30 Sept 2025

1-) Retour sur Dev Modèle. ✓

2-) Trade off Biais-variance. ✓

3-) PCA - ACL ✓

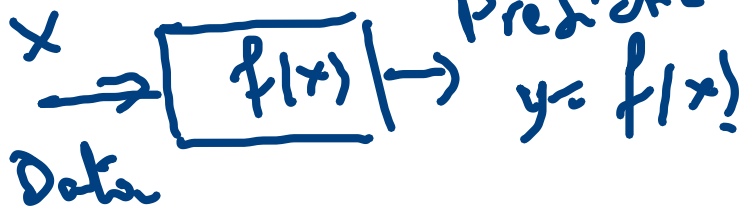
---

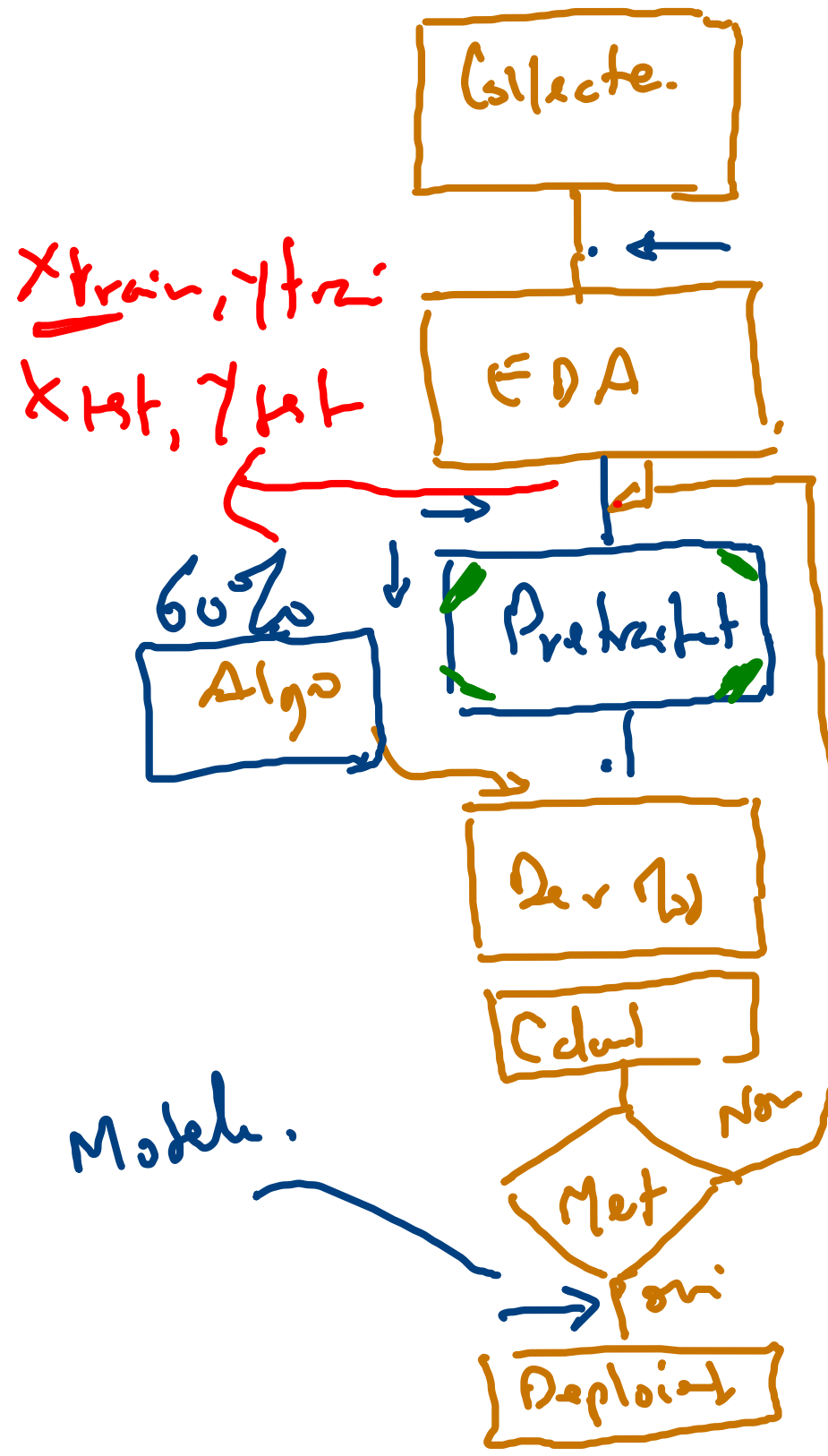
4-) TP progression

# Algos Apprentissage Supervisé

## Algo ≠ Modèle

Algo
- Prédiction (Théorie)
- Implémentation (Pratique)



# A-S

Approche Base

Regression — RL

Classification — KNN
DT
NV
...

CLASSIFICAT

Regression.

