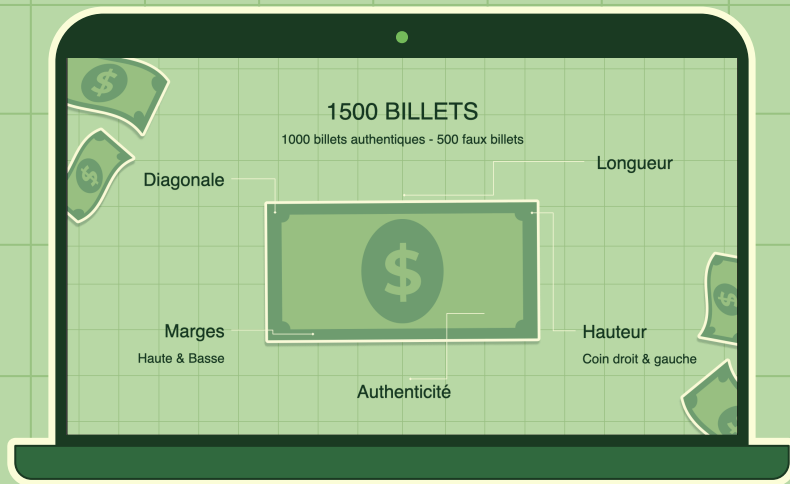


DETECTION DE FAUX BILLETS

ICART MATTHIEU

INTRODUCTION

ORGANISATION NATIONALE DE LUTTE CONTRE LE FAUX-MONNAYAGE



**Méthode d'identification des faux
billets pour lutter contre la
contrefaçon.**

*L'objectif est de mettre en place une
application de machine Learning capable
de prédire l'authenticité des billets.*

