



Lena Verboom



Détection de faux billet

Projet 6 : Détectez des faux billets



Contexte

- ❑ Mission pour le ministère de l'intérieur



- ❑ **Lutte contre la criminalité** en réprimant la fausse monnaie



- ❑ **Stratégie** : Créer un **algorithme** de détection de faux billets



- ❑ Analyse des caractéristiques **billets** (Longueur, hauteur, diagonal, marge...)



Quelles sont les caractéristiques des vrais et faux billets ?



Analyse
données

ACP

K-means

Régression
logistique

Conclusion

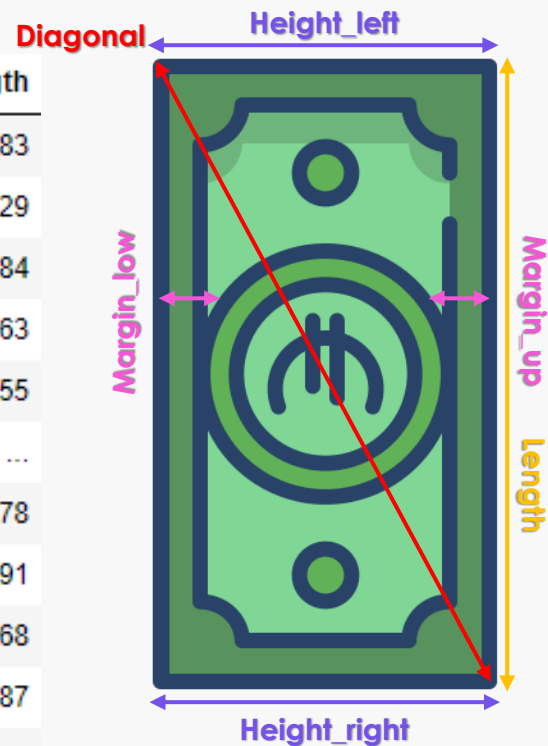
Données à ma disposition :

| | 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 | 171.67 | 103.74 | 103.70 | 4.01 | 2.87 | 113.29 |
| 2 | True | 171.83 | 103.76 | 103.76 | 4.40 | 2.88 | 113.84 |
| 3 | True | 171.80 | 103.78 | 103.65 | 3.73 | 3.12 | 113.63 |
| 4 | True | 172.05 | 103.70 | 103.75 | 5.04 | 2.27 | 113.55 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 165 | False | 172.11 | 104.23 | 104.45 | 5.24 | 3.58 | 111.78 |
| 166 | False | 173.01 | 104.59 | 104.31 | 5.04 | 3.05 | 110.91 |
| 167 | False | 172.47 | 104.27 | 104.10 | 4.88 | 3.33 | 110.68 |
| 168 | False | 171.82 | 103.97 | 103.88 | 4.73 | 3.55 | 111.87 |
| 169 | False | 171.96 | 104.00 | 103.95 | 5.63 | 3.26 | 110.96 |

100 : True
↓
Vrai billet

70 : False
↓
Faux billet

170 rows x 7 columns



Analyse
données

ACP

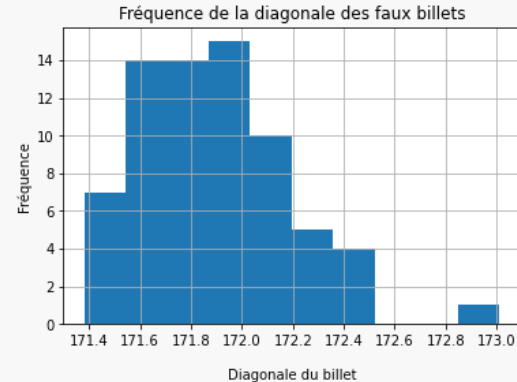
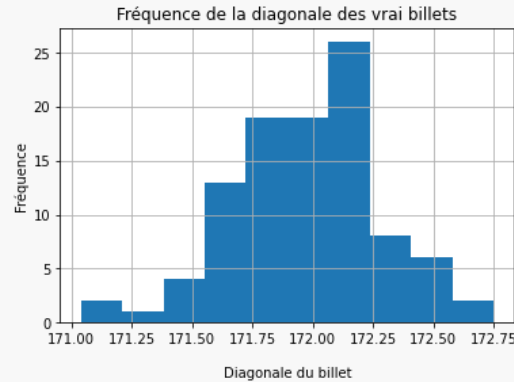
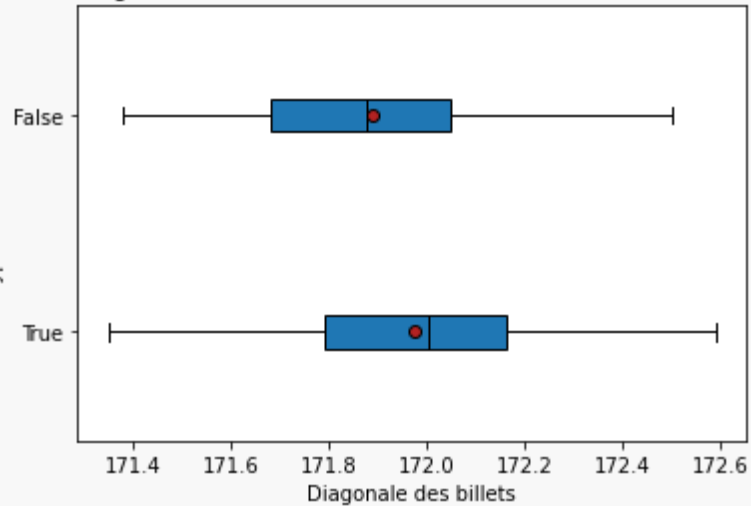
K-means

Régression
logistique

Conclusion

Diagonale du billet (mm) :

Diagonale des billets en fonction des billets vrai ou faux



Test d'Anova Welch

p-value = 0.068

H0 : Les moyennes des échantillons sont égales

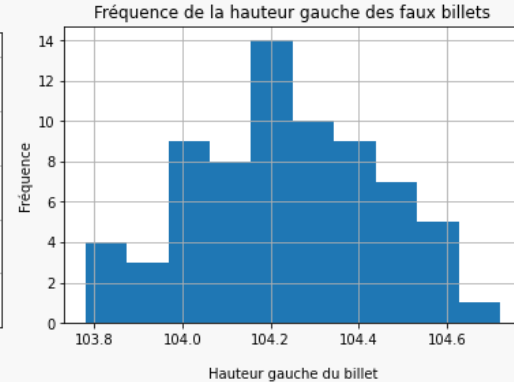
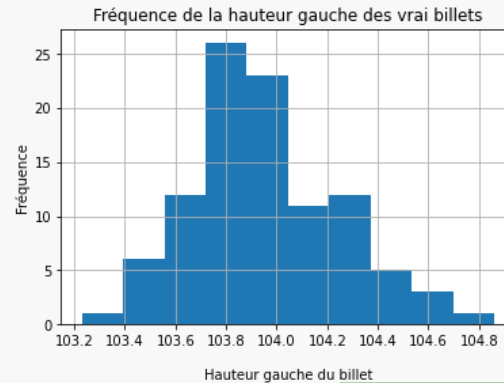
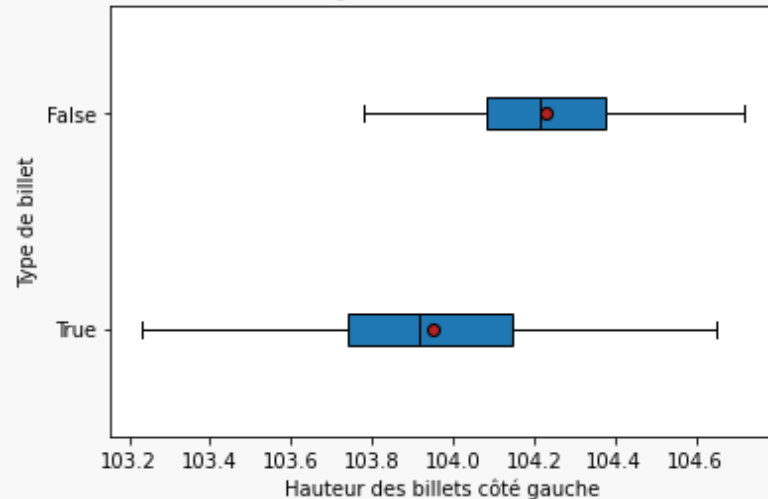
Ha : Au moins une des moyennes est différentes

p value > 0.05 : Les moyennes de diagonal sont égales entre les différents billets



Hauteur du billet coté gauche (mm) :

