

Détectez des faux billets

Analyse des données

Analyse en composante principale

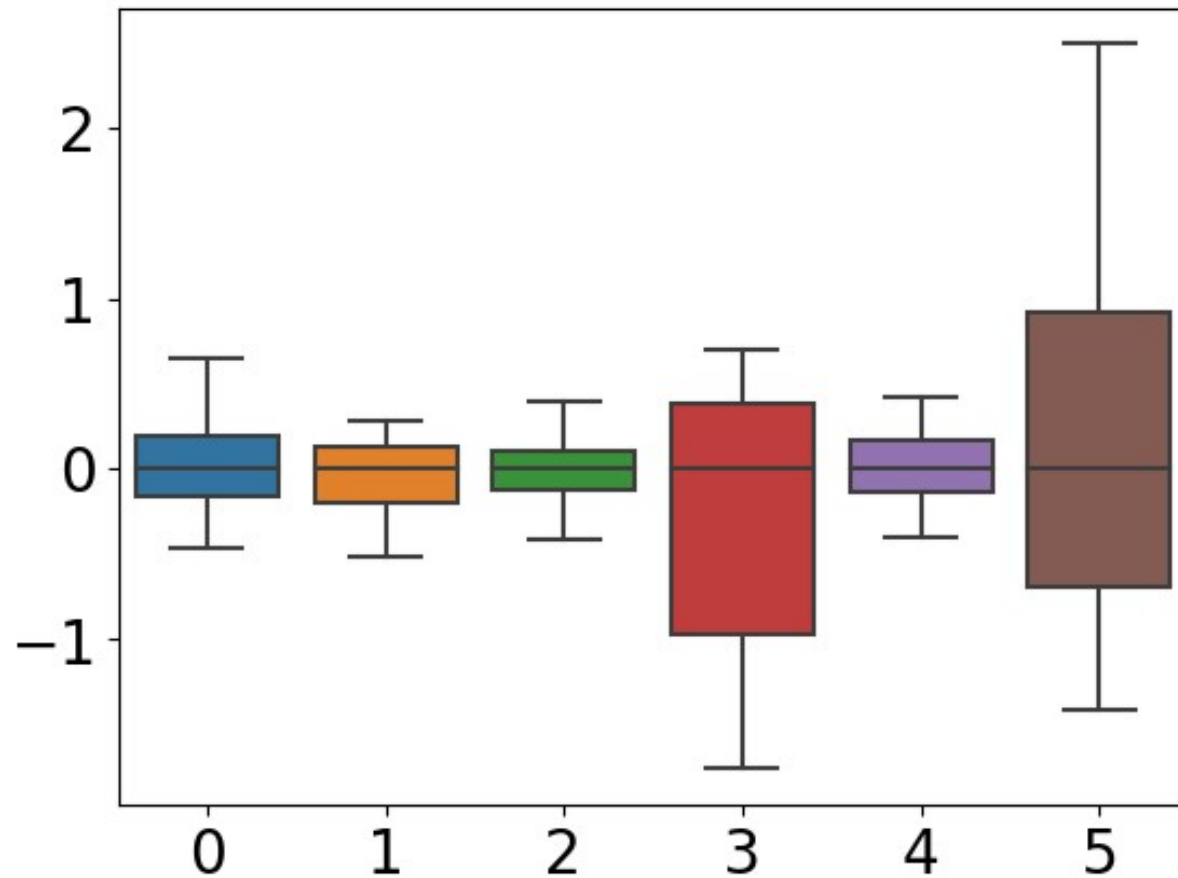
Classification

Modèle prédictif

Analyse des données

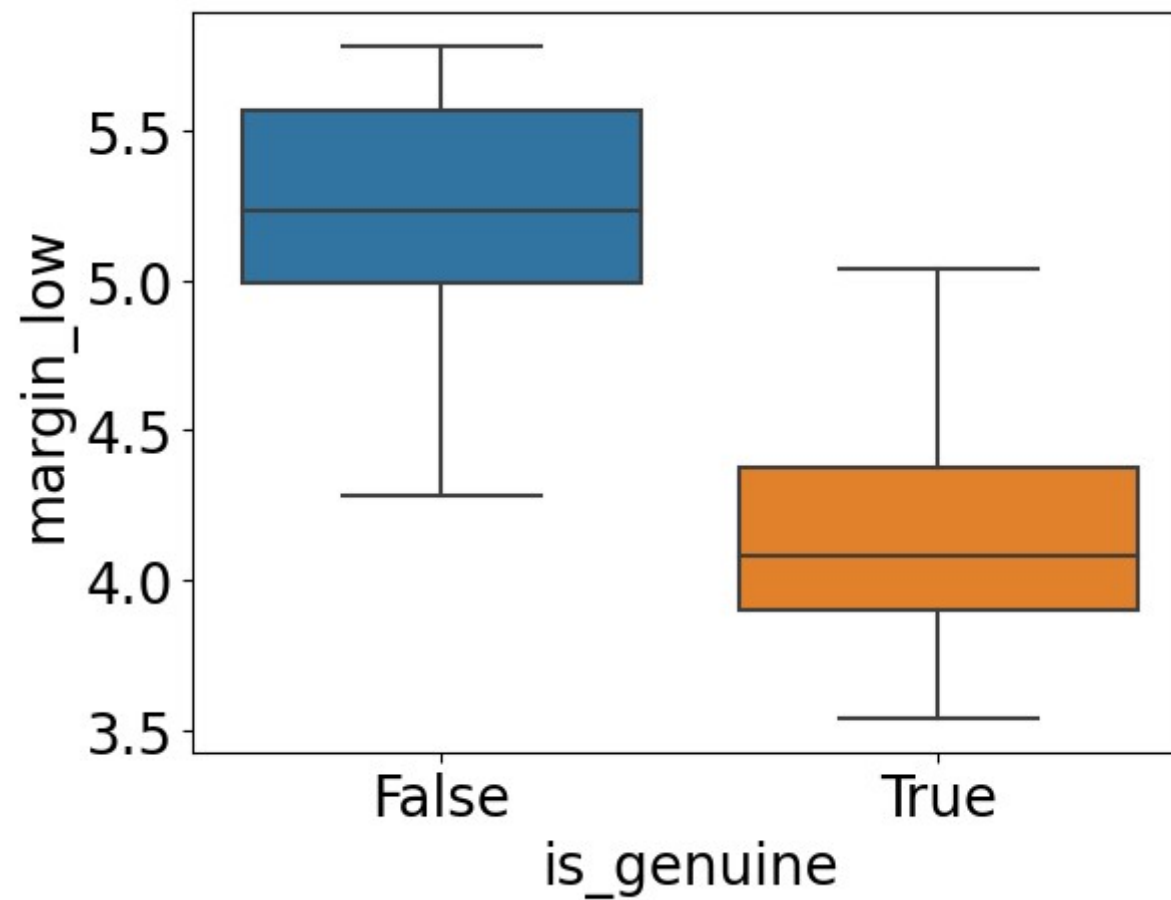
diagonal	1	0.1	0	0	0	0
height_left	0.1	1	0.5	0.2	0.1	0.2
height_right	0	0.5	1	0.3	0.1	0.2
margin_low	0	0.2	0.3	1	0	0.4
margin_up	0	0.1	0.1	0	1	0.3
length	0	0.2	0.2	0.4	0.3	1
	diagonal	height_left	height_right	margin_low	margin_up	length