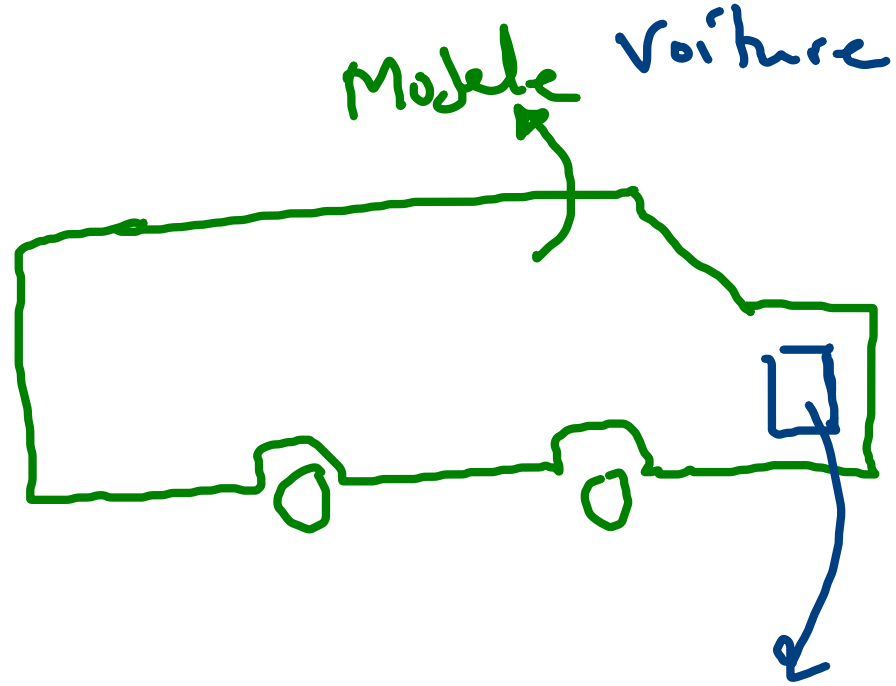
Regularisation — Ridge.
— Lasso.

Approche Ensemble.

Random Forest

Bagging

Boosting




Test Route.

Voiture
- Transport Quotidien
- Sport
- Tourisme
- ⋮

Algo Moteur
- 6 cv
- 12 cv
- 24 cv
- Ev
- ⋮

Notebook.

```
df = pd.read_csv('adresse....csv')
:
:
```

—

```
df.describe()
df.corr()
```

} EDA.

—

```
df['age'] ....
```

} prétraitement

—

Développement modèle.

```
clr = LinearRegression()
clr.fit(----)
```

—

. mse. Metrics

clr ← modèle.

clr.predict(—)

tout.write(clr) ↙ fichier Binaire

$X_{Mac}$

$X_{10}$ (Age)

Iteration 1

$n_{an} \rightarrow X_{10}$ (mac)

imputing

25

35

→ R

40

40

moy =

$Clr_1 \rightarrow$

25
35
35
40
40

Met1
$R^2$
= 79%

iteration 2 méthode :

25 35 40 40

37.5

Dist Prob

Gauss

25
35
37.5 →
40
40

Met2

$R^2 = 82\%$

35
36
36
36
38

109
3
36

$R^2$ 95%

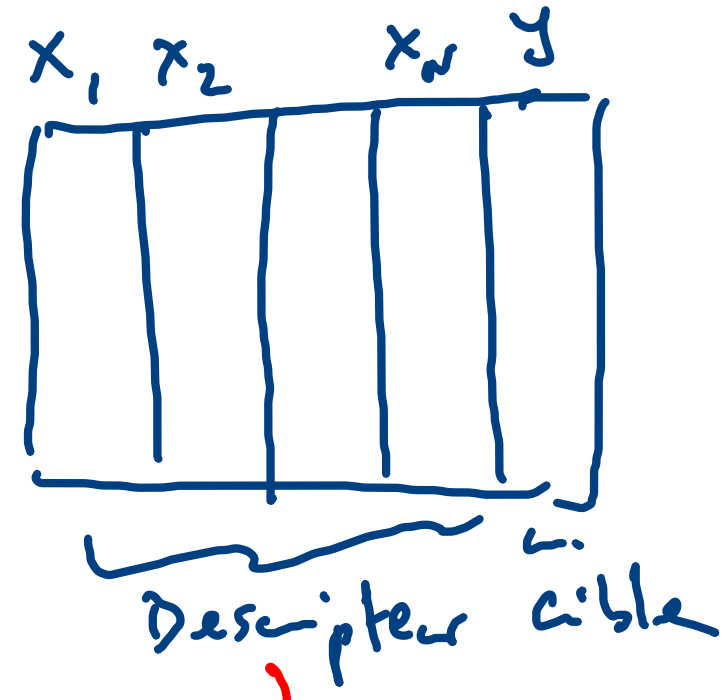
predire $y$

Collecte $x_1 \dots x_N$

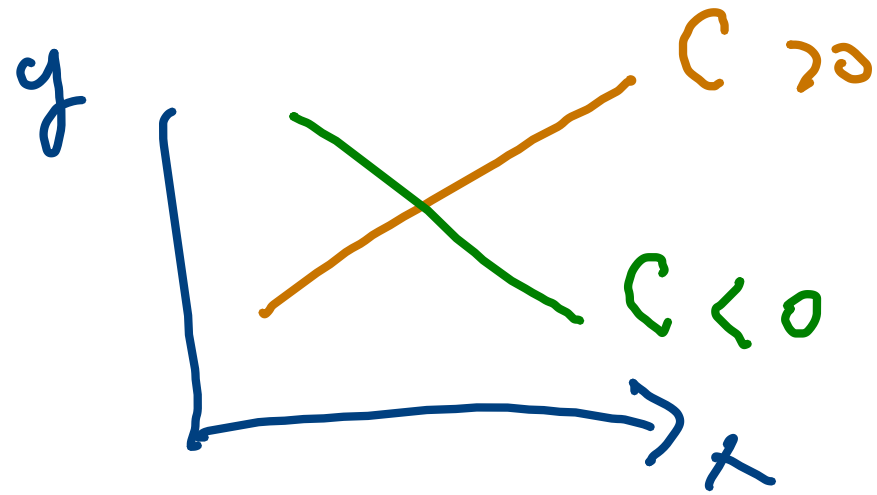
Influence sur $y$ $\Rightarrow$ Predicteurs

$(X_i, y)_{i=1,\dots N} \xrightarrow[1]{\text{Corr}}$ Influence ?

