ONCFM

Algorithme de détection automatique de faux billets

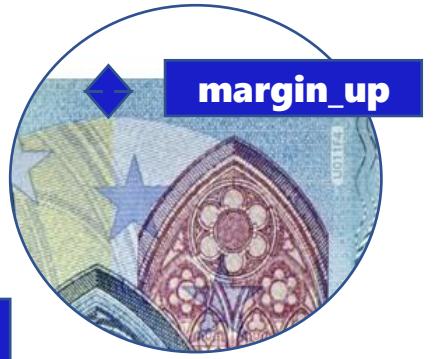
Analyse descriptive des données

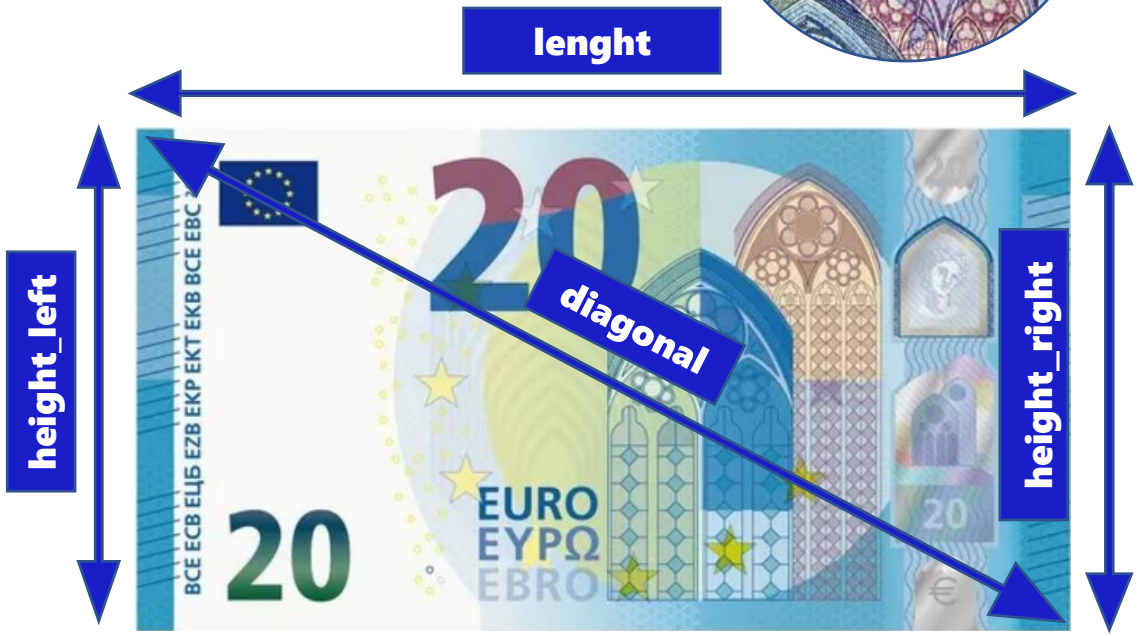
Authenticité du billet

Dimensions du billet

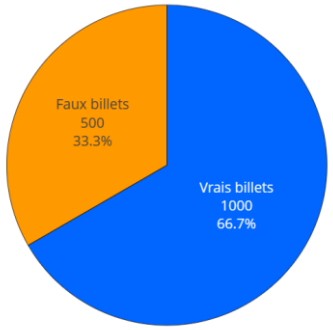
	is_genuine	diagonal	height_left	height_right	margin_low	margin_up	length
0	True	171.81	104.86	104.95	4.52	2.89	112.83
1	True	171.46	103.36	103.66	3.77	2.99	113.09
2	True	172.69	104.48	103.50	4.40	2.94	113.16
3	True	171.36	103.91	103.94	3.62	3.01	113.51
4	True	171.73	104.28	103.46	4.04	3.48	112.54
...	...	...	...	...	...	...	...
1495	False	171.75	104.38	104.17	4.42	3.09	111.28
1496	False	172.19	104.63	104.44	5.27	3.37	110.97
1497	False	171.80	104.01	104.12	5.51	3.36	111.95
1498	False	172.06	104.28	104.06	5.17	3.46	112.25
1499	False	171.47	104.15	103.82	4.63	3.37	112.07