Analyse des données

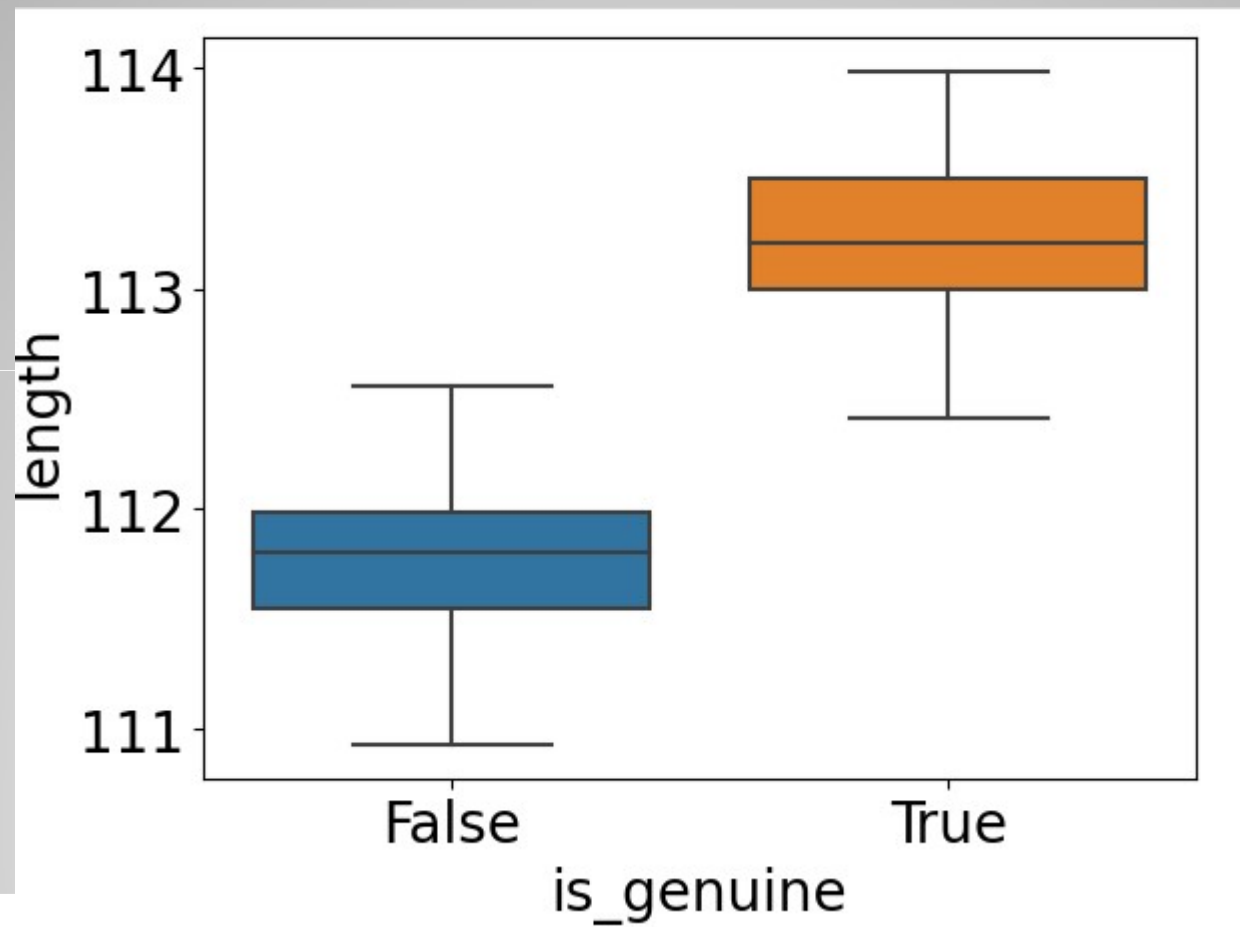


0 diagonal
1 height_left
2 height_right
3 margin_low
4 margin_up
5 length

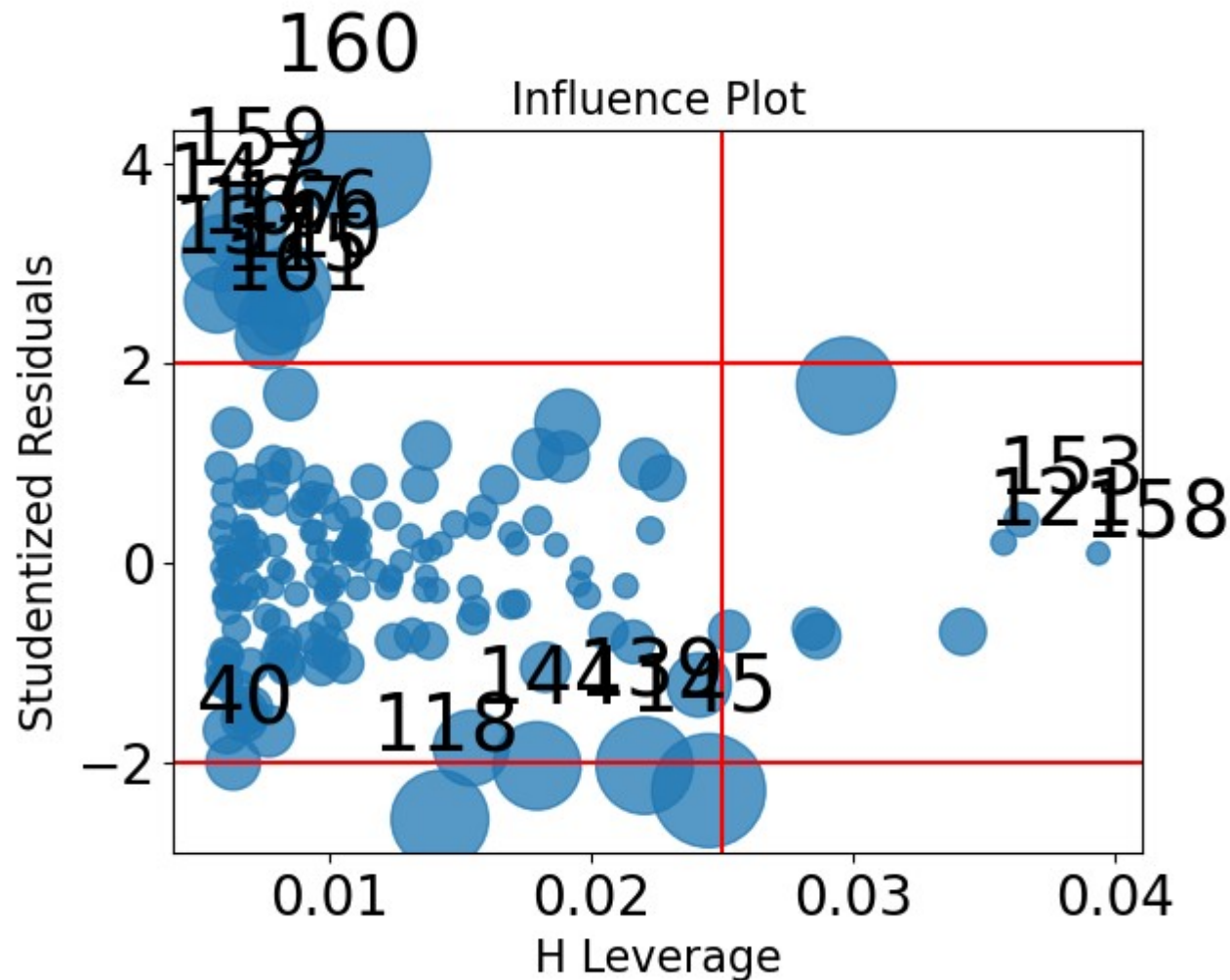
Analyse des données



