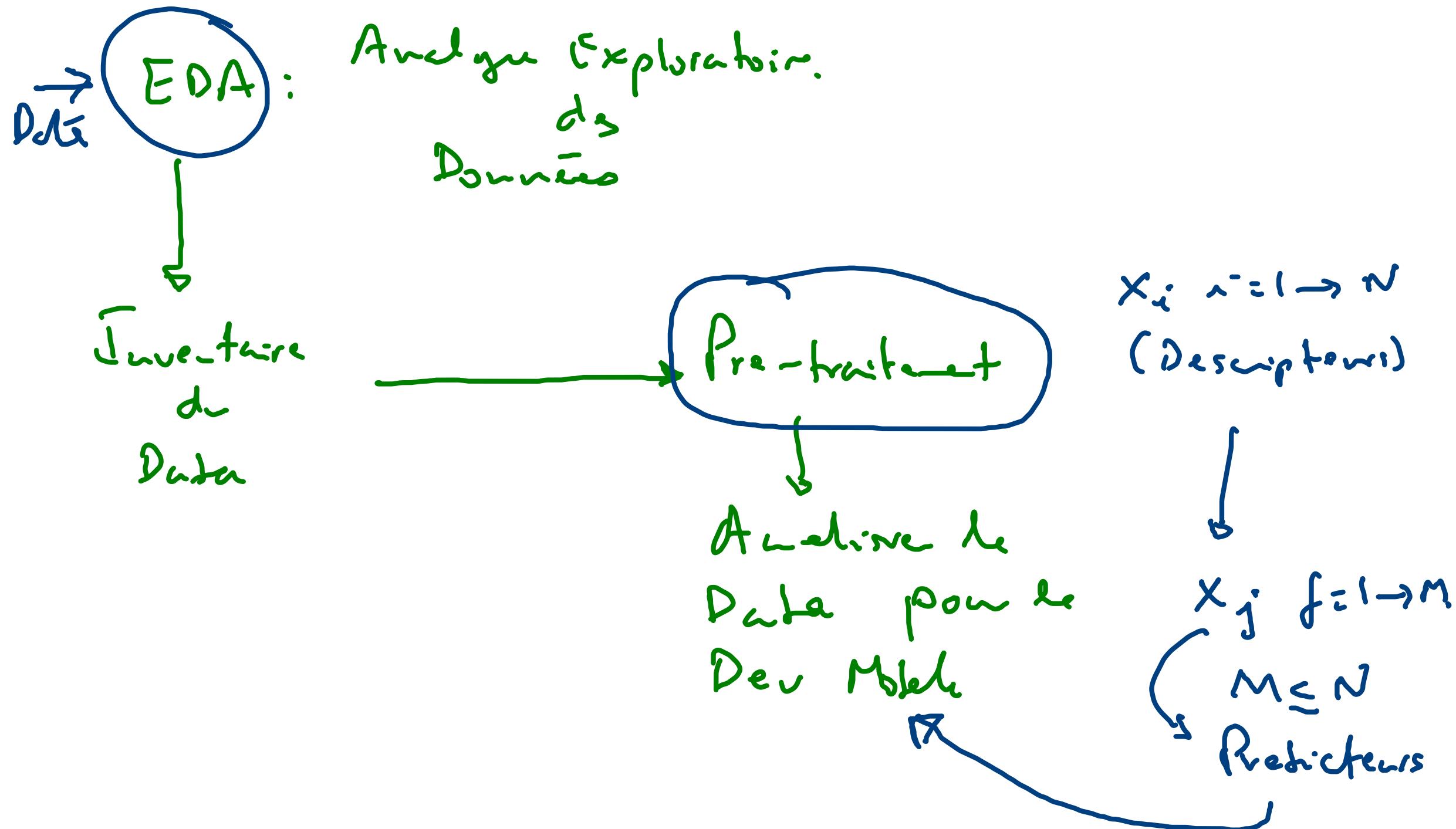
- Obligation:
- 1) Séparation du Data collecté en deux sous-ensembles
 - Data training : Pour l'entraînement
 - Data Test (Hidden) : validation, caché
 - 2) Séparation doit se faire aléatoirement



Reangu: Le date de test doit être caché
(ne doit pas être utilisé pour le Dv)



objectif: Dev modèles pour
prédictre la note final
du A52



y : Continue

5-8

1250-75

:

f

A.S

Regression

- Regression linéaire
- Regression Non linéaire
- Regression avec fact. Bur.

