

9 Sept 2025

Intro A52
Algorithmes d'apprentissage
supervisé

- 1-) Présentation du plan de cours.
- 2-) Workflow (Processus) pour le Dev Modèle.
- - 3-) Intro : Régression Linéaire .

Processus de Dev Modèle (Prediction).

1-) Statistique: Estimation d'un paramètre
d'une population.

2-) Data Scientist: Prediction \rightarrow Modèle

sur la base 
utiliser des Données

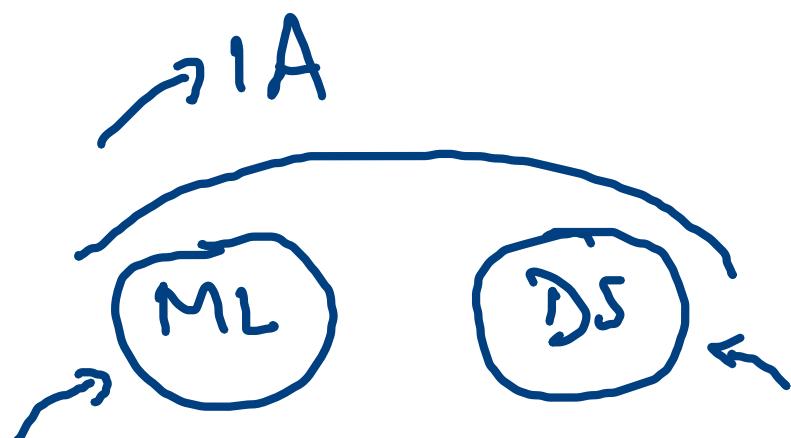
Buzz :

Terminologie

FRAUDEURS

1-) Intelligence Artificielle

2-) Machine Learning
(Apprentissage automatique)



3-) Data Science.
(Sciences de données)

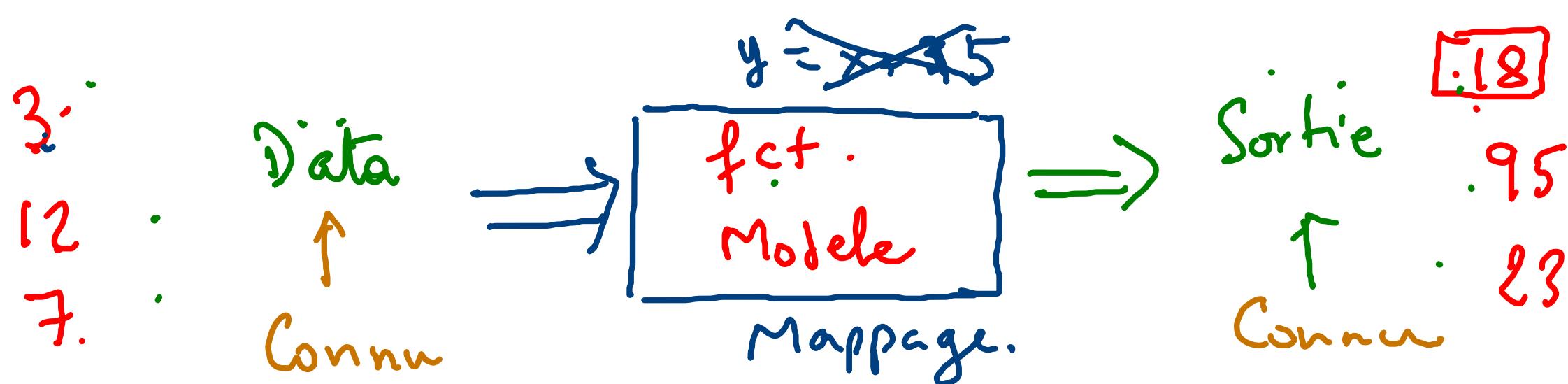
1 - Intelligence Artificielle.