1500 rows × 7 columns

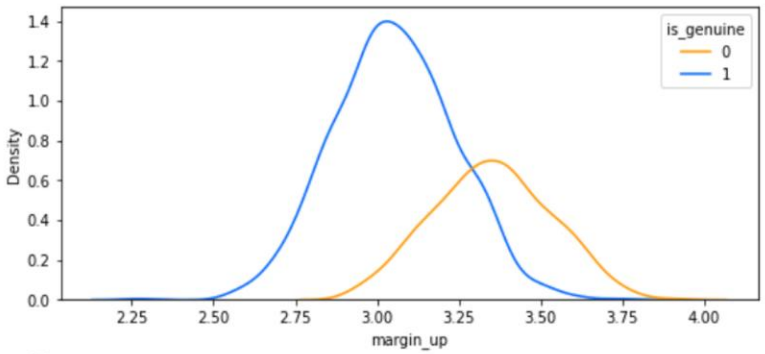


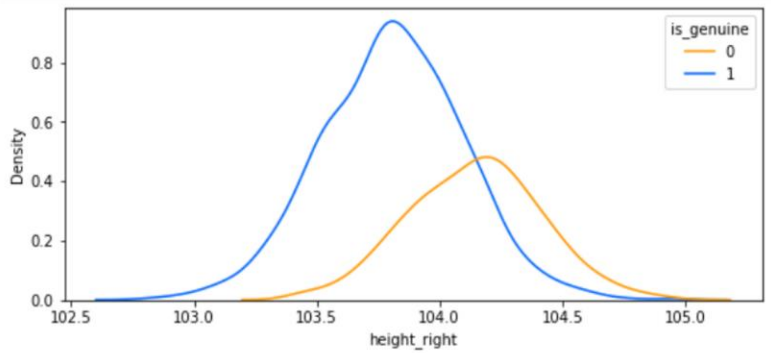


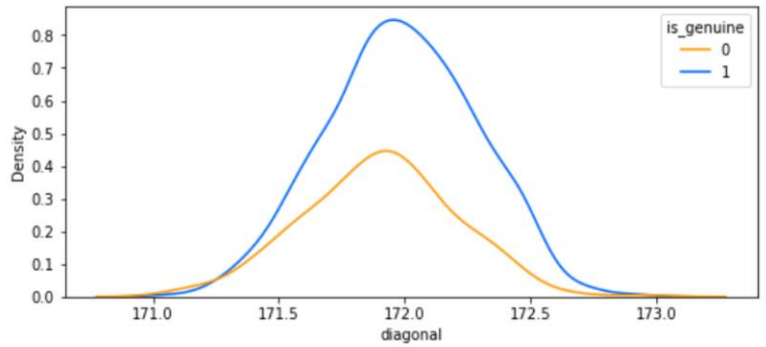
- 1500 billets dont :
  - 1000 billets authentiques
  - 500 faux billets
- Aucune dimension en double
- 37 marges inférieures manquantes

