

# **DÉTECTION DE FAUX BILLETS**

Création d'un algorithme de détection de faux billets de banque

# SOMMAIRE

## **Partie I**

Contexte et enjeux

## **Partie II**

Comparatif entre vrais et faux billets

## **Partie III**

ACP et classification

## **Partie IV**

Algorithme

# **ENJEUX**

lutte contre propagation de faux billets

# ÉMISSION ET CIRCULATION DE BILLETS : FILIÈRE FIDUCIAIRE EUROPÉENNE

Les billets suivent un parcours déterminé dans l'économie. Afin d'éviter une pénurie de billets, ces transferts sont coordonnés de manière centralisée et financés par la BCE.

## 28 milliards de billets

ont été émis dans la **zone euro** depuis 2002

Seules les banques centrales émettent les billets ; en France, il s'agit de la **Banque de France**, de l'**IEDOM** et de l'**IEOM**.

La production des billets en euros est donc le fruit d'un travail en commun des **banques centrales nationales (BCN)** et de la **Banque centrale européenne (BCE)**.

Les billets émis dans la zone euro se distinguent entre eux par des **couleurs différentes** mais aussi par les **dimensions**.

Pour les billets de **5 € à 100 €**, les dimensions augmentent, d'une coupure à l'autre, de 5 millimètres en hauteur et de 6 ou 7 millimètres en longueur.

Les coupures de **100 €, 200 € et 500 €** ont la même hauteur et se distinguent entre elles par la longueur.



## CIRCULATION DE FAUX BILLETS DANS LA ZONE EUROPÉENNE

Les flux de billets en euros au sein de la zone euro sont très importants. Les pouvoirs publics ont donc mis en place des grands moyens afin de stopper les flux de faux billets.

De nos jours, les criminels bénéficient d'un accès relativement facile aux technologies, au matériel et au savoir-faire nécessaires pour s'adonner à la contrefaçon. Il s'agit d'une activité lucrative qui permet généralement de financer d'autres activités illégales comme **la traite d'êtres humains, le trafic de stupéfiants** ou encore **le terrorisme**.

### CONSÉQUENCE

Ces dernières années, la BCE a progressivement doté la zone euro de nouveaux billets avec des dispositifs de sécurité renforcés, les derniers renouvelés ayant été les coupures de 100 et 200 euros en 2019.

*Seule l'émission de 500 euros s'est arrêtée, le billet mauve étant accusé de faciliter les activités illégales.*

La contrefaçon ou la falsification des pièces de monnaie ou des billets de banque ayant cours légal en France ou émis par les institutions étrangères ou internationales habilitées à cette fin est punie de **trente ans de réclusion criminelle** et de **450000 euros d'amende**.

(Loi n° 2004-204 du 9 mars 2004 art. 6 XIII Journal Officiel du 10 mars 2004)

**700 000**

faux billets en circulation dans la zone euro

**IS GENUINE?**

comparatif entre vrais et faux billets

# COMPARATIF

## 02

## COMPARATIF ENTRE VRAIS ET FAUX BILLETS

Afin de mettre en place notre algorithme de détection de faux billets, il nous a fallu comparer les données disponibles de vrais et faux billets.

**170 billets avec 7 variables**

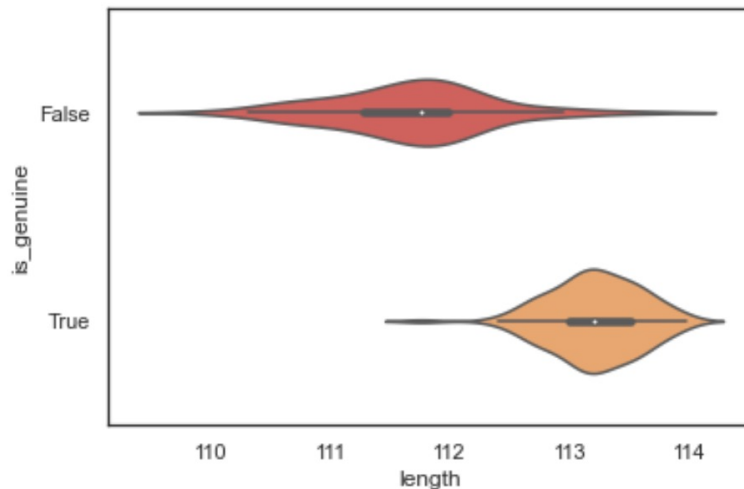
- 100 vrais billets
- 70 faux billets

1 variable qualitative : **is\_genuine**

6 quantitatives : **length**, margin low, margin up, diagonal, height left et height right