Hauteur des billets côté gauche en fonction des billets vrai ou faux



Test d'Anova Welch

p-value = 2.7e-11

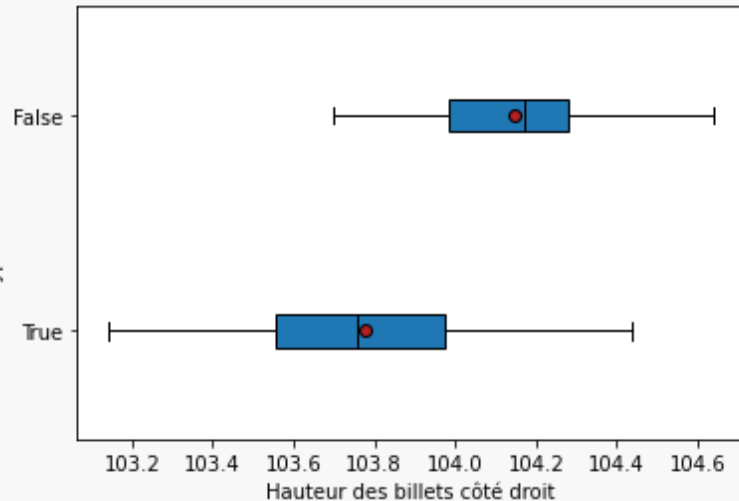
H0 : Les moyennes des échantillons sont égales

Ha : Au moins une des moyennes est différentes

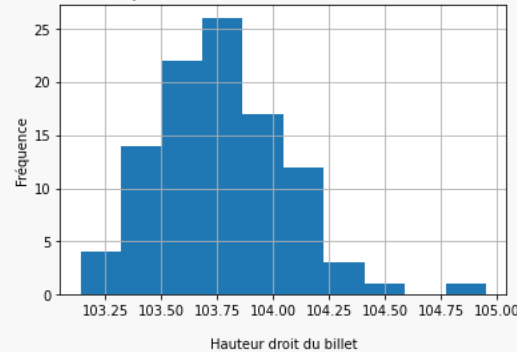
p value < 0.05 : Il y a une différence significative des moyennes de l'hauteur coté gauche entre les différents billets

Hauteur du billet coté droit (mm) :

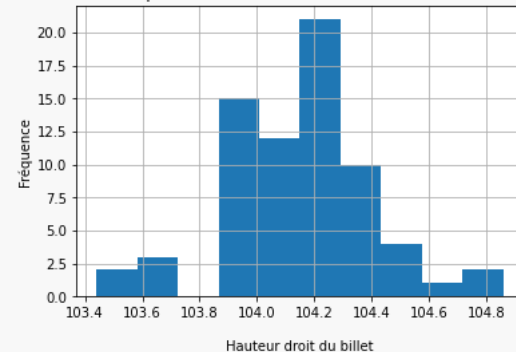
Hauteur des billets côté droit en fonction des billets vrai ou faux



Fréquence de la hauteur droit des vrai billets



Fréquence de la hauteur droit des faux billets



Test d'Anova Welch

p-value = 2.2e-15

H0 : Les moyennes des échantillons sont égales

Ha : Au moins une des moyennes est différentes

p value < 0.05 : Il y a une différence significative des moyennes de l'hauteur coté droit entre les différents billets

Analyse
données

ACP

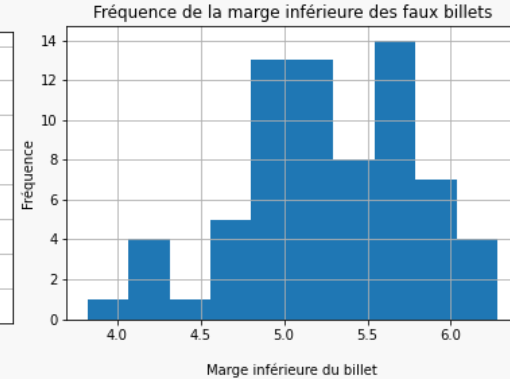
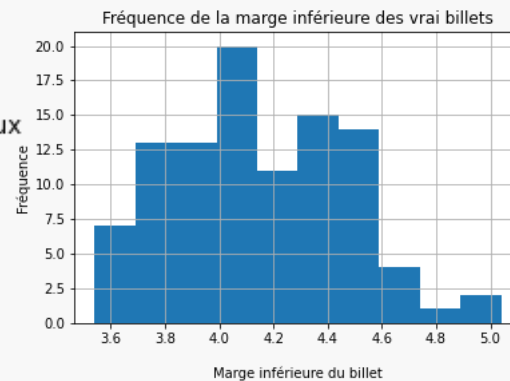
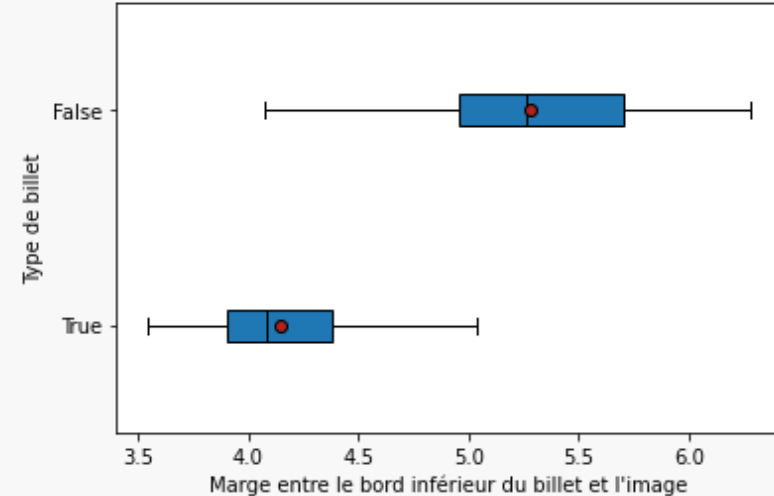
K-means

Régression
logistique

Conclusion

Marge entre le bord inférieure du billet et de l'image (mm) :

Marge bord inférieur du billet et l'image en fonction des billets vrai ou faux



Test d'Anova Welch

p-value = 3.5e-29

H0 : Les moyennes des échantillons sont égales

Ha : Au moins une des moyennes est différentes

p value < 0.05 : Il y a une différence significative des moyennes de la marge bord inférieure entre les différents billets



Analyse
données

ACP

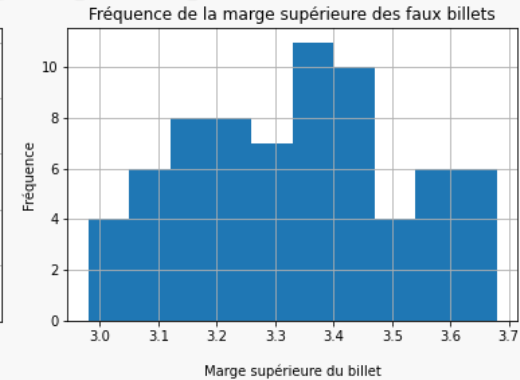
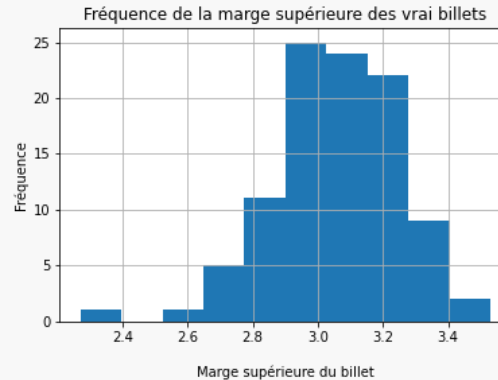
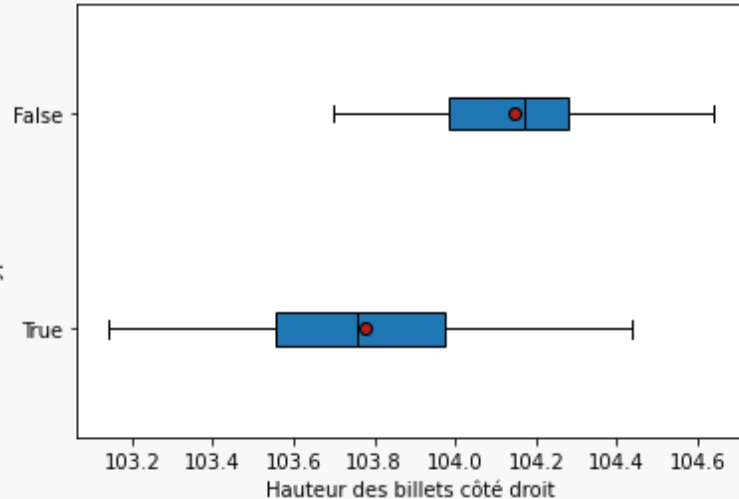
K-means

