

# DÉTECTION DE FAUX BILLETS

ROMAIN RAMAT 18/08/2025



# AGENDA

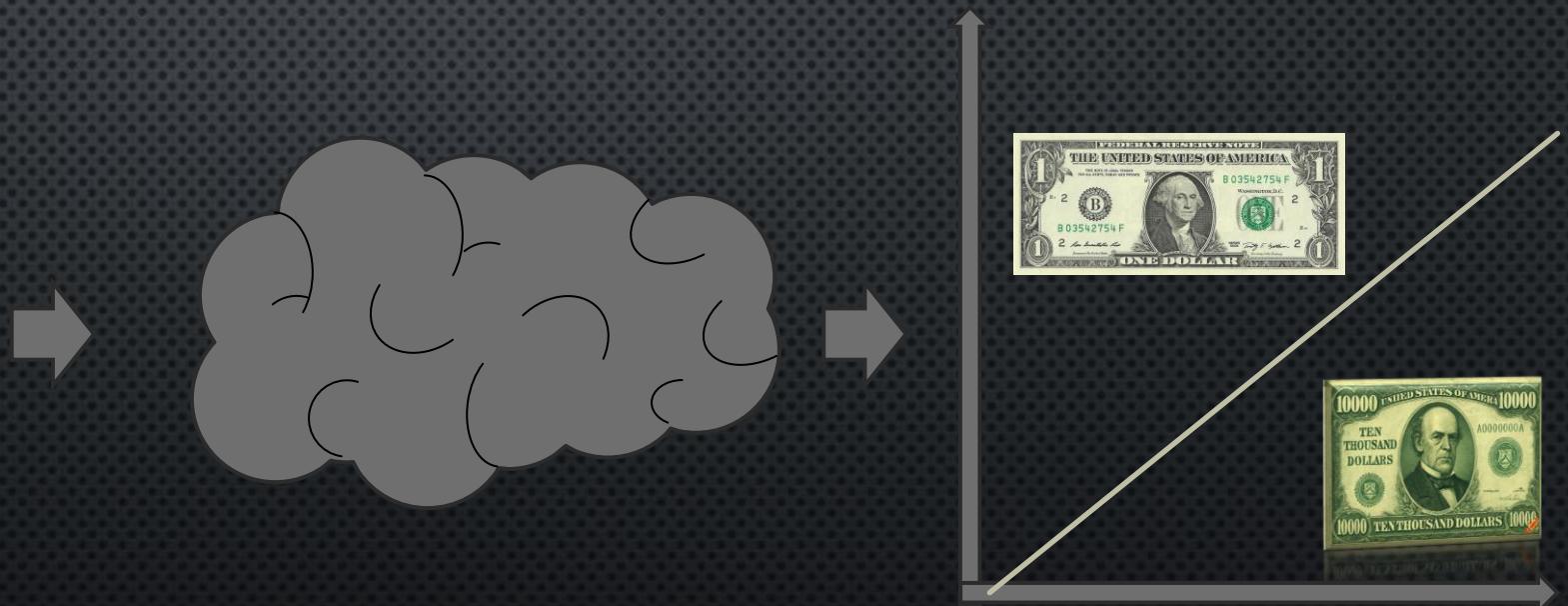
- 
- 01** Objectif
  - 02** Etape 1 : Correction des valeurs manquantes
  - 03** Etape 2 : Identification des vrais / faux billets en régression logistique
  - 04** Etape 3 : Identification des vrais / faux billets en K-Means
  - 05** Etape 4 : Identification des vrais / faux billets en KNN
  - 06** Etape 4 : Identification des vrais / faux billets en Random Forest

# OBJECTIF : IN MACHINE LEARNING WE TRUST

6 caractéristiques permettent de déterminer si il s'agit d'un vrai ou faux billet

L'objectif est de déterminer par les modèles de Machine Learning à travers des caractéristiques techniques si le billet est vrai ou faux

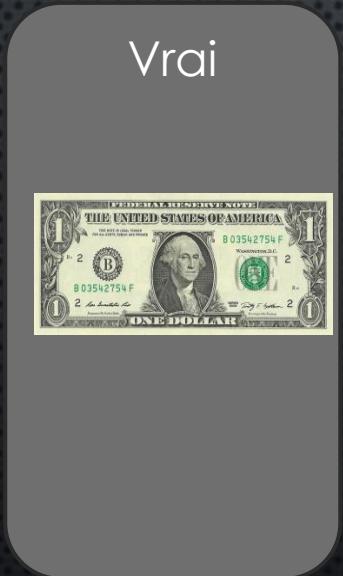
## True or not ?



# ETAPE 1 : CORRECTION DES VALEURS MANQUANTES

Le fichier test contient 37 valeurs manquantes de Margin Low

- Utilisation de la régression linéaire pour prédire la valeur des données manquantes



Margin low

A grey rounded rectangle containing two horizontal bars. To its right is the text "Margin low".