### EXEMPLE

Comparatif entre la longueur des vrais et faux billets



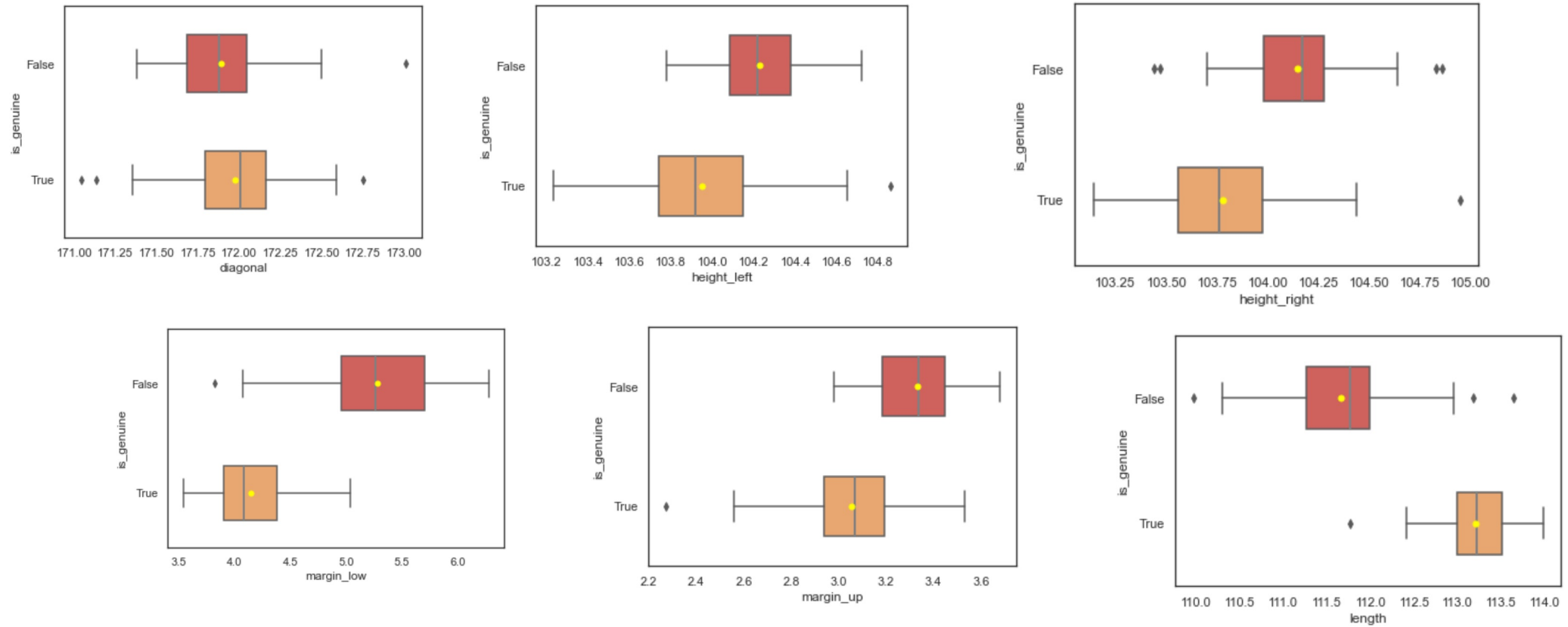
Les longueurs des vrais billets sont homogènes, ce qui *n'est pas le cas* des faux.

## COMPARATIF 02

# COMPARATIF ENTRE VRAIS ET FAUX BILLETS

Afin de mettre en place notre algorithme de détection de faux billets, il nous a fallu comparer les données disponibles de vrais et faux billets.

Comparatif entre les données des variables quantitatives des vrais et faux billets



On constate ici que les valeurs les plus discriminantes sont **length** et **margin\_low**

