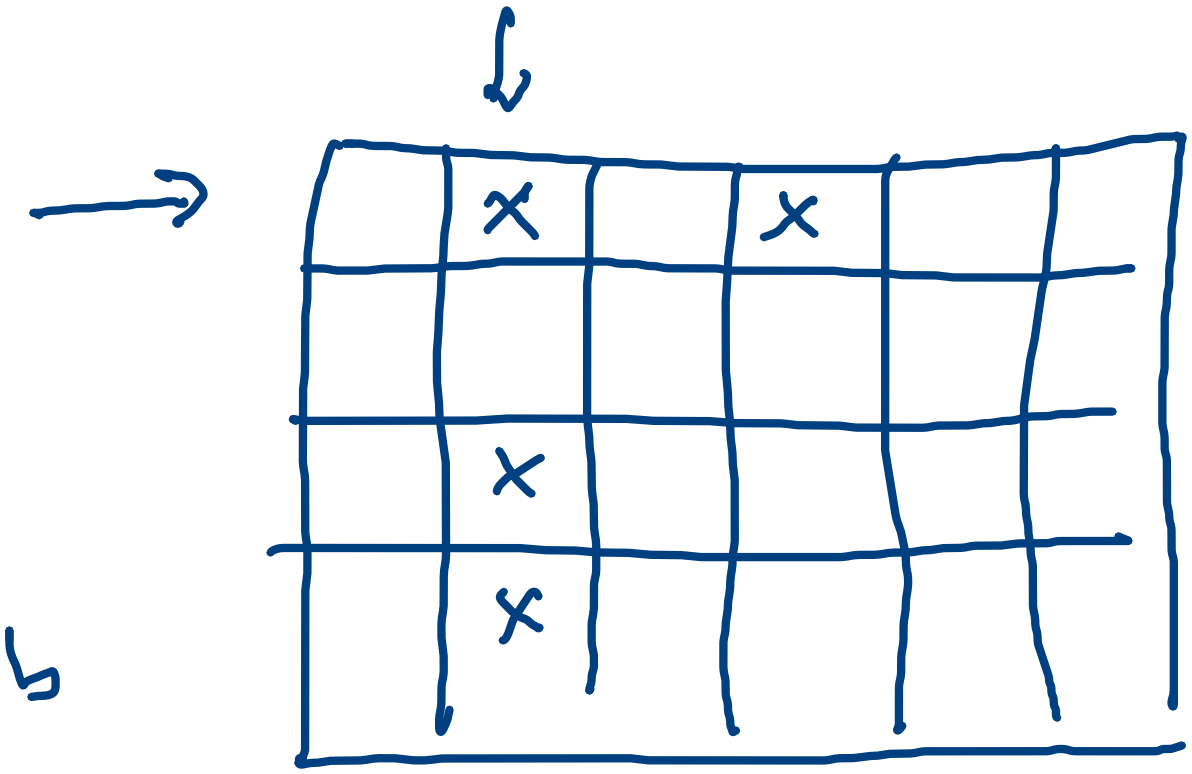
26 Sept 2025

1-) Operations de pre-traite_ents

2-) Contenu de la presentation.

TP: EDA, pre-traite_ent.

Questions & Reponses



Valeurs manquantes

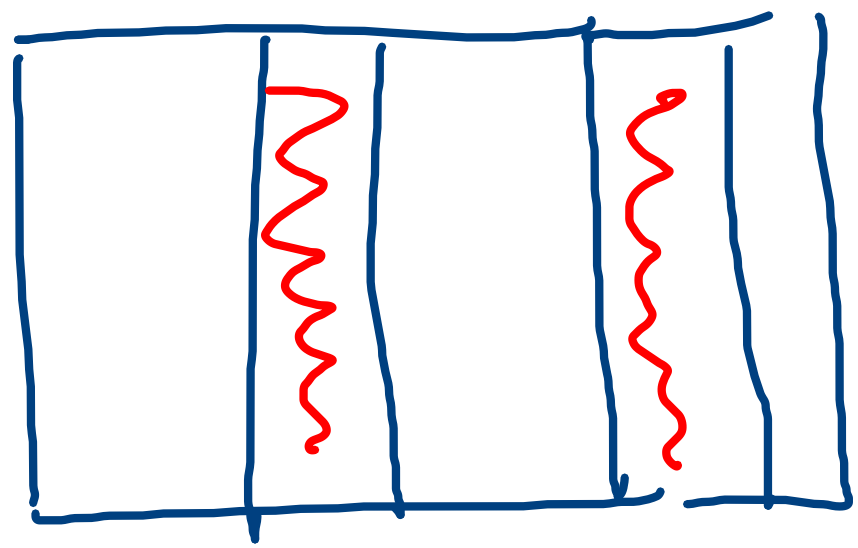
1-) Series

Enregistrer

↳ Imputing ⇒ Stratégies de Remplacement

Suppression ⇒ Retrait ⇒ Perte d'information

Age ← salaire .