# TEST DE COLINÉARITÉ DES VARIABLES

Utilisation du modèle Variance Inflection factor

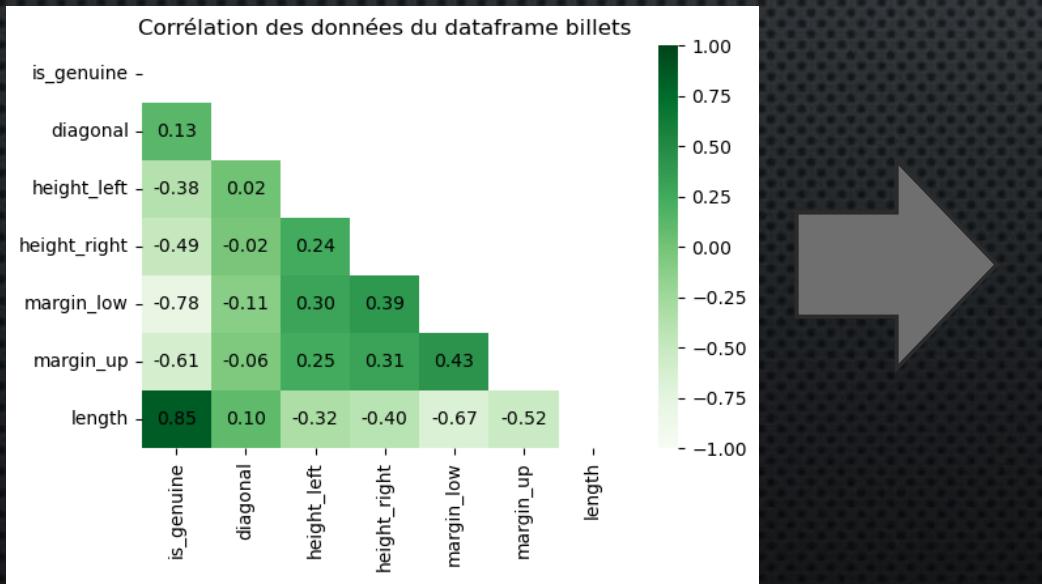
$$VIF(X_i) = \frac{1}{1-R^2_i}$$

VIF = 1 : Pas de multi-colinéarité (Idéal)

VIF < 5 : Multi-colinéarité faible (acceptable)

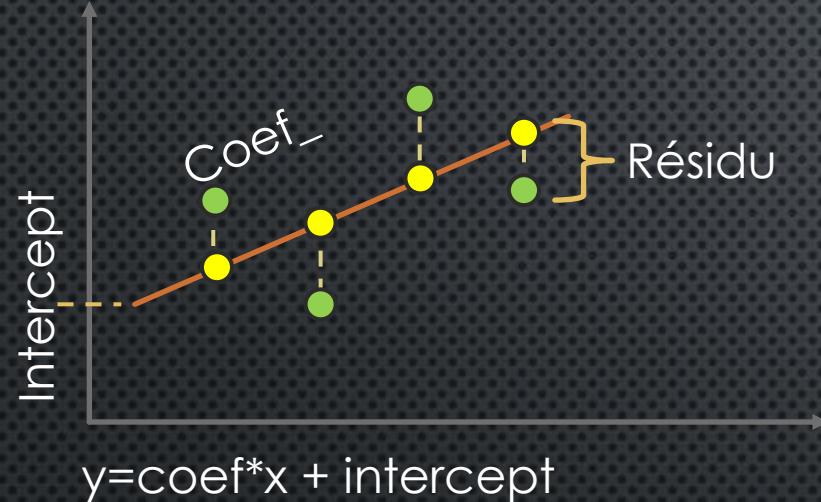
VIF > 5 : Forte multi-colinéarité (Solution : Supprimer ou combiner variables corrélées)

VIF > 10 : Multi-colinéarité sévères (Redondance des variables)



Suppression des données  
redondantes  
(Fake / True)

# ANALYSE DES MÉTRIQUES



Erreurs quadratiques moyennes

$$\text{MSE} = \frac{1}{m} \sum (\text{ytest} - \text{ypred})^2$$

Erreurs absolues Moyennes

$$\text{MAE} = \frac{1}{m} \sum |\text{ytest} - \text{ypred}|$$

Root Mean Square Error (mise à l'échelle)

$$\text{RMSE} = \sqrt{\left( \frac{1}{m} \sum (\text{ytest} - \text{ypred})^2 \right)}$$

La MSE accorde une importance aux grandes erreurs

La MAE importance d'une erreur linéaire avec son amplitude

$$R^2 = \frac{\sum (\text{ytest} - \text{ypred})^2}{\sum (\text{ytest} - \text{moyenne } (\text{ytest}))^2}$$

soit : somme des carrés des résidus (**erreur du modèle**)