$(X_i, X_j) \xrightarrow[0]{\text{Corr}}$ ?




Redicteurs

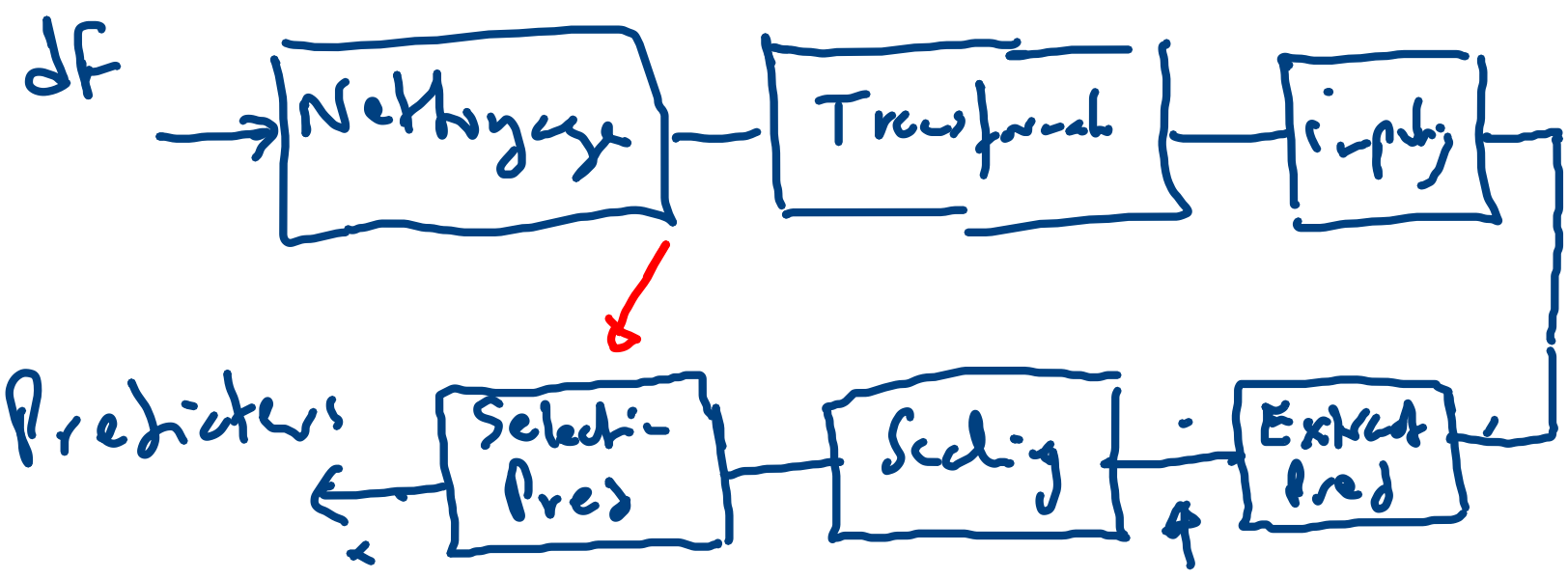


→ Combien de prédicteurs sont nécessaires } Prétraitement

→ Descripteurs

→ Générer des prédicteurs additionnels

⇒ Feature Engineering




Action  Prix_début  Prix_fin

FE ← $x_{14} = \frac{Prix\_fin - Prix\_début}{Prix\_fin + Prix\_début}$

axis 1

$X_{12}$ $X_{13}$ $X_{15}$ — y

axis 0

Skewness - Kurtosis

v.A

0

Descriptors — Predict

$X_{12}$ $X_{15}$ — Pred

$X_{15}$ — Predict

v. observ.

QI [500]

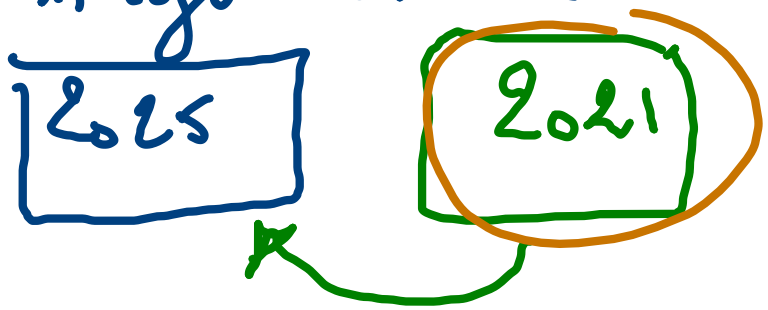
| | $E_1$ | $E_2$ | $E_{100}$ |
|---|---|---|---|
| Nb Her | 3 | 5 | 7 |