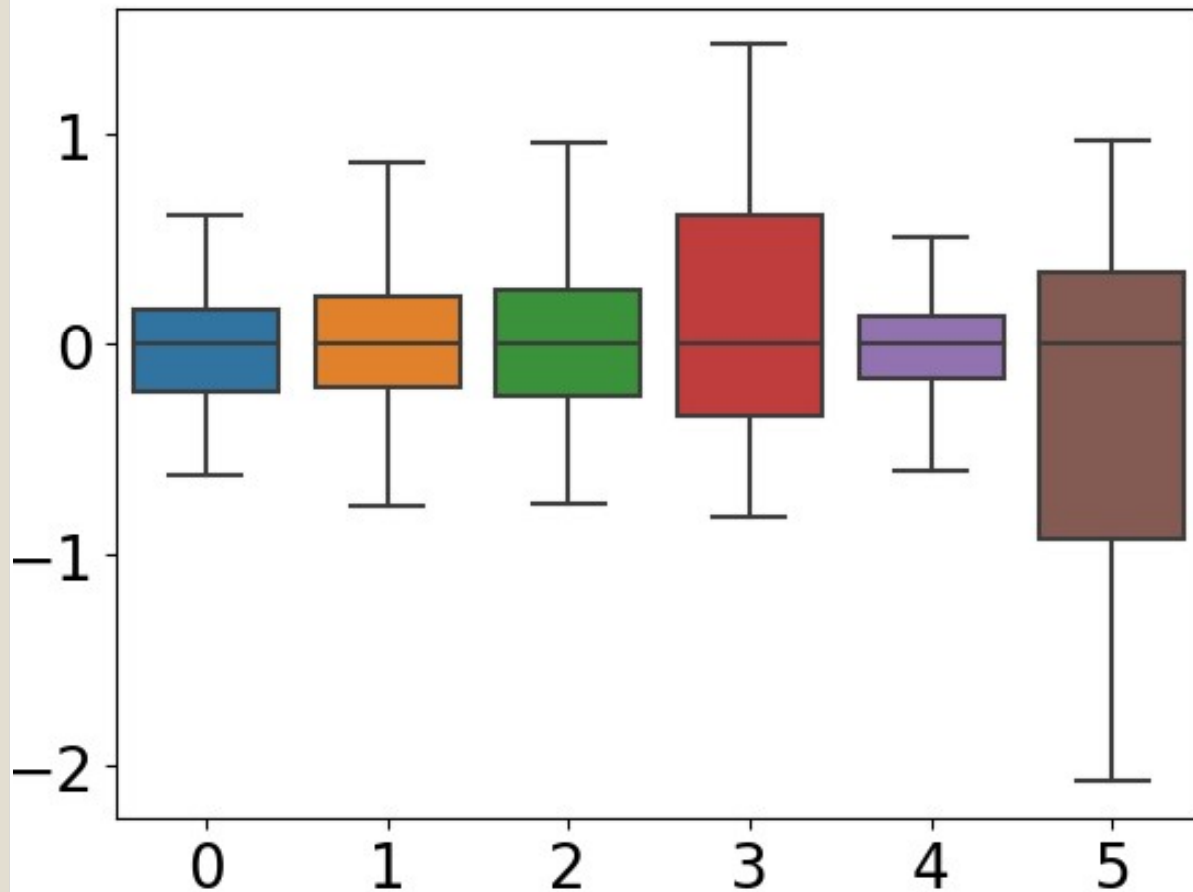
Analyse des données



Nettoyage des données

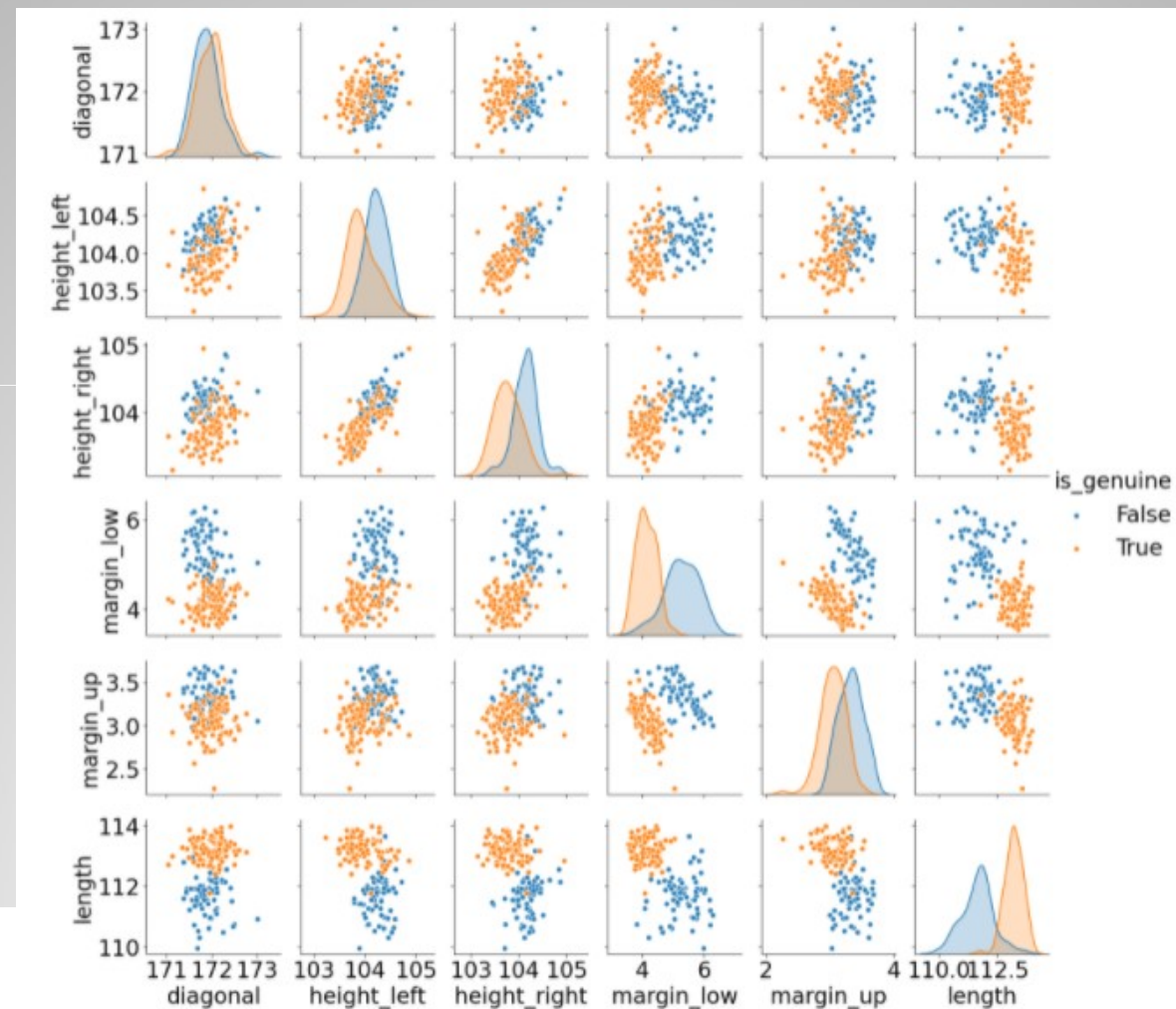


Analyse des données



0 diagonal
1 height_left
2 height_right