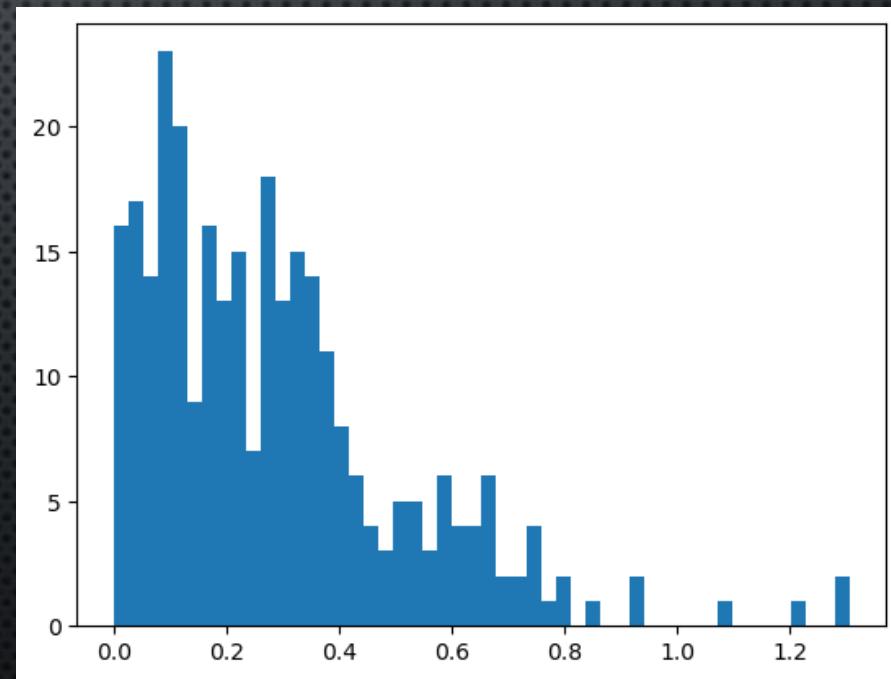
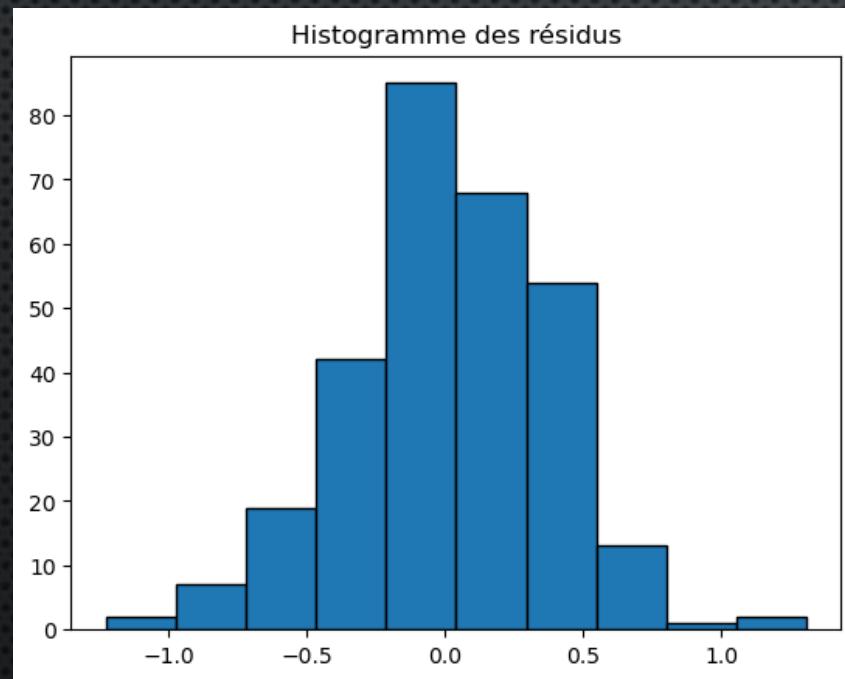
sur somme des carrés totaux (**variance totale de la donnée**)