Nbre find
y

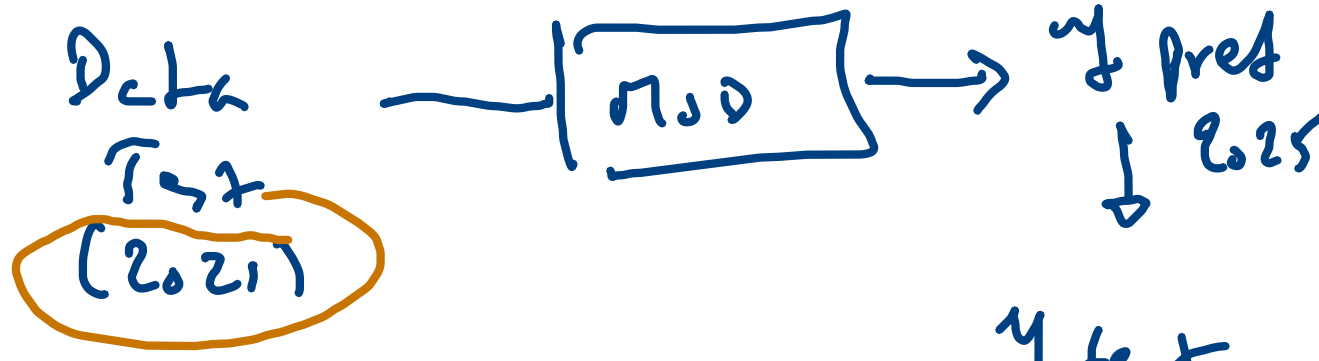
Corrélation ≠ Causalité

**Objectif** : Sachant les $x_i$, prédire le $y$.

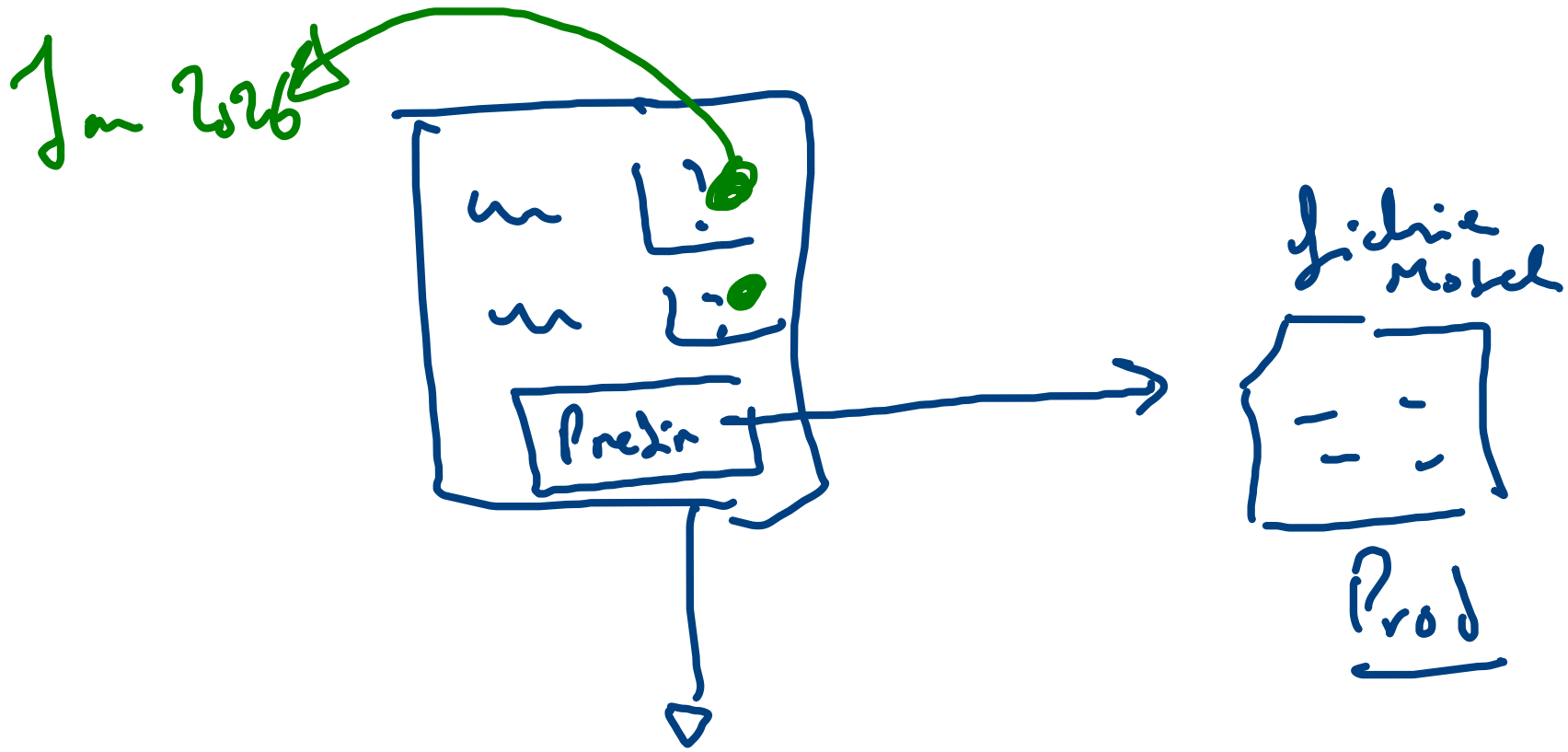
**Phase de Training** : Développer le modèle en utilisant un Algo et la Data Training