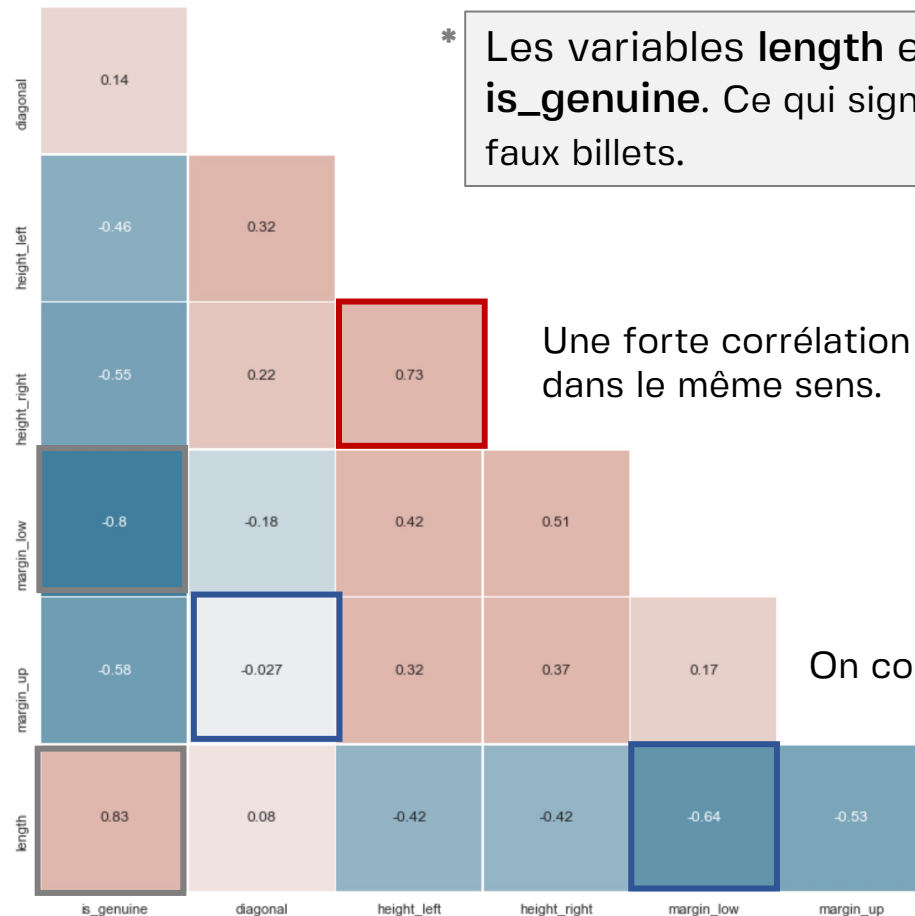


## COMPARATIF 02

### COMPARATIF ENTRE VRAIS ET FAUX BILLETS

Afin de mettre en place notre algorithme de détection de faux billets, il nous a fallu comparer les données disponibles de vrais et faux billets.

\* Les variables **length** et **margin\_low** sont les variables les plus corrélées à la variable **is\_genuine**. Ce qui signifie que ce sont ces 2 variables qui sont les plus importantes lors de la détection de faux billets.



Une forte corrélation positive est constatée entre **height\_right** et **height\_left** : les deux variables varient dans le même sens.

On constate une très faible corrélation entre **margin\_up** et **diagonal**.

Une corrélation négative existe entre **length** et **margin\_low** : lorsqu'une des variables augmente, l'autre diminue.