Test	Résultat
MSE	0.14
RMSE	0.37
MAPE	0.06
MAE	0.29
R <sup>2</sup>	0.67

# TEST DE NORMALITÉ DES RÉSIDUS

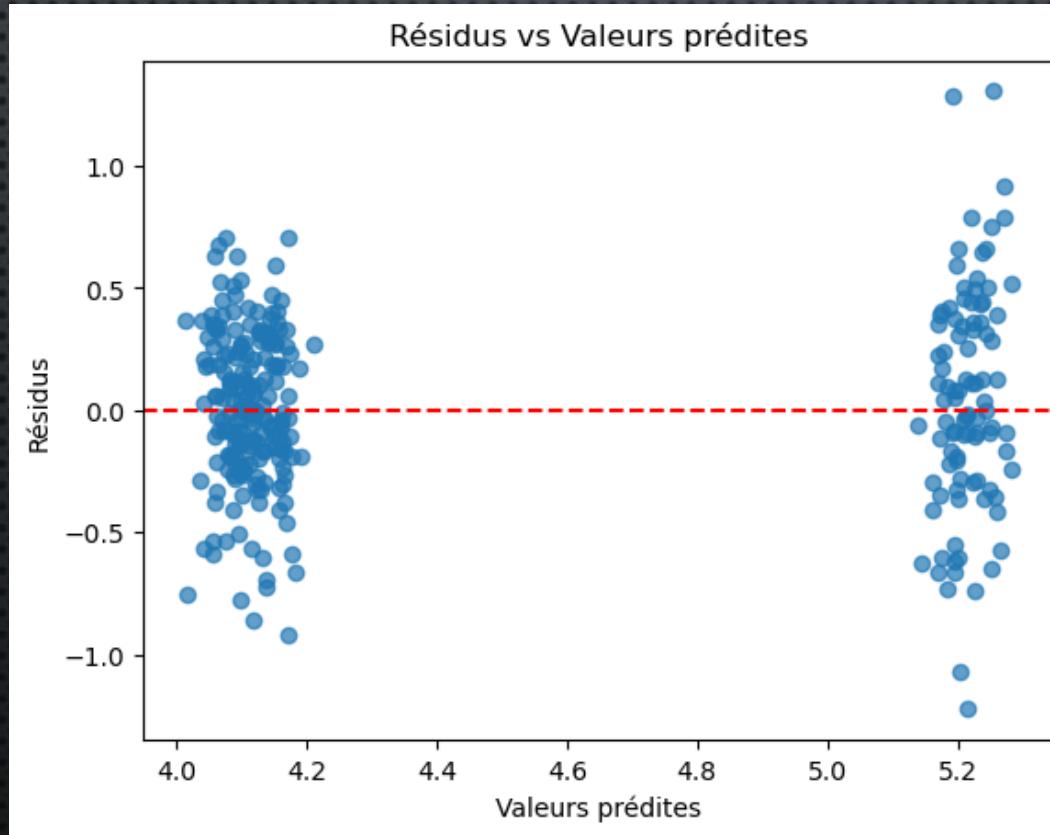
résidus =  $y_{\text{test}} - y_{\text{pred}}$

Test de Shapiro wild confirme que les résidus suivent une loi normale  $p\_value = 0.058$  les résidus suivent une loi normale



# TEST D'HOMOSCÉDASTICITÉ

Test de Breusch-Pagan rejette l'hypothèse nulle, il faut donc considérer l'hypothèse alternative (hétéroscédasticité). P\_value = 3,46e-33



# ETAPE 2 : IDENTIFICATION VRAI VS FAUX BILLETS

Entrainement des jeux des données en **régession logistique** : Vrai billet / Faux Billet  
Exclusion des valeurs initialement manquantes

Entrainement

X\_train

y\_train



Evaluation

X\_test

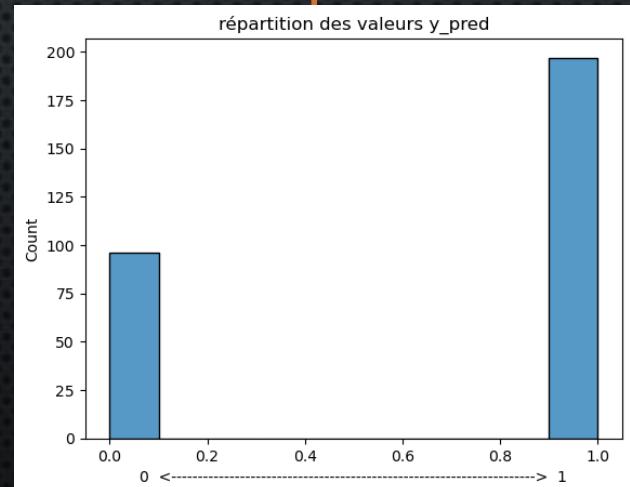
y\_test



Accuracy...