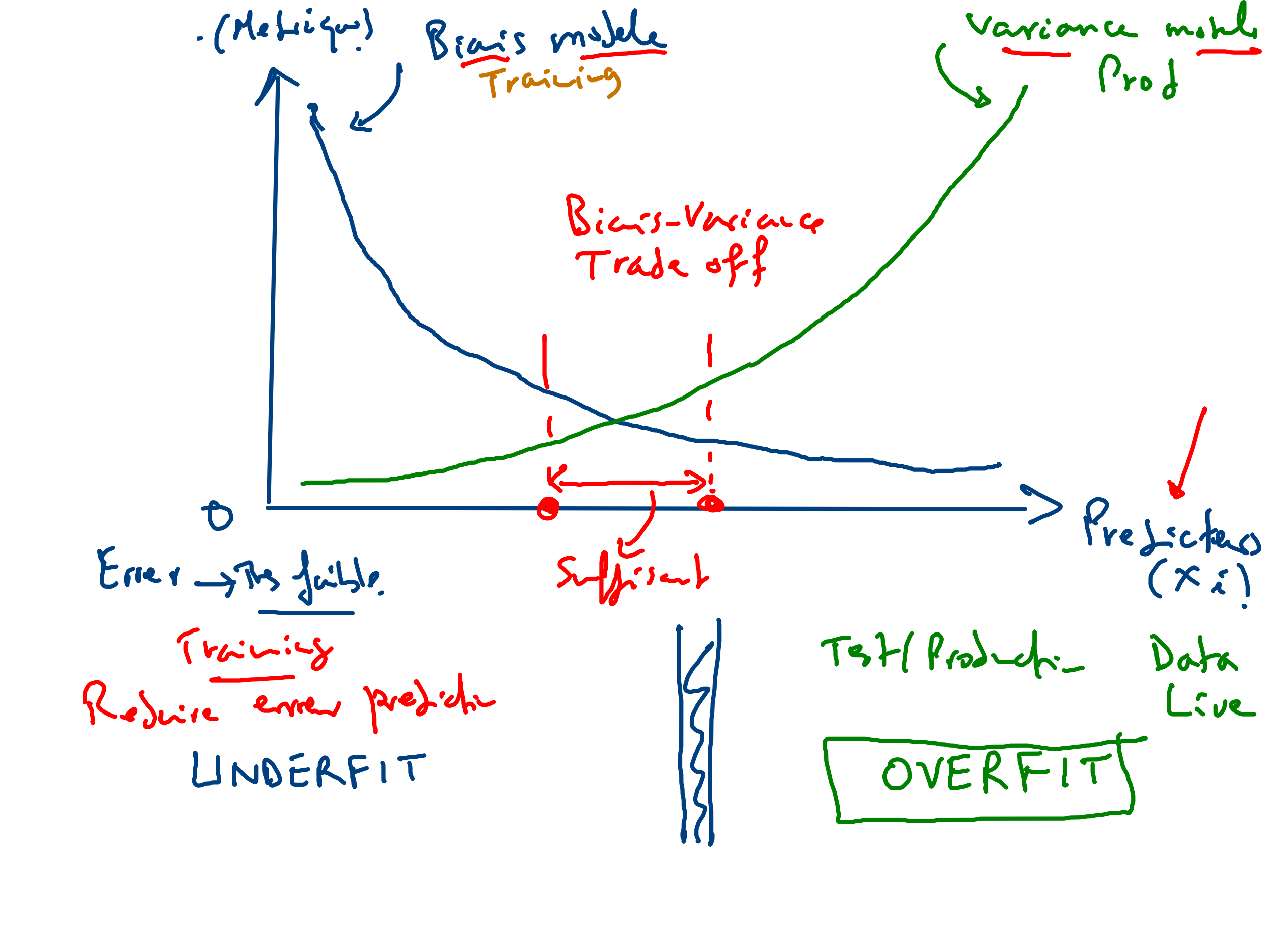

**Phase de Test** :



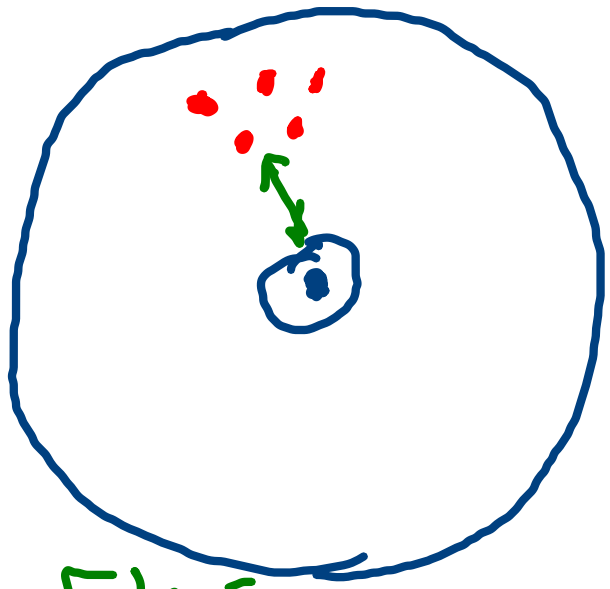

**Doc 2025** ⇒ Modèle Correct ⇒ Déploiement


Jan 2026
Predict
Live Model
Prod


Result

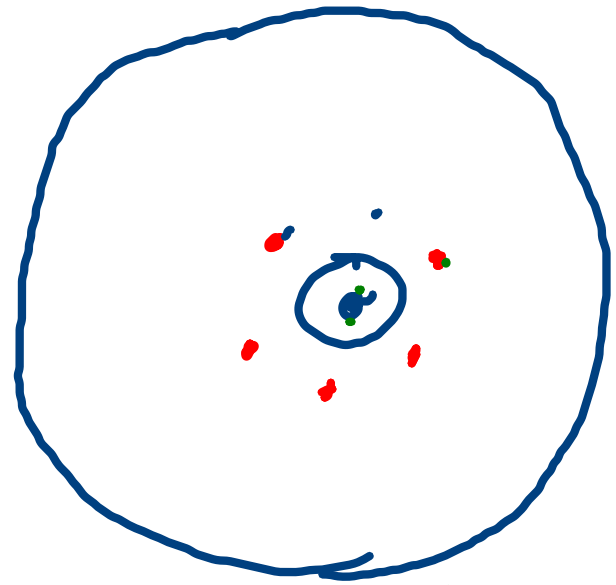

Erreur (Metrique)
Biais modele
Training
Variance modele
Prod
Biais-Variance
Trade off
0
Predictors
(Xi)
Error → Too faible.
Training
Reduire erreur prediction
UNDERFIT
Sufficient
Test/Production
Data
Live
OVERFIT