Matrice de corrélation

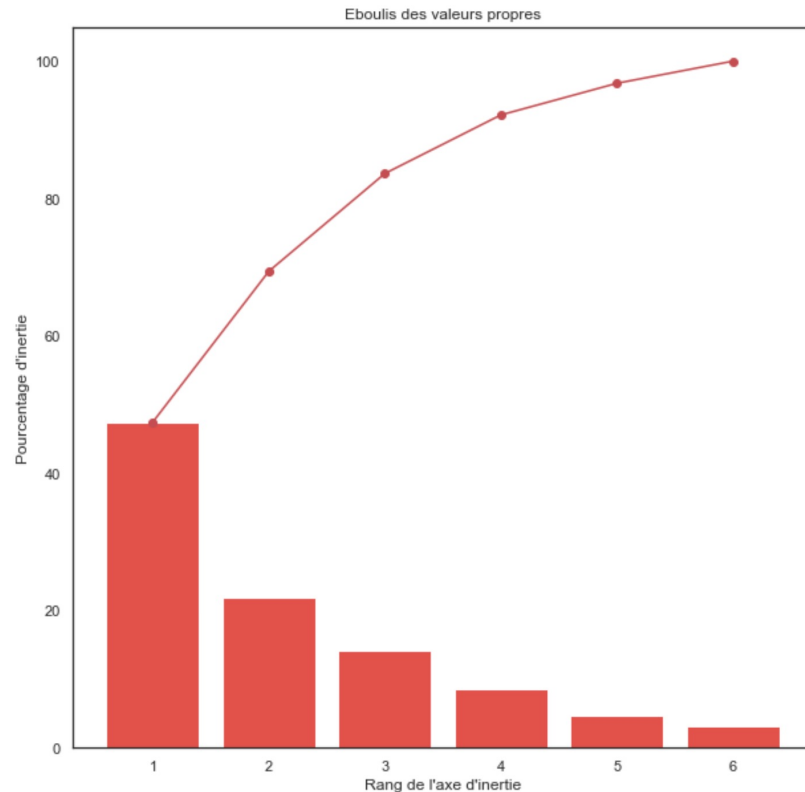
# **ANALYSE**

ACP et classification

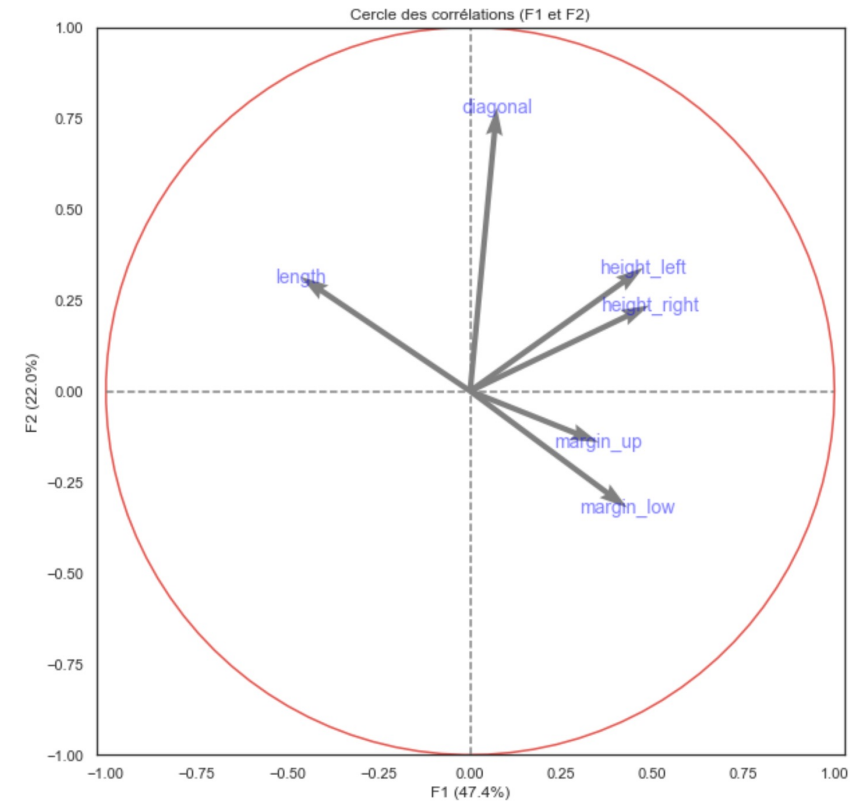
# ANALYSE 03

## ANALYSE EN COMPOSANTES PRINCIPALES

Analyse en composantes principales : éboulis des valeurs propres et cercles des corrélations.



Le critère du Kaiser nous conduit à retenir les **deux premiers axes**. En effet le premier axe retient 47.4% de l'inertie totale quant à l'axe 2 retient tout de même 22% de l'inertie, ce qui n'est pas négligeable. Et qui conduit à un taux d'inertie expliquée de 69.4%, ce qui est un bon résultat.

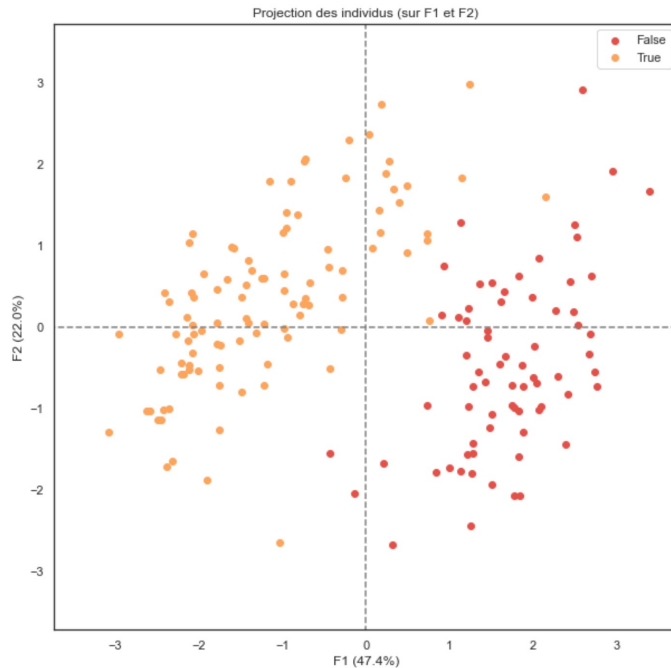


**F1** correspond aux informations concernant aux marges et hauteurs  
**F2** correspond aux informations concernant les longueurs