Prédiction

Y\_pred



# EVALUATION DU MODÈLE DE RÉGRESSION LOGISTIQUE

Rapport de classification	precision	recall	f1-score	support
0	1.00	1.00	1.00	96
1	1.00	1.00	1.00	197
accuracy			1.00	293
macro avg	1.00	1.00	1.00	293
weighted avg	1.00	1.00	1.00	293

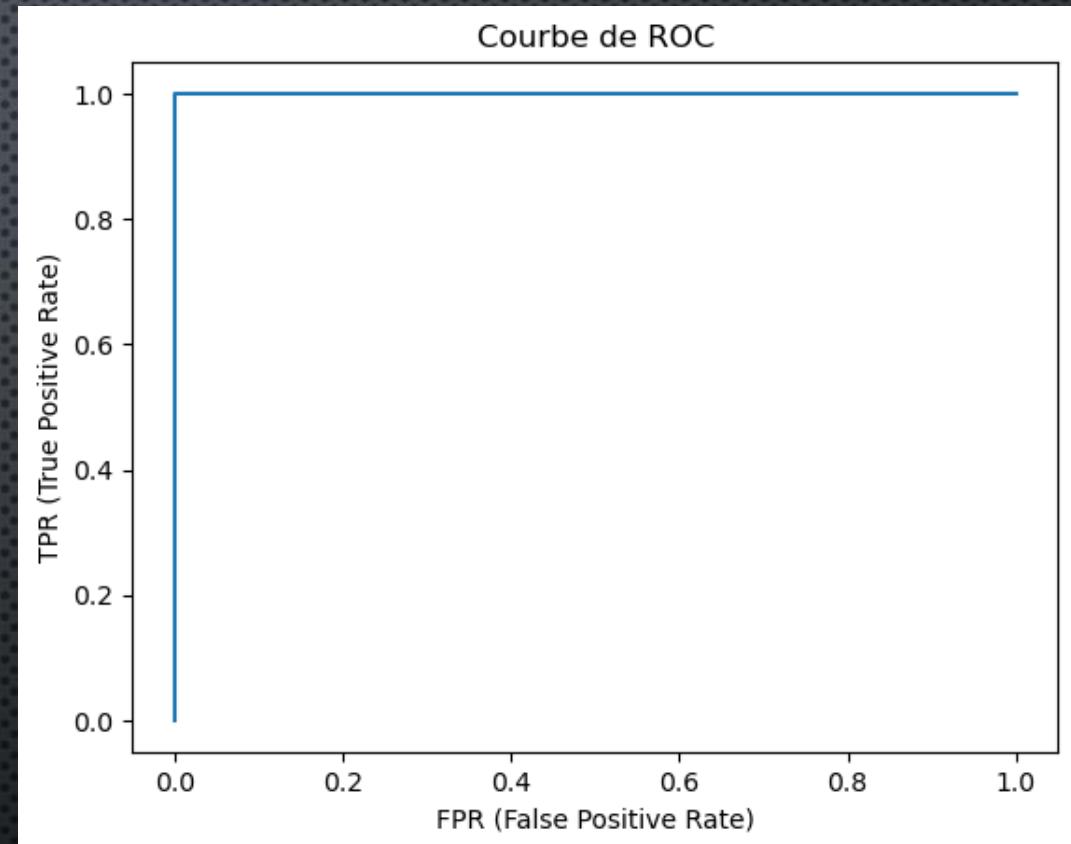
Matrice de confusion	
0	96
1	0
0	197
1	0

POSITIF / NEGATIF

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

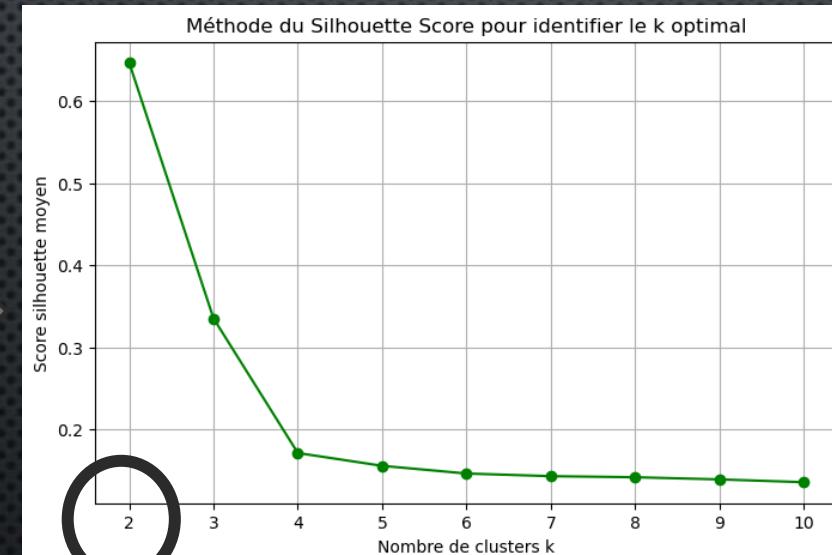
$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