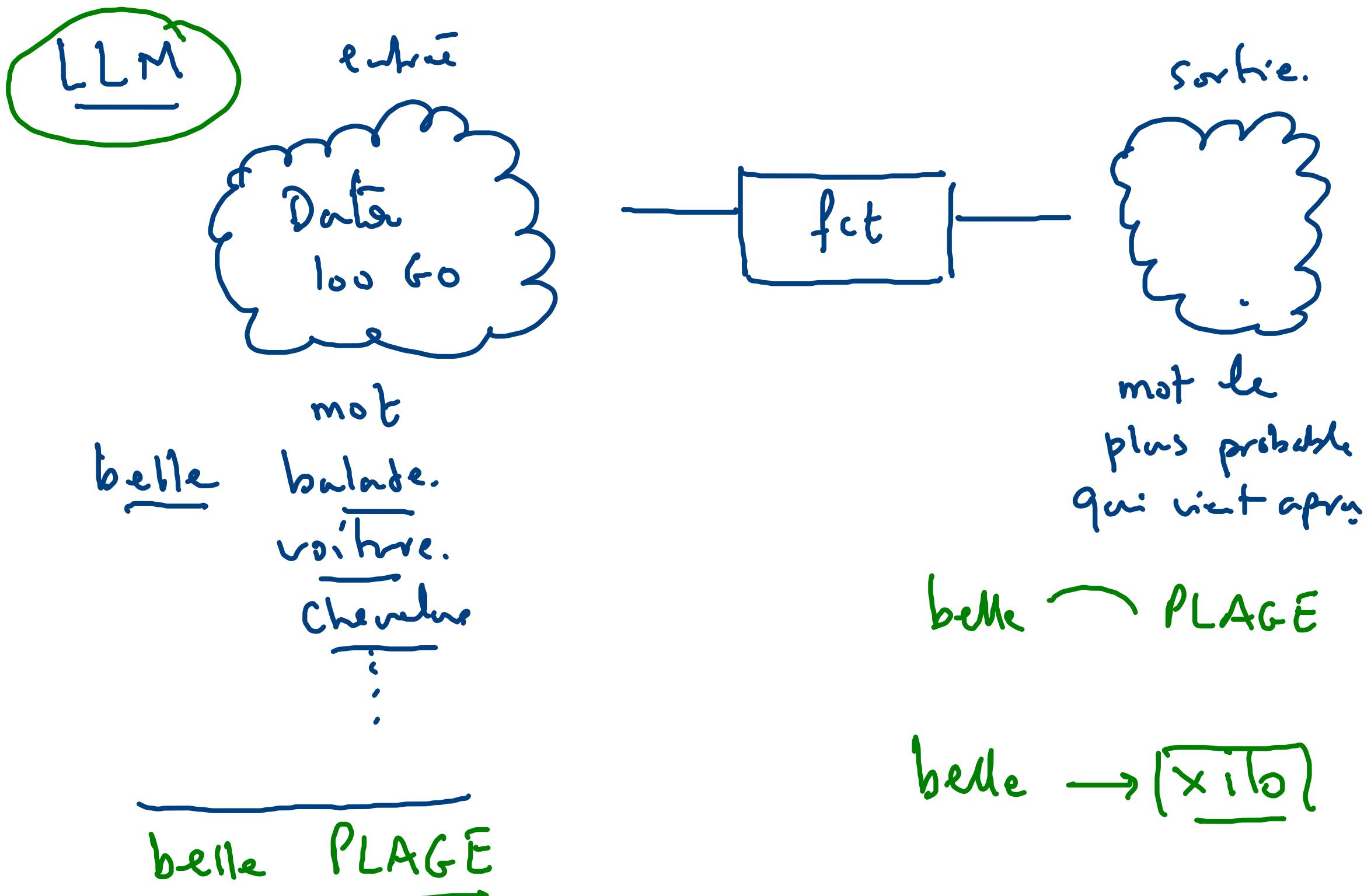
C'est la CONCEPTION et le DEVELOPPEMENT d'agents intelligents (IA Agent) [processus informatiques / processus biologique] qui sont capables d'APPRENDRE (par eux m^{es}), d'INTERAGIR avec leur environnement (avec des capteurs) et de PRENDRE des decisions.

Compétence. Développer des Agents IA.
n8n Make

2-) Machine learning (ML)

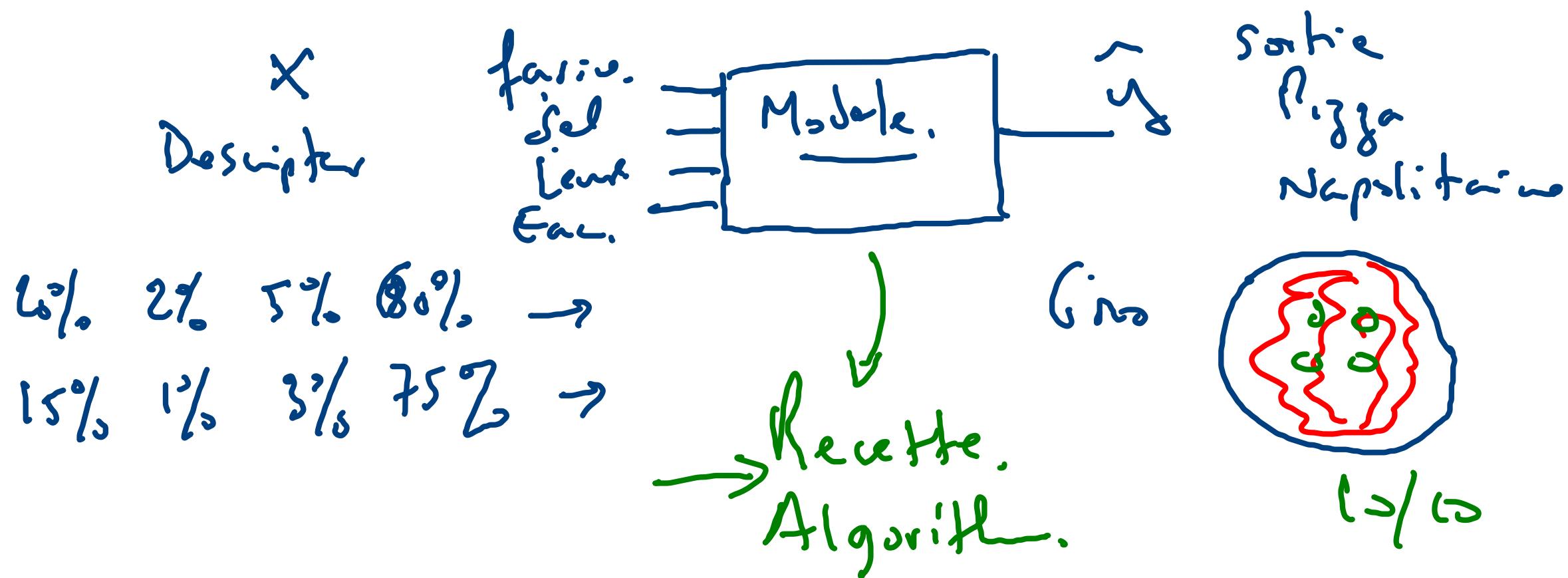
- Développer un processus informatique.
- Prend des données en entrée.
- Apprend sur la base de ces données (incluant les sorties correspondantes aux entrées).
- Prédit sur la base de son apprenant.