Régression
logistique

Conclusion

Marge entre le bord supérieure du billet et de l'image (mm) :

Hauteur des billets côté droit en fonction des billets vrai ou faux



Test d'Anova Welch

p-value = 6.8e-17

H0 : Les moyennes des échantillons sont égales

Ha : Au moins une des moyennes est différentes

p value < 0.05 : Il y a une différence significative des moyennes de la marge bord supérieure entre les différents billets



Analyse
données

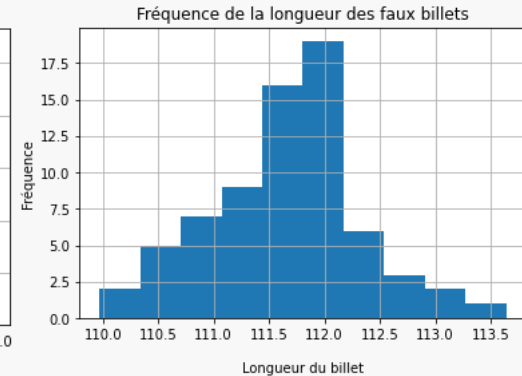
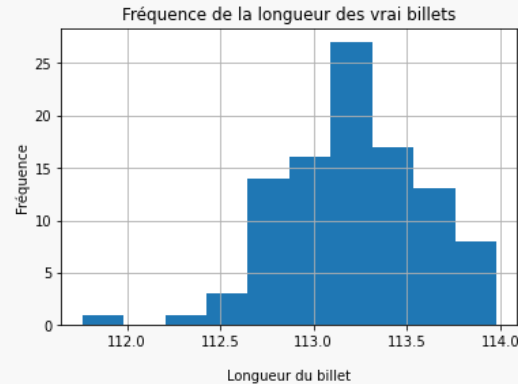
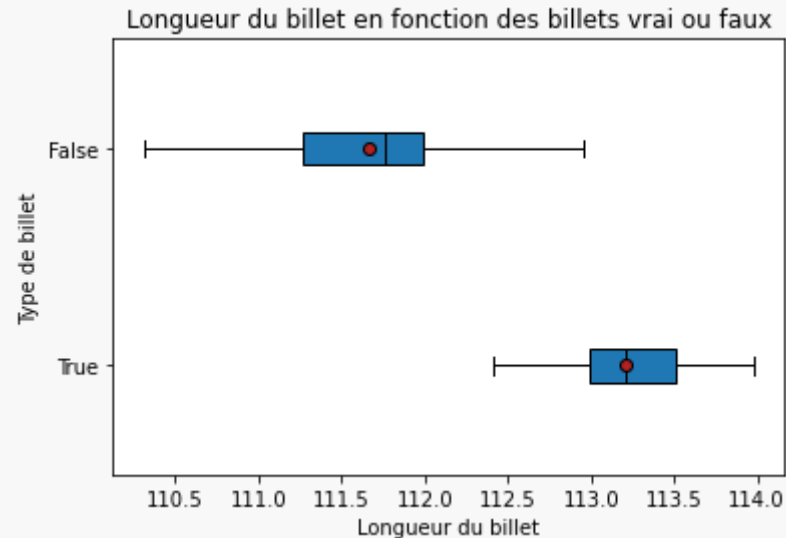
ACP

K-means

Régression
logistique

Conclusion

Longueur du billet (mm) :



Test d'Anova Welch

p-value = 9.4e-32

H0 : Les moyennes des échantillons sont égales

Ha : Au moins une des moyennes est différentes

p value < 0.05 : Il y a une différence significative des moyennes de longueur entre les différents billets



Analyse
données

ACP

K-means

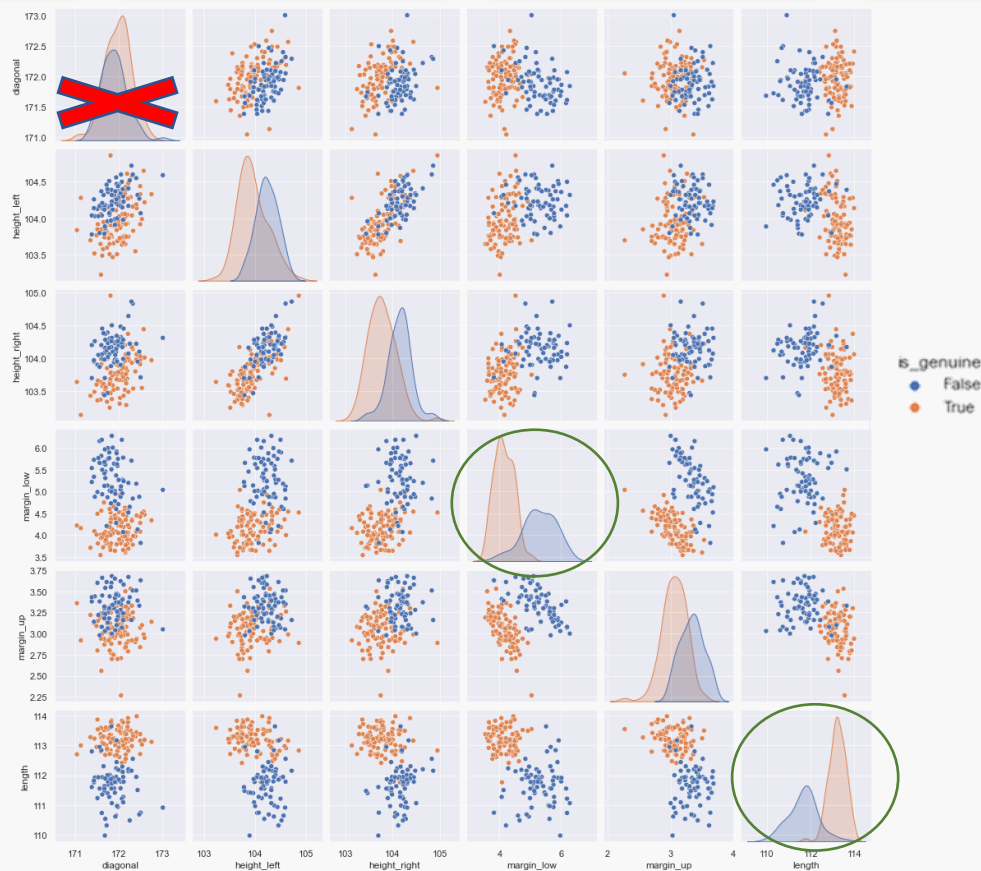
Régression
logistique

Conclusion

Pair plot :

La diagonale entre les vrais et
faux billets est très proches

Grande différence entre la
marge inférieure et la
longueur des billets



Analyse
données

ACP

K-means

Régression
logistique

Conclusion

Centrage et réduction des données avec StandardScaler :

```
array([[ -0.42834385,  2.6694537 ,  3.09656321, -0.13159013, -1.18987367,  
         0.28163308],  
       [-0.88755932, -1.09769709, -0.69125444, -0.86012554, -1.27473985,  
         0.78069736],  
       [-0.36274164, -1.03042654, -0.50943919, -0.30301023, -1.23230676,  
         1.37740465],  
       ...,  
       [ 1.73652911,  0.68497248,  0.52084721,  0.38267015,  0.67718214,  
        -2.05094997],  
       [-0.39554274, -0.32408577, -0.1458087 ,  0.16839504,  1.61071005,  
        -0.75989238],  
       [ 0.06367273, -0.22317994,  0.06630909,  1.45404575,  0.38015054,  
        -1.74717171]])
```