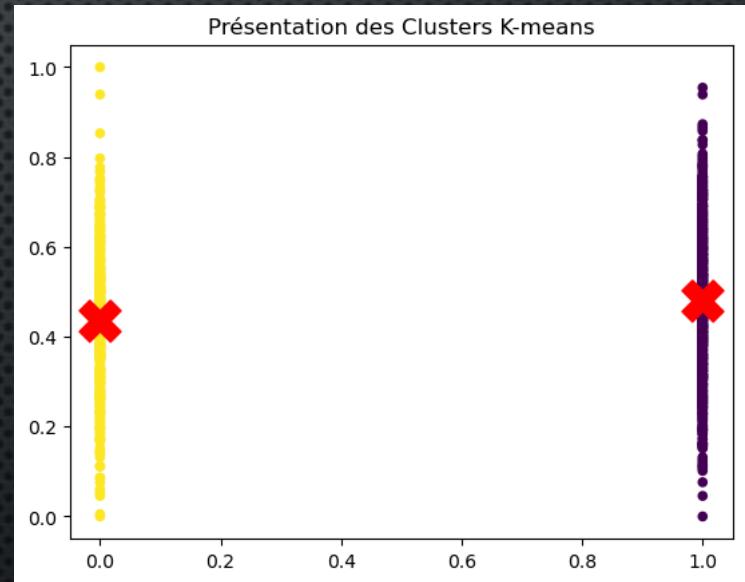


# ETAPE 3 : IDENTIFICATION VRAI VS FAUX BILLETS

Entrainement des jeux des données en **K-Means**: Vrai billet / Faux Billet  
Exclusion des valeurs initialement manquantes

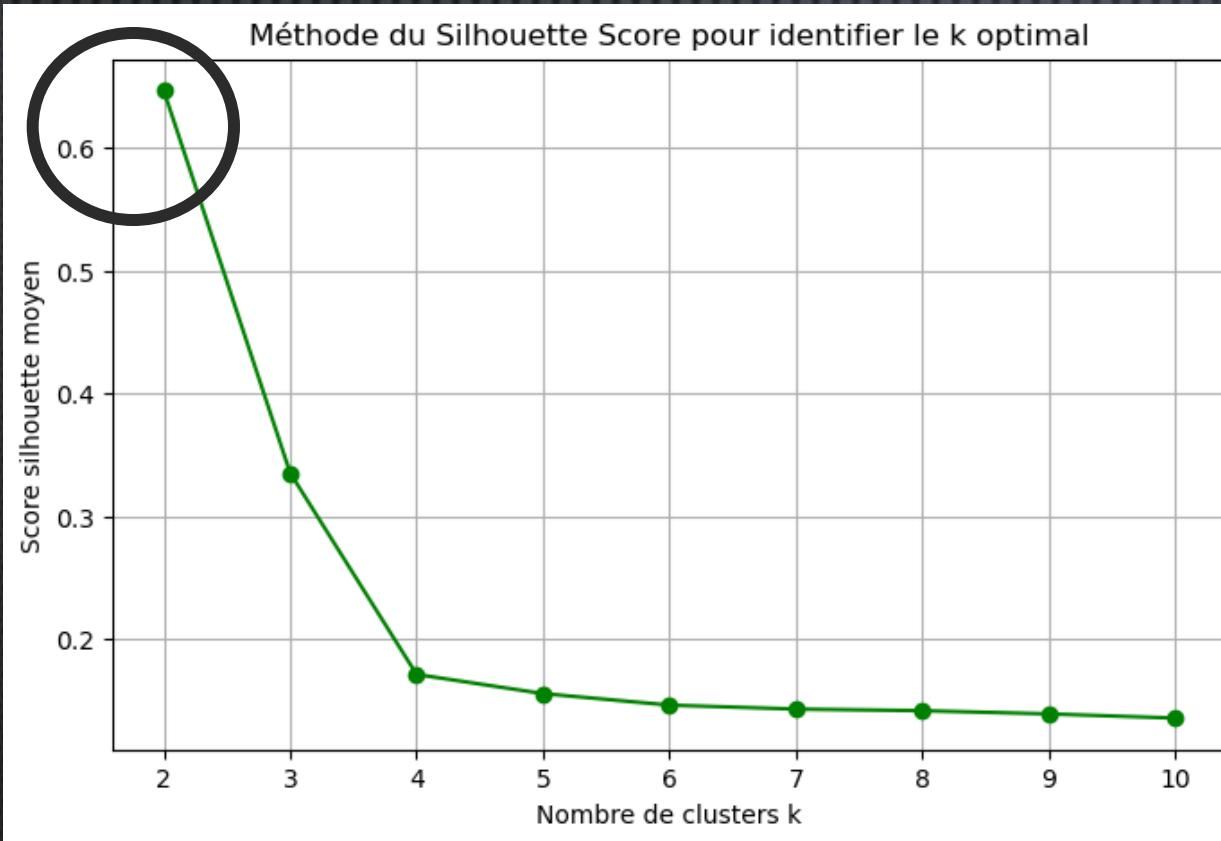


Test de silhouette pour identifier le nombre de cluster ici « 2 »