3-) Data Scientist

- Développer le workflow (processus scientifique et Technique) qui permet d'arriver à un modèle en exploitant les données.



- 1-) Comprendre le processus Data Science.
- 2-) Bonne maîtrise des Algorithmes principaux (théorique). ✓
-
- 3-) Bonne maîtrise de l'implémentation des algorithmes, dans le sens de modèles.

Role.

1-) Spécialiste en Machine Learning:

- Très Bonne connaissance des algo's ML
- Implementation (Dev. Modèle).

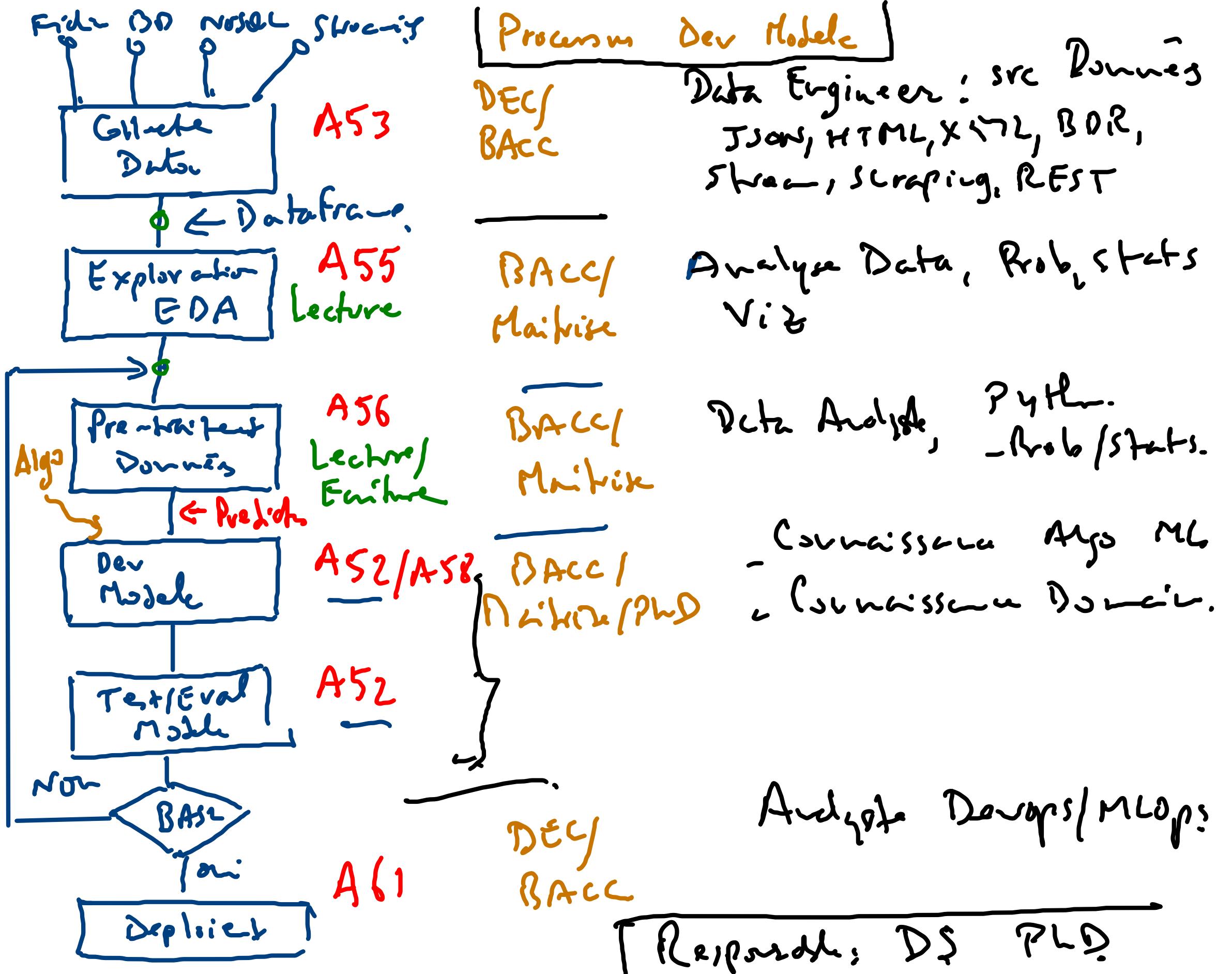
2-) Data Scientit



- Très Bonne connaissance du workflow
(proc de dev du modèle)
- Evaluation et Mise en prod du modèle.

3-) Scientifique/ Spécialiste en Algo ML

- Dev d'algorithmes.
- Tuning d'algorithmes.



Comment réussir dans le processus DS ?

- Garder une trace de tout ce qui est expérimenté pendant le Dev du Modèle.
Jupyter NoteBook → Documentation + Code.
- Rigoureux au niveau de chaque étape. !!
- Questionner !!
- Esprit Ouvert

Salaire.

1000 Fr/heure.

Collecte : 80 - 100 Fr/heure

EDA : 60 - 70 Fr/heure.

Pré-traitement : 80 - 120 Fr/heure.