3 margin_low
4 margin_up
5 length

Analyse des données



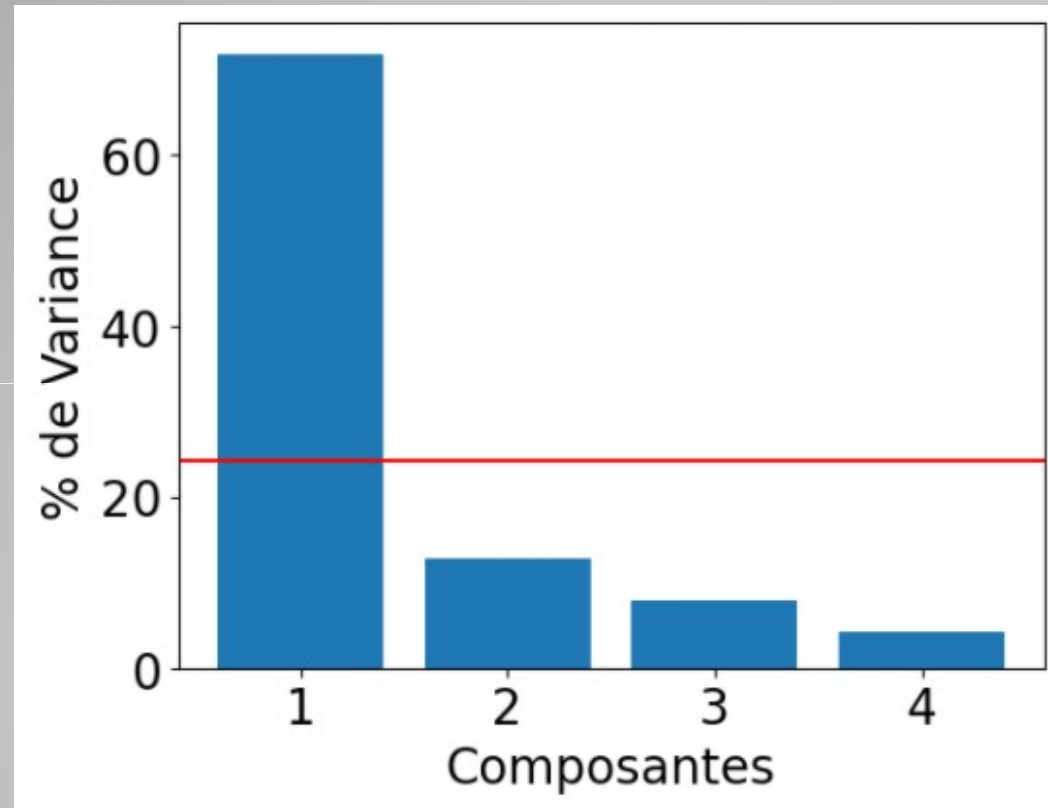
ACP

Objectif : Classer les billets

Comment : Créer un synthèse

Pourquoi : Comprendre les différences