Représentation K-Means

# EVALUATION DU MODÈLE DE K-MEANS



Silhouette score = 0,65

Matrice de confusion K-Means

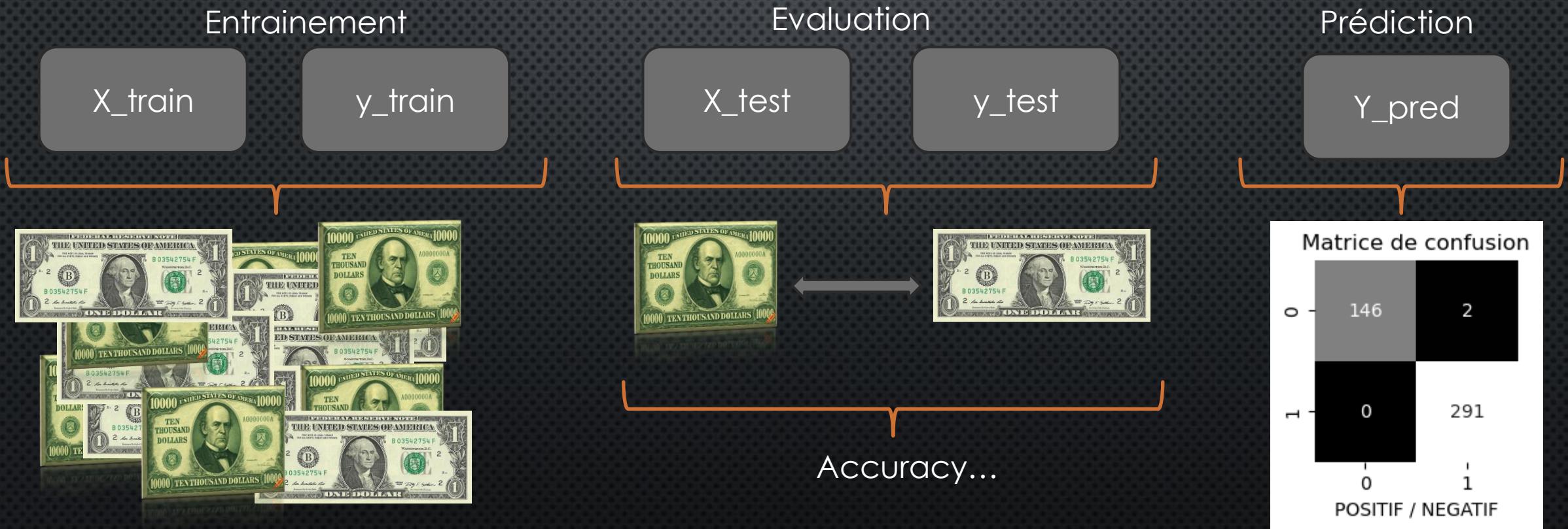
		Cluster1 / Cluster2	
		0	1
Cluster1	0	0	971
	1	492	0

Faux Billet - Vrai Billet

Adjusted Rang Index = 1

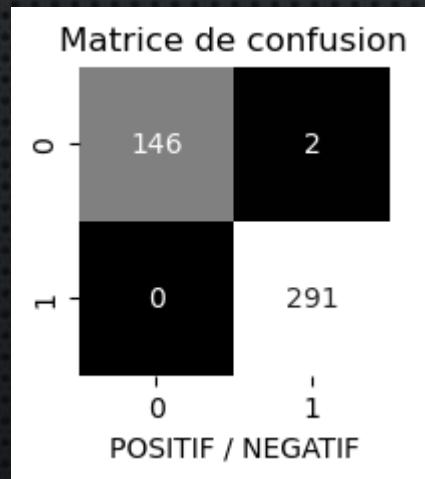
# ETAPE 4 : IDENTIFICATION VRAI VS FAUX BILLETS

Entrainement des jeux des données en **KNN**: Vrai billet / Faux Billet  
Exclusion des valeurs initialement manquantes



# EVALUATION DU MODÈLE KNN

Rapport de classification	precision	recall	f1-score	support
0	1.00	0.99	0.99	148
1	0.99	1.00	1.00	291
accuracy			1.00	439
macro avg	1.00	0.99	0.99	439
weighted avg	1.00	1.00	1.00	439



$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

# ETAPE 5 : IDENTIFICATION VRAI VS FAUX BILLETS

Entrainement des jeux des données en **Random Forest**: Vrai billet / Faux Billet  
Exclusion des valeurs initialement manquantes

Entrainement

X\_train

y\_train



Evaluation

X\_test

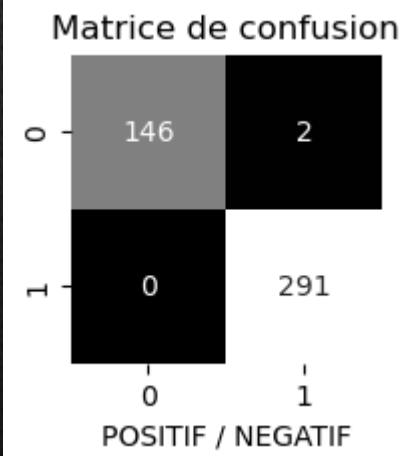
y\_test



Accuracy...