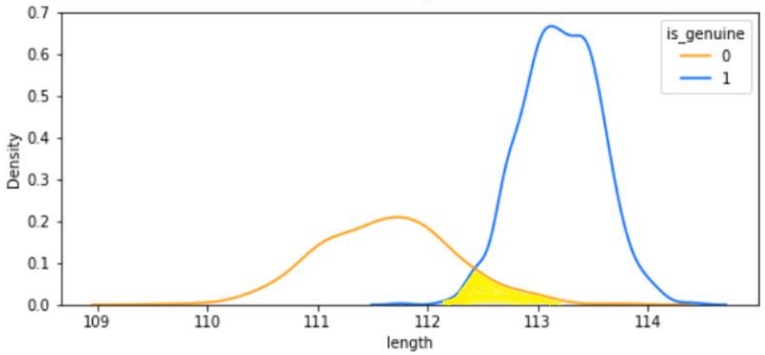


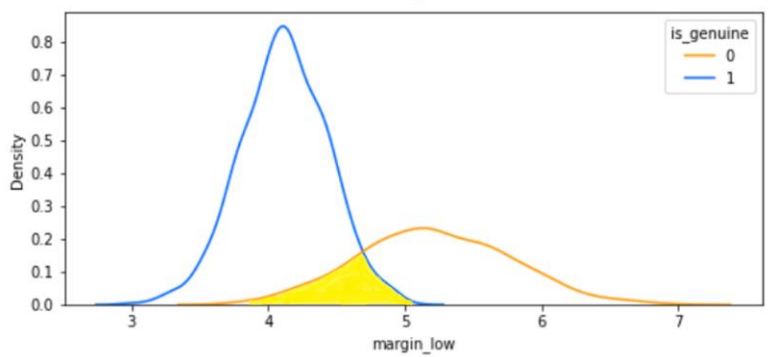
Dimensions / Authenticité

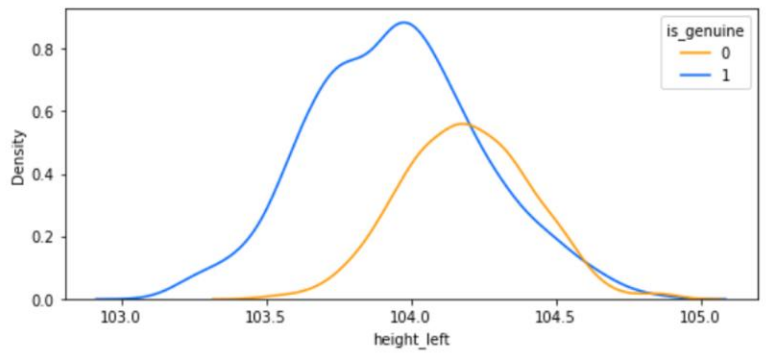






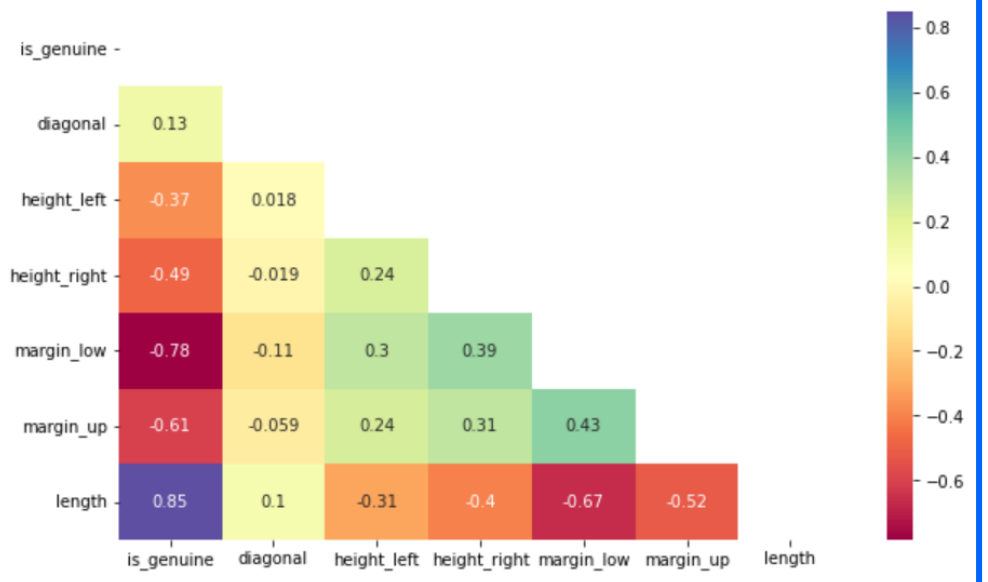