Regression avec Regularis.

- Ridge
- Lasso
- Elastic Net

Classification

y Discret

oui Right 10

Non vert 20

Blanc 30

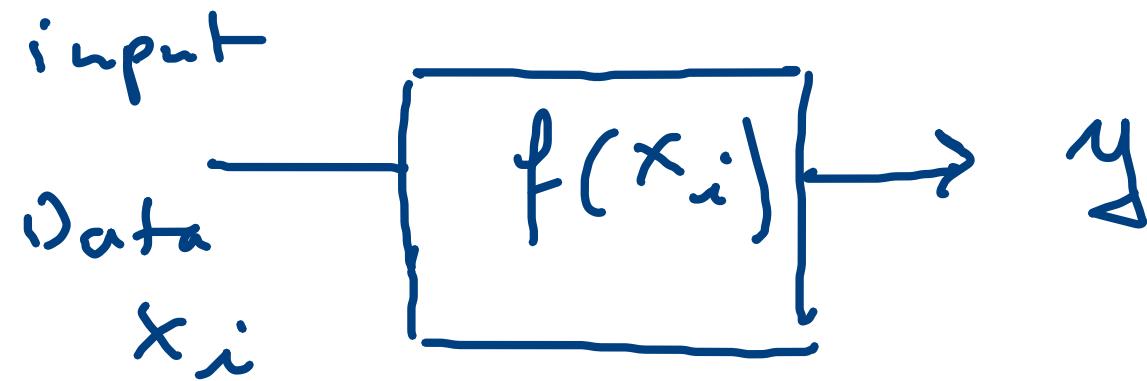
- Naive Bayes

- KNN.

- Logistic
Regression

- Decision
Tree.

;



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

Modelo.

EDA

Matrice de Correlation
 ↴
 HeatMap

↳ Matplotlib
 Seaborn..

- (X_i, y) Correlation (dependencia)
 ↴ effet
- (X_i, X_j) independance
 ↴ Relative celui qui est dependent

$$\left. \begin{array}{l} C(x_1, y) = 75\% \\ C(x_3, y) = 55\% \end{array} \right\}$$

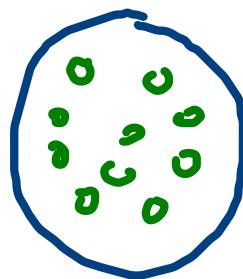
$x_1 = \text{Oignon.}$

$x_2 = \text{Sel.}$

$x_3 = \text{Pouvre.}$

$x_4 = \text{Pois. did.}$

$x_5 = \text{Persil.}$



plat

Pois.

Savon fin.