- $y$ réel.
- prédiction



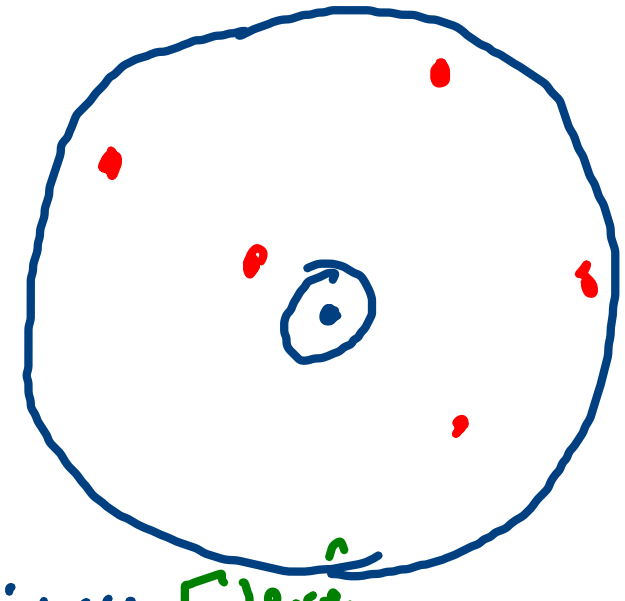
1

Biais : Elevé
Variance : Faible



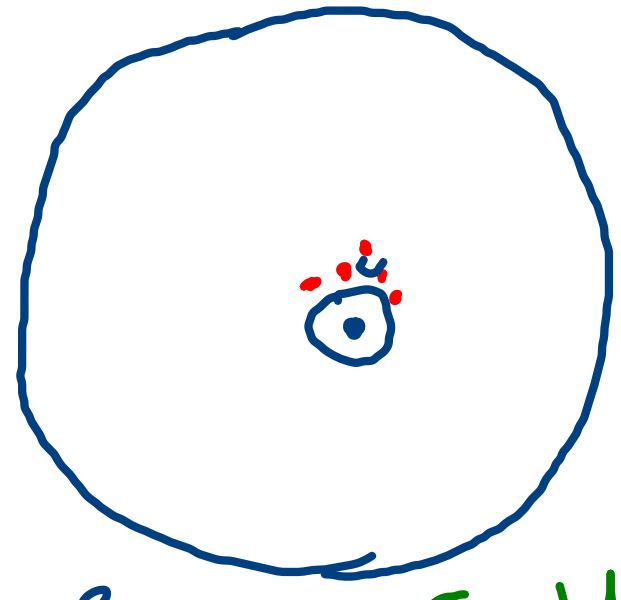
2

Biais : Faible
Variance : Elevé



3

Biais : Elevé
Variance : Elevé



4

Biais : Faible
Variance : Faible.

# Analyse de Composants principales (ACP)
(P.C.A)

~~ID~~ $X_1$ $X_2$ ~~$X_3$~~ $X_4 \dots X_{10}$ $y \leftarrow$ note_finale

$X_1$ : QI
$X_2$ : Note Extra
~~$X_3$~~ : Poids
~~$X_4$~~ : Taille
$X_4 \dots X_{10}$ : Note TP