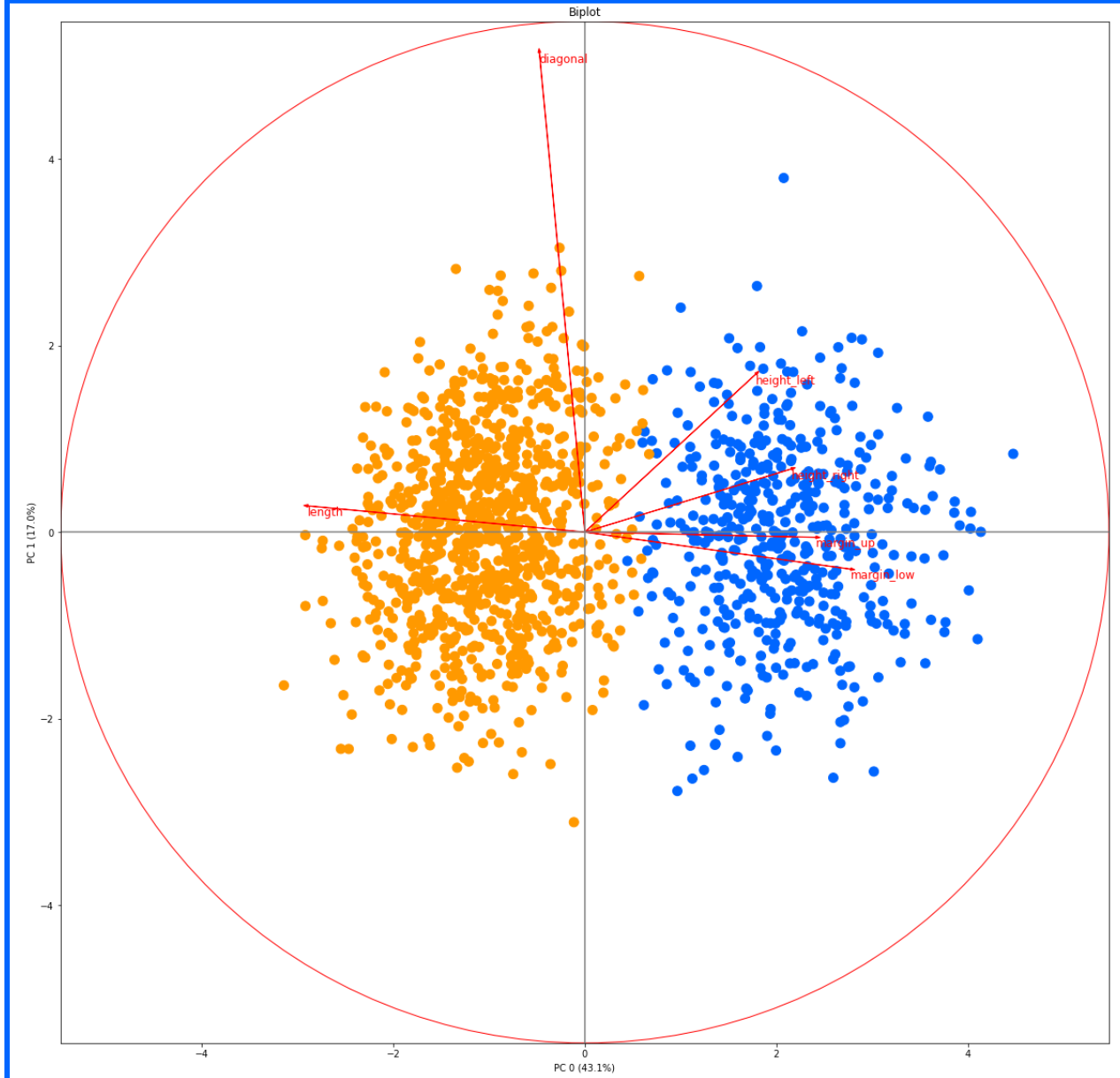
- Les billets authentiques sont plut longs que les faux.
- La marge entre le bord inférieur du billet et l'image de celui-ci (en mm) est plus grande dans les faux billets.