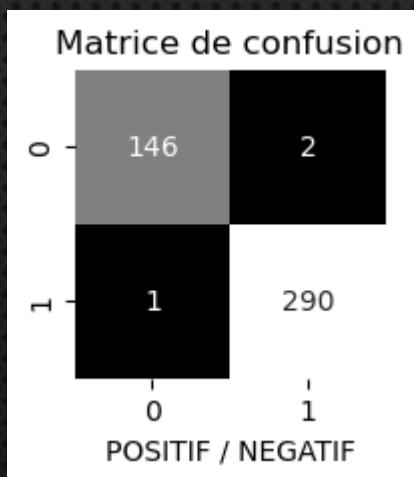
Prédiction

Y\_pred



# EVALUATION DU MODÈLE RANDOM FOREST

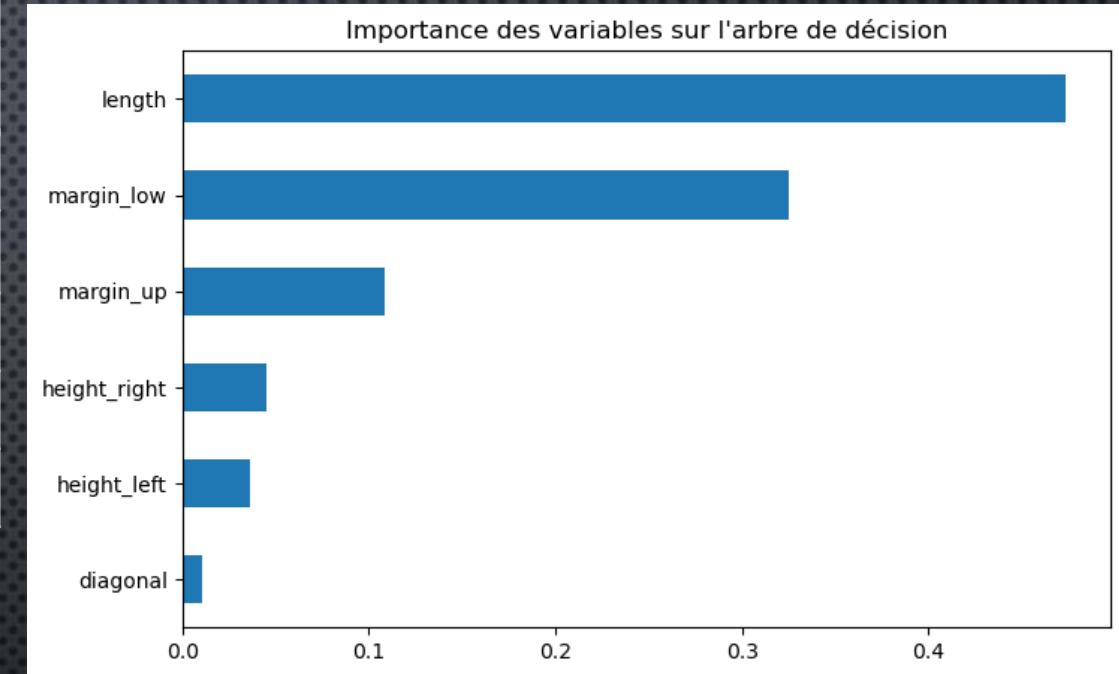
Rapport de classification	precision	recall	f1-score	support
0	0.99	0.99	0.99	148
1	0.99	1.00	0.99	291
accuracy			0.99	439
macro avg	0.99	0.99	0.99	439
weighted avg	0.99	0.99	0.99	439



$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$



# CONCLUSION :

- DANS NOTRE CAS, LES TESTS DE RÉGRESSION LOGISTIQUE ET DE K-MEANS S'AVÈRENT LES MODÈLES LES PLUS PERFORMANTS AVEC DES SCORES ACCURACY À 1
- LE JEU DE TEST A ÉTÉ SOLUTIONNÉ :

	is_genuine	diagonal	height_left	height_right	margin_low	margin_up	length	id
0	Faux_billet	171.76	104.01	103.54	5.21	3.30	111.42	A_1
1	Faux_billet	171.87	104.17	104.13	6.00	3.31	112.09	A_2
2	Faux_billet	172.00	104.58	104.29	4.99	3.39	111.57	A_3
3	Vrai_billet	172.49	104.55	104.34	4.44	3.03	113.20	A_4
4	Vrai_billet	171.65	103.63	103.56	3.77	3.16	113.33	A_5