

# PROJET 10 - DÉTECTEZ DES FAUX BILLETS AVEC R OU PYTHON

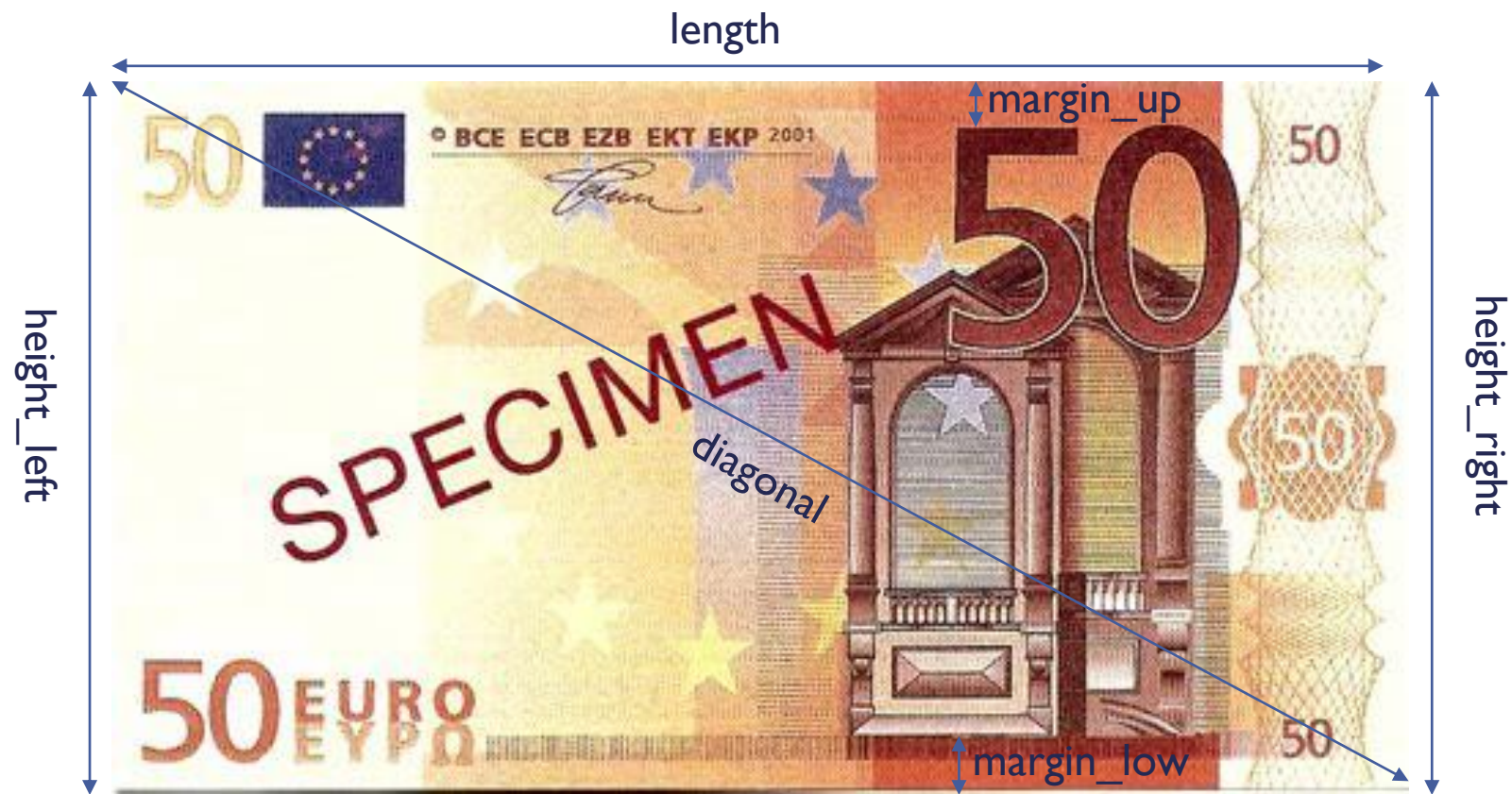
CAUDAL ADRIEN



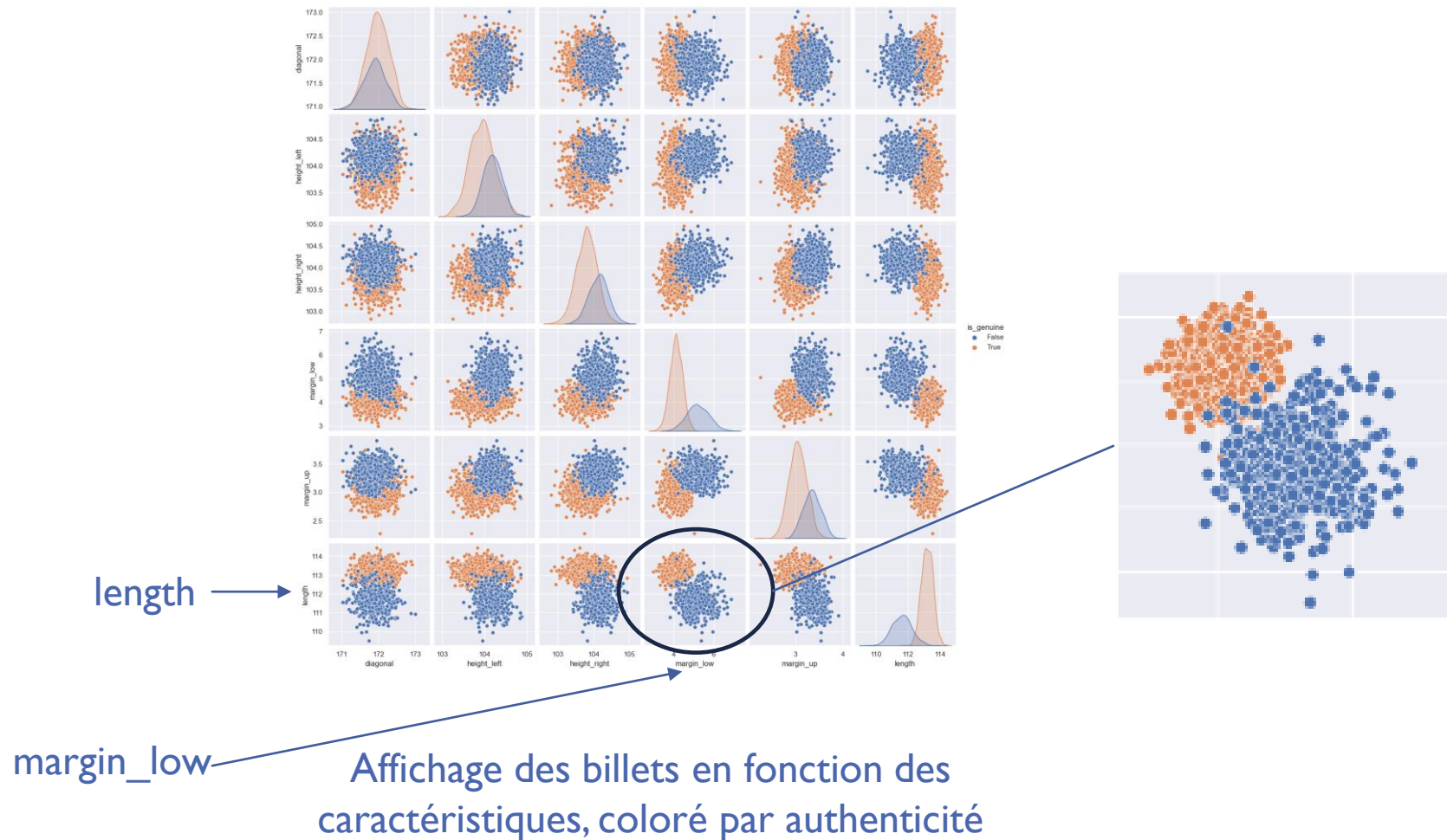
# CONTEXTE

- Méthodes d'identifications des contrefaçons
  - Créer un algorithme pour différencier les vrais des faux billets.
- Les différences de dimensions permettent de distinguer les faux billets