Matrice des corrélations

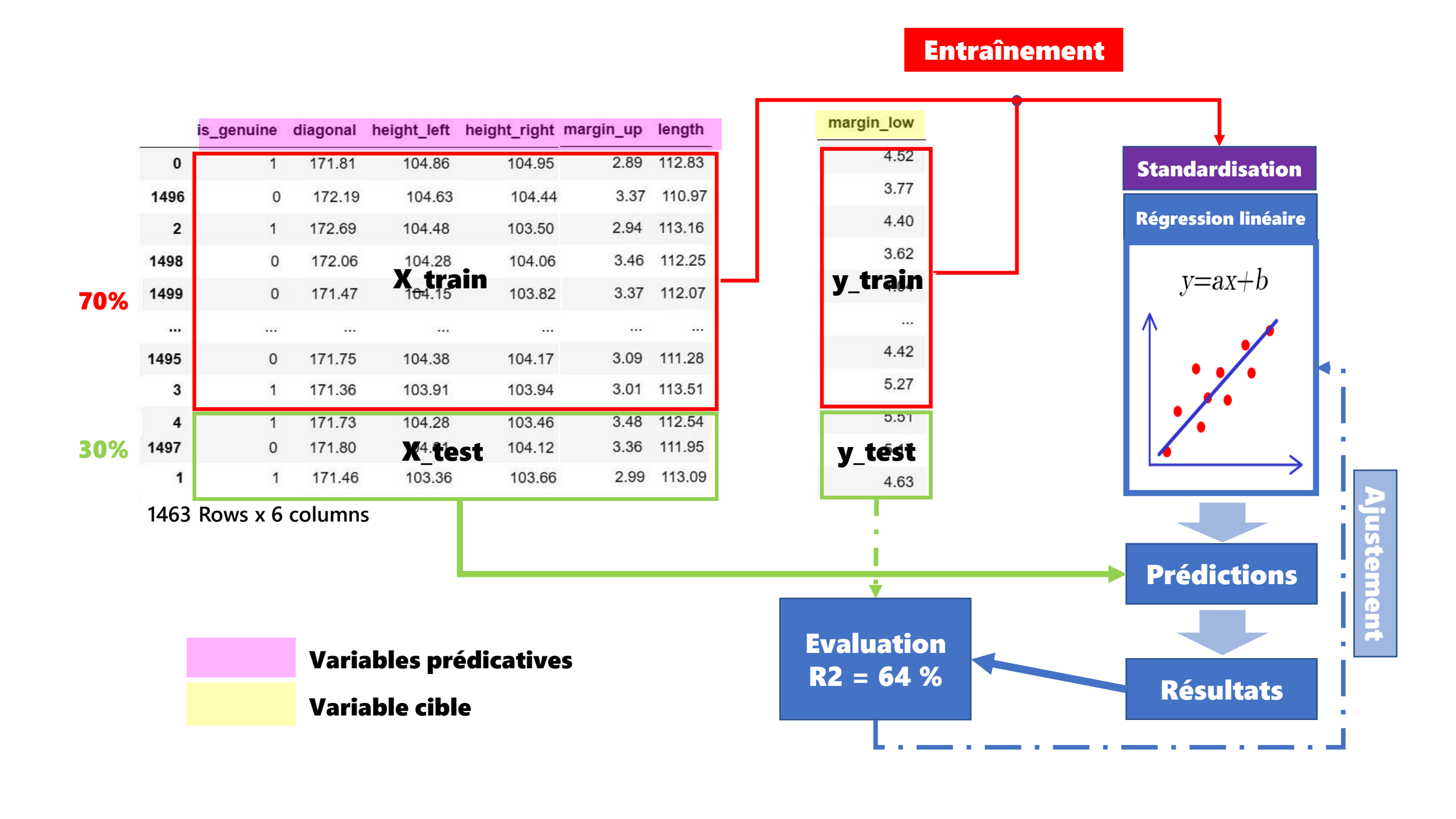


- La longueur et la marge basse sont respectivement corrélées positivement et négativement à l'authenticité du billet.
- La longueur est une dimension distinctive pour l'authenticité d'un billet.