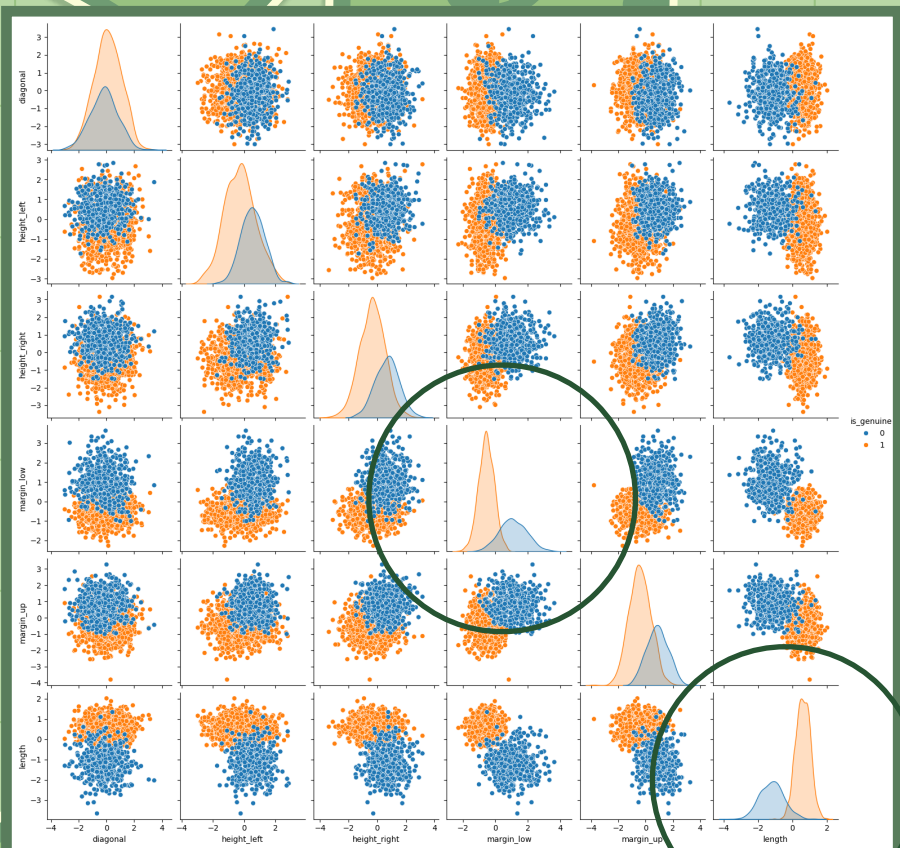
ANALYSE EXPLORATOIRE

IDENTIFICATION DES VARIABLES PERTINENTES

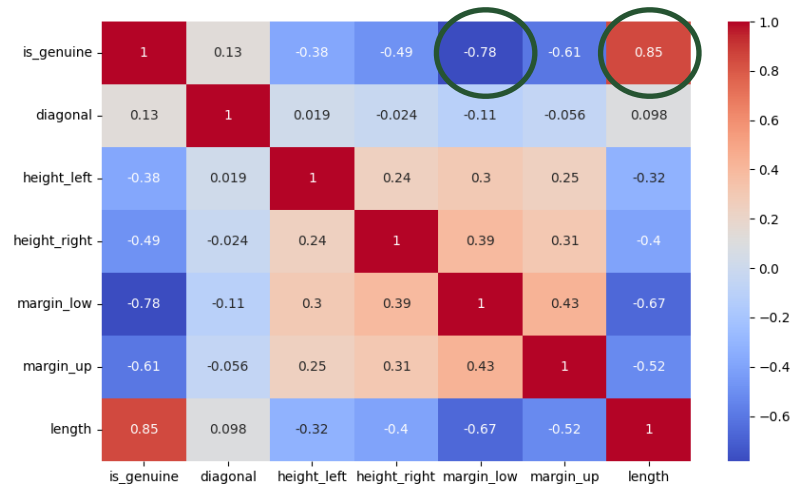
	diagonal	height_left	height_right	margin_low	margin_up	length
Ecart median (%)	0.05	-0.22	-0.34	-26.28	-9.84	1.39

MARGIN_LOW : 37 VALEURS MANQUANTES SOIT ENVIRON 2,5% DES DONNEES

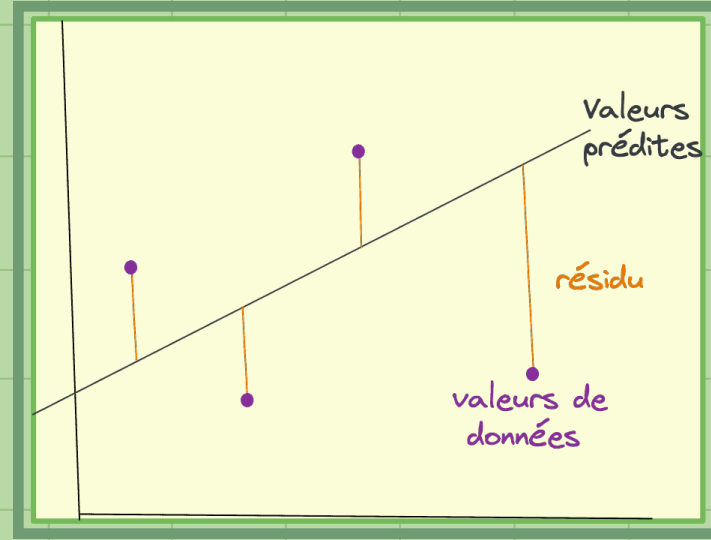
ANALYSE DES CORRELATIONS



Certaines variables de notre jeu de données sont *statistiquement* liées à l'authenticité du billet, notamment la longueur et la marge haute.