18

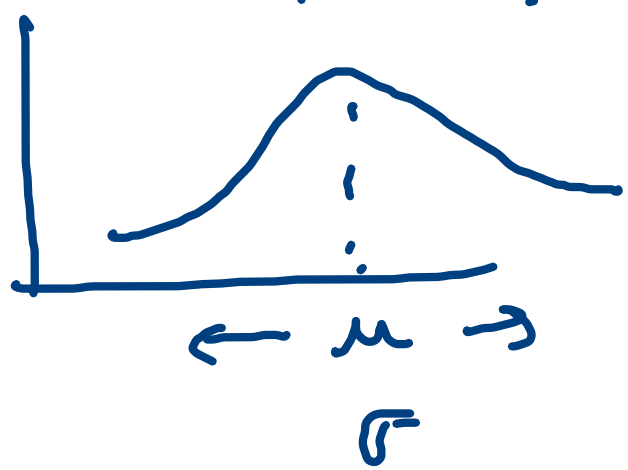
75

25 000

150 000

Distribution Prob

Normal.



StandardScaler. → $X \sim Normal(\mu, \sigma^2)$

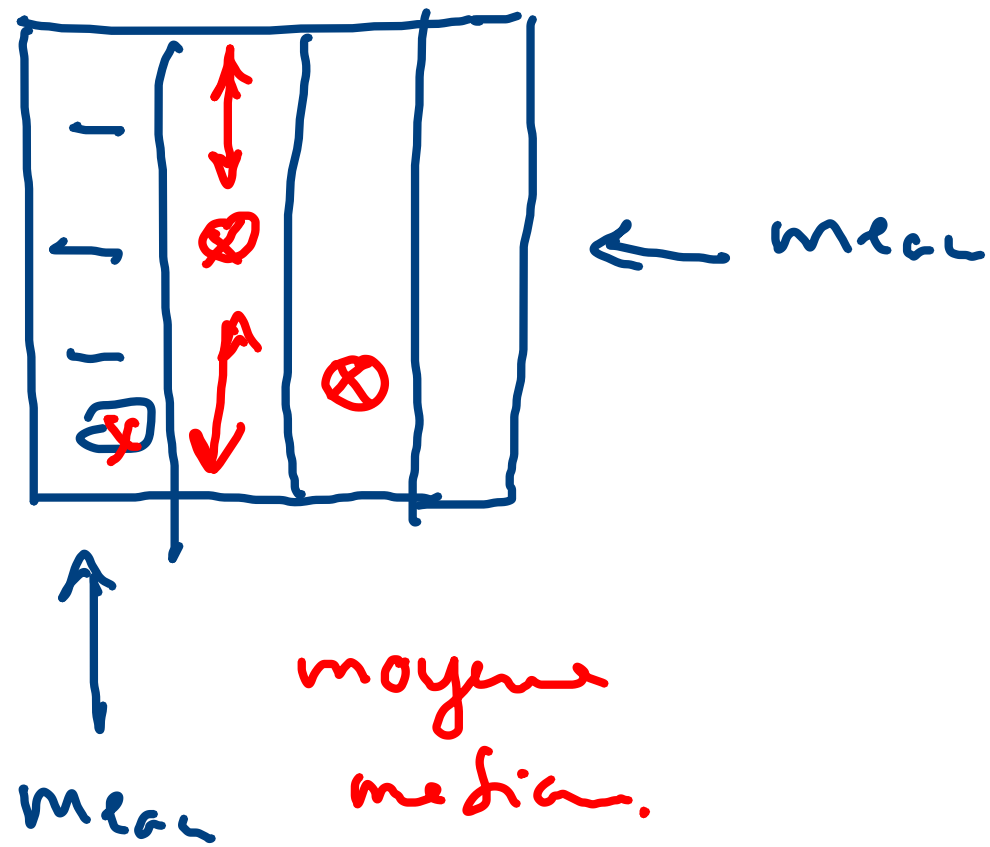
RobustScaler → En presencia Valores Aberrantes.