Modèle / proc : 120 - 200 Fr/heure.

→ Collecte

Minerai
(Ov)

EDA

Expert
Minerai

→ Pré-traitement

Lingot
Ov.

→ Dev
Modèle

Bijou
(orfèvre)

Techique:

- { - Statistiques. EDA, pre-traitement.
- { - Probabilités
- { - Algèbre linéaire
- { - Calcul Matriciel.

Python : {

- sklearn -
- pandas.
- numpy .
- matplotlib / seaborn -
- statslib

Apprentissage ML.

A52

Supervisé

- Training
- Test

Cible/ Target

- Continue : Regression
- Discrete : Classification

Range vert Bleu

A58

Non Supervisé

Par étape de
Training

Découvrir des
tendances / Patterns

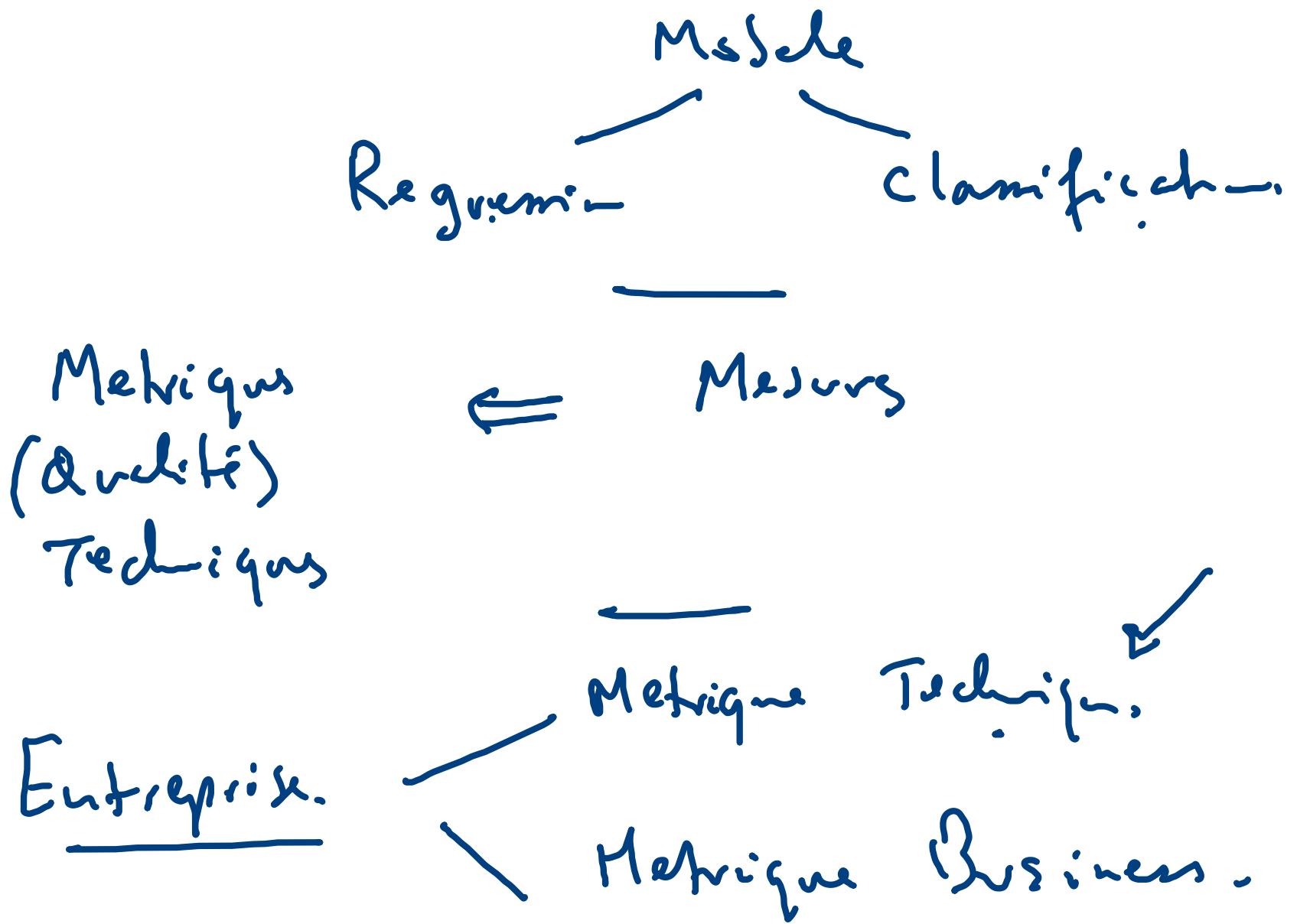
- Clustering

Non supervisé →
une étape préliminaire
