

PRESENTATION

Détecteur : faux billets

SOMMAIRE

MISSION

PRE - TRAITEMENT

Description

Données retirées

ACP : Synthèse

Vérification : Kmeans

MODELISATION

Préparation

Echantillonnage

Réglage du modèle

Vérification

MISSION

OBJECTIF	Détecter des faux billets
MISSION	Algorithme de détection
CONDITION	si probabilité $\geq 50\%$ alors VRAI
DONNEES	Police Judiciaire

DESCRIPTION

is_genuine

Etiquette authenticité d'un billet :

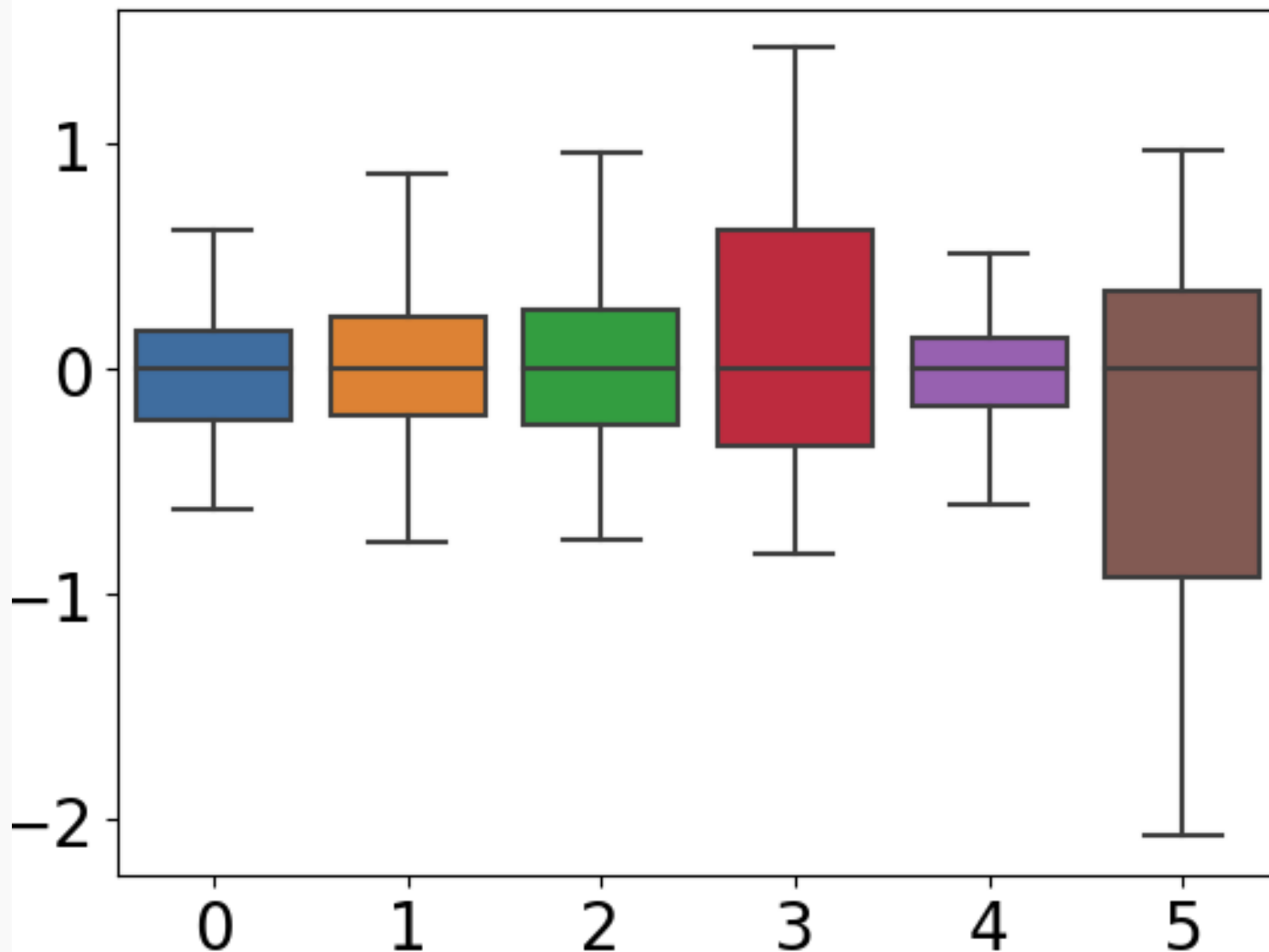
- 98 Vrai, 46 Faux

6 caractéristiques numériques

Dimensions exprimées en mm

DESCRIPTION

Différences de variation



Grande variabilité

3 margin_low

5 length

Variabilité similaire

0 diagonal

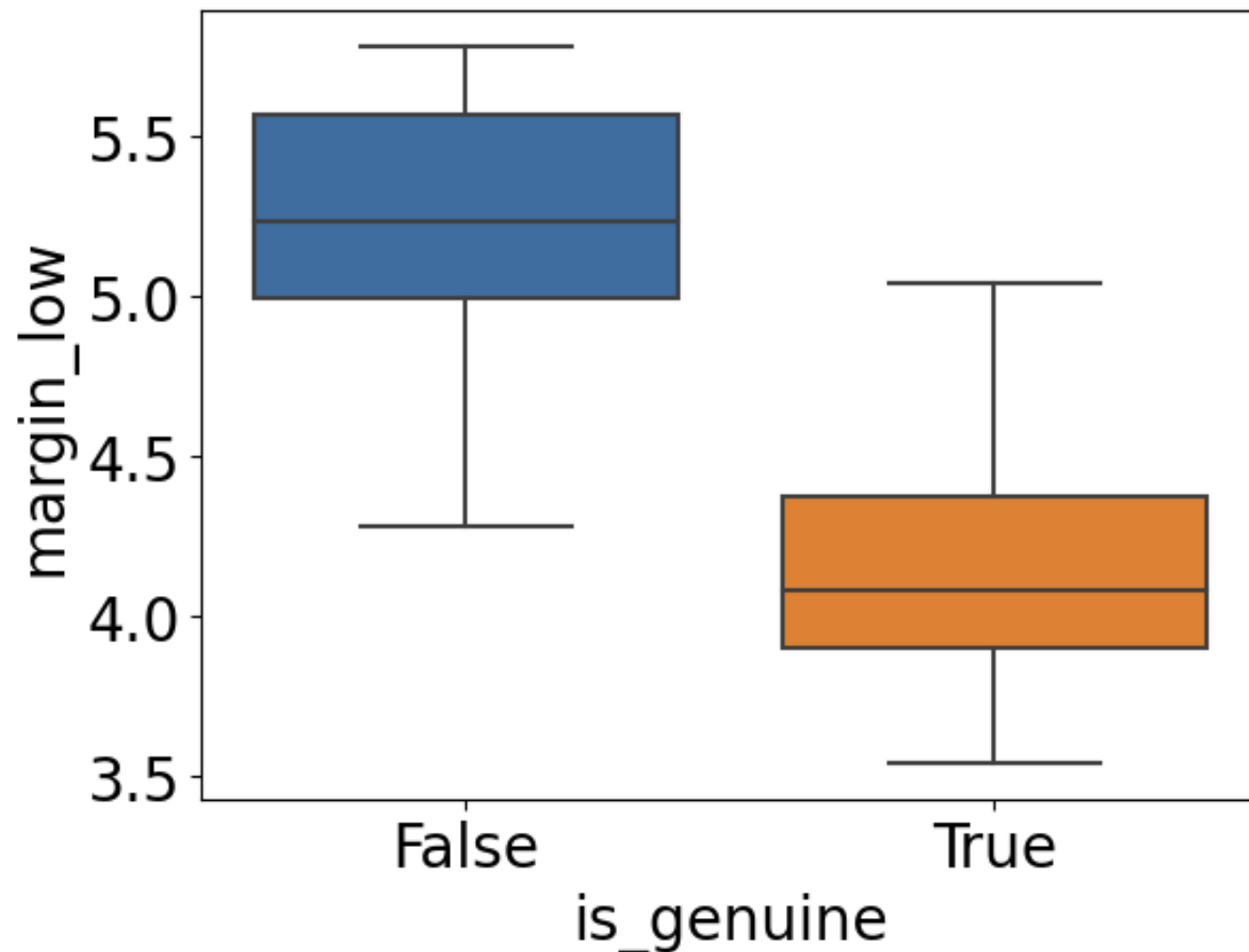
1 height_left

2 height_right

4 margin_up

DESCRIPTION

margin_low par catégorie