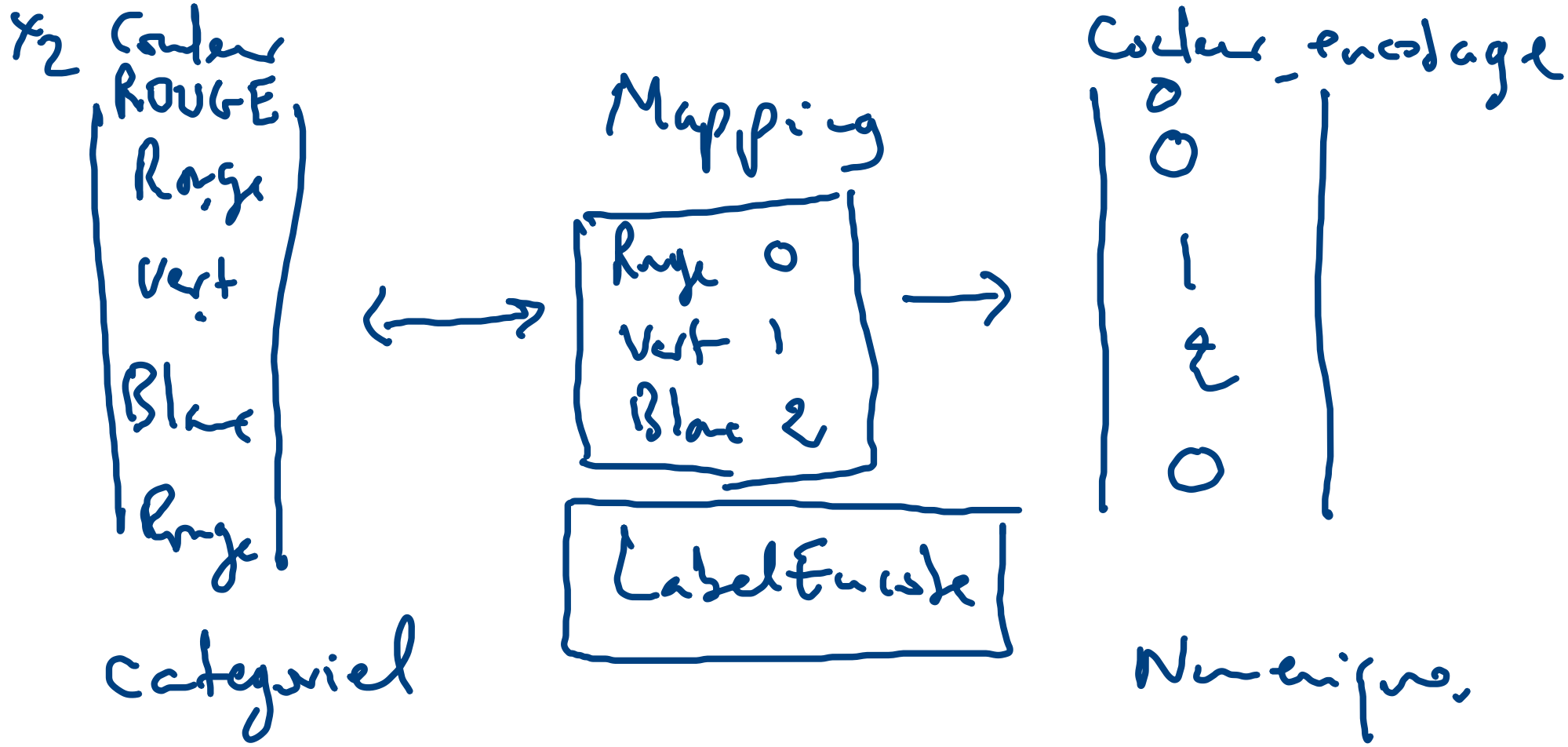
MinMaxScaler → Distribución es otra que Normal

Preliminares: Determinar la Distribución de X

Simple Input

inputing






```
df['X2'].map(lambda r: r.upper())

    for tmp in X2:
        tmp.upper()
```

OHE

one hot encoder

x_2
Rouge
Rouge
vert
Blanc
Blanc
Rouge

→

$x_{2\text{-rouge}}$	$x_{2\text{-vert}}$	$x_{2\text{-Blanc}}$
1	0	0
1	0	0
0	1	0
0	0	1
0	0	1
1	0	0

Plein de Zero.