Transformation des
variables [0 ; 1]

$$x_{scaled} = \frac{x - x_{moyenne}}{\sigma_{\text{écart_type}}}$$

Analyse
données

ACP

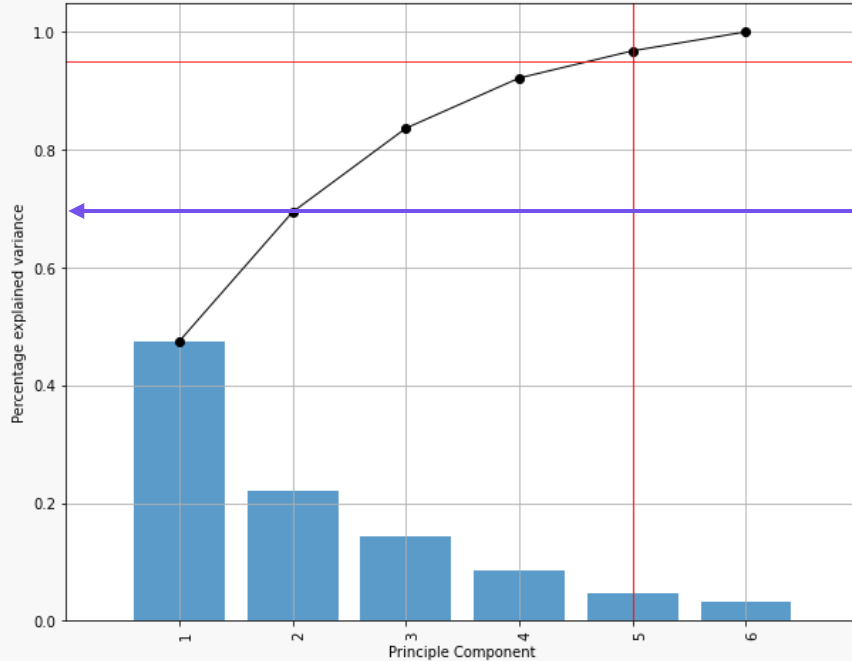
K-means

Régression
logistique

Conclusion

Détermination du nombre de composantes :

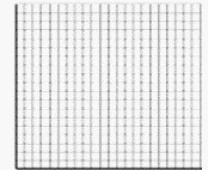
Cumulative explained variance
5 Principal Components explain [95.0%] of the variance.



Courbe de variance accumulée

70% avec 2
composantes

2 axes



Variance pour chaque axe principal

Analyse
données

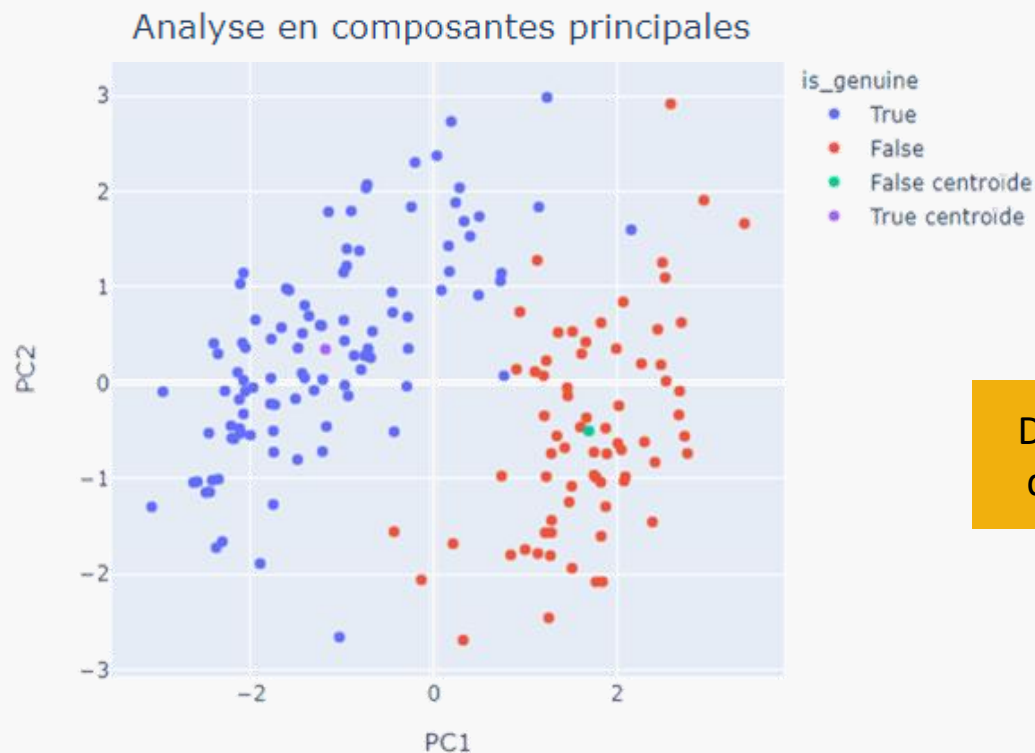
ACP

K-means

Régression
logistique

Conclusion

Analyse en composantes principales :



Deux groupes distincts avec
quelques chevauchements

Analyse
données

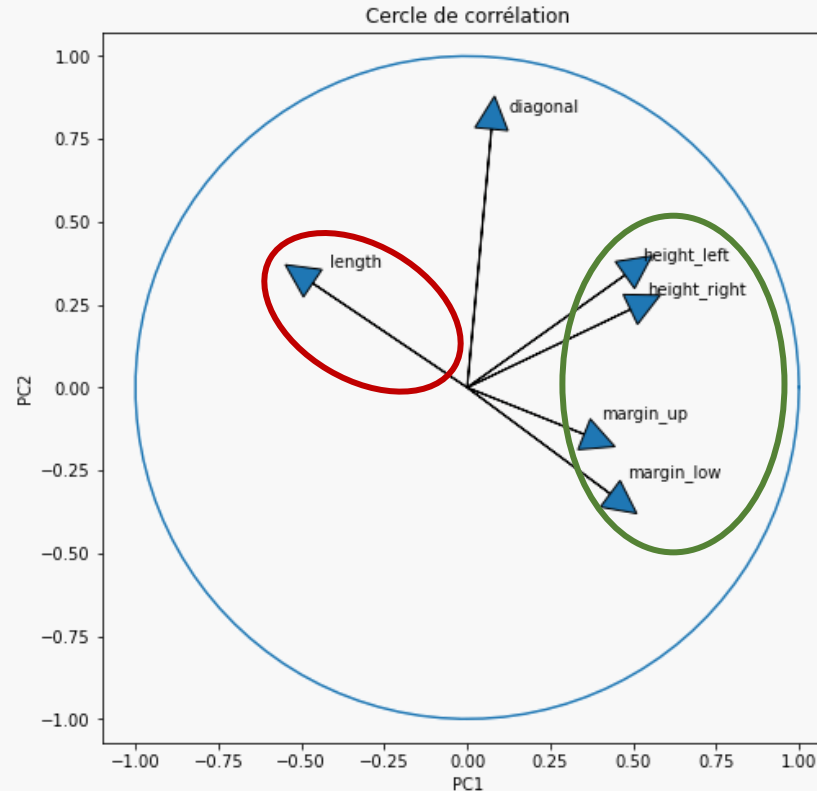
ACP

K-means

Régression
logistique

Conclusion

Cercle de corrélation :



Corrélation positive entre la
hauteur droite et gauche,
entre la marge inférieure et
supérieure

Corrélation négative entre la
longueur et la marge
inférieure

Analyse
données

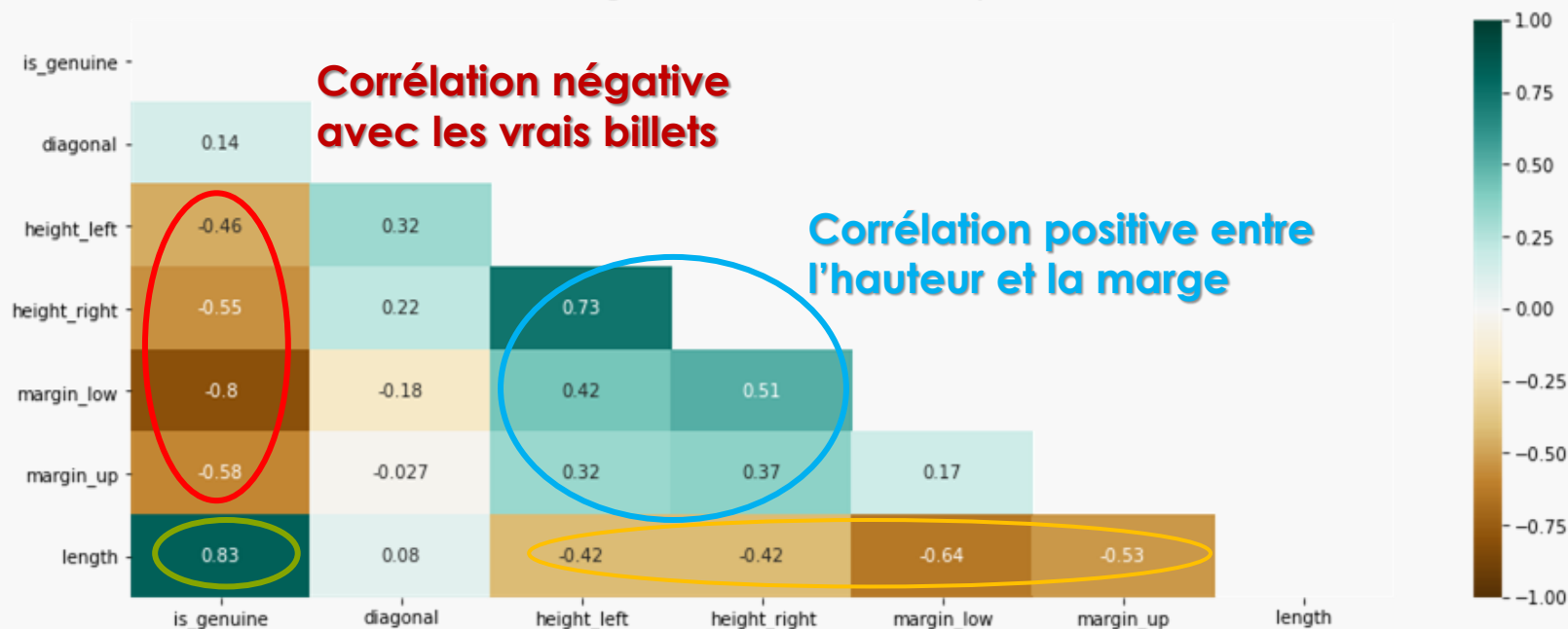
ACP

K-means

Régression
logistique

Conclusion

Heatmap de corrélation :