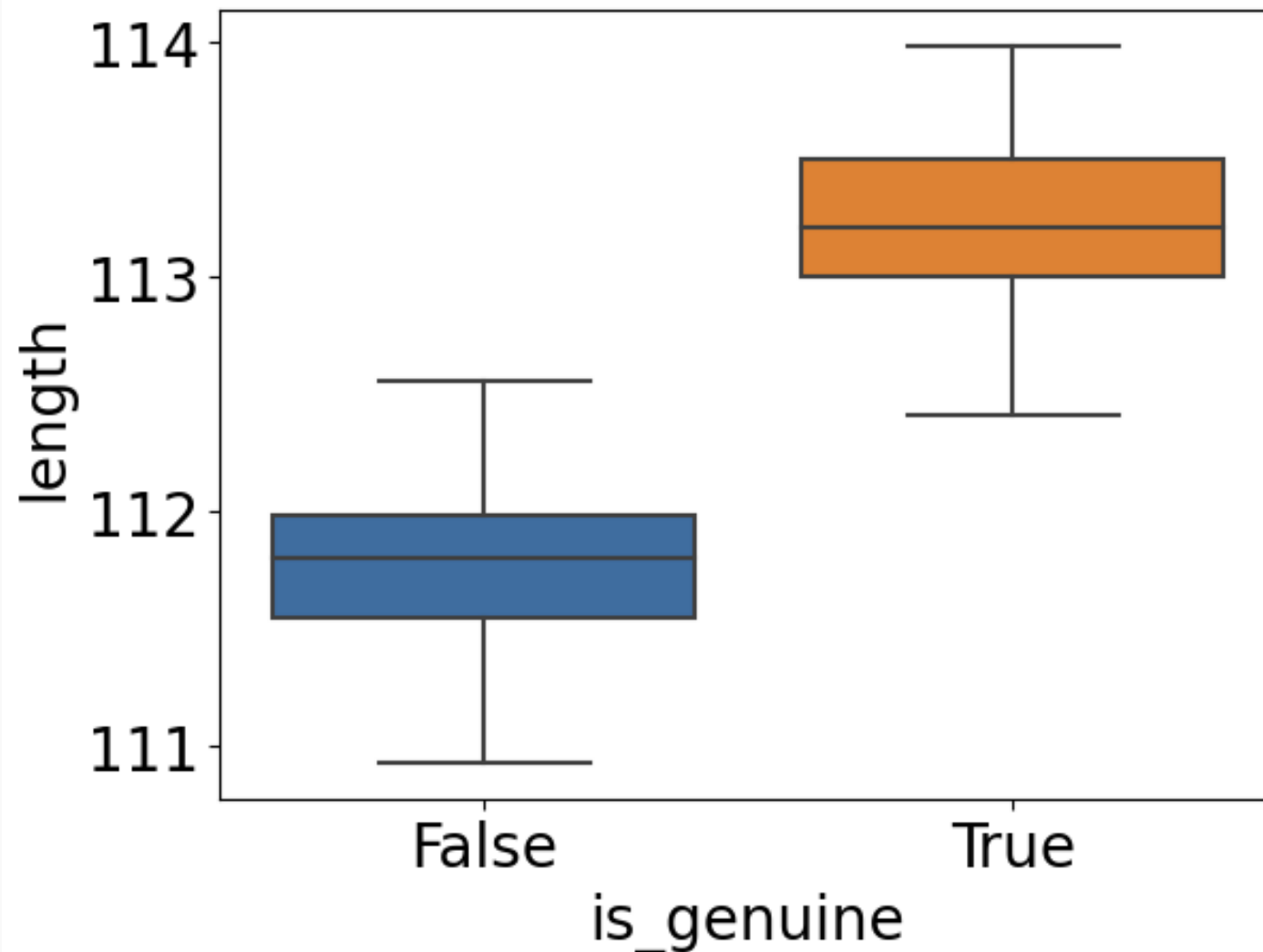


Interprétation

faux billets plus longs

DESCRIPTION

length par catégorie



Interprétation

faux billets plus courts

DESCRIPTION

Relations entre caractéristiques

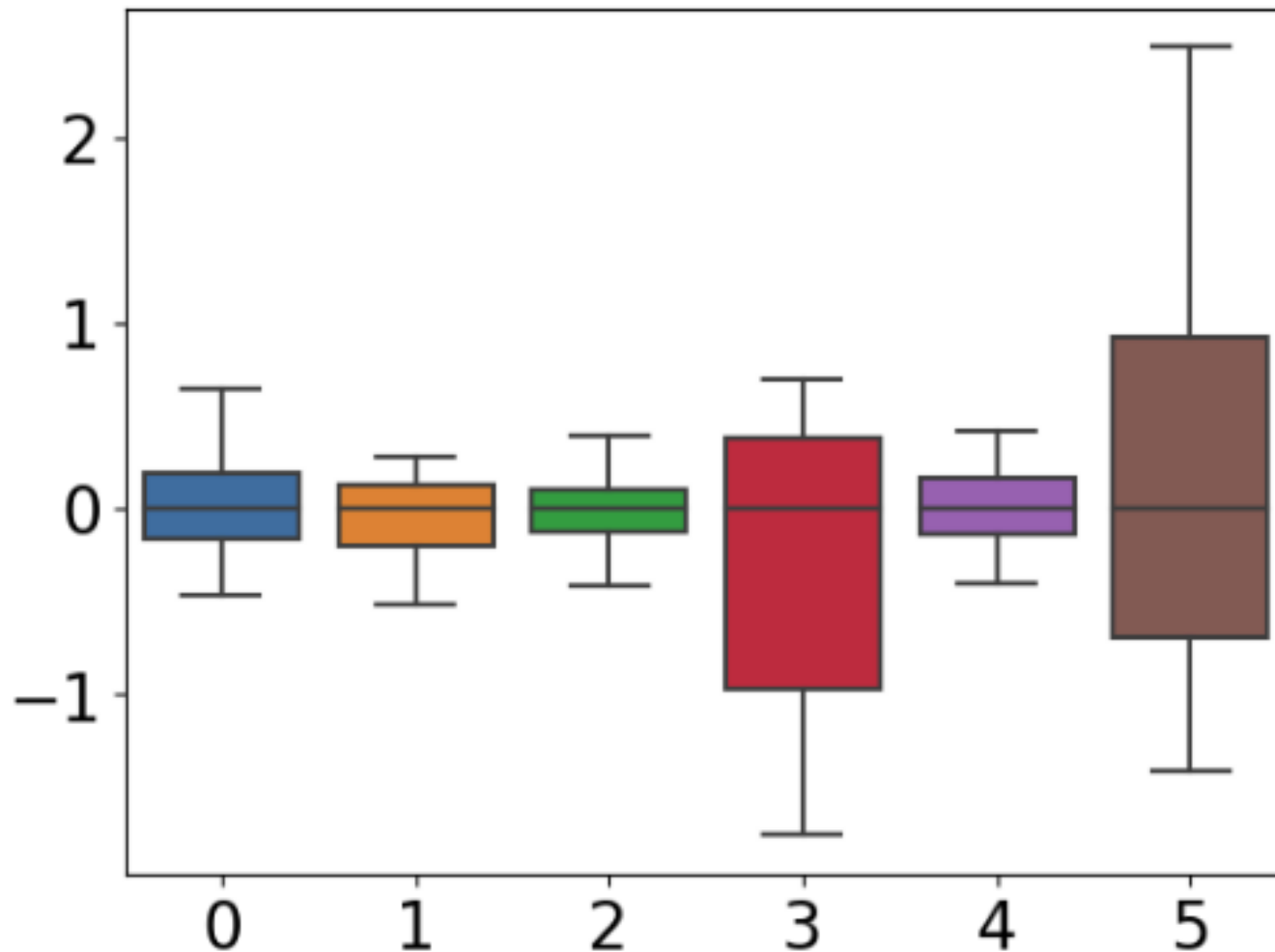
diagonal : pas de relation linéaire ($R^2=0$)

is_genuine : margin_low et length (R^2 : 0.7 et 0.8)

height_left/height_right : relation ($R^2=0.5$)