ACP

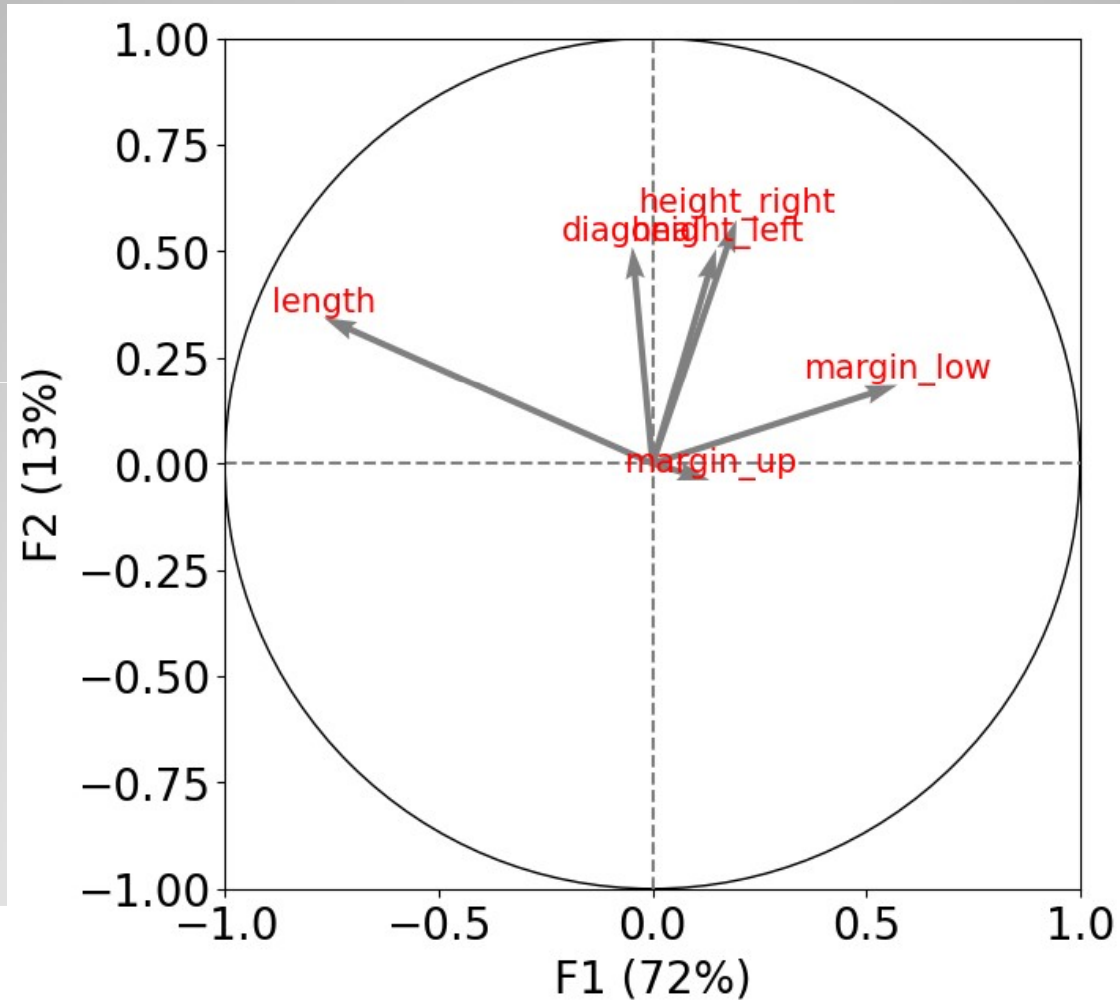


Avec l'ensemble des valeurs

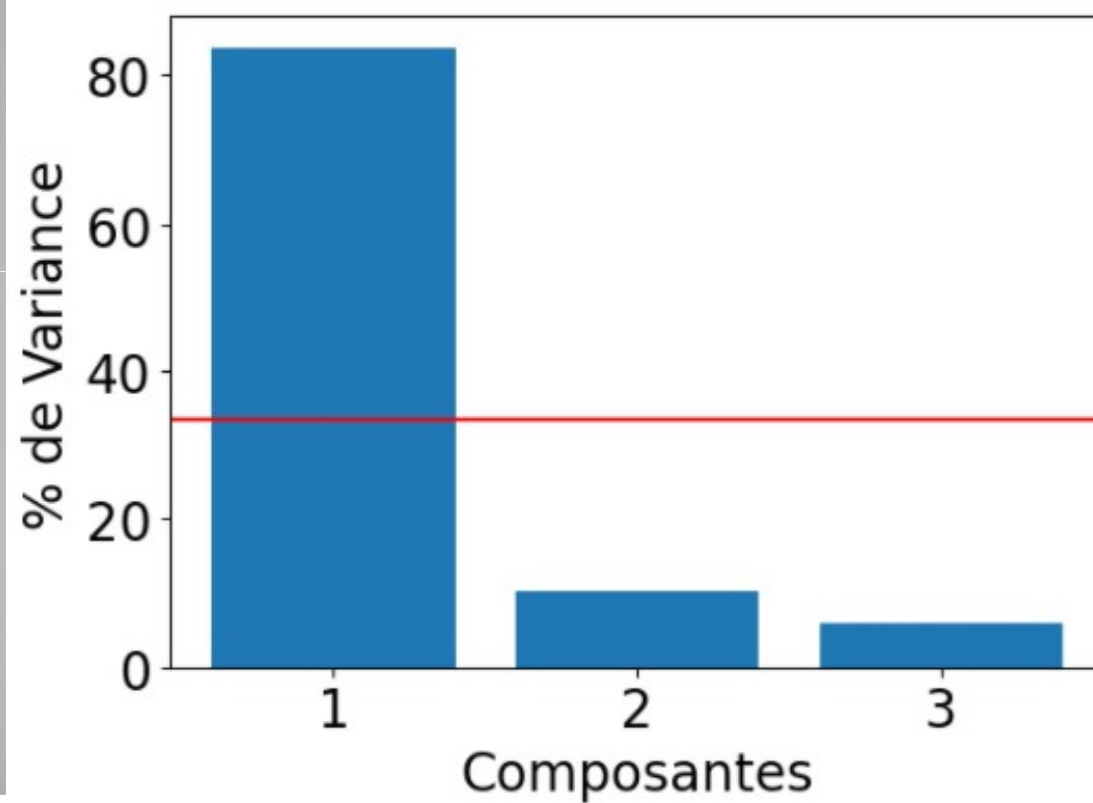
Choix du nombre de composante :

- Méthode du coude
- Kaiser (moyenne)