Corrélation positive
avec les vrais billets

Corrélation négative
avec la longueur

Analyse
données

ACP

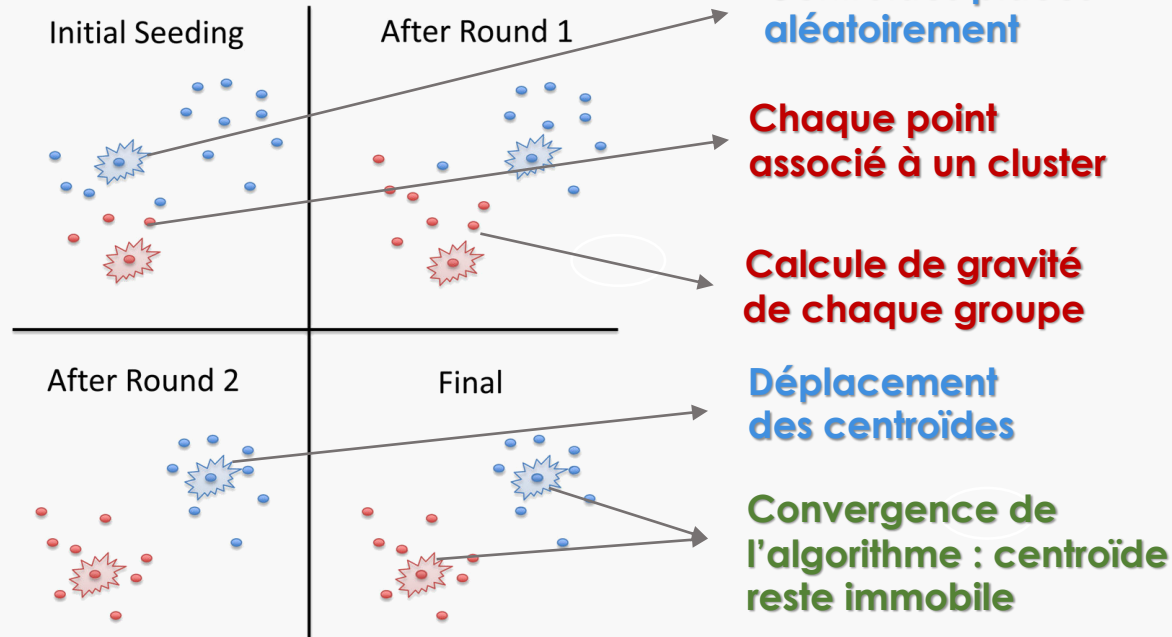
K-means

Régression
logistique

Conclusion

Méthode du clustering K-means :

Détermination de groupes en
minimisant l'inertie interclasse



Analyse
données

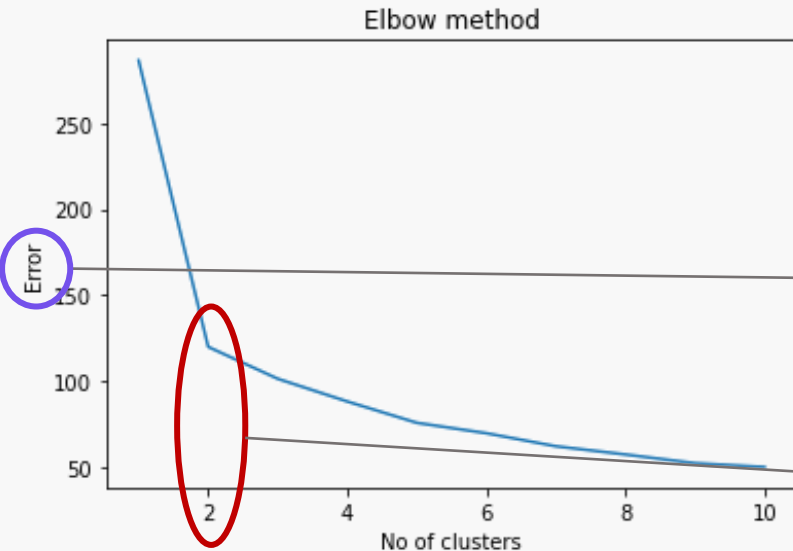
ACP

K-means

Régression
logistique

Conclusion

Détermination du nombre de classes :



Méthode du coude :

Lancement du k-means
avec différentes valeurs
de nombre de groupe (K)

Détermination de
l'inertie interclasse

Valeur de K au-delà de
laquelle la diminution est
plus faible

Analyse
données

ACP

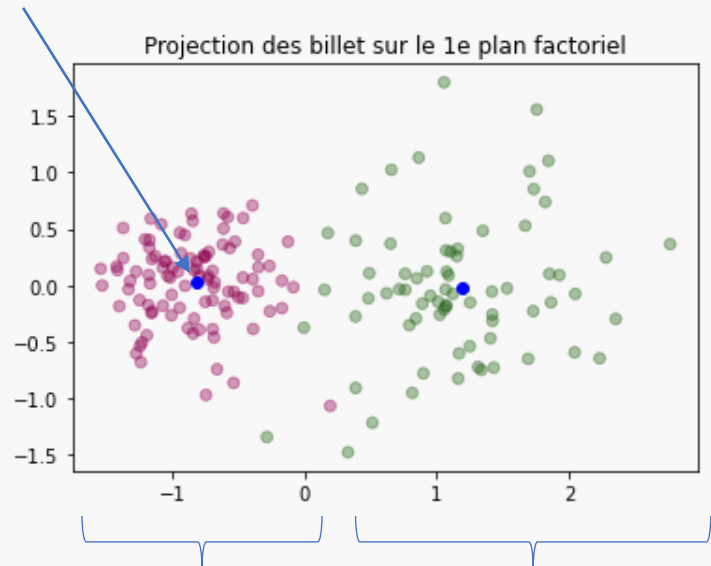
K-means

Régression
logistique

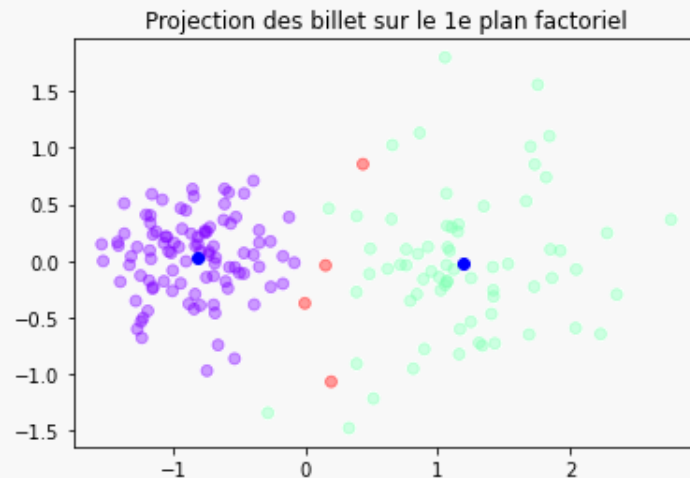
Conclusion

Clustering k-means sur les données brutes :