REGRESSION LINEAIRE



METHODE DES MOINDRES CARRES - LEAST SQUARE

Minimise la somme du carré des écarts

REGRESSION PAR ELIMINATION SUCCESSIVE - BACKWARD REGRESSION

```
OLS Regression Results
=====
Dep. Variable:      margin_low  R-squared:          0.617
Model:              OLS        Adj. R-squared:         0.615
Method:             Least Squares  F-statistic:        390.7
Date:               Mon, 17 Jun 2024  Prob (F-statistic):    4.75e-299
Time:               10:58:43     Log-Likelihood:     -774.14
No. Observations:   1463        AIC:                  1562.
Df Residuals:       1456        BIC:                  1599.
Df Model:            6
Covariance Type:    nonrobust
=====
               coef    std err          t      P>|t|      [0.025    0.975]
-----
Intercept      2.8668      8.316      0.345    0.730     -13.445    19.179
diagonal      -0.0130      0.036     -0.364    0.716      -0.083    0.057
is_genuine     -1.1406      0.050    -23.025    0.000      -1.238    -1.043
height_left    0.0283      0.039      0.720    0.468      -0.048    0.105
height_right   0.0267      0.038      0.700    0.484      -0.048    0.102
margin_up     -0.2128      0.059    -3.621    0.000      -0.328    -0.098
length        -0.0039      0.023     -0.166    0.868      -0.050    0.042
=====
Omnibus:            21.975  Durbin-Watson:      2.038
Prob(Omnibus):      0.000  Jarque-Bera (JB):    37.993
Skew:               0.061  Prob(JB):            5.62e-09
Kurtosis:           3.780  Cond. No.             1.95e+05
=====
```

61% DE LA VARIANCE
TOTAL DE NOTRE
VARIABLE DEPENDANTE
EST EXPLIQUEE PAR
NOTRE MODELE DE
REGRESSION LINEAIRE.

IDENTIFICATION DES VARIABLES
PREDICTRICES SIGNIFICATIVES

```
OLS Regression Results
=====
Dep. Variable:      margin_low  R-squared:          0.617
Model:              OLS        Adj. R-squared:         0.616
Method:             Least Squares  F-statistic:        1174.
Date:               Mon, 17 Jun 2024  Prob (F-statistic):    1.24e-304
Time:               10:58:43     Log-Likelihood:     -774.73
No. Observations:   1463        AIC:                  1555.
Df Residuals:       1460        BIC:                  1571.
Df Model:            2
Covariance Type:    nonrobust
=====
               coef    std err          t      P>|t|      [0.025    0.975]
-----
Intercept      5.9263      0.198    30.003    0.000      5.539    6.314
is_genuine     -1.1632      0.029   -40.477    0.000      -1.220    -1.107
margin_up     -0.2119      0.059    -3.612    0.000      -0.327    -0.097
=====
Omnibus:            22.365  Durbin-Watson:      2.041
Prob(Omnibus):      0.000  Jarque-Bera (JB):    39.106
Skew:               0.057  Prob(JB):            3.22e-09
Kurtosis:           3.793  Cond. No.             65.0
=====
```

LIMITES DE LA REGRESSION LINEAIRE

01

Absence d'outliers

Aucun outliers significatifs selon l'étude d'un diagramme des leviers, de la distance de Cook et des résidus studentisés.



02

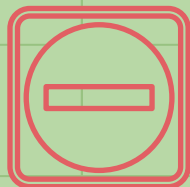
Colinéarité des variables

Aucune colinéarité significatives entre les variables indépendantes selon la méthode VIF (Variance Inflation Factor).

03

Homoscédasticité

Le modèle ne respecte pas le seuil d'homoscédasticité des résidus selon les tests de Test de Breusch-Pagan et de White.



04

Normalité des résidus

Les résidus de notre modèle ne suivent pas une distribution normale selon le test de Shapiro-Wilk.

PREDICTIONS DES VALEURS MANQUANTES

	is_genuine	diagonal	height_left	height_right	margin_low	margin_up	length
72	1	-0.060441	-0.466103	-1.444792	-0.620446	0.425168	0.127803
99	1	-0.093217	0.135176	0.797784	-0.585314	-0.049510	0.460204
151	1	0.365659	-0.766742	1.412188	-0.546987	-0.567341	0.288272
197	1	-1.666506	-1.234404	-0.369585	-0.738619	2.021814	0.677984
241	1	-0.420986	0.369007	0.429141	-0.546987	-0.567341	-0.365069

	Without_regression	With_regression	Difference
count	1463.000	1500.000	37.000
mean	0.000	-0.005	-0.005
std	1.000	0.994	-0.006
min	-2.269	-2.269	0.000
25%	-0.710	-0.691	0.019
50%	-0.265	-0.265	0.000
75%	0.579	0.579	0.000
max	3.638	3.638	0.000

RESULTAT FINAL

Les statistiques descriptives de notre jeu de données avant et après l'intégration des valeurs prédites par notre modèle de régression linéaire.