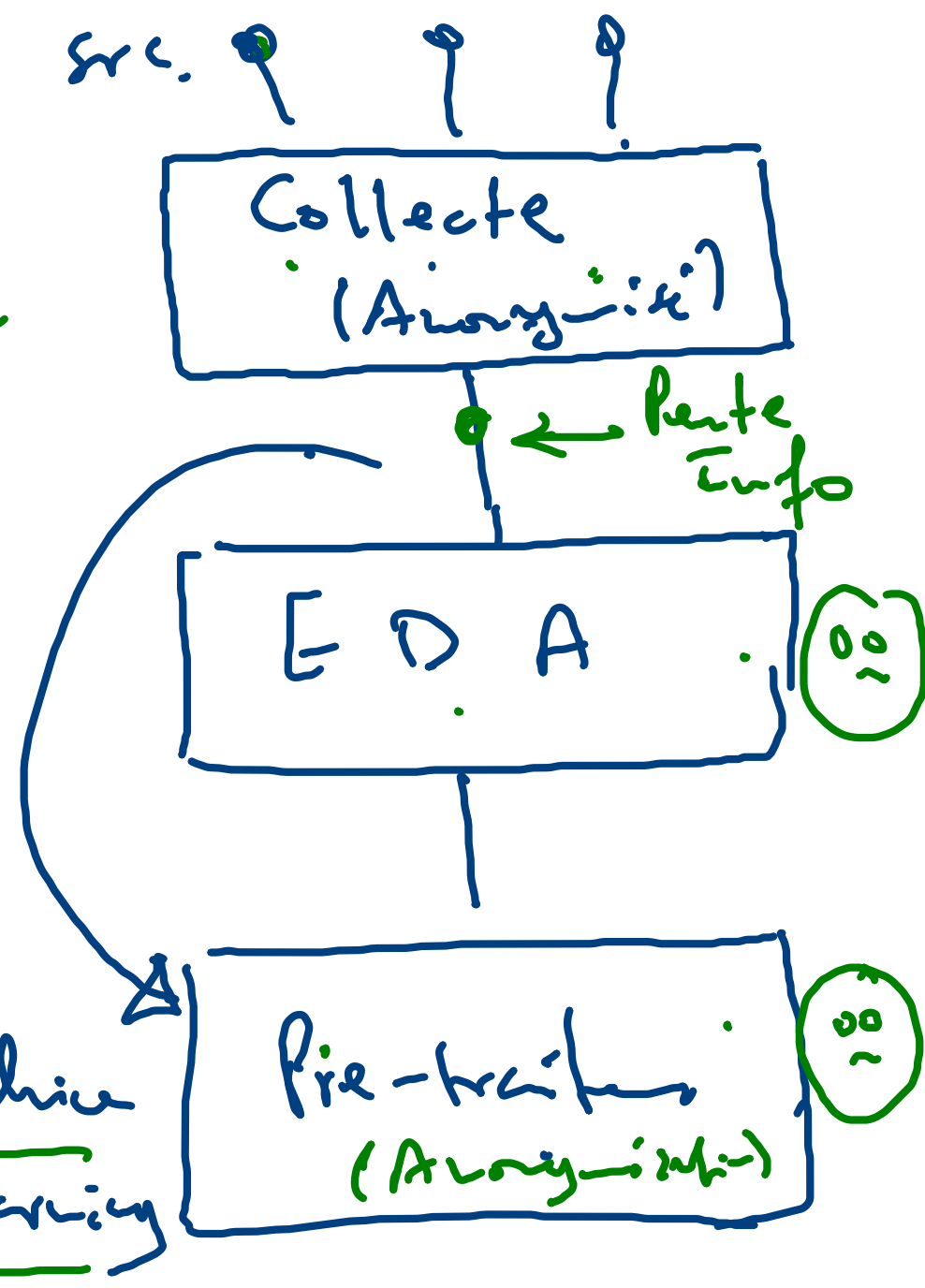
1-) Eliminer les $X_i$ qui n'ont pas d'influence $y$.

2-) Diminuer le nombre de prédicteurs en effectuant une orthogonalisation.

1

Anonymisation ← Privé ← Data Engineer

Réunion Coordination



$x_1$ — $Q_I$

$x_2$ — Nulle — Infini

$x_{10}$ — Nulle — Fini

Représente la même entité physique pour $x_2$ et $x_{10}$

$Z_2$

Combiner

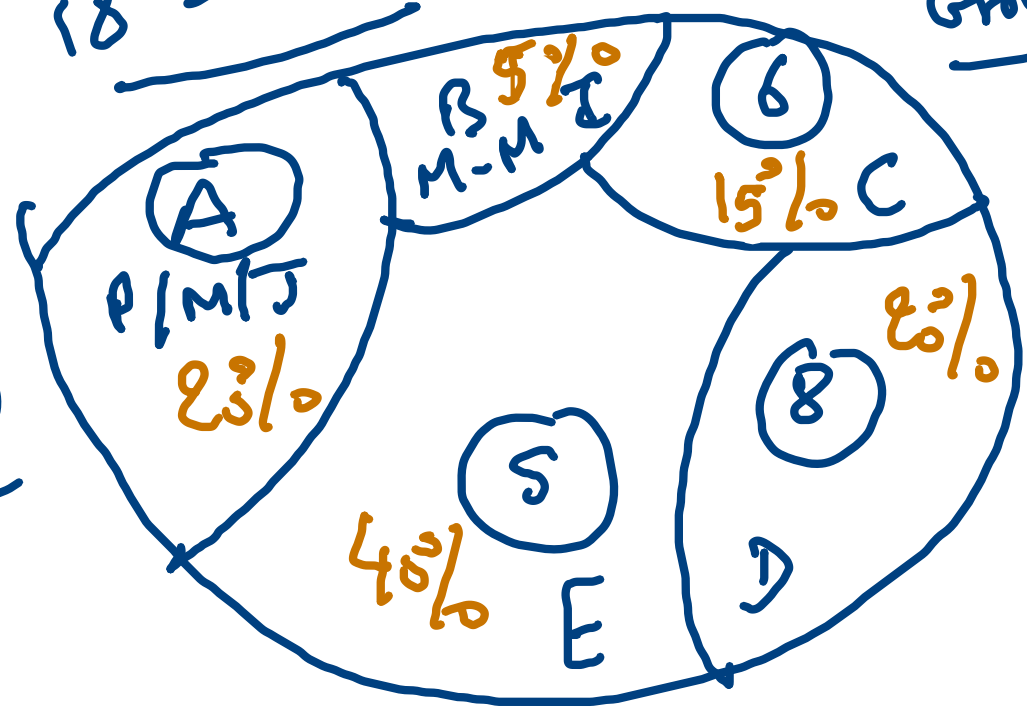
$Z_2$ : n'a pas d'interprétation physique.