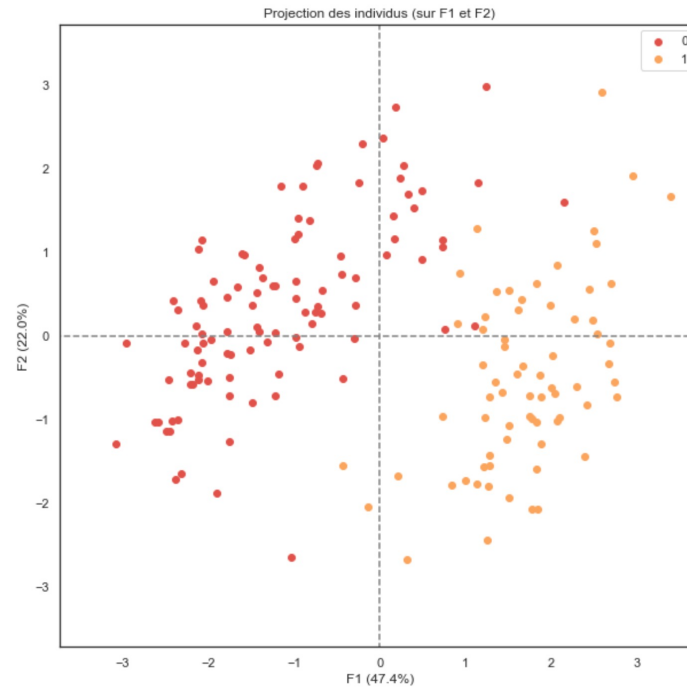
# ANALYSE 03

## CLASSIFICATION : K-MEANS

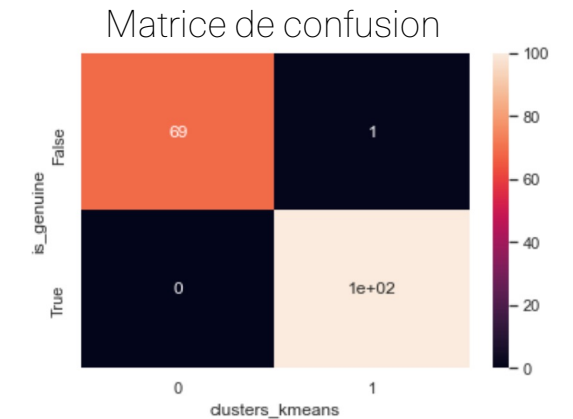
L'algorithme non supervisé de clustering non hiérarchique, k-means, a été appliqué à notre cas. Il permet de regrouper en clusters distincts les observations du data set. Ainsi, les données similaires se retrouveront dans un même cluster.



Projection des données des vrais et faux billets



Projection du résultat du clustering k-means



Matrice qui mesure la qualité d'un système de classification

100 Vrais positifs  
0 Faux positif  
1 Faux négatif  
69 Vrais négatifs

Contiennent ~70% de l'information.  
On observe très distinctement les deux groupes.

# **ALGORITHME**

Modélisation et programme

# ALGORITHME

## 04

## ALGORITHME ET MODÉLISATION

Nos données ont été modélisées à l'aide d'une régression logistique permettant de créer un programme capable d'effectuer une prédiction sur un billet, c'est-à-dire s'il s'agit d'un vrai ou d'un faux billet.

### Division des données

- Apprentissage : 80%
- Test : 20% → `X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20)`

```
# Evaluation du modèle
print(classification_report(y_test, predictions))
print(accuracy_score(y_test, predictions))
```

	precision	recall	f1-score	support
False	1.00	0.83	0.91	12
True	0.92	1.00	0.96	22
accuracy			0.94	34
macro avg	0.96	0.92	0.93	34
weighted avg	0.95	0.94	0.94	34

0.9411764705882353 → pourcentage de bonnes prédictions

	Predicted Negative	Predicted Positive
Actual Negative	14	1
Actual Positive	1	18

Matrice de confusion

**Precision** : proportion de prédictions correctes parmi les points que l'on a prédits positifs

**Accuracy** : pourcentage de bonnes prédictions

**Recall** : proportion de positifs que l'on a correctement identifiés (et de négatifs que l'on a correctement identifiés)

# **MERCI**

**Anissa MANSOUR**

parcours Data Analyst (2020/2021)