ACP



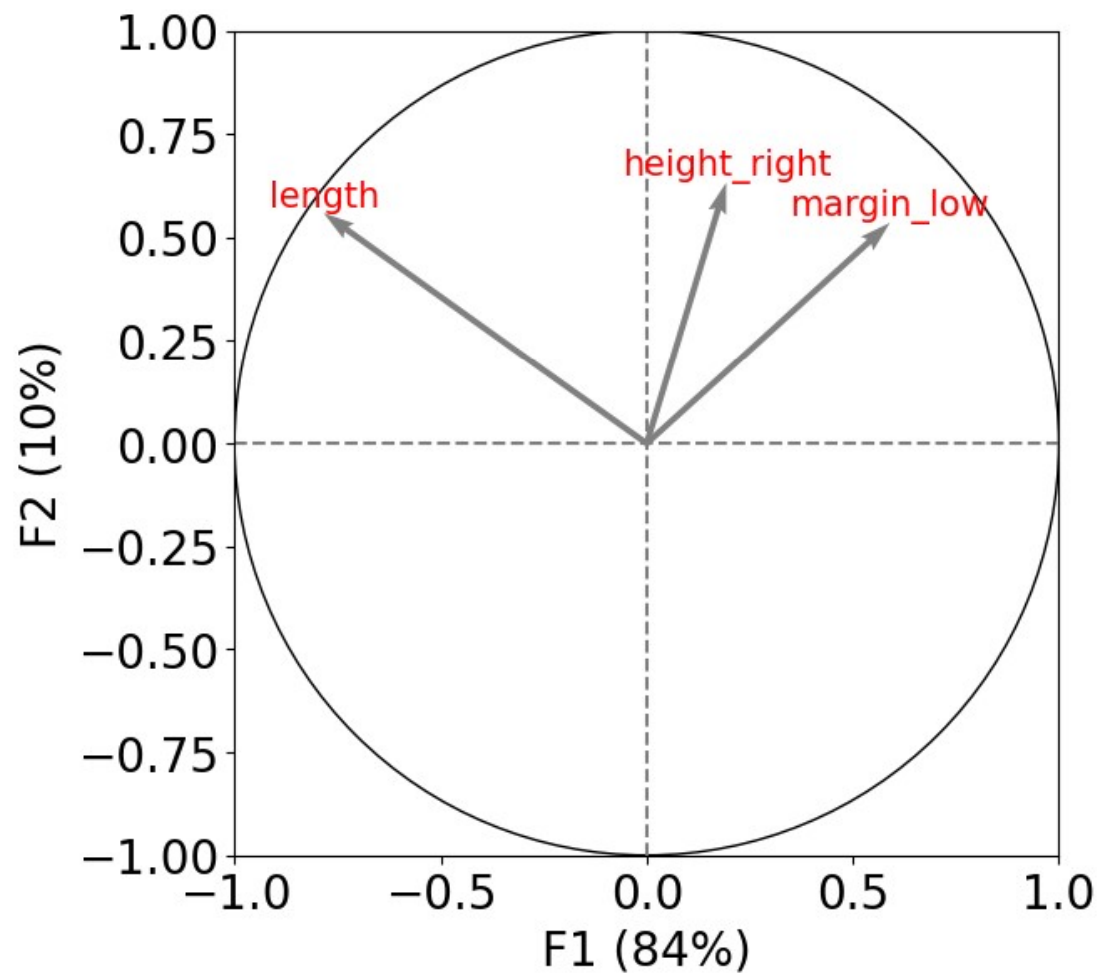
ACP



Avec les variables
Lenght, margin_low et
height_right

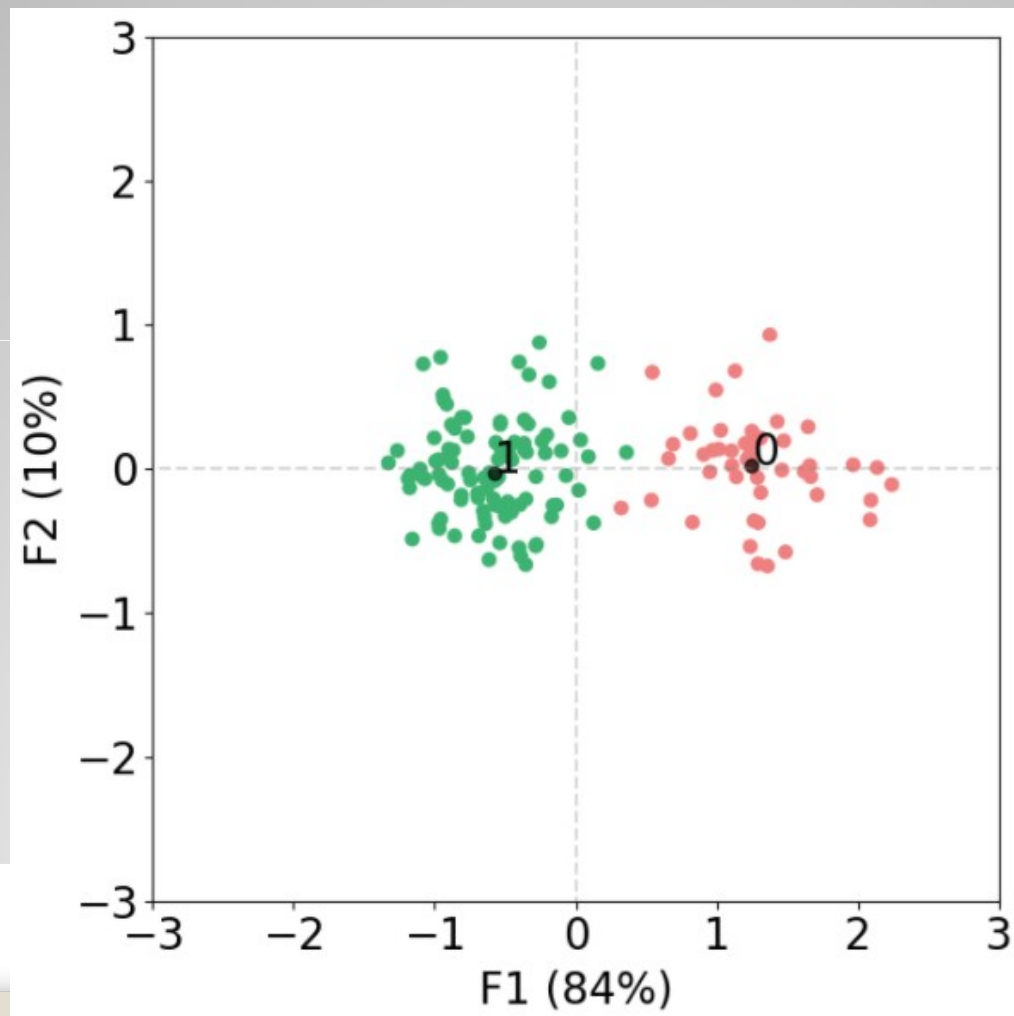
Choix du nombre de
composante :
-Méthode du coude
-Kaiser (moyenne)