Contient l'information contenue dans $x_2$ et $x_{10}$

18H33

18H - 20H

Optional Groupe




Zoom

$\sum I$ = Total

ACP

E → 40% ✓

A → 20% ✓

D → 20 ✓

C → 15%

B → 5%

Marche :

3 Prefic → 18H - 19H

→ 19H - 20H

Panel

Maths

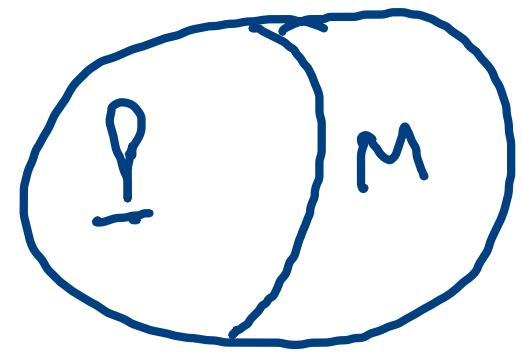
22 etudiants

Info 18H - 22H

18H - 19H Panel

19H - 20H Maths

orthogonal.



R-1 R-2 RMSE

Classif : Accuracy 95%
Precision 80%

Mise en production Modèle

Evaluation & Suivi de Performance, Drift

Data Live

Data Training

Baseline

Métiers

UNDERFIT

Outils de Mesure de Performance

Stale (Modèle) Demodé

Entrainer MAJ Modèle

Dev (ACL)

Prod

Model + Model Serving Application

30/09

CI/CD

11 Dec 2025

$X_{10}$

Training

$X_{10}$ (live)

Prod

Mesure

le changement

Chaine de traitement automatique

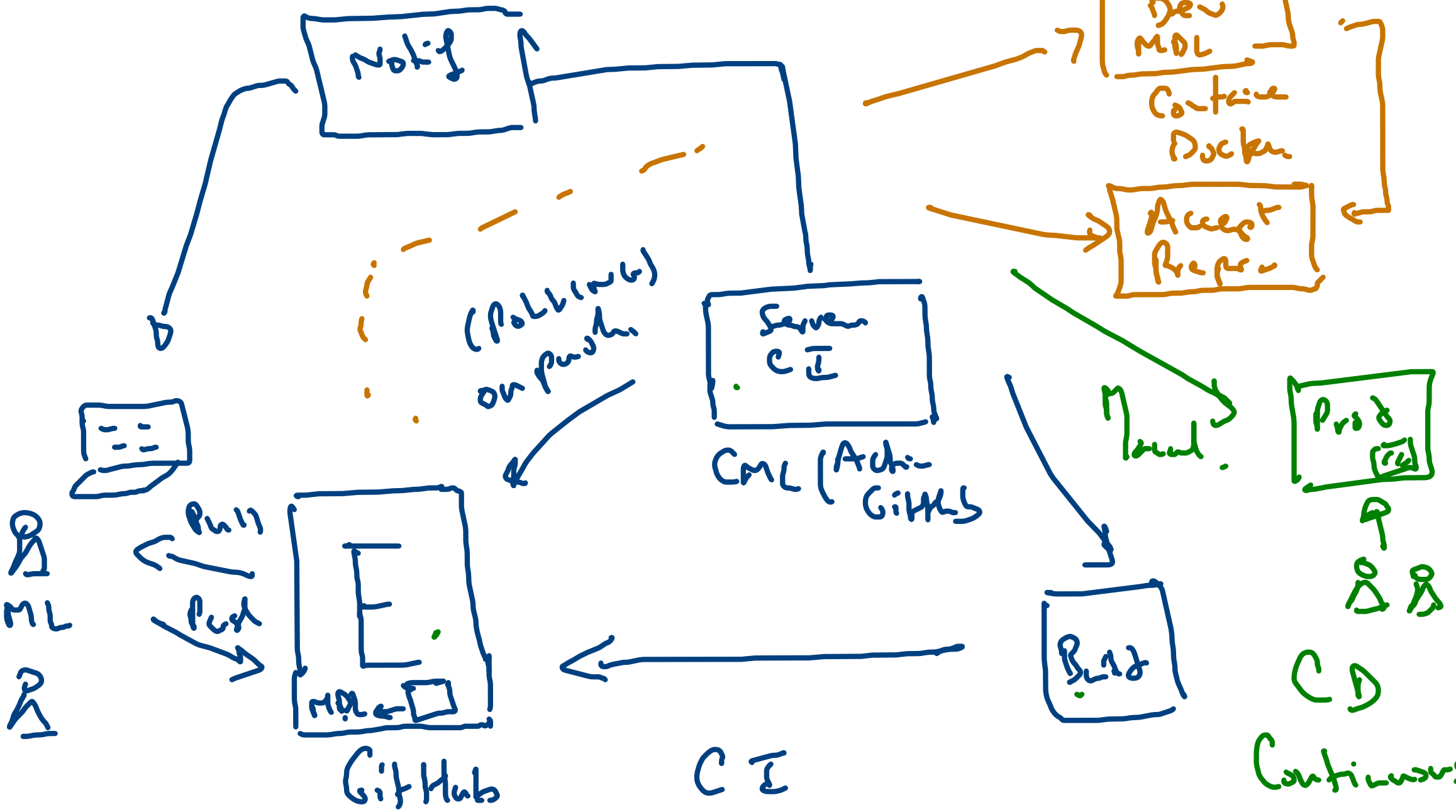
D.D.

Caractéristiques statistiques

D.D.

Caractéristiques statistiques

Risque de Drift si écart


Notif
Dev
MDL
Contains
Docker
Accept
Repo
(Polling)
on push
Server
CI
CML (Actions
GitHub)
Manual
Prod
CD
Pull
Push
ML
MDL
GitHub
Build
CI
Continuous
Integration
CI/CD
Continuous
Delivery