# CARACTÉRISTIQUES

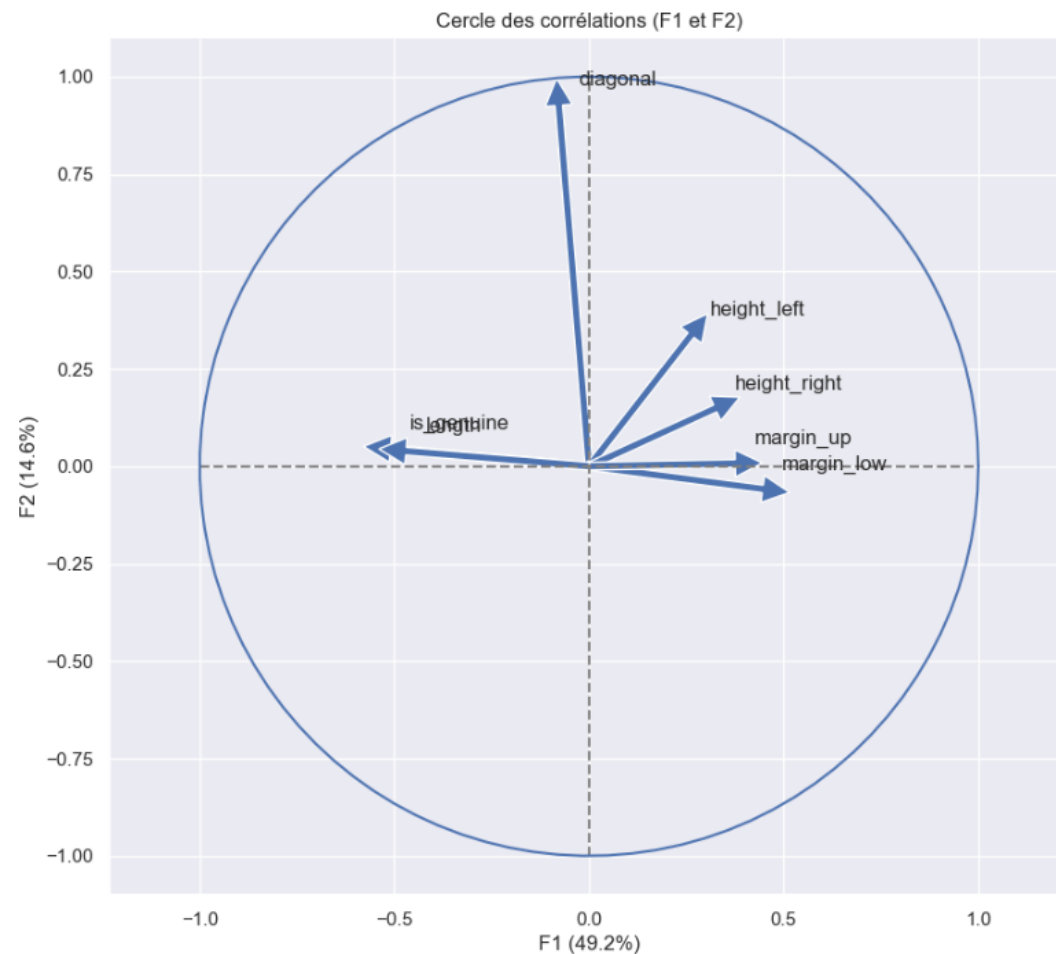


# VISUALISATIONS DES DONNÉES D'ENTRAÎNEMENT



# ANALYSE PAR COMPOSANTES

- Faire des liens entre les variables :
  - F1 résume bien is\_genuine, margin\_up, margin\_low et length
  - Les autres variables synthétiques ne sont associées qu'à une variable
- Aide à répondre à la problématique qui suit