Zero

Presentation. 15 min - 20 min.

12-15 slides

1- Presentation du Besoin.

2- Chaine de Dev utilisée.

3-) Type model et Algo(s) utilisés.

4-) Metriques obtenus.

5-) Comparaison et Interpretation des models.

6-) Difficultés rencontrés -

7-) Competences acquises.

8-) Conclusion

Regression

RL.
Ridge.
Lasso.

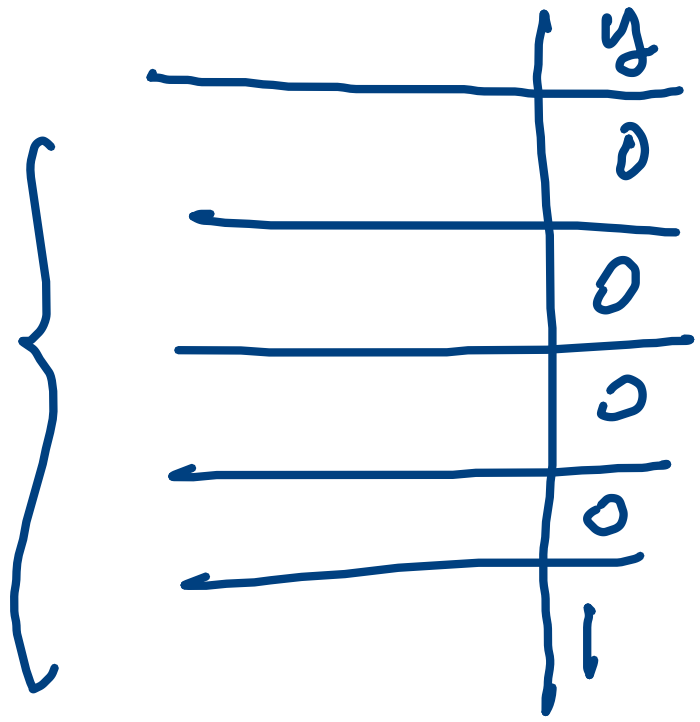
RMSE
R-2

Classification

KNN.
DT
R.F

Accuracy.
Precision.

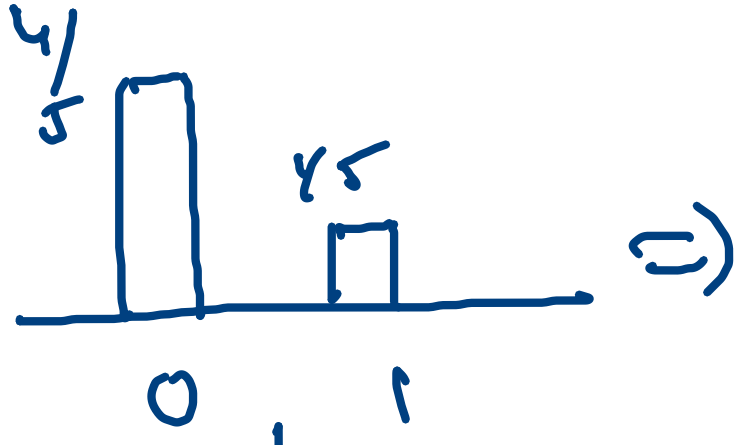
Training



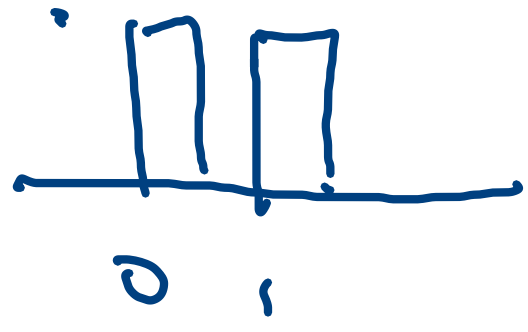
$0 \rightarrow 4/5$?

$1 \rightarrow \frac{1}{5}$.

→ prédire 0 ou 1



⇒ Resampling du Dataset



machine learning mastery