pour l'apprentissage
supervisé

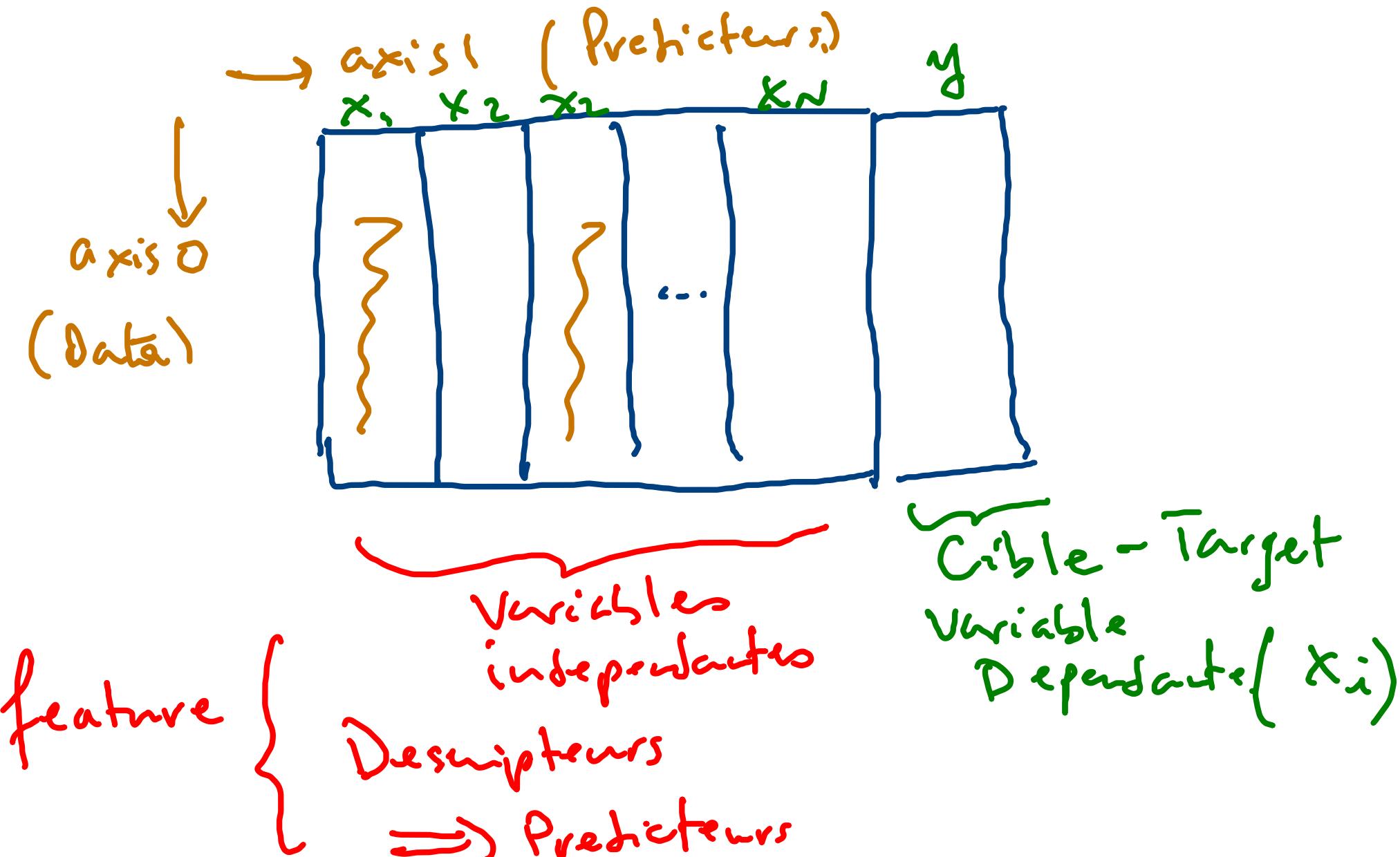
A59

Pas de
réforcement

Approche
Basé sur
les
Rewards
(Bonnes!)



Data se présente sous forme Tabulaire
DataFrame.



Etudiants

ID, nom, prénom, cheveux, DN, Poids, taille, Note., QI, Précédent, [y]

Io Flavio Ma
— —
X X X

Descripteurs
X

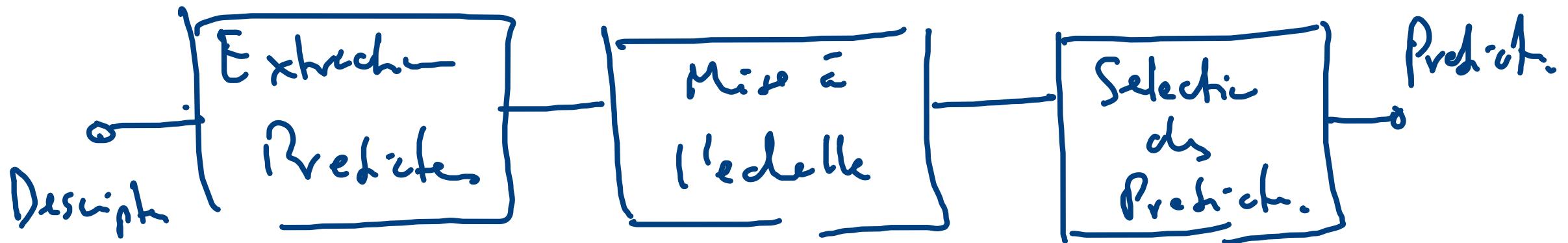
Note A 52

[y]



Predicteurs { DN
Note Pre., → Modèle →

y
Note
finale



M : Continue

17-62

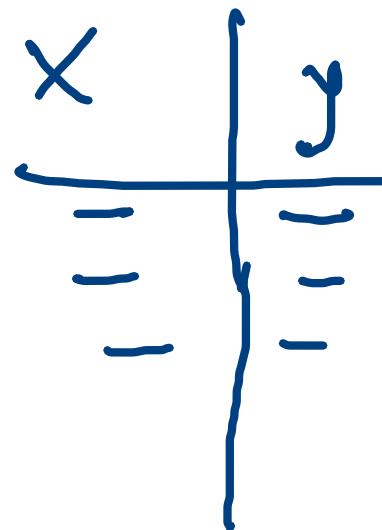
2500-3

Regressin

- Regression Linéaire. ✓
- Regression Polynomiale.
- Regression avec fact de Bnx.
- Regularisation.
 - Ridge ✓
 - Lasso ✓
 - Elastic Net