The background of the slide is a repeating pattern of stylized green dollar bills. The bills are depicted in various orientations and are slightly offset from each other, creating a sense of depth and movement. Each bill features a large, bold dollar sign (\$) in the center. The overall color palette is a range of greens, from light to dark.

ALGORITHMES D'APPRENTISSAGE

MESURES DE PERFORMANCES

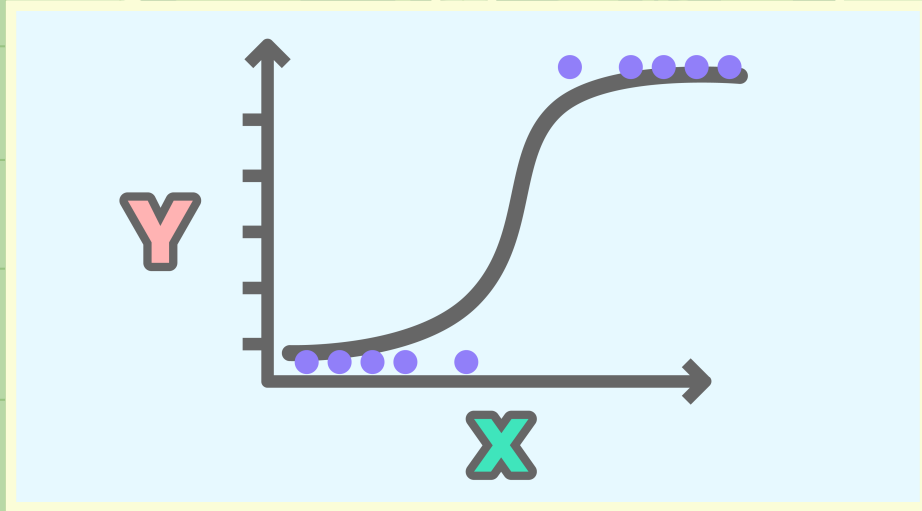
La précision mesure la proportion des billets que notre modèle a identifiés comme faux et qui sont réellement faux.

Le rappel mesure la proportion des faux billets réels que notre modèle a correctement identifiés comme faux.

Le F1-score est la moyenne harmonique de la précision et du rappel, offrant un équilibre entre les deux.

L'accuracy mesure la proportion totale de billets correctement identifiés (vrais ou faux) par le modèle.

REGRESSION LOGISTIQUE



Modéliser la probabilité qu'un événement binaire se produise, en fonction de valeurs d'une ou plusieurs variables indépendantes.