ACP – KMeans

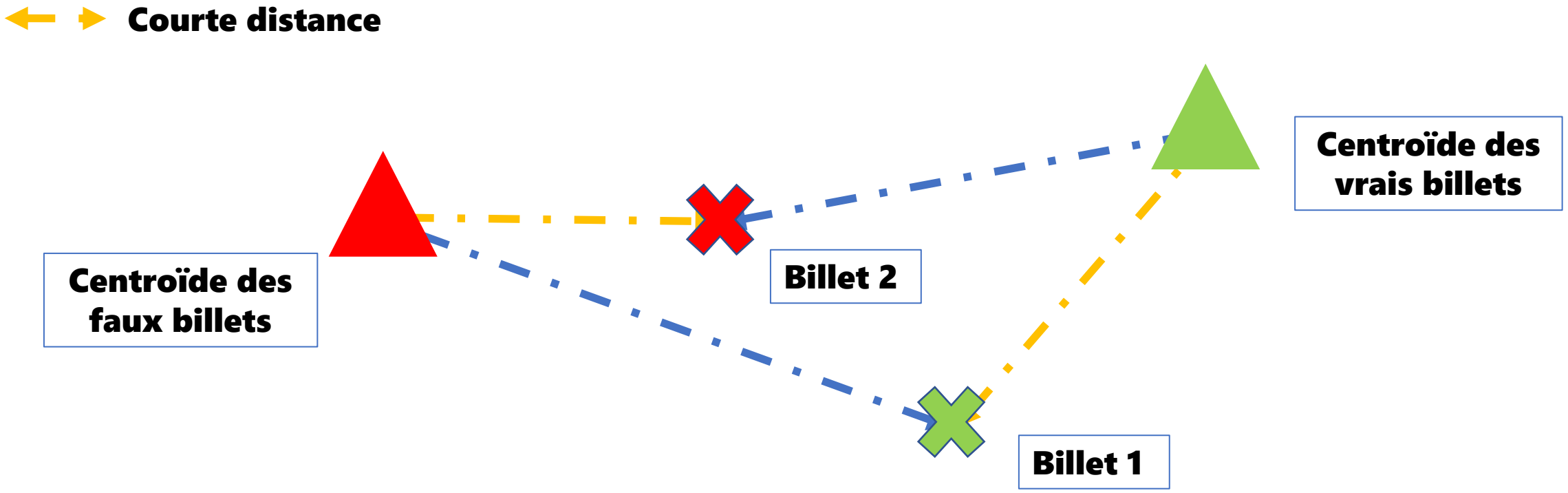


Traitement des valeurs manquantes


Entraînement
is_genuine diagonal height_left height_right margin_up length
0 1 171.81 104.86 104.95 2.89 112.83
1496 0 172.19 104.63 104.44 3.37 110.97
2 1 172.69 104.48 103.50 2.94 113.16
1498 0 172.06 104.28 104.06 3.46 112.25
1499 0 171.47 104.15 103.82 3.37 112.07
... ...
1495 0 171.75 104.38 104.17 3.09 111.28
3 1 171.36 103.91 103.94 3.01 113.51
4 1 171.73 104.28 103.46 3.48 112.54
1497 0 171.80 104.12 3.36 111.95
1 1 171.46 103.36 103.66 2.99 113.09
1463 Rows x 6 columns
70%
30%
X_train
X_test
margin_low
4.52
3.77
4.40
3.62
y_train
...
4.42
5.27
y_test
4.63
Standardisation
Régression linéaire
y=ax+b
Prédictions
Résultats
Evaluation
R2 = 64 %
Ajustement
Variables prédicatives
Variable cible

Algorithme de classification
KMeans

Algorithme Classification KMeans

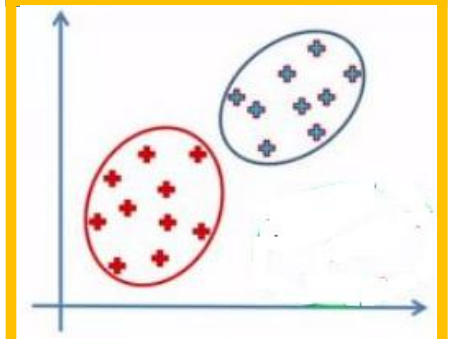


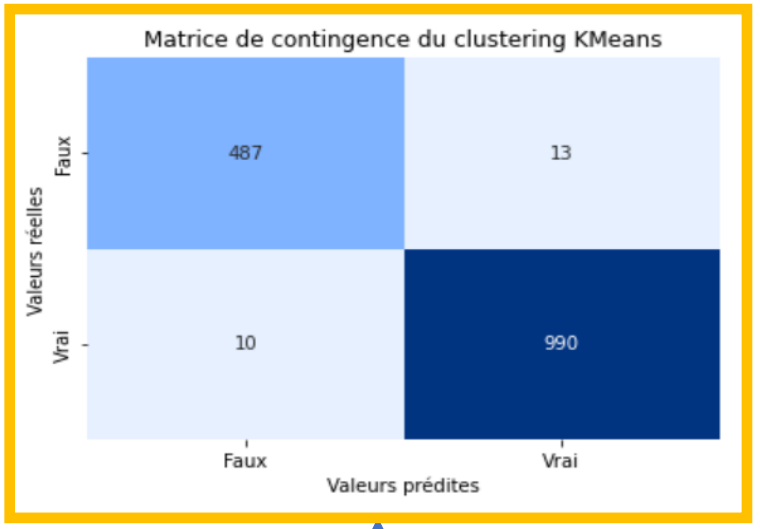
	is_genuine	diagonal	height_left	height_right	margin_low	margin_up	length
0	1	171.81	104.86	104.95	4.52	2.89	112.83
1	1	171.46	103.36	103.66	3.77	2.99	113.09
2	1	172.69	104.48	103.50	4.40	2.94	113.16
3	1	171.36	103.91	103.94	3.62	3.01	113.51
4	1	171.73	104.28	103.46	4.04	3.48	112.54
...	...	...	...	...	...	...	...
1495	0	171.75	104.38	104.17	4.42	3.09	111.28
1496	0	172.19	104.63	104.44	5.27	3.37	110.97
1497	0	171.80	104.01	104.12	5.51	3.36	111.95
1498	0	172.06	104.28	104.06	5.17	3.46	112.25
1499	0	171.47	104.15	103.82	4.63	3.37	112.07