Centroïde



En réalité



● Billets mal classés

Analyse
données

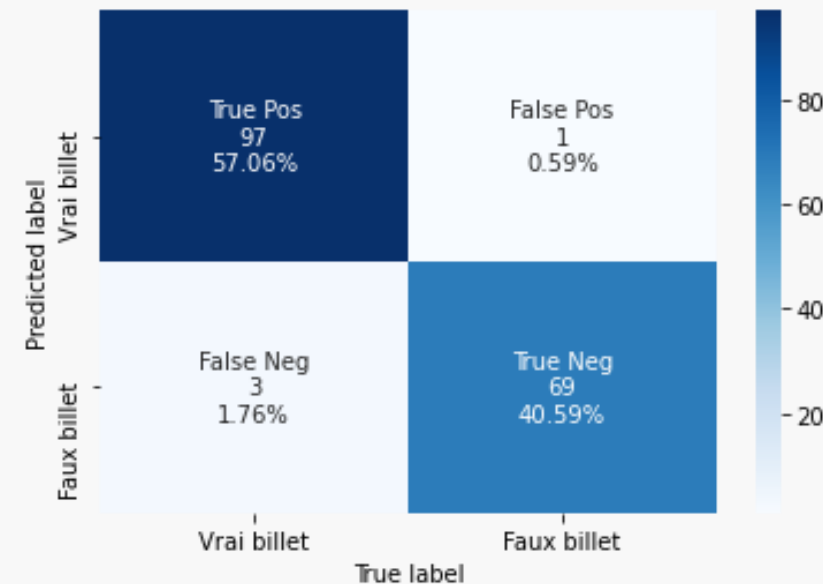
ACP

K-means

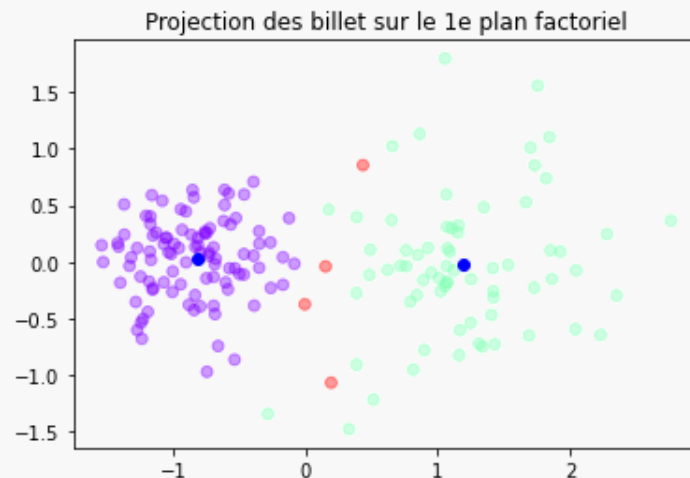
Régression
logistique

Conclusion

Clustering k-means sur les données brutes :



Efficacité du
clustering



● Billets mal classés

Analyse
données

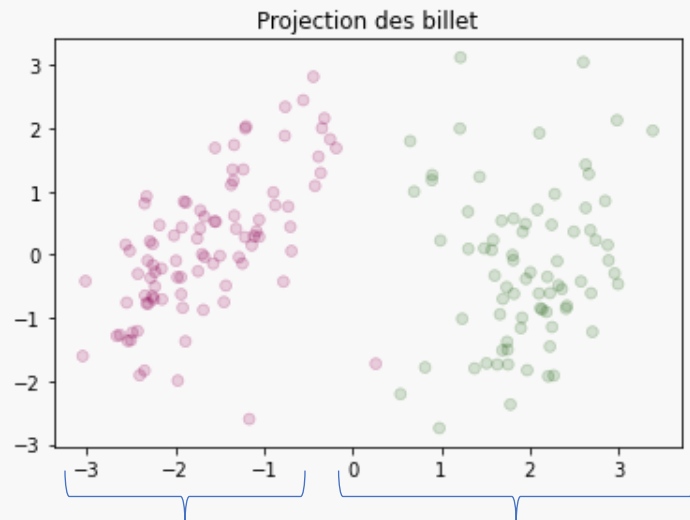
ACP

K-means

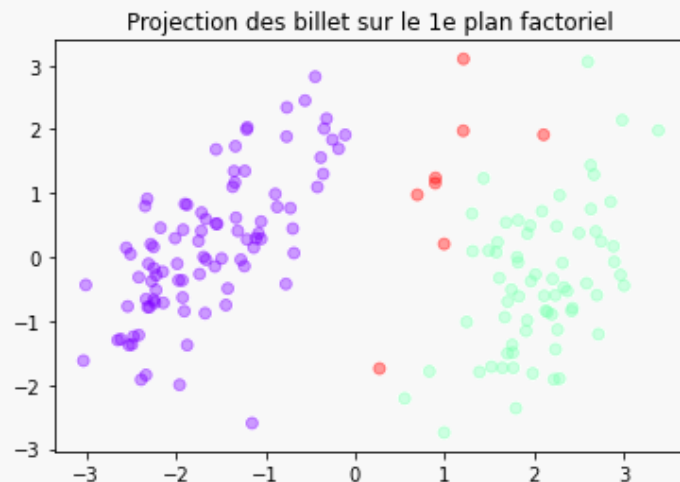
Régression
logistique

Conclusion

Clustering k-means sur les données PCA :



En réalité



● Billets mal classés

Cluster k-means True Cluster k-means False

Analyse
données

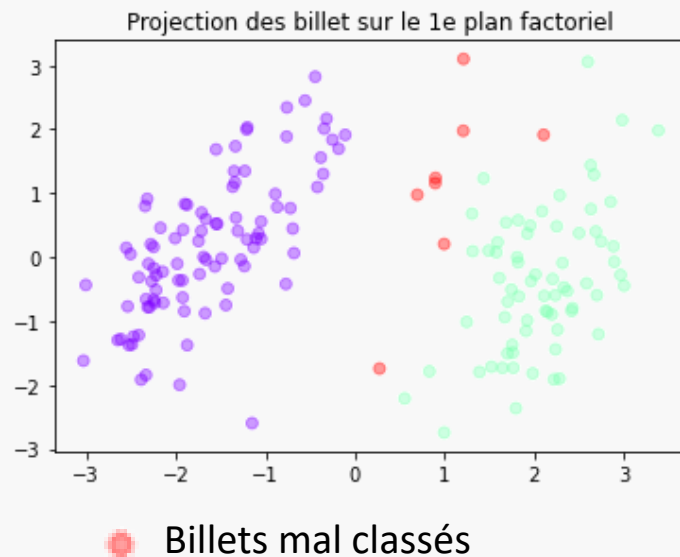
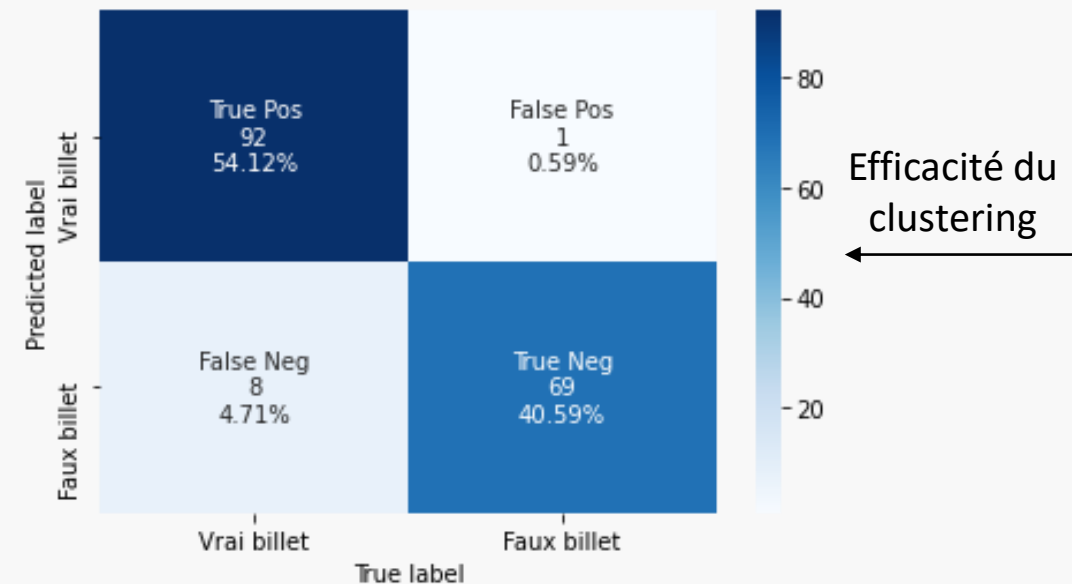
ACP

K-means

Régression
logistique

Conclusion

Clustering k-means sur les données PCA :