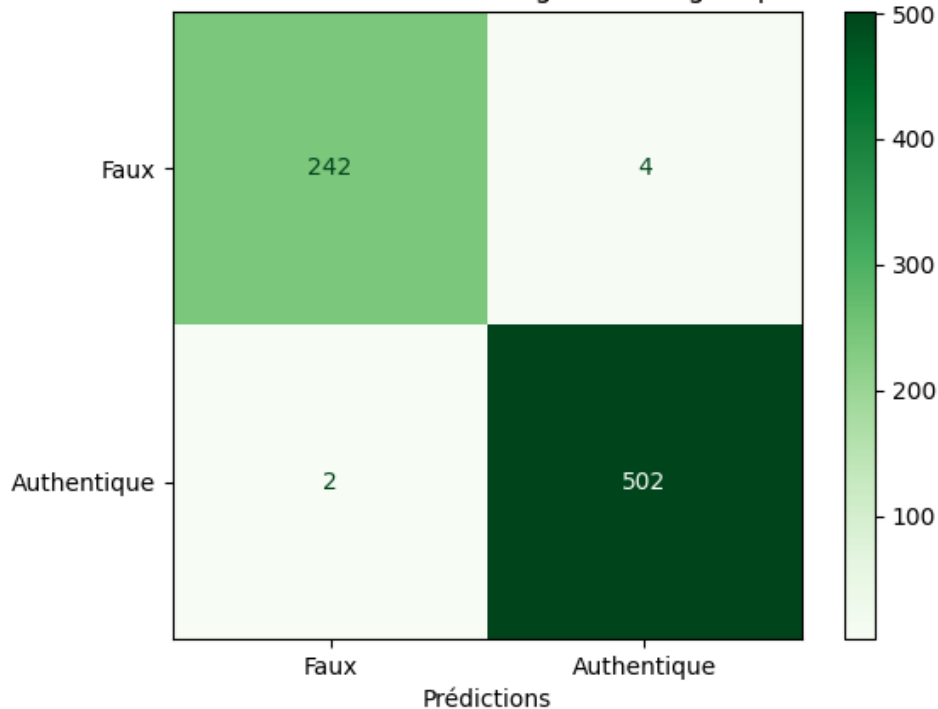
Prédit la probabilité qu'un événement se produise

Sigmoïde transforme les valeurs en probabilités

Résultat en fonction d'un seuil, généralement 0.5

MATRICE DE CONFUSION

Matrice de confusion - Régression logistique



<i>PRECISION</i>	99,21
<i>RECALL</i>	99,6
<i>F1 SCORE</i>	99,41
<i>ACCURACY GENERALE</i>	99,2

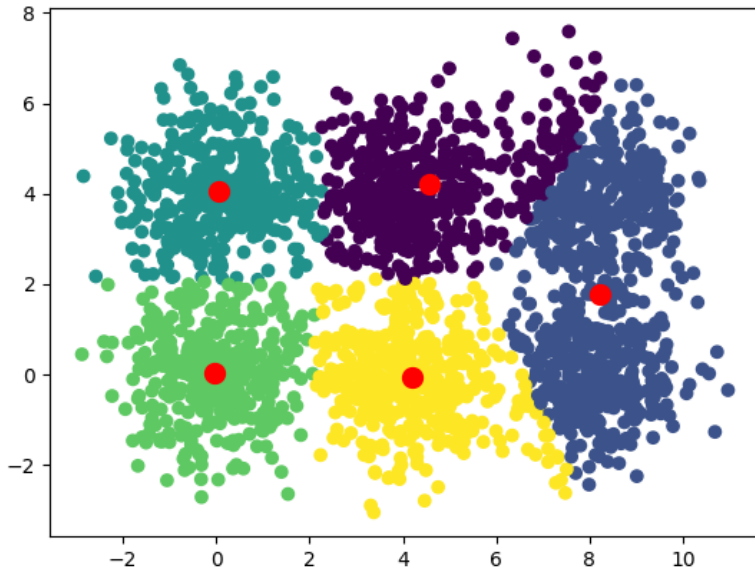
CLUSTERING KMEANS

Diviser un ensemble de données en plusieurs "clusters" distincts, de manière à ce que les individus d'un même groupe soient plus similaires entre eux qu'avec ceux des autres groupes.

Divise les données en un nombre fixe de groupe, appelés clusters.