# COMPLÉTION DU JEU DE DONNÉES

- 37 valeurs manquantes (margin\_low)
  - Régression linéaire
- Normalité des résidus :
  - Aderson-Darling > OK
- Indépendance des résidus :
  - Durbin-Watson > OK
- Homocédasticité :
  - White > OK

```
=====
                        OLS Regression Results
=====
Dep. Variable:          margin_low    R-squared:                0.477
Model:                  OLS           Adj. R-squared:           0.476
Method:                 Least Squares F-statistic:                266.1
Date:                   Wed, 27 Dec 2023 Prob (F-statistic):        2.60e-202
Time:                   17:30:35      Log-Likelihood:           -1001.3
No. Observations:       1463         AIC:                     2015.
Df Residuals:           1457         BIC:                     2046.
Df Model:                5
Covariance Type:        nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	22.9948	9.656	2.382	0.017	4.055	41.935
diagonal	-0.1111	0.041	-2.680	0.007	-0.192	-0.030
height_left	0.1841	0.045	4.113	0.000	0.096	0.272
height_right	0.2571	0.043	5.978	0.000	0.173	0.342
margin_up	0.2562	0.064	3.980	0.000	0.130	0.382
length	-0.4091	0.018	-22.627	0.000	-0.445	-0.374

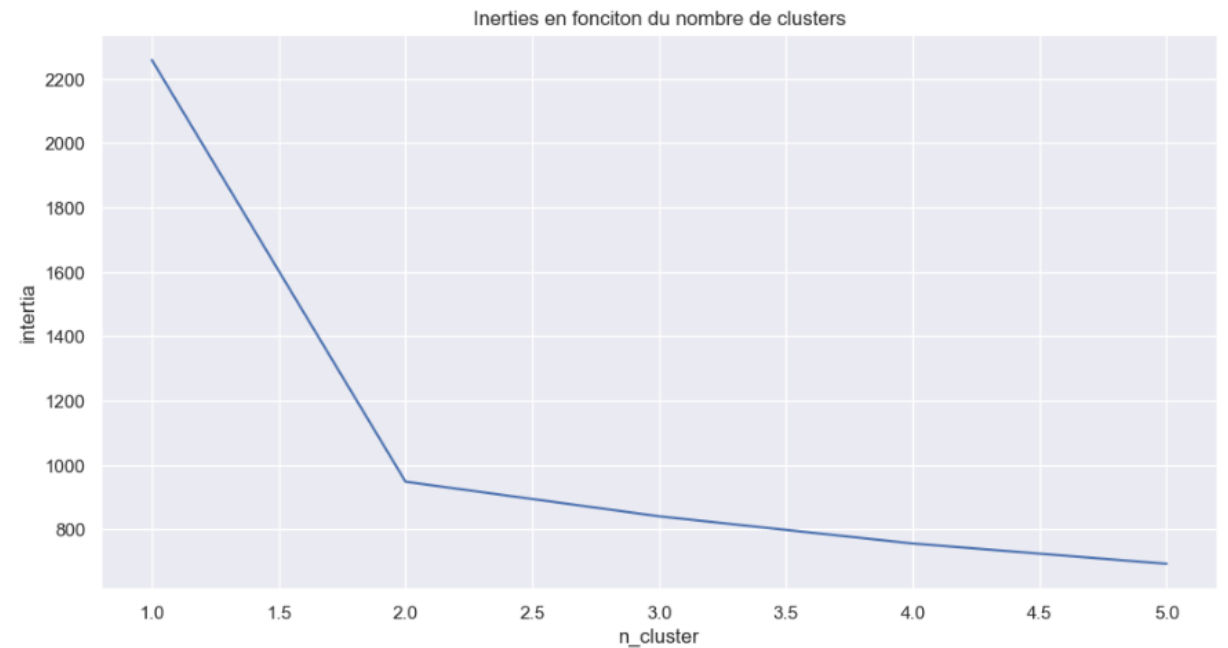
```
=====
Omnibus:                73.627      Durbin-Watson:            1.893