Analyse
données

ACP

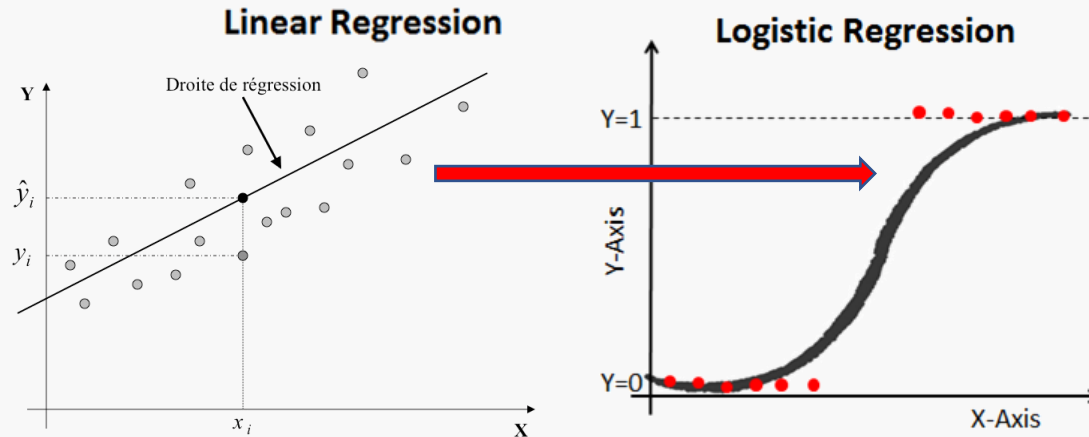
K-means

Régression
logistique

Conclusion

Régression logistique :

Méthode statistique de prédiction de classes binaires



Transformation régression linéaire en logistique
pour une prédiction de probabilité

Analyse
données

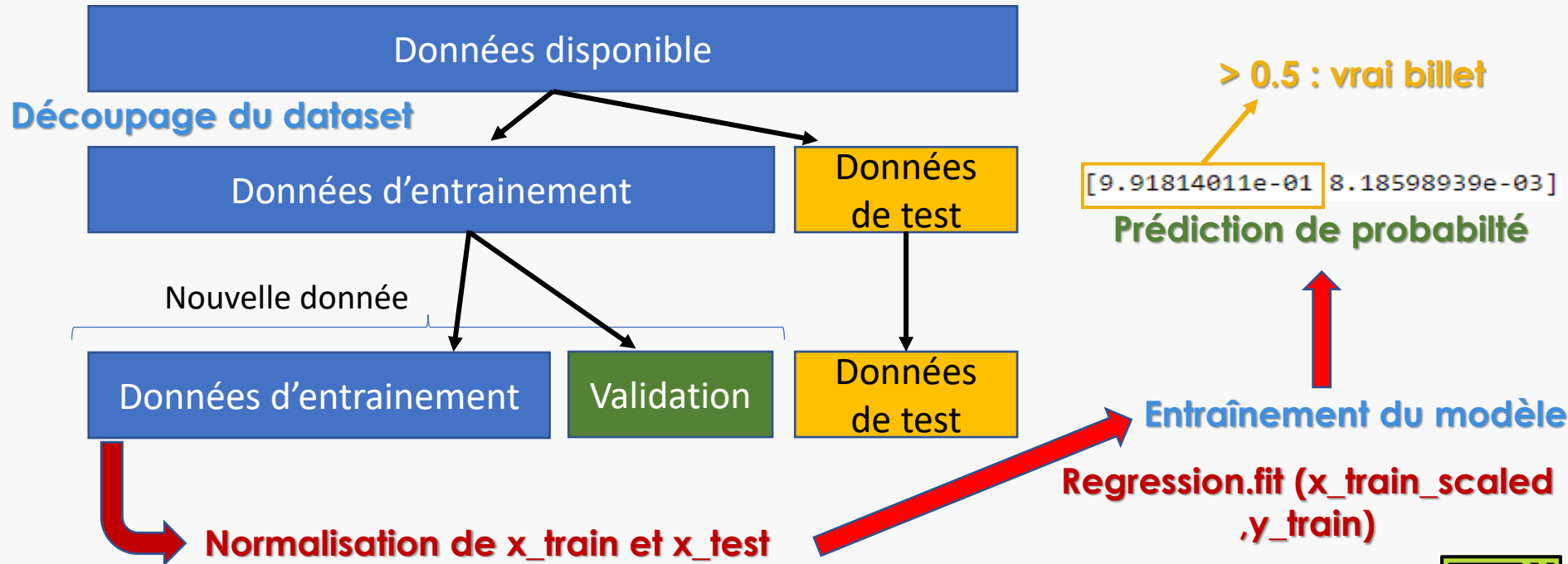
ACP

K-means

Régression
logistique

Conclusion

Train test split et prédiction :



Analyse
données

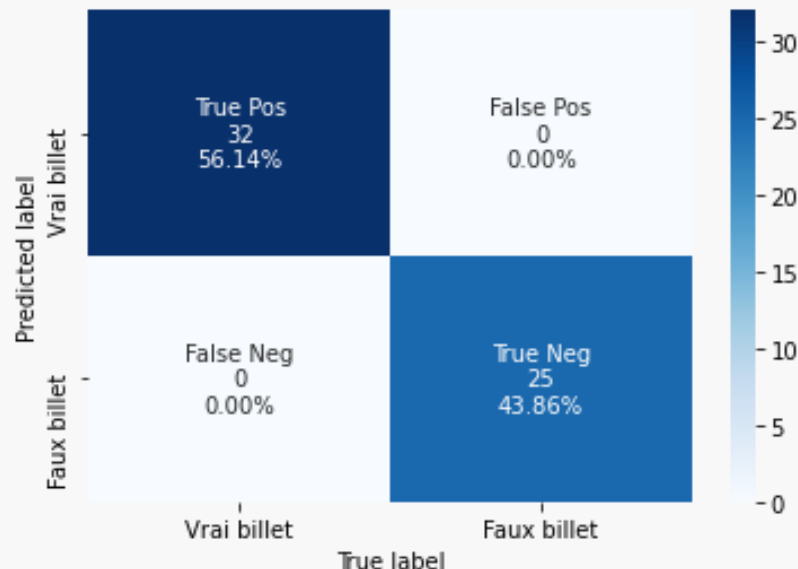
ACP

K-means

Régression
logistique

Conclusion

Matrice de confusion et classification report :



| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| Vrai billet | 1.00 | 1.00 | 1.00 | 32 |
| Faux billet | 1.00 | 1.00 | 1.00 | 25 |
| accuracy | | | 1.00 | 57 |
| macro avg | 1.00 | 1.00 | 1.00 | 57 |
| weighted avg | 1.00 | 1.00 | 1.00 | 57 |