DONNEES RETIREES

Différences de variations



Interprétation

3 margin_low

5 length

Variations inverses

ACP : SYNTHÈSE

OBJECTIF : Classer les billets

COMMENT : créer une synthèse

POURQUOI : comprendre les différences

ACP : SYNTHÈSE

CENTRER par la médiane

COMMENT

données - médiane des données

POURQUOI

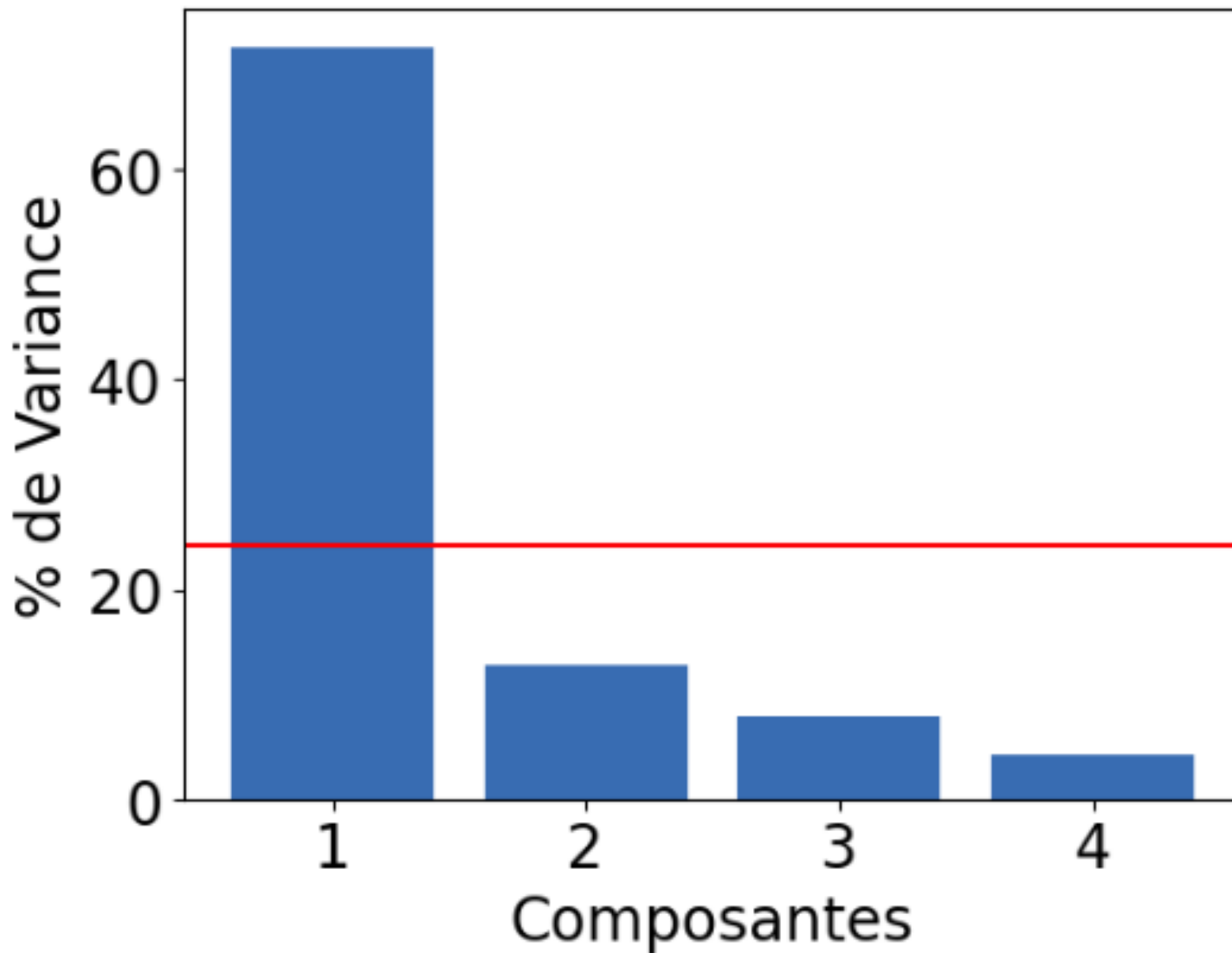
non sensible aux données atypiques

PAS DE REDUCTION ?

**même unité de mesure : comparaison possible
garder la variabilité des caractéristiques**

ACP : SYNTHÈSE

EBOULIS DES VALEURS PROPRES (aperçu initial)



Création des axes

- distances
- perpendiculaires

Choix du nombre

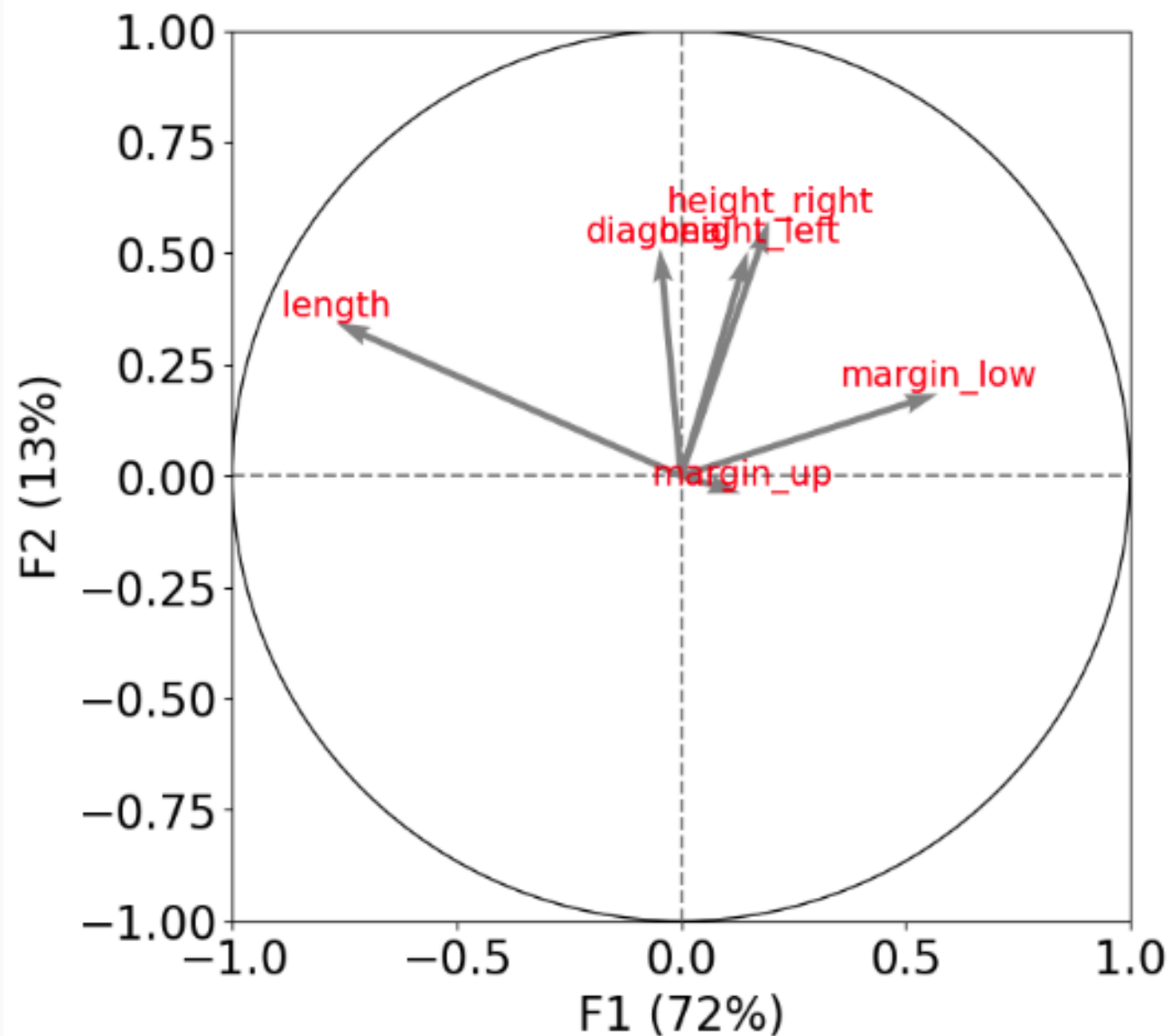
- méthode du coude
- kaiser (moyenne)

