

PROJET A3

ISEN

ALL IS DIGITAL

QUEST

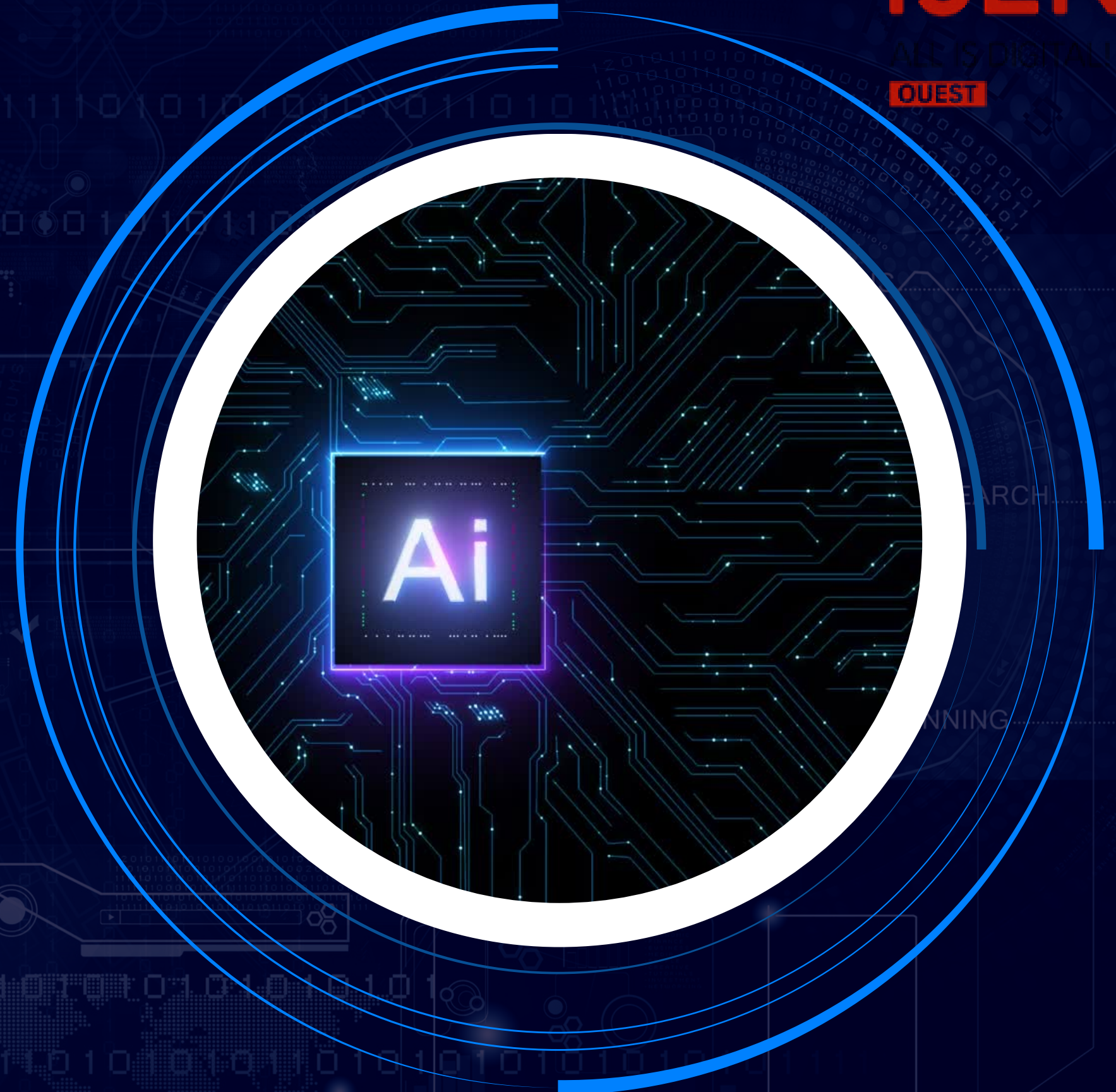
INTELLIGENCE ARTIFICIELLE

Nicolas Capelle, Thomas Catros, Hermine Paul

02 - 06 - 2023

SOMMAIRE

- *Gestion de projet*
- *Réduction de dimension*
- *Clustering*
- *Répartition des données*
- *Classification avec KNN*
- *Classification avec algorithme de haut niveau*
- *Métriques*
- *Création de scripts*



GESTION DE PROJET

ISEN

ALL IS DIGITAL!