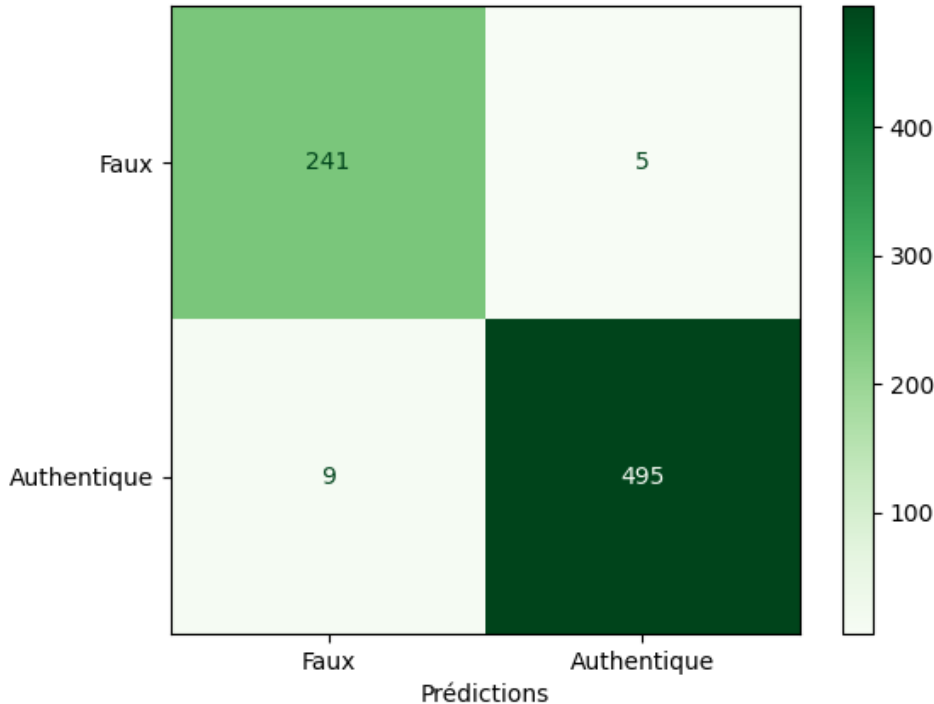
Assigne chaque individu au cluster dont le centre est le plus proche.

Ajuste les centres des clusters jusqu'à ce que les points soient bien regroupés.



MATRICE DE CONFUSION

Matrice de confusion - Kmeans



PRECISION

99,00

RECALL

98,21

F1 SCORE

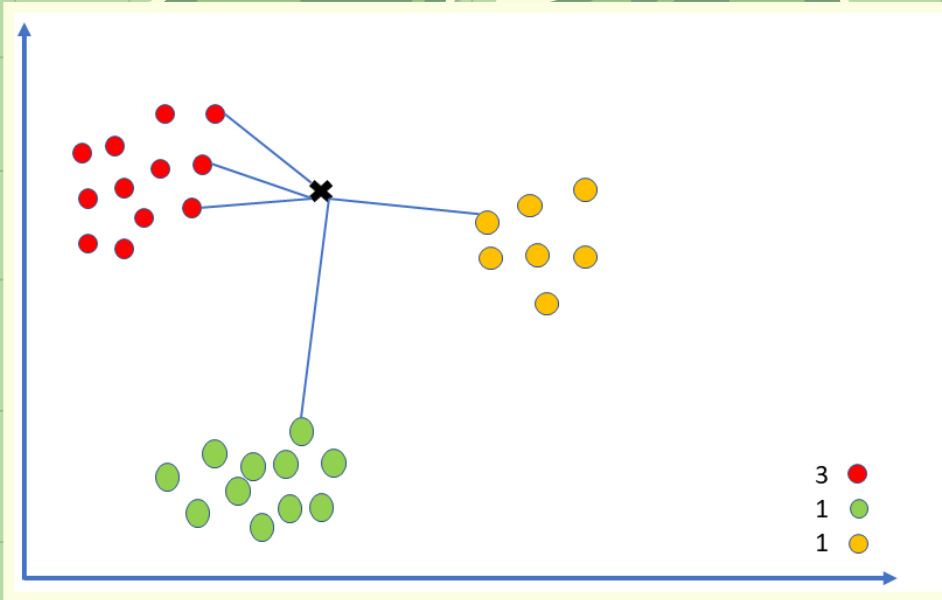
98,61

***ACCURACY
GENERALE***

98,13

K-NEAREST NEIGHBORS

CLASSIFIER LES DONNEES EN
FONCTION DES DONNEES LES
PLUS PROCHES DANS L'ESPACE
DES VARIABLES INDEPENDANTES.