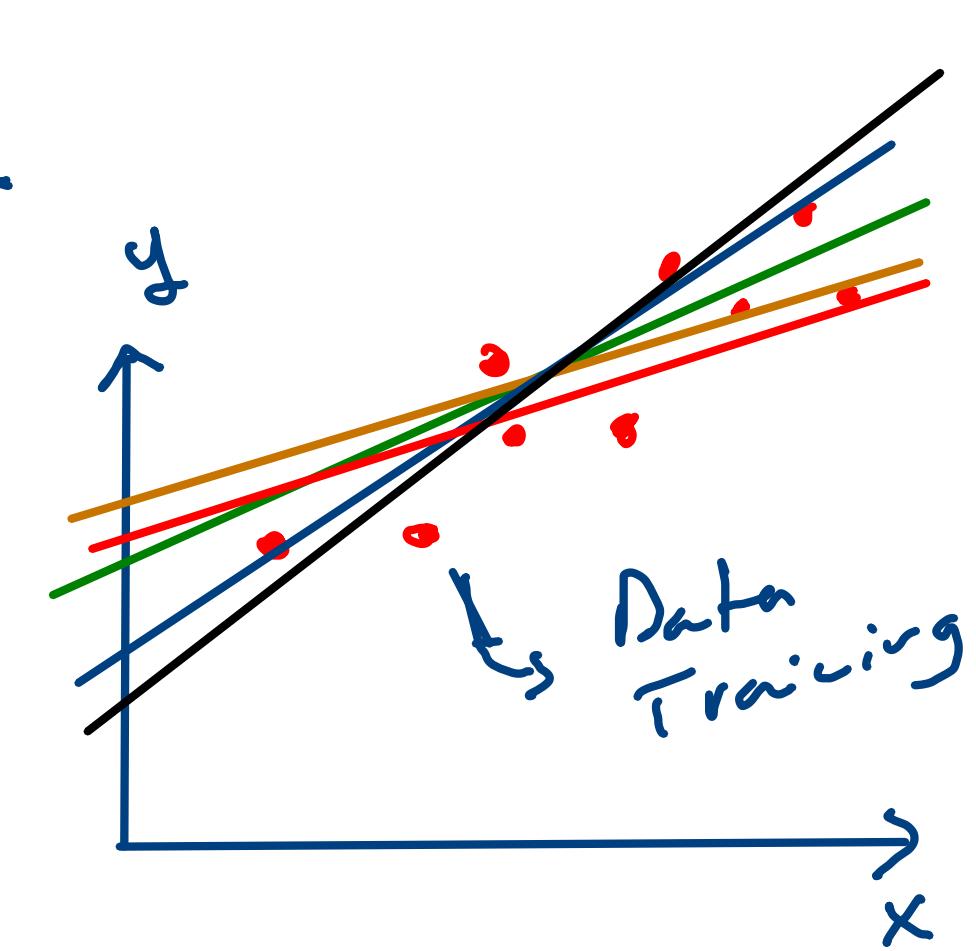
Univariée.