1500 rows × 7 columns

Standardisation

KMeans





Comparaison

Labels 0 et 1

Correspondance

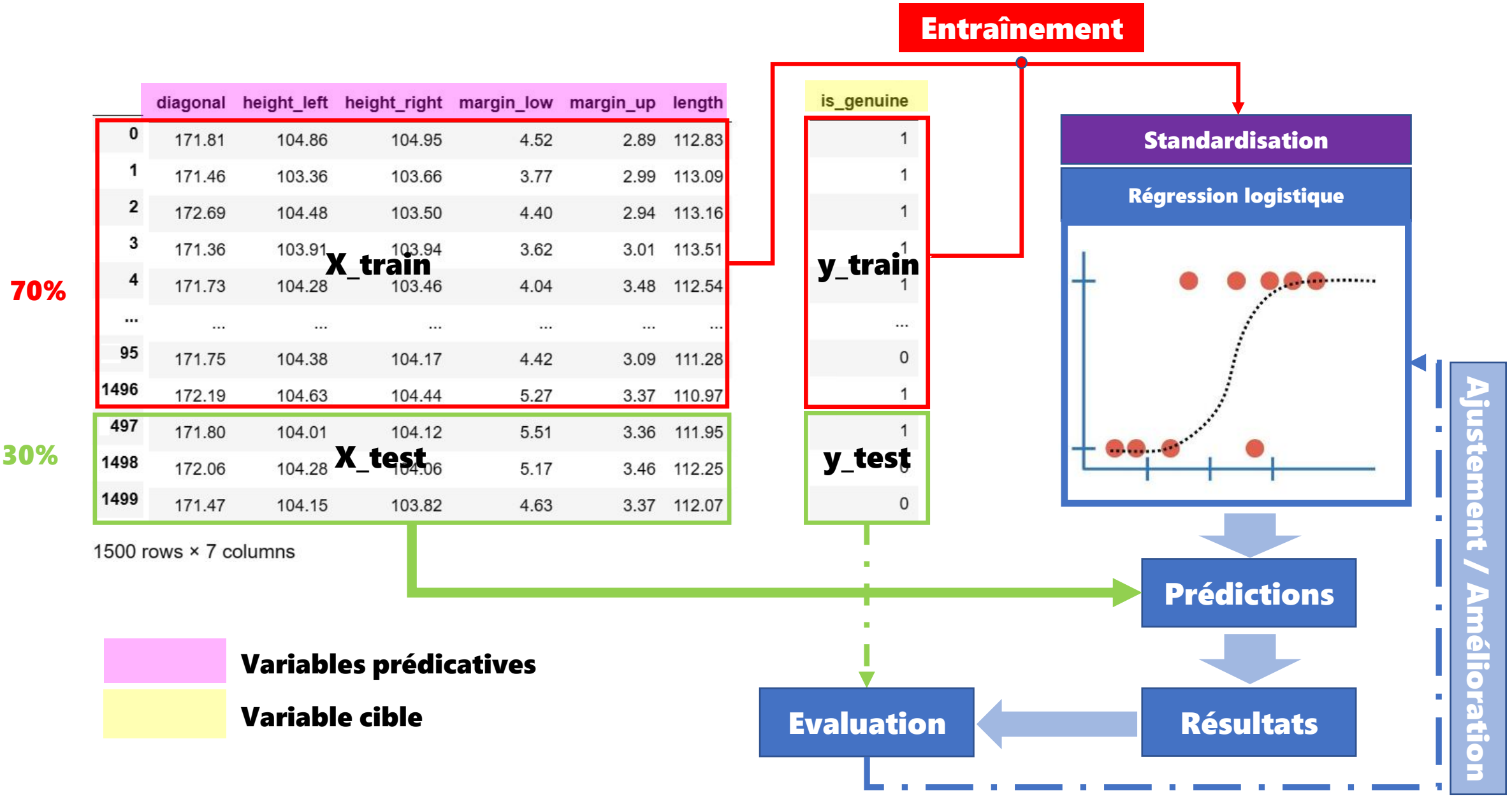
Centroïdes 0 et 1

Centroïdes Vrai et Faux

Algorithme Classification

Comparaison

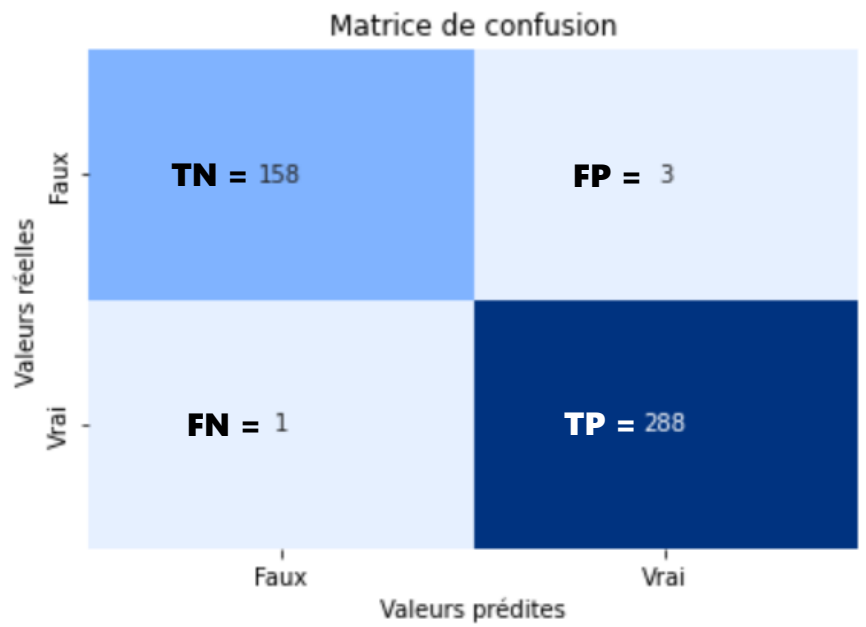
Algorithme de classification
Régression logistique



	diagonal	height_left	height_right	margin_low	margin_up	length	is_genuine
0	171.81	104.86	104.95	4.52	2.89	112.83	1
1	171.46	103.36	103.66	3.77	2.99	113.09	1
2	172.69	104.48	103.50	4.40	2.94	113.16	1
3	171.36	103.91	103.94	3.62	3.01	113.51	1
4	171.73	104.28	103.46	4.04	3.48	112.54	1
...	...	...	...	...	...	...	...
95	171.75	104.38	104.17	4.42	3.09	111.28	0
1496	172.19	104.63	104.44	5.27	3.37	110.97	1
497	171.80	104.01	104.12	5.51	3.36	111.95	1
1498	172.06	104.28	104.06	5.17	3.46	112.25	0
1499	171.47	104.15	103.82	4.63	3.37	112.07	0

Evaluation du modèle

Matrice de confusion



Rapport de classification

	precision	recall	f1-score	support
Faux billets	0.9937	0.9814	0.9875	161
Vrais billets	0.9897	0.9965	0.9931	289
accuracy			0.9911	450
macro avg	0.9917	0.9890	0.9903	450
weighted avg	0.9911	0.9911	0.9911	450

Erreurs de classification :

- Trois faux billets classifiés comme étant authentiques (Faux Positifs).
- Un vrai billet classifié comme étant faux (Faux Négatifs).