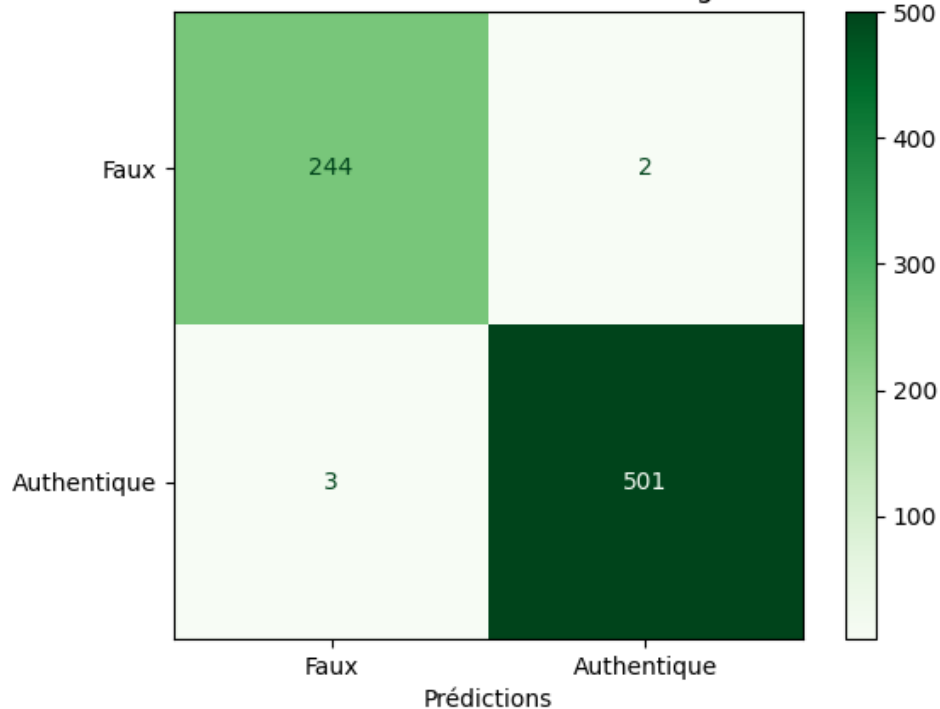
Trouve les k points les plus proches du point à classer.

Calcule la classe majoritaire parmi ces k voisins.

Assigne la classe majoritaire au point à classer.