ACP



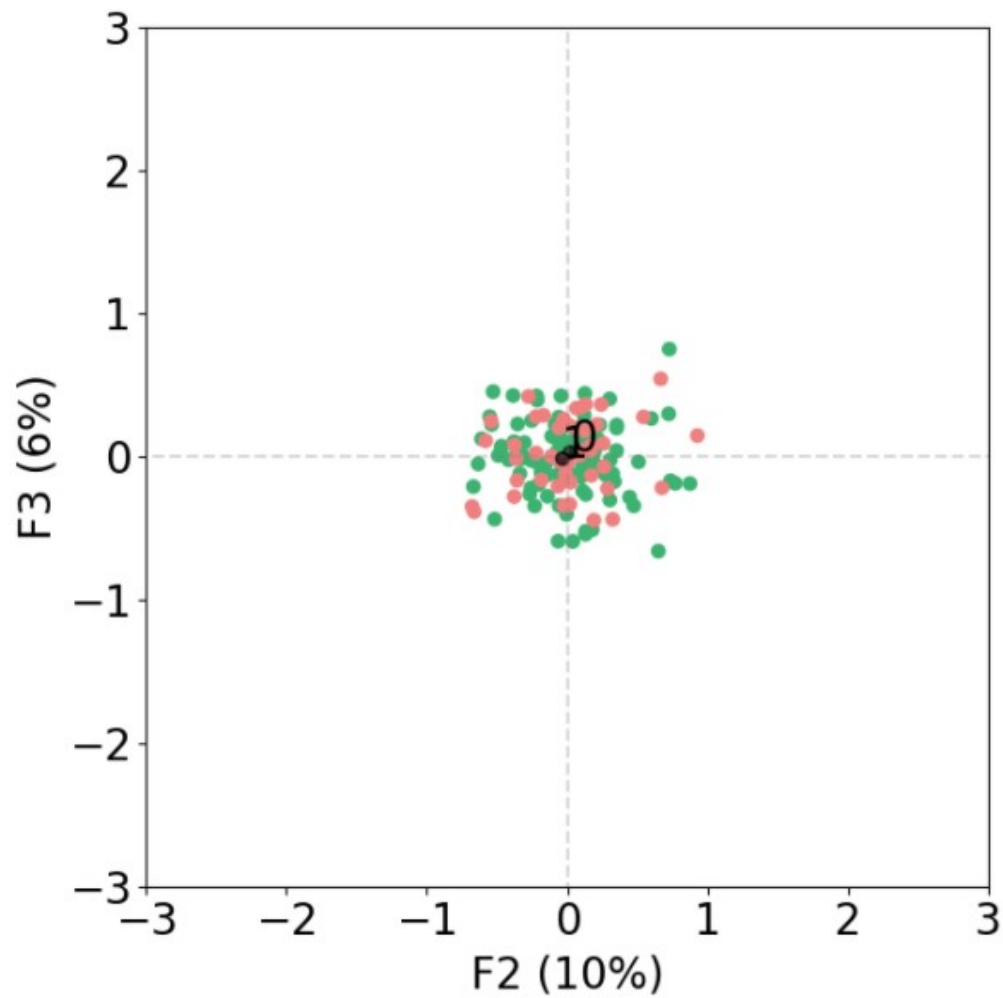
ACP

Premier plan Factoriel

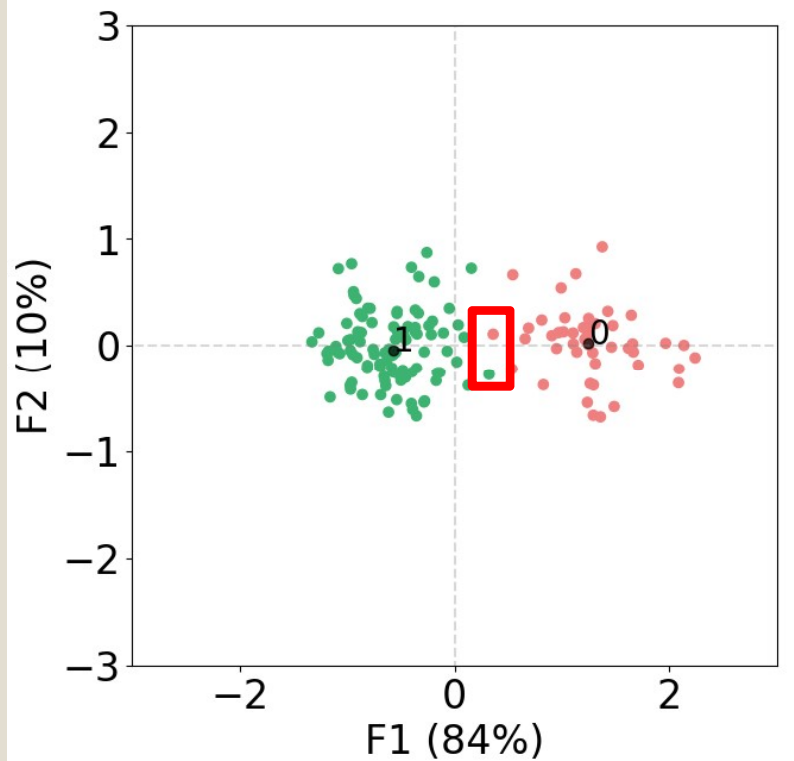


ACP

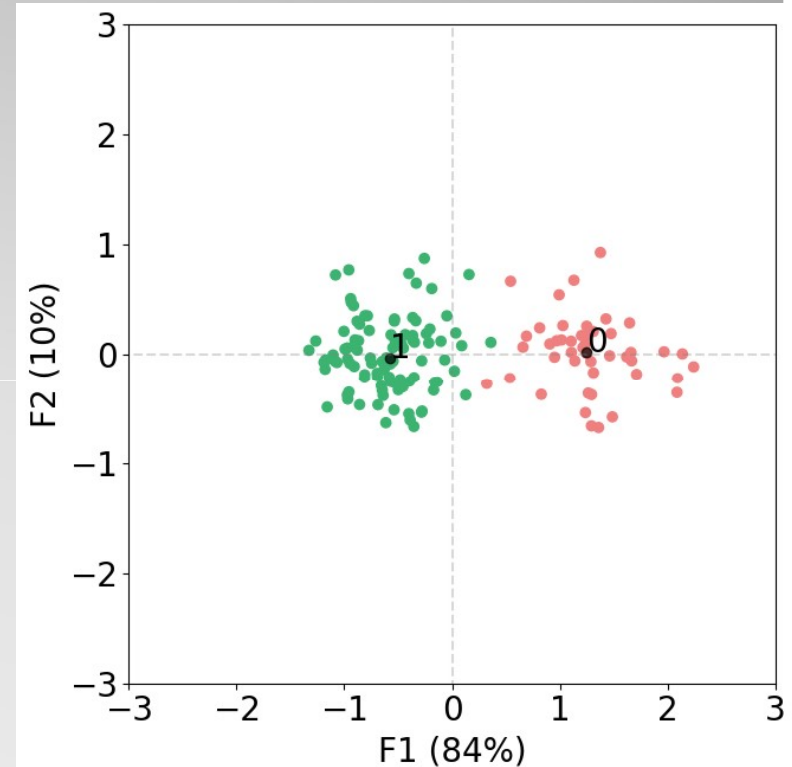
Second plan Factoriel



Classification

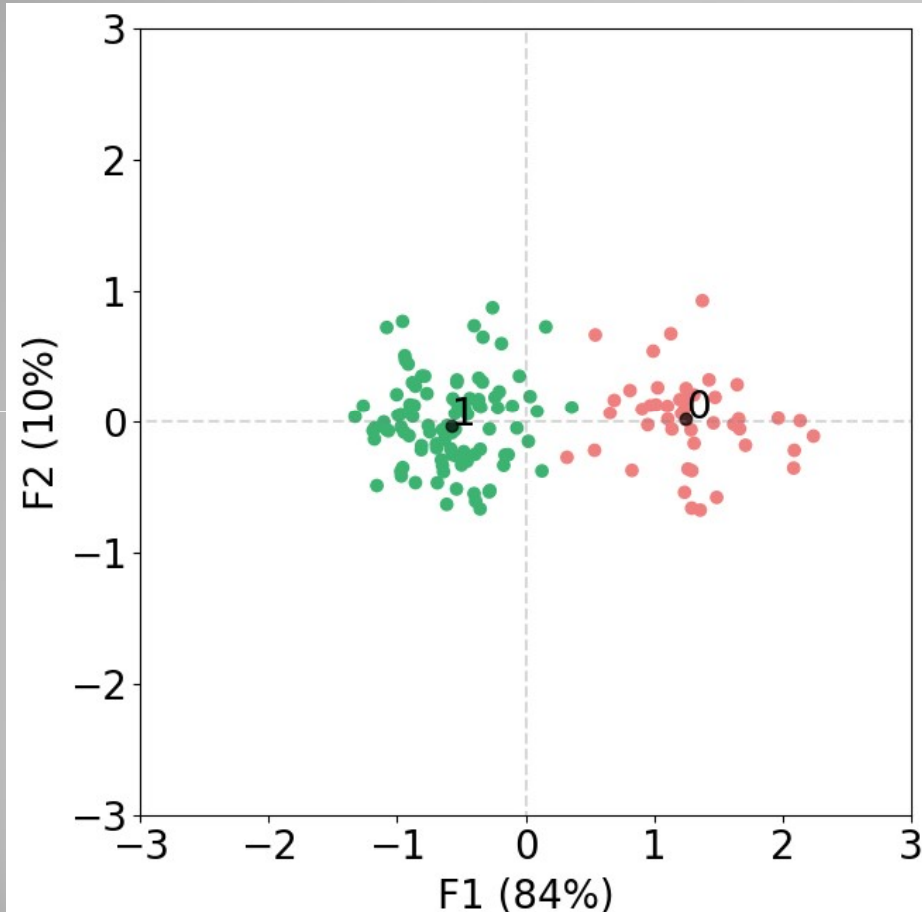


Représentation des Faux et des vrais billets avec la Prédiction Kmean (98% accuracy) dans le 1^{er} plan factoriel