Amélioration du modèle

↓

Baisse des faux positifs

↓

$$\uparrow Precision = \frac{TP}{TP + FP} \downarrow$$

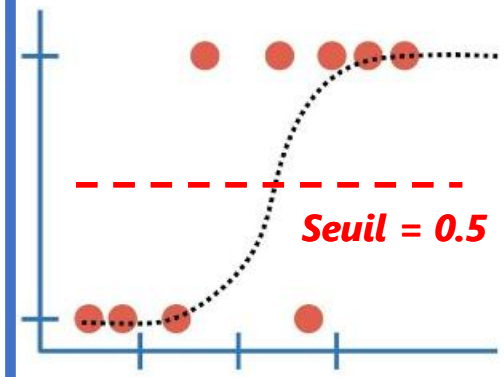
Amélioration du modèle

Changement du seuil

GridSearchCV
Optimisation des Hyperparamètres

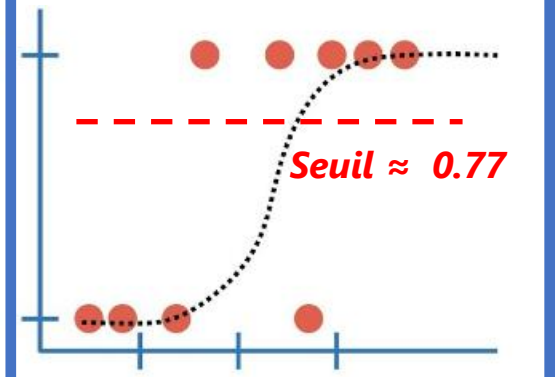
Standardisation

Régression logistique

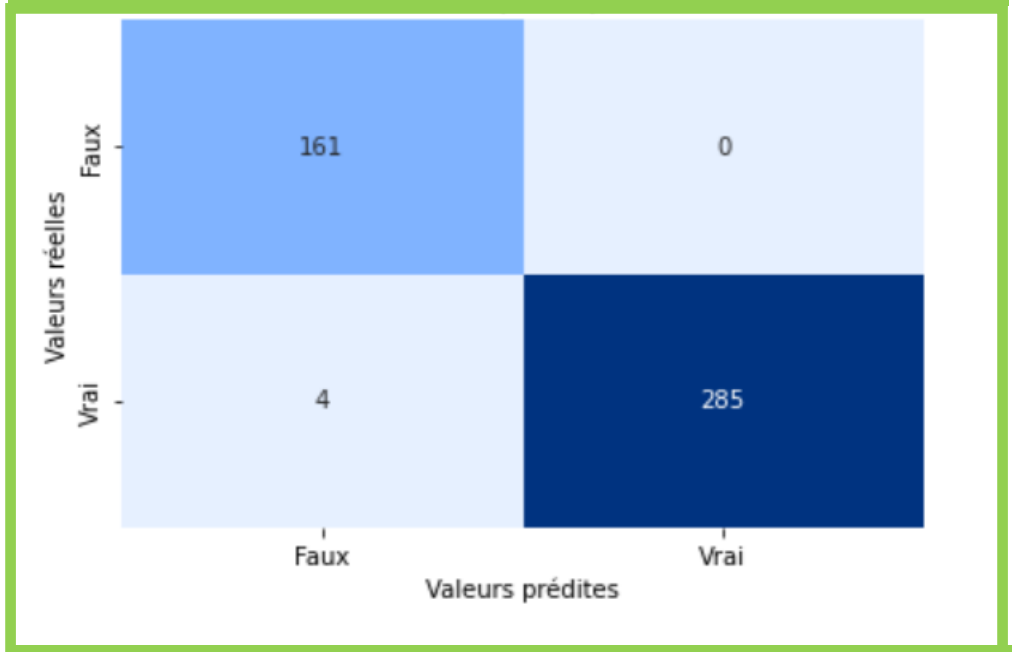


Standardisation

Régression logistique paramétrée



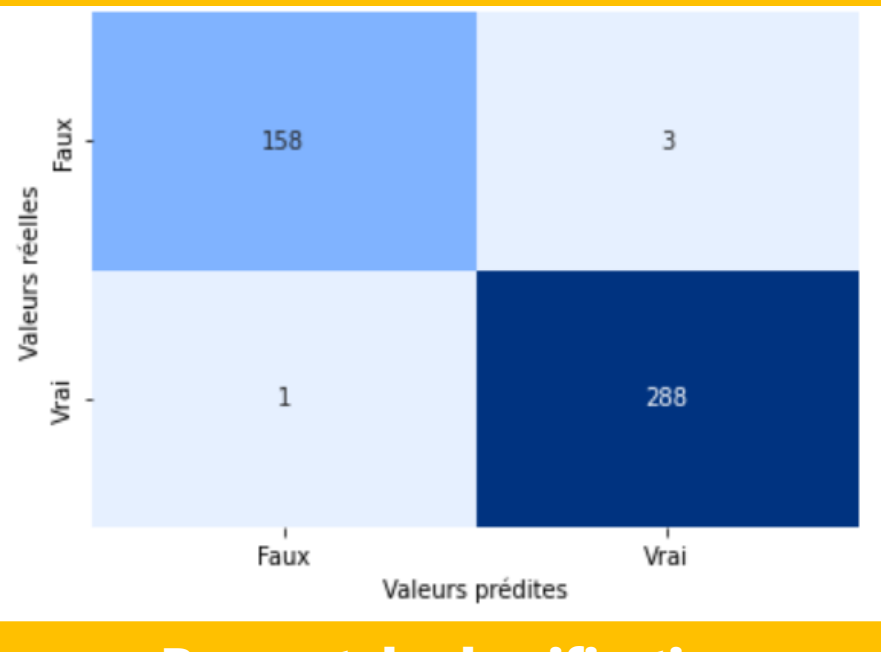
Matrice de confusion après amélioration



Rapport de classification

	precision	recall	f1-score	support
Faux billets	0.9758	1.0000	0.9877	161
Vrais billets	1.0000	0.9862	0.9930	289
accuracy			0.9911	450
macro avg	0.9879	0.9931	0.9904	450
weighted avg	0.9913	0.9911	0.9911	450

Matrice de confusion initiale



Rapport de classification

	precision	recall	f1-score	support
Faux billets	0.9937	0.9814	0.9875	161
Vrais billets	0.9897	0.9965	0.9931	289
accuracy			0.9911	450
macro avg	0.9917	0.9890	0.9903	450
weighted avg	0.9911	0.9911	0.9911	450

Test des algorithmes

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

Classification KMeans

Classification Logistique

id	diagonal	height_left	height_right	margin_low	margin_up	length	authenticité
A_1	171.76	104.01	103.54	5.21	3.30	111.42	Faux
A_2	171.87	104.17	104.13	6.00	3.31	112.09	Faux
A_3	172.00	104.58	104.29	4.99	3.39	111.57	Faux
A_4	172.49	104.55	104.34	4.44	3.03	113.20	Vrai
A_5	171.65	103.63	103.56	3.77	3.16	113.33	Vrai

ACP – KMeans