MATRICE DE CONFUSION

Matrice de confusion - K-Nearest Neighbors



PRECISION

99,60

RECALL

99,40

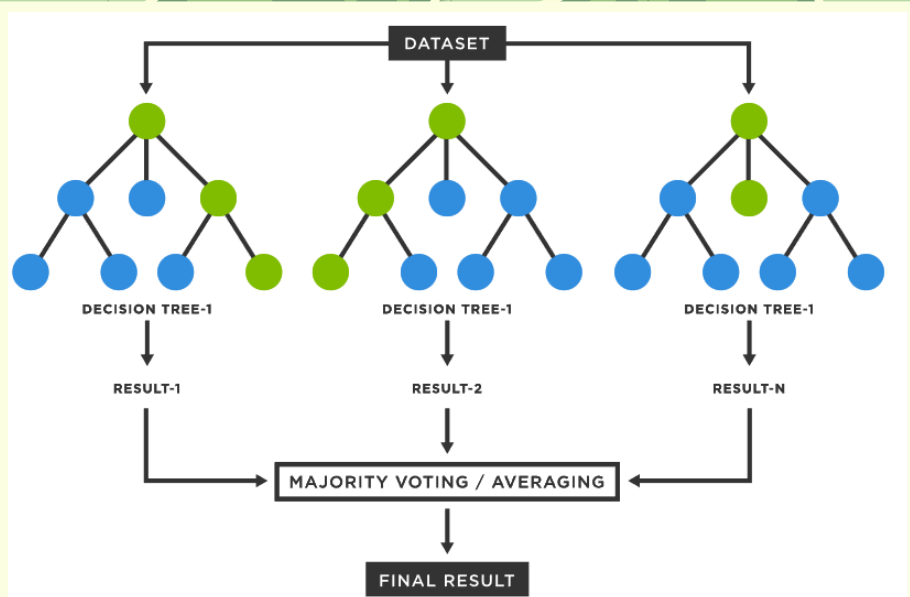
F1 SCORE

99,50

***ACCURACY
GENERALE***

99,33

RANDOM FOREST



AGREGER LES PREDICTIONS DE PLUSIEURS ARBRES DE DECISIONS FORME SUR DES SOUS ENSEMBLE ALEATOIRES.

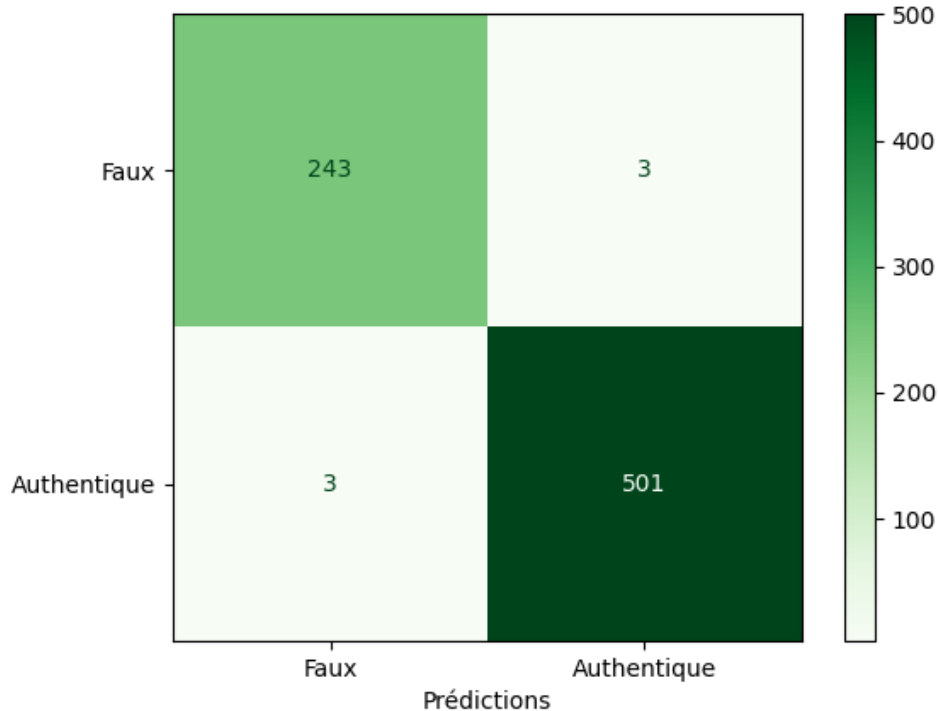
Plusieurs arbres de décision sur des sous-ensembles.

Chaque arbre donne une prédiction pour la donnée.

La prédiction finale est la majorité des votes.

MATRICE DE CONFUSION

Matrice de confusion - Random Forest



PRECISION

99,40

RECALL

99,40

F1 SCORE

99,40

***ACCURACY
GENERALE***

99,20

CONCLUSION & DEMONSTRATION

	Modèle	Précision	Rappel	F1 Score	Précision Globale
0	Régression Logistique	99.21	99.60	99.41	99.20
1	KNN	99.60	99.40	99.50	99.33
2	Random Forest	99.40	99.40	99.40	99.20
3	Kmeans	99.00	98.21	98.61	98.13

The background is a light green grid. Scattered across the grid are several stylized green dollar bills with white outlines and a large green dollar sign in the center. The bills are of various sizes and orientations, some appearing to float or fall.

MERCI

**SI VOUS AVEZ DES
QUESTIONS, N'HESITEZ PAS !**

matthieuicartpro@gmail.com