Représentation des Faux et des vrais billets réel dans le 1^{er} plan factoriel

Classification



Classification avec les valeurs F1 et F2 et en utilisant la méthode de KNeighborsClassifier : Un faux positif dans les outliers éliminés

Mais la classification n'est pas capable de donner la probabilité d'un vrai/faux billet

Représentation des Faux et des vrais billets avec la Prédiction KNeighborsClassifier

(99.3% accuracy) dans le 1^{er} plan factoriel

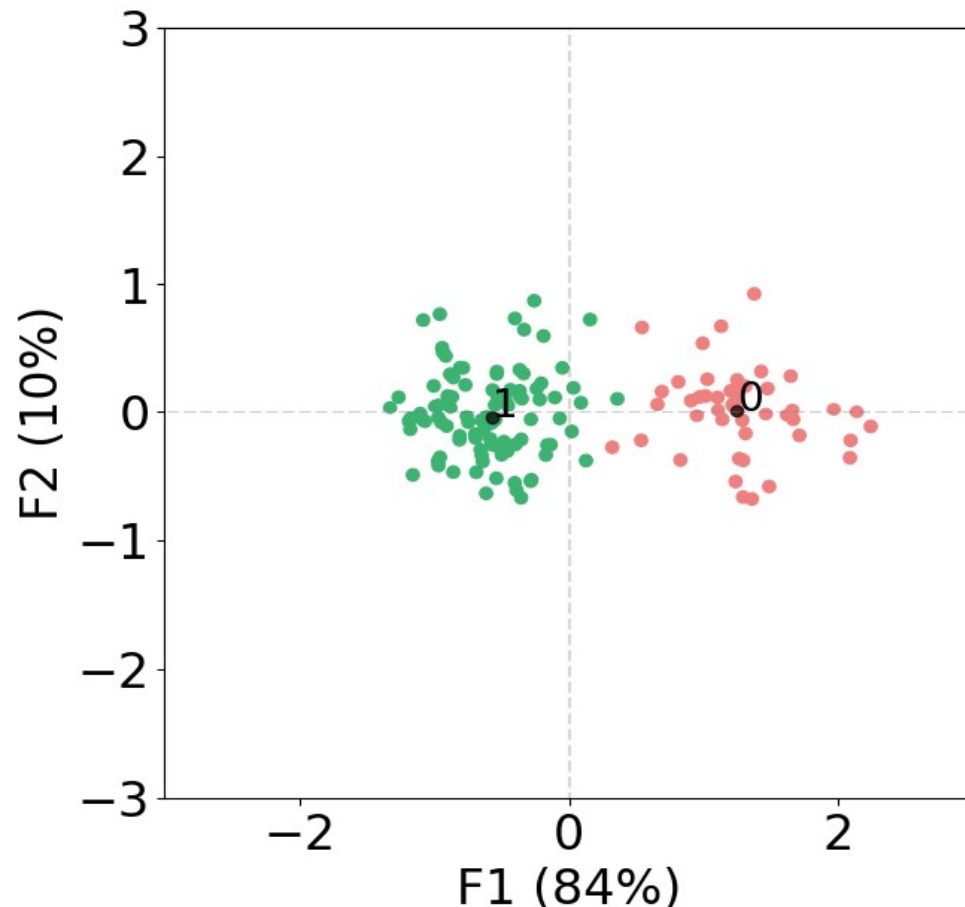
Modèle prédictif

Intérêt : Prédire qu'un événement soit vrai ou faux sous forme de probabilité

Comment : Utiliser les données numériques, les données étiquetées (Vrai/Faux)

Determiner : Si une probabilité est supérieur (ou égal) à 50%, il est prédit que c'est un vrai billet

Modèle prédictif



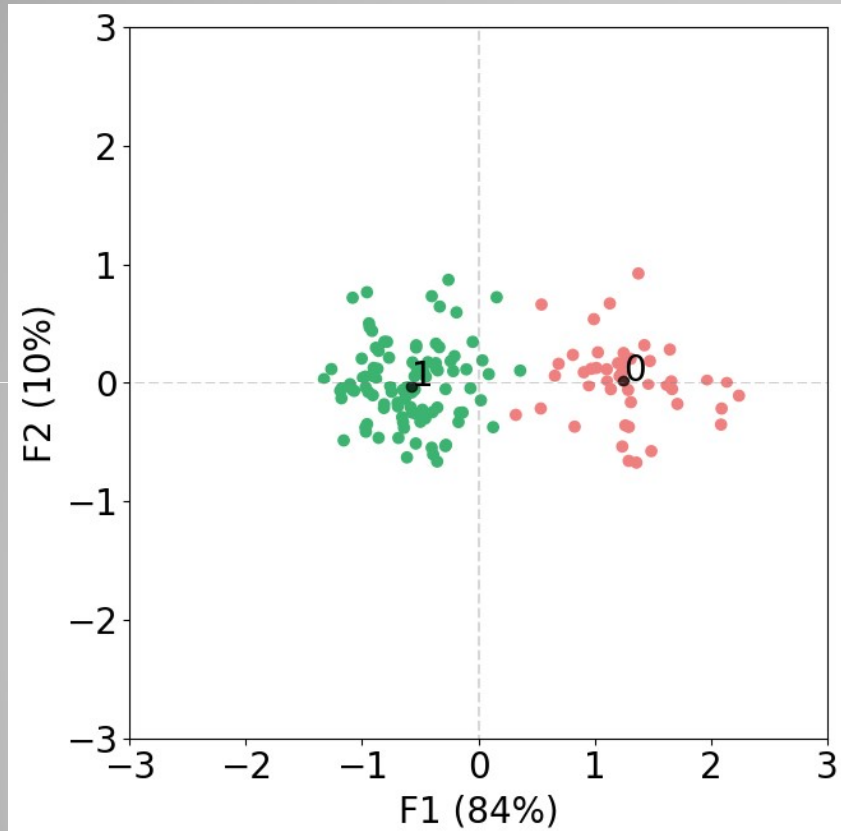
J'ai utilisé Gridsearch pour déterminer les meilleurs paramètres à utiliser dans la regression :

```
fit_intercept=True  
class_weight='balanced',  
penalty='l1',  
solver='liblinear'
```

Un faux positif

Représentation des Faux et des vrais billets avec la Régression logistique avec les valeurs F1 et F2 (99,3% accuracy) dans le 1^{er} plan factoriel

Modèle prédictif



Régression logistique
LogisticRegression
"margin_low", "diagonal",
"length", "margin_up"

100% accuraccy

Représentation des Faux et des vrais billets
avec la Régression logistique avec les
valeurs réelles des billets dans le 1^{er} plan
factoriel