$$y = \beta_0 + \beta_1 x$$



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

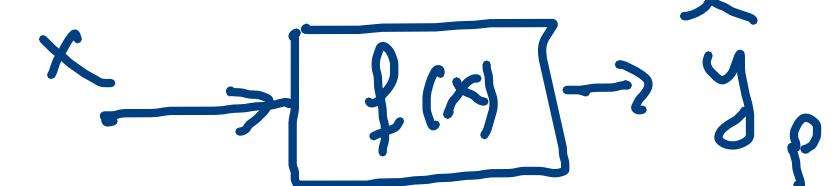
optimal



objectif : Trouve $\hat{\beta}_0$ et $\hat{\beta}_1$
coeff du modèle

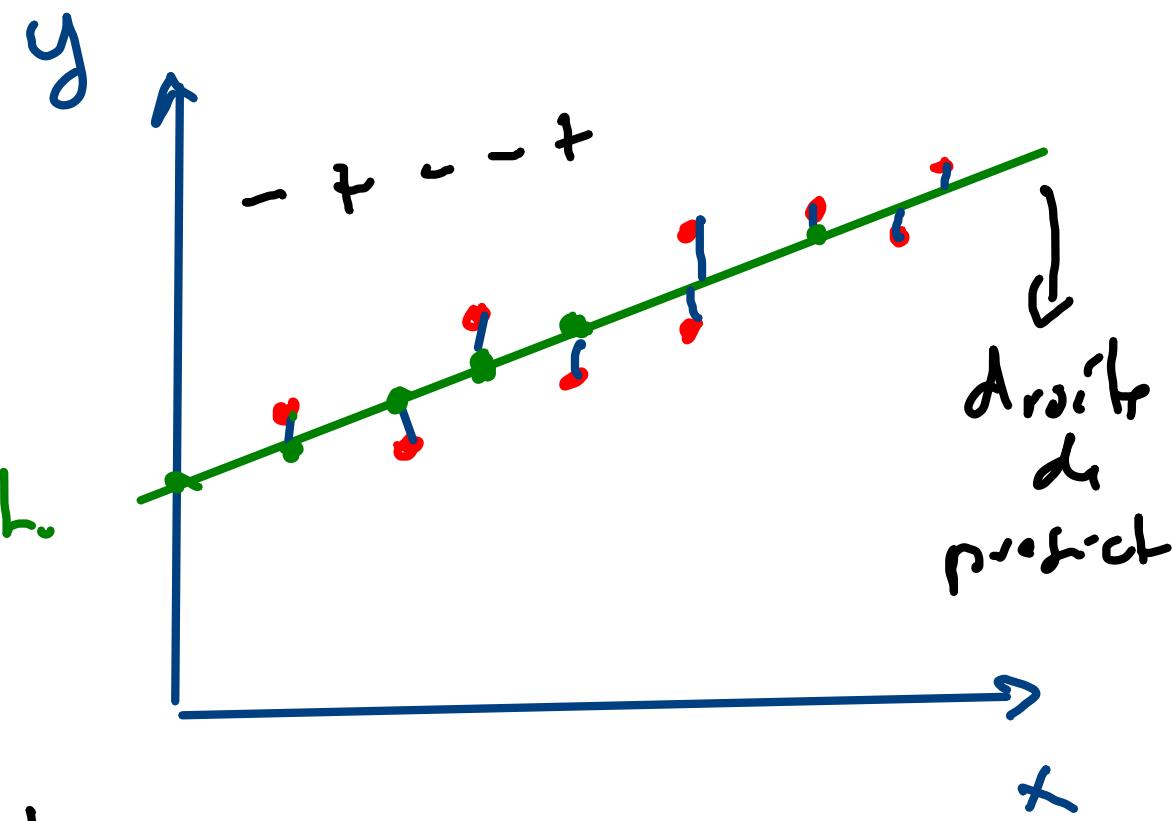
{ valeurs estimées à partir
du Data !!

Estimateurs



• Data Training Red

- vert: Droite Regres-
valeur prédictive.
- ⇒ distance: erreur prédictive



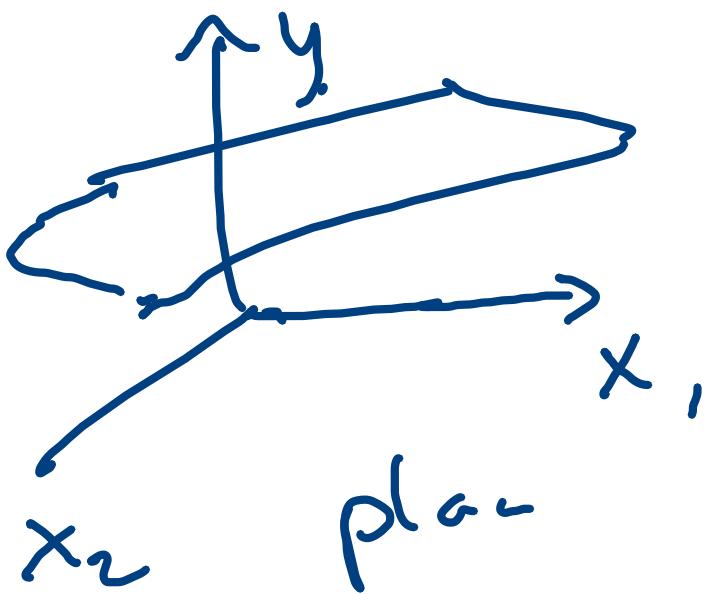
$$\sum_{i=1}^N e_i = \sum_{i=1}^N (y_i^r - y_i^p) \rightarrow \text{minimiser la somme}$$

$$\frac{1}{N} \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (y_i^r - y_i^p)^2 \leftarrow \begin{array}{l} \text{fct loss} \\ \text{Minimiser} \end{array}$$

$$\hat{y} : y_i^p \\ = \hat{\beta}_0 + \hat{\beta}_1 x$$

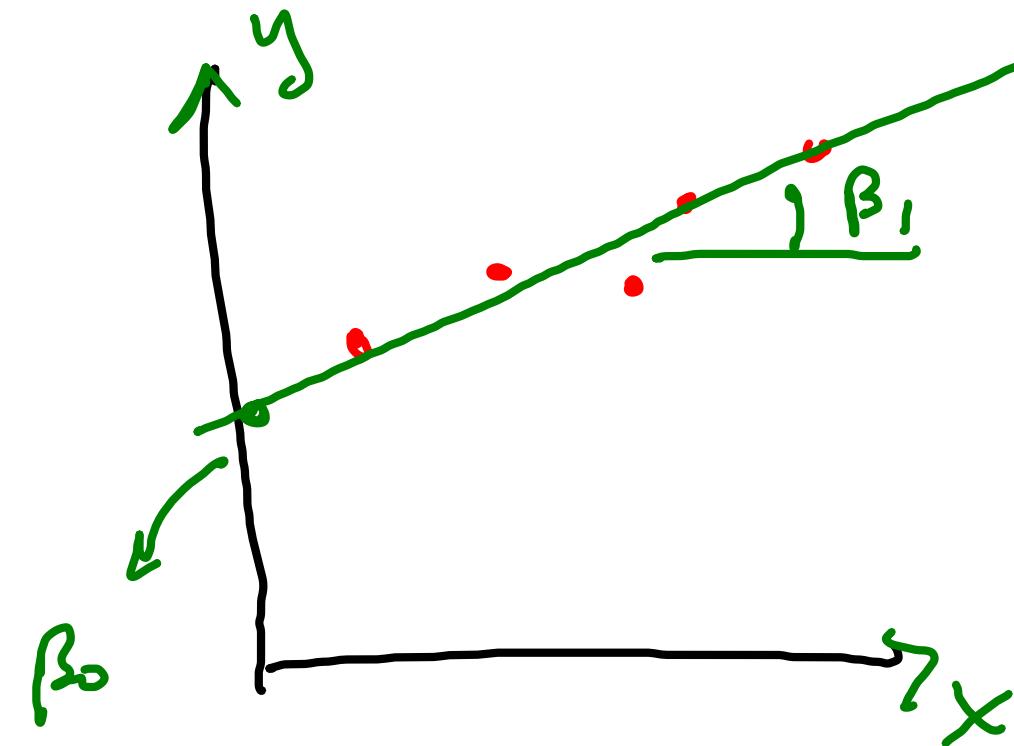
Training $\leftarrow \hat{\beta}_0 \text{ et } \hat{\beta}_1$