Conclusion

- F1 et F2

ACP : SYNTHÈSE

CERCLE DES CORRELATIONS (aperçu initial)



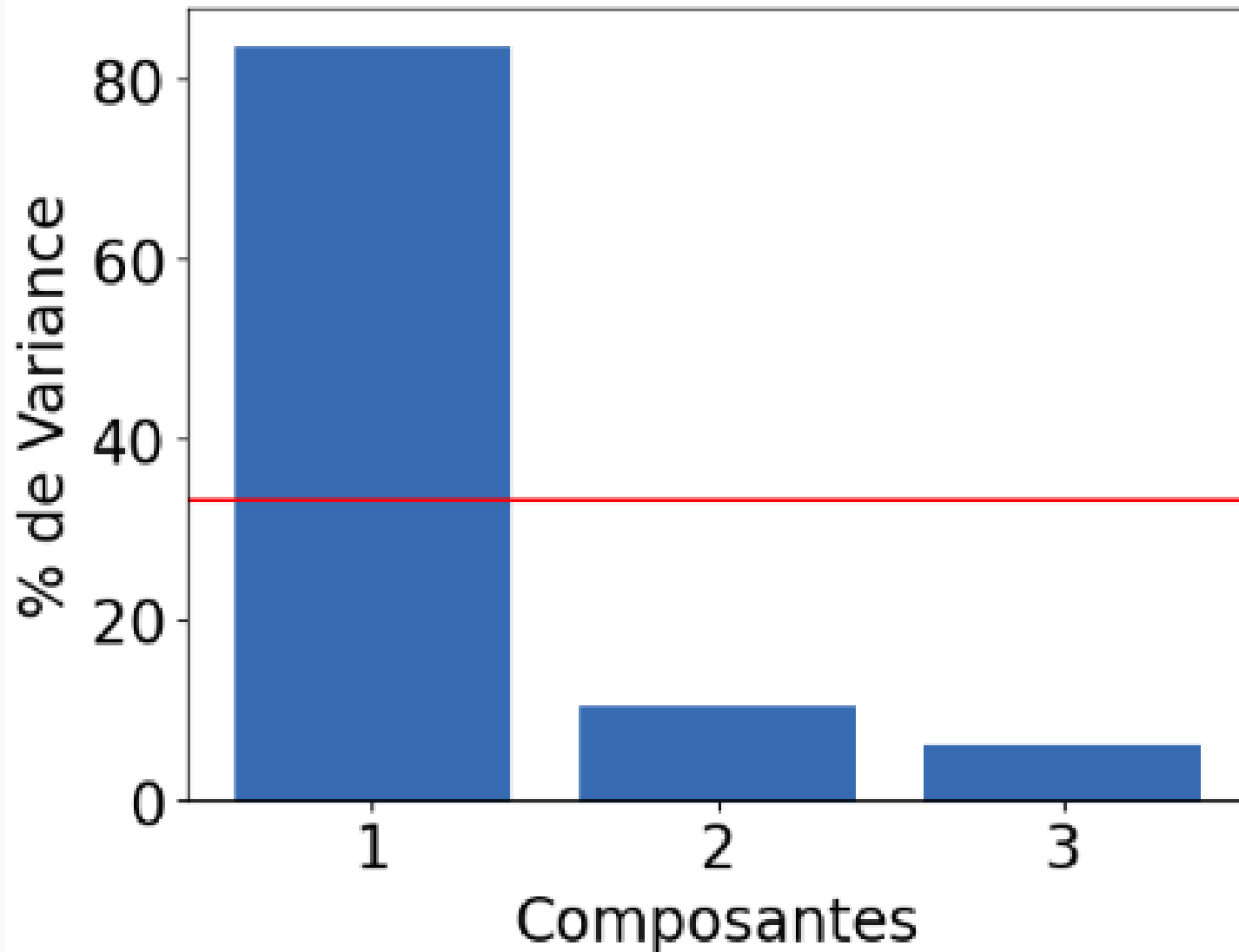
Mauvaise projection
margin_up

Redondance
height_left

Non corrélé
diagonal

ACP : SYNTHÈSE

EBOULIS DES VALEURS PROPRES

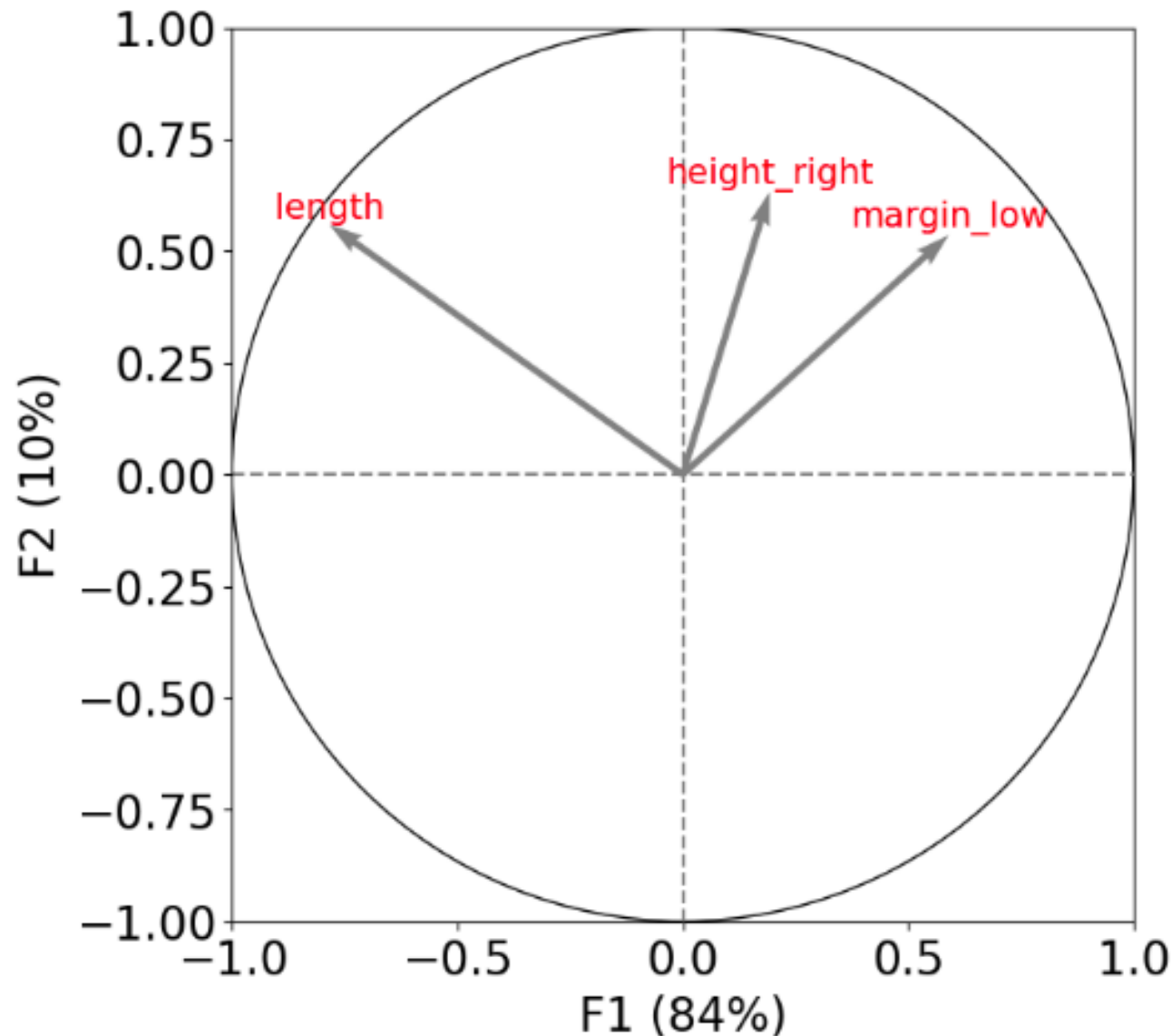


Choisir le nombre
- méthode du coude
- kaiser (moyenne)