Evaluation des performances du modèle de classification

Analyse
données

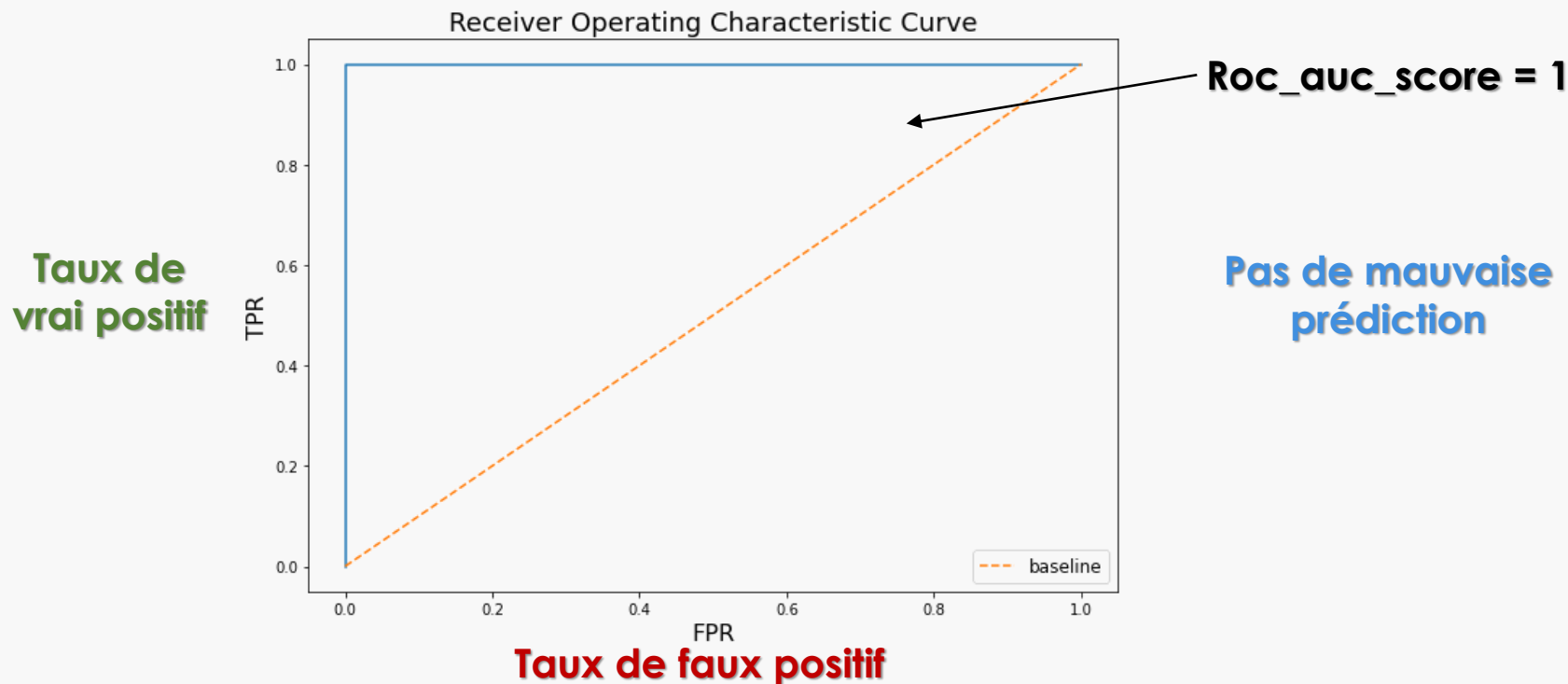
ACP

K-means

Régression
logistique

Conclusion

Courbe ROC :



Analyse
données

ACP

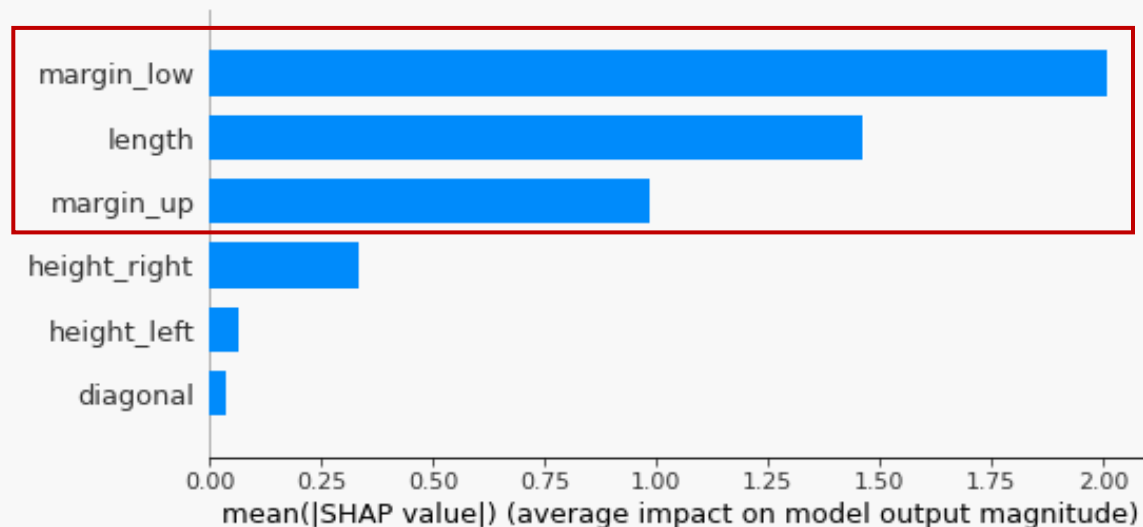
K-means

Régression
logistique

Conclusion

Interprétation des variables avec SHAP :

Prédiction de l'impact de chaque variables



**3 variables qui ont le plus
d'impact pour la prédiction**

Analyse
données

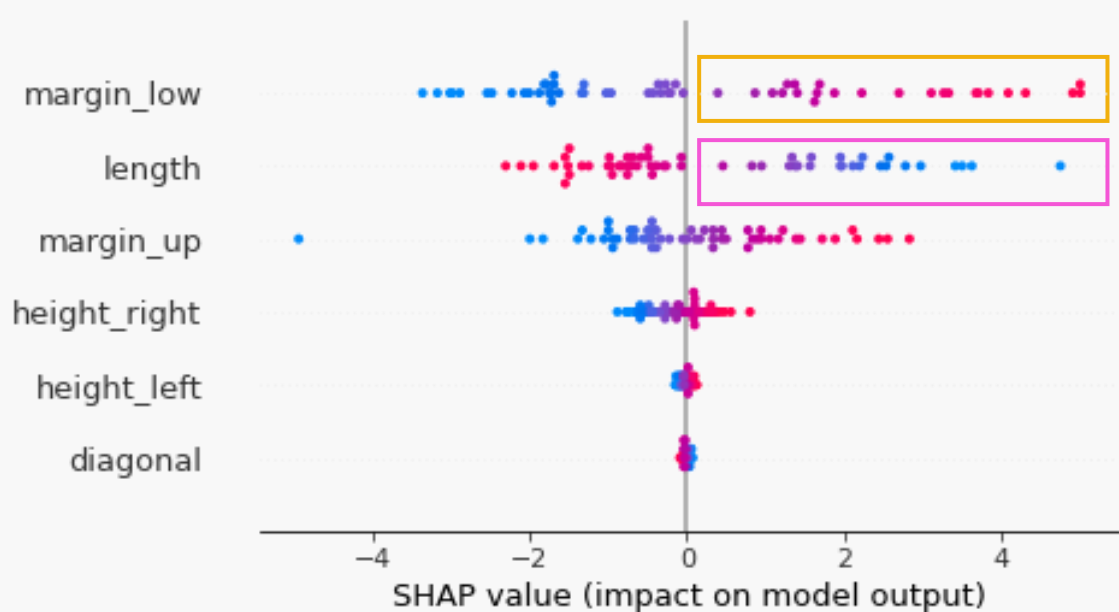
ACP

K-means

Régression
logistique

Conclusion

Interprétation des variables avec SHAP :



High Plus c'est élevé plus les billets ont tendance à être faux

Feature value Plus c'est élevé plus les billets ont tendance à être vrai

Low

Analyse
données

ACP

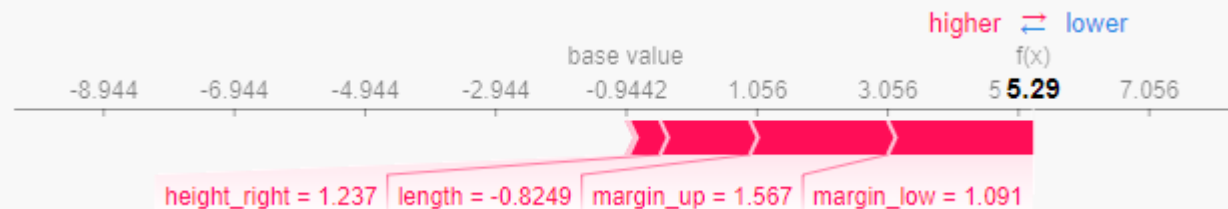
K-means

Régression
logistique

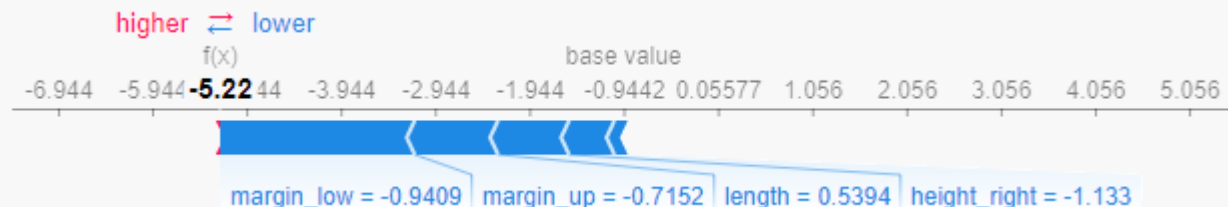
Conclusion

Interprétation de prédiction local avec SHAP :

Faux billet



Vrai billet



Analyse
données

ACP

K-means

Régression
logistique

Conclusion

Test de la fonction de prédiction :



Analyse
données

ACP

K-means

Régression
logistique

Conclusion

❑ La régression logistique a permis **de prédire la probabilité** du billet à être vrai



❑ Les variables **les plus impactant** sont : **margin_low**, **margin_up** et **length**

❑ Plus la **margin_low et up sont élevées** plus les billets ont tendance à **être faux**



❑ Plus la **longueur est grande** plus les billets ont tendance à **être vrai**



❑ **La fonction de prédiction** possède **une précision de 1** (il n'y a pas eu de mauvaises classifications)





Merci pour votre attention