Prob(Omnibus):           0.000      Jarque-Bera (JB):         95.862
Skew:                    0.482      Prob(JB):                 1.53e-21
Kurtosis:                3.801      Cond. No.                 1.94e+05
=====
```

# ALGORITHMES DE MACHINE LEARNING UTILISÉS

- K-means (non supervisé)
- Random Forest (supervisé)
- Régression logistique (supervisé)
- k-Neighbors Classifier (supervisé)

# K-MEANS

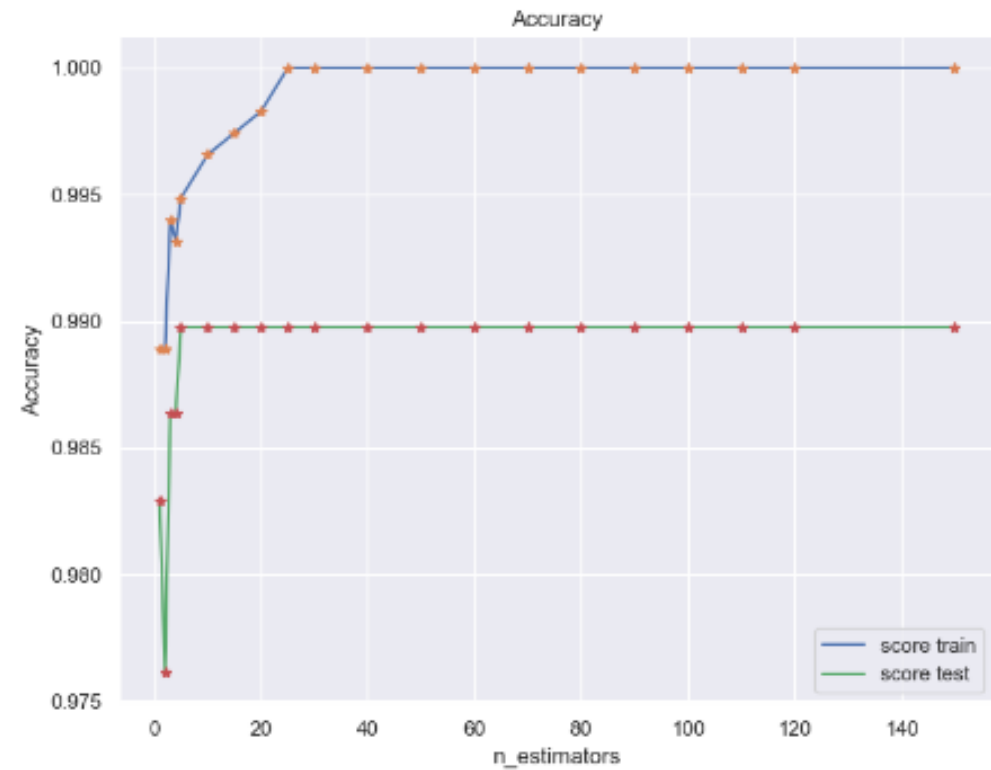
- Principe : regrouper de proche en proche
- Résultats :
  - 19 faux positifs et 2 faux négatifs sur les 1463 billets





# RANDOM FOREST

- Principe : Combinaison de prédicteurs faibles pour former le modèle d'ensemble.
- Résultats :
  - 2 faux positifs et 1 faux négatif sur 293 données test



# RÉGRESSION LOGISTIQUE

- Principe : chercher une relation (calcul) entre les variables.
- Résultats :
  - 2 faux positifs et 0 faux négatifs pour 293 données test

# K-NEIGHBORS CLASSIFIER

- Principe : classifier en fonction des voisins
- Résultats :
  - 2 faux positifs et 0 faux négatifs sur 293 données test

## COMPARAISON ET CHOIX FINAL

	<b>K-means</b>	<b>Random forest</b>	<b>Régression logistique</b>	<b>k-Neighbors Classifier</b>
Faux positifs	1,3%	0,7%	0,7%	0,7%
Faux négatifs	0,1%	0,3%	0%	0%