Conclusion
- F1 et F2

ACP : SYNTHÈSE

CERCLE DES CORRELATIONS

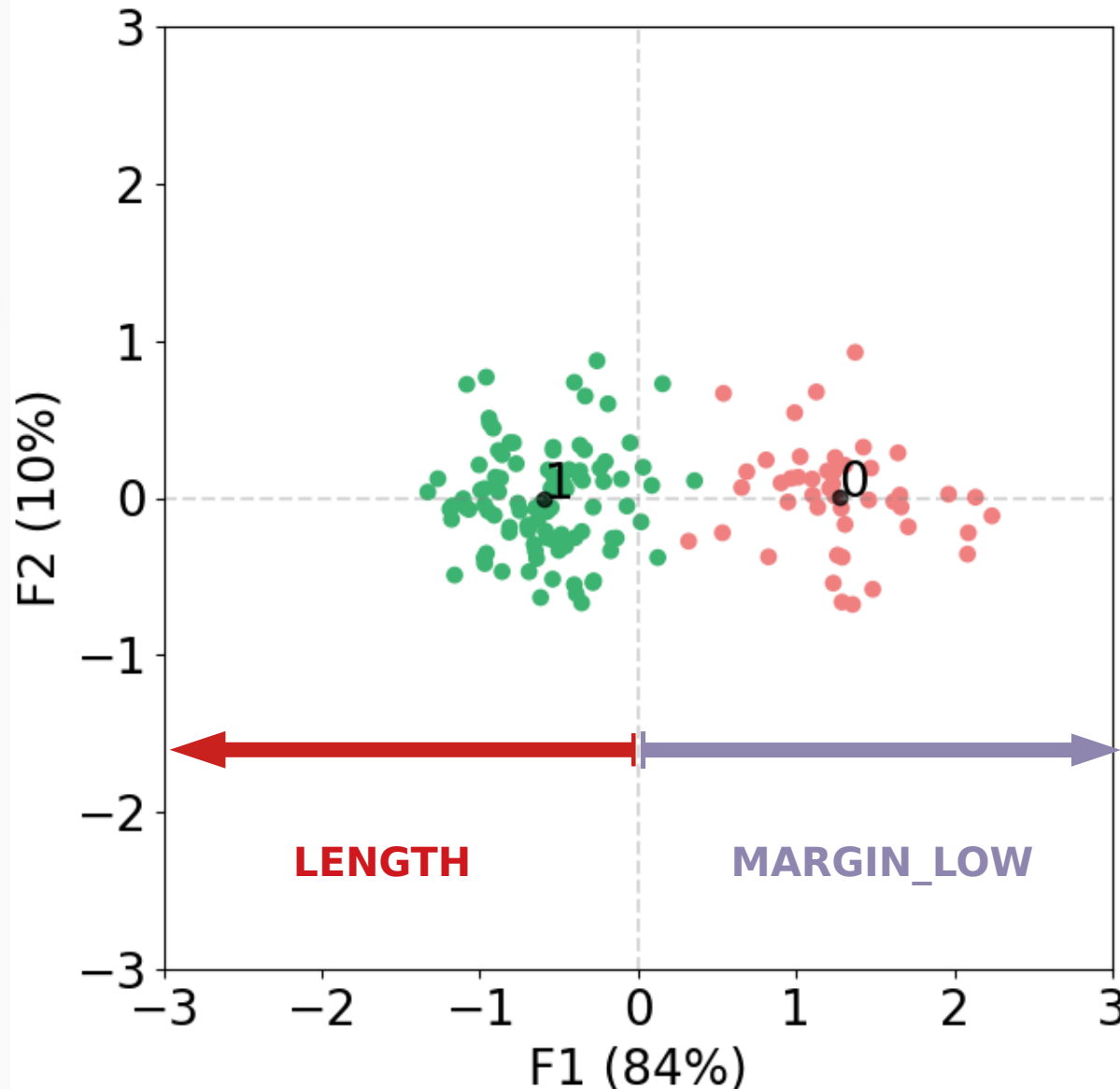


Variables retirées
margin_up
height_left
diagonal

Projection
Angle (axe,
caractéristique)
=
Coefficient de
corrélacion

ACP : SYNTHÈSE

PREMIER PLAN FACTORIEL



CONTRIBUTION

1 Vrais Billets

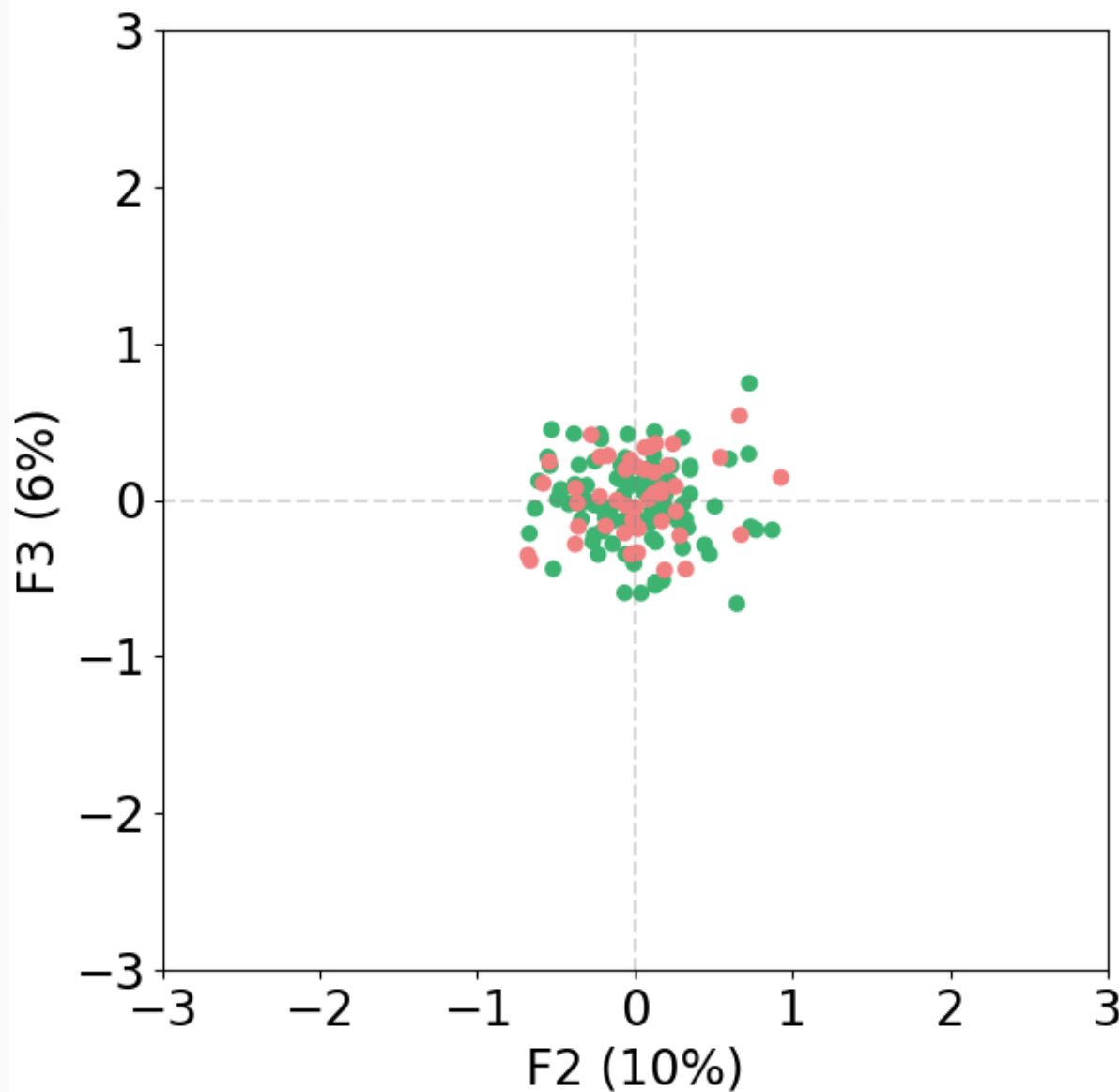
0 Faux Billets

QUALITE

94 % différences

ACP : SYNTHÈSE

SECOND PLAN FACTORIEL



QUALITE
6 % de ressemblances

VERIFICATION : KMEANS

Classer k groupes avec leurs moyennes

COMMENT

Sélectionner et positionner k moyennes

Affecter à la moyenne la plus proche

Déplacer la moyenne

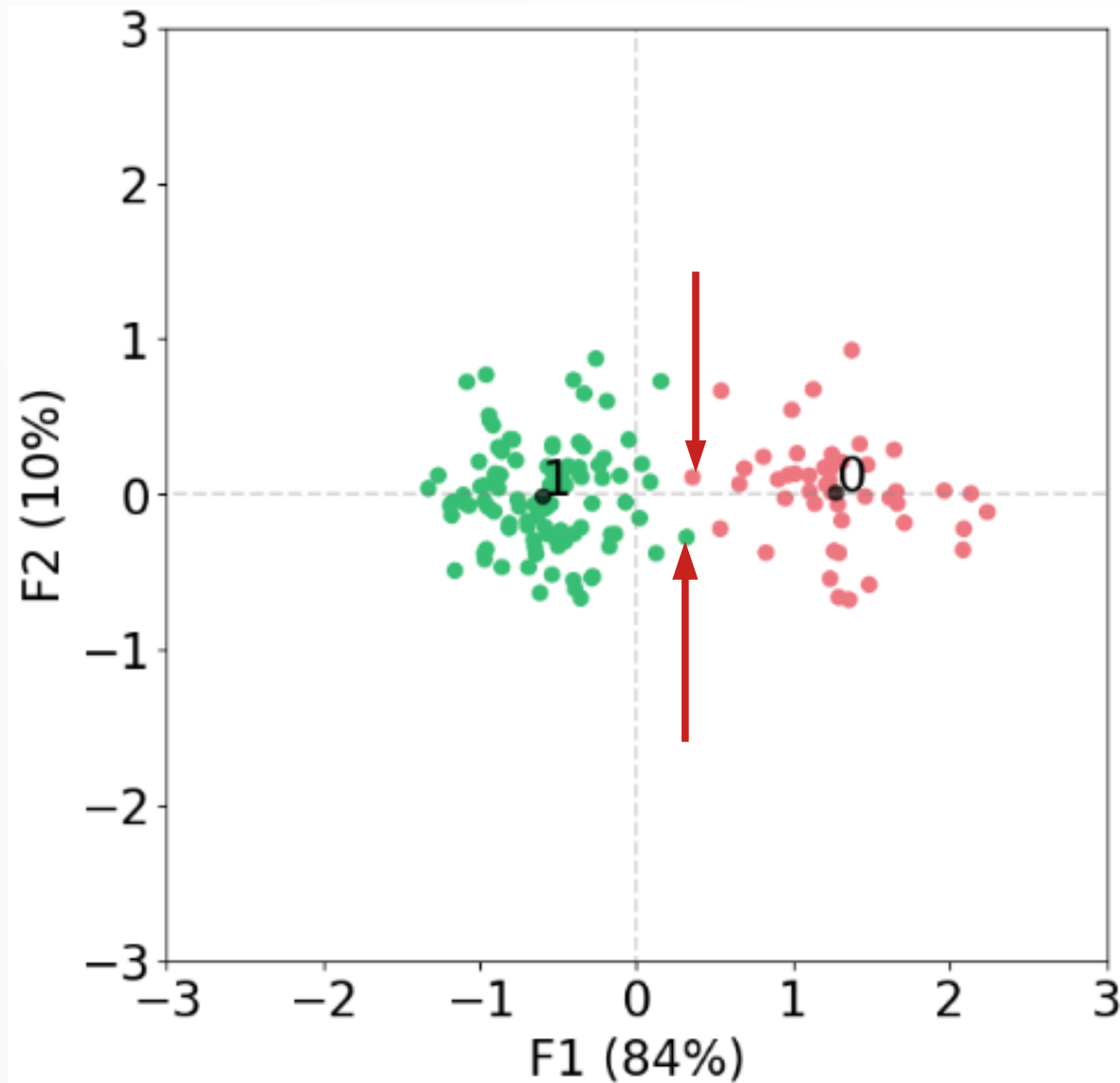
Choisir la meilleure fois

POURQUOI

Vérifier l'affectation point-moyenne

VERIFICATION : KMEANS

PREMIER PLAN FACTORIEL



Résultats

1 Faux Positif

1 Faux Négatif

Interprétation

**Moyenne plus
proche**

VERIFICATION : KMEANS

PROBLEME RENCONTRE

PROBLEME

Résultat du kmeans aléatoire :

- classement identique**
- le numéro de groupe varie**

SOLUTION

- Fixer l'état aléatoire**
- Affecter le numéro de groupe**

REGRESSION LOGISTIQUE

INTERET

- **Prédire qu'un événement soit Vrai ou Faux sous forme de probabilité**
- **Donner l'utilité prédictive d'une caractéristique**

DETERMINATION

- **données numériques**
- **données étiquetées Vrai ou Faux**
- **décision : Vrai si probabilité $\geq 50\%$**

PREPARATION

INTERET

- résultats du modèle = reflet des données
- Lisser F2 pour faciliter l'apprentissage

QuantileTransformer

- transformer en une distribution normale
- utilisation des quantiles

ECHANTILLONAGE

RESPECTER LA CONDITION probabilité 50 %

- compatibilité à une loi de Bernoulli
- classes homogènes

SEPARER LES DONNEES

- 80 % pour l'apprentissage
- 20 % pour la vérification

FAIRE UNE VALIDATION CROISEE

- faire n échantillons du jeu d'apprentissage
- faire une moyenne des scores

REGLAGE DU MODELE

BUT

- améliorer la qualité de la prédiction

OPTIMISER LES HYPERPARAMETRES

- créer une liste de paramètres
- validation croisée des résultats

VERIFICATION

matrice de confusion

sensibilité

taux de positifs correctement classés

spécificité

taux de faux billets correctement classés

score AUROC

capacité à classer correctement les billets

CONCLUSION

SYNTHESE DES DONNEES

CLASSIFICATION KMEANS

REGRESSION LOGISTIQUE

ESTIMATION DU MODELE

QUESTIONS