$$y: 0 \rightarrow 1$$



$$C(x_1, x_2) = 10\%$$

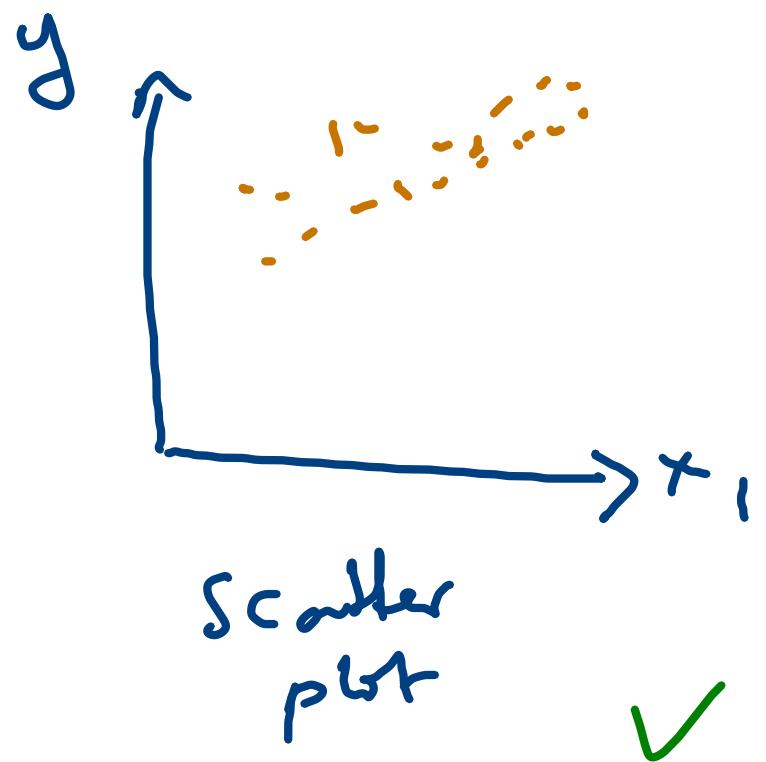
$$C(x_1, x_3) = 85\% \rightarrow$$

x_1 et ~~x_3~~ dépendants

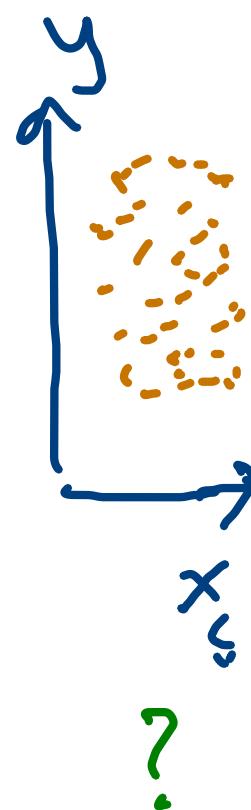
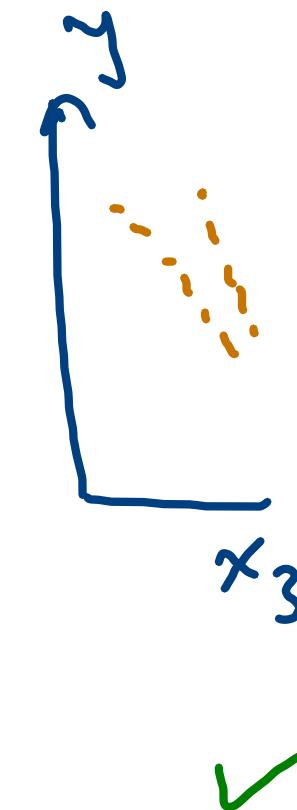
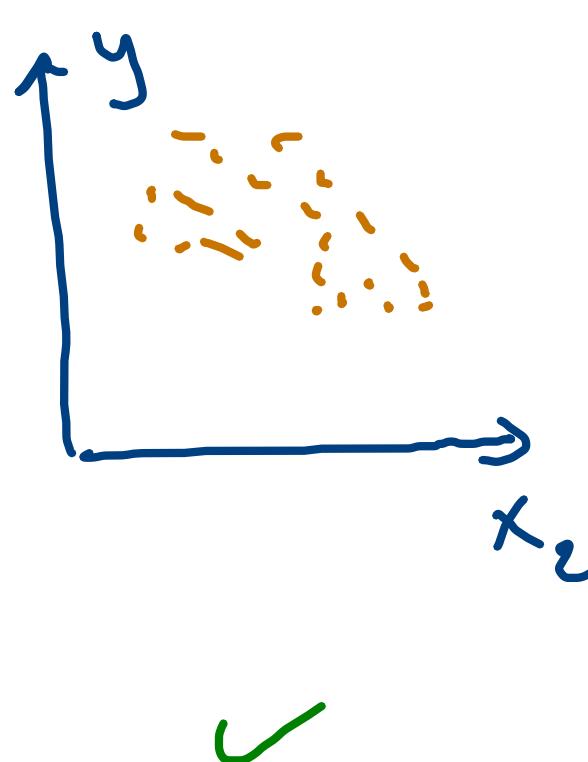
;

$$0 \rightarrow 2\%$$

$x_1 \ x_2 \ x_3 \ x_4$
Nur ein p.
Nur ein p.



y
Nur ein p.
Nur ein p.

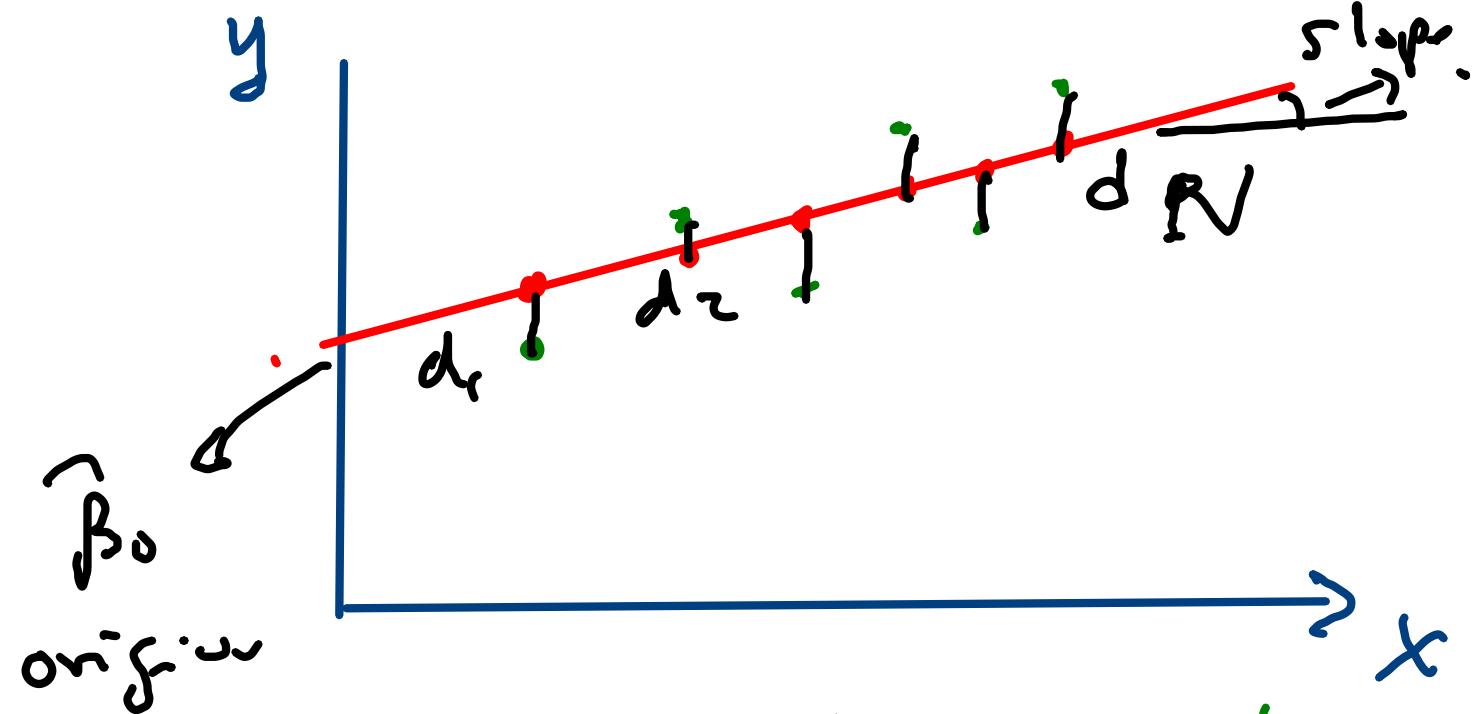


$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

\|

Transer Is
estimator $\hat{\beta}_i$



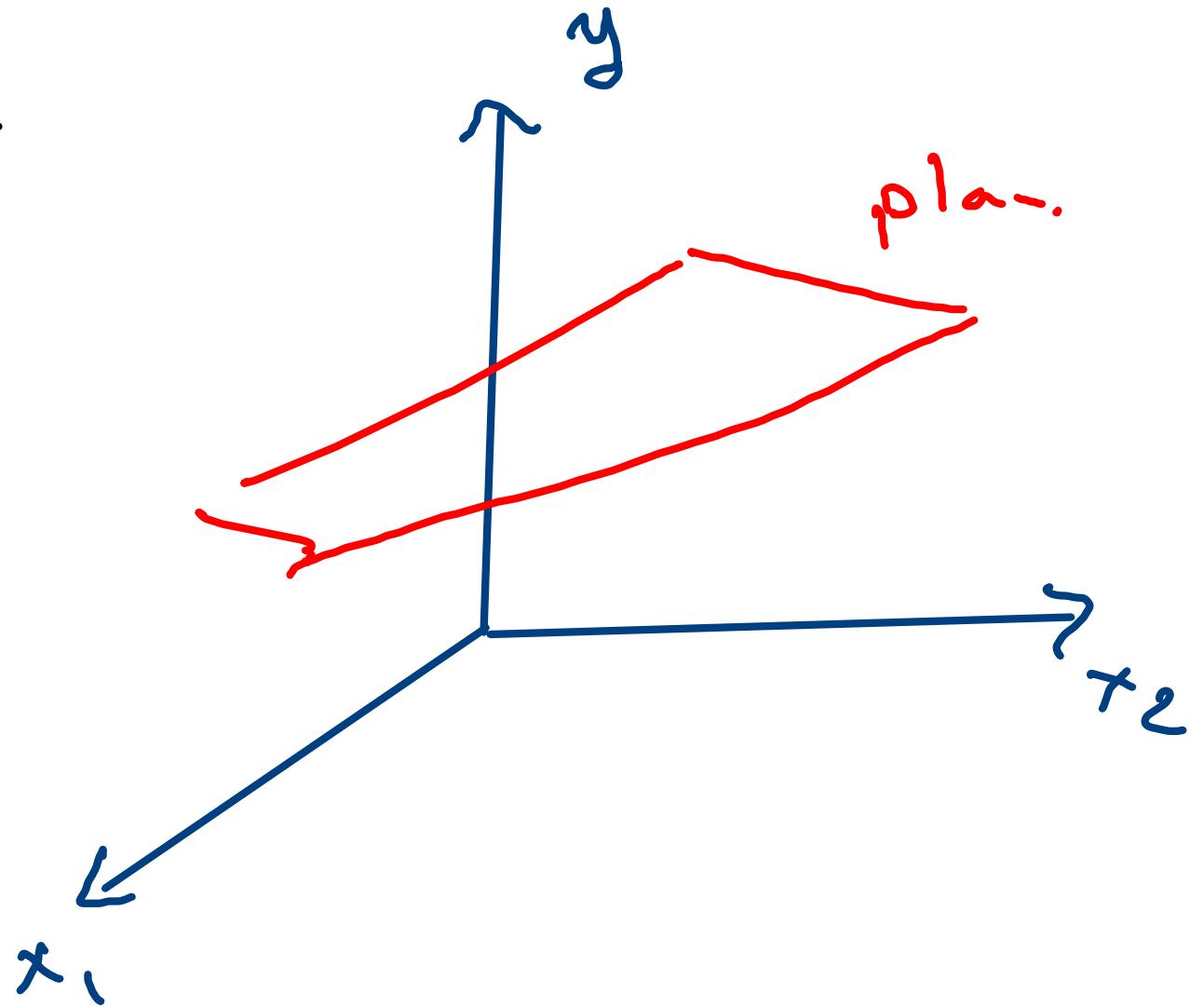
- y_{fitting} (red)
- y_{pred}

Transse b min fcf los:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_{xi}^{\text{fit}} - y_i^{\text{pred}})^2 \rightarrow \min.$$

$$\Rightarrow \hat{\beta}_0 \quad \hat{\beta}_1$$

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$



Algo: utilise des opérations
arithmétq. $\Rightarrow X$ NUMÉRIQUE

Qu'enfin: Une partie des X est catégorielle
 \Rightarrow Transf en numérique: ENCODAGE

χ^2 genre

genre_nurig.

	0	2
Male	0	2
Female	1	5
Income	2	15

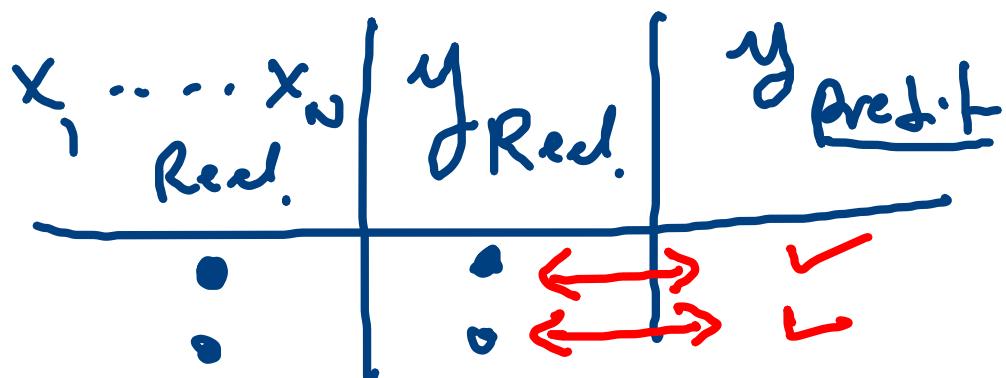
Test et Evaluation:

Evaluer les métriques du modèle (utilise le data de test)

Regressio: y continu.

1) Erreur (Distance)

RMSE: ? refence: 0



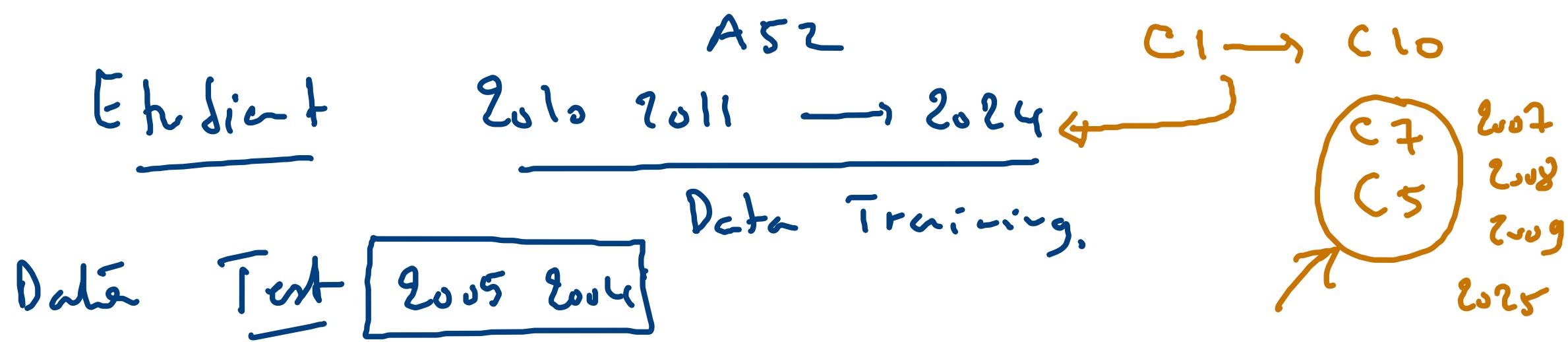
2-) R-squared: $0 \xrightarrow{70\%} 1$

R^2 ?

45% Train
b
Prof

95%

Capacité du modèle à approcher la variance contenue dans le data.
(Résistance à la variation dans le data de Prof)



Prod (joue Exem)

3/4 Questions sont similaires
en partie aux Ex 2010 → 2024.

1/4 Questions sont dans les examens
2025 2009, 2008, 2007

Niveau Confiance 60% ? → Sans métrique.

R² 98%

Pratique RL

- 1) LR processus simple. ✓
- 2-) PDF Regression logistique avec EDA ✓
- ✓ 3-) RL avec données incomplètes: techniques simple de correction

Entreprise de Télécom 100 000 x 60 \$. 20 / 80.000
client arrête l'abonnement → Service clients.

Objectif: Identifier les clients à risque, C1 oui
⇒ EVITER leur départ !! .

Modèle 1: Prédire les clients
qui vont quitter!

QUI
—
oui
Non

Type: CLASSIFICATION

objectif: Identifier le CHURN

→ C1: oui
C2: non
;

Modèle 2: QUAND va-t-il quitter?

Type: Regression

objectif: Identifier la date

→ Cgouv: oui.

Modèle 3: Pour qui va-t-il quitter WHY

$x_1, x_2, x_3, \dots, x_n$

Pourquoi

Identifie les x qui causent prob.

Corrélation avec le CHURN.

Modèle 4: Quels Cadeaux à Proposer.

Prédire le cadeau à donner pour tous les offres dispos

Type; Classification

{ Modèle 5: Programme de Fidélisation
Prédiction d'offres à votre client
Type; classification

TP de Session

objectif: Développer un modèle de prédiction.
—
Regression Classification.

Data: Validation du Data 

- Etapes:
- 1-) Apprendre à faire un EDA Sommaire
 - 2-) Faire des opérations de pré-traitement
Buniques.
 - 3-) Utiliser 3 Algos du Cours
 - 4-) Évaluer le modèle.
 - 5-) Interpréter et Comparer les modèles
 - 6-) Présenter votre travail

EDA

Summary:

- Describe du Data
- Determiner quelles sont les series numériques et Catégorielles,
- Visualiser les relations entre x_i et y .
- Corrélation. (Matrice de Corrélation).
- Déterminer les valeurs manquantes.
- Déterminer les valeurs aberrantes.
- Encodage à faire.
- etc ...