Extending in planar Dimensions



.....

Hyperplane



Demande d'un process R.L

1 -> Envir / Collecte Data.

2 -> Évaluer la data

3 -> Separer Data en Training / Test.

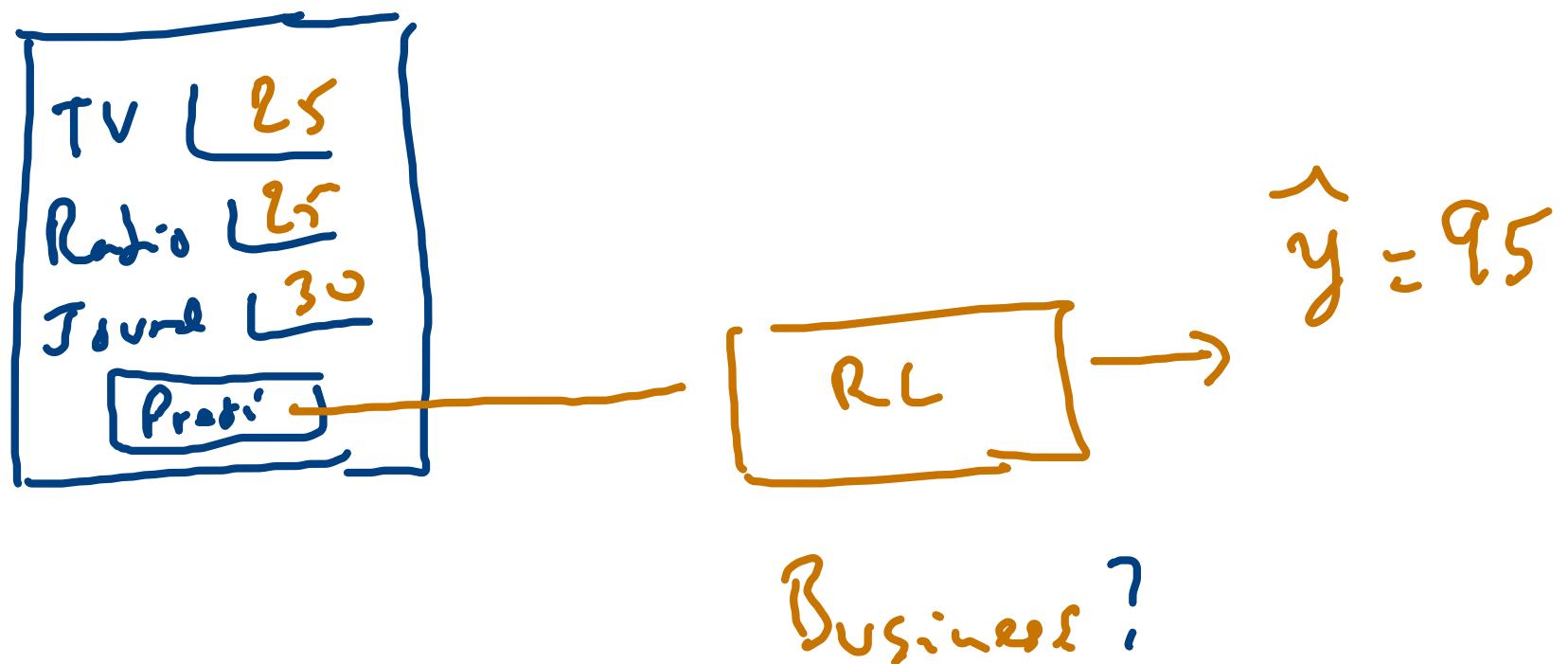
4) Dev Model.

5) Évaluer le modèle

- RMSE \rightarrow 0

- R-squared $0 \rightarrow 1$
Mauris Boz.

Input	TV	Radio	Journal	y Revenue
F13	Cust	Tv		
	10	30	50	80
	100	125	75	15
	:	:	:	:



Searc2 : TP Session

- Open Data.

- Domaine

- KAGGLE

- Thème (Besoin).

- Équipe: 2 ou 3 max.

Santé : . Prédire l'admission Urgee.

- Prédire les spécialistes

CHUAN.

Servis :- Prédire qui sont les clients qui vont quitter.

Assurus - Prédire à quel moment ils vont quitter

Support Technique

Predict le nombre de
Billets avec un certain (Urgent
niveau)
Niveau :
Urgent
Normal

Université

Predict les spécialistes qui
sont les plus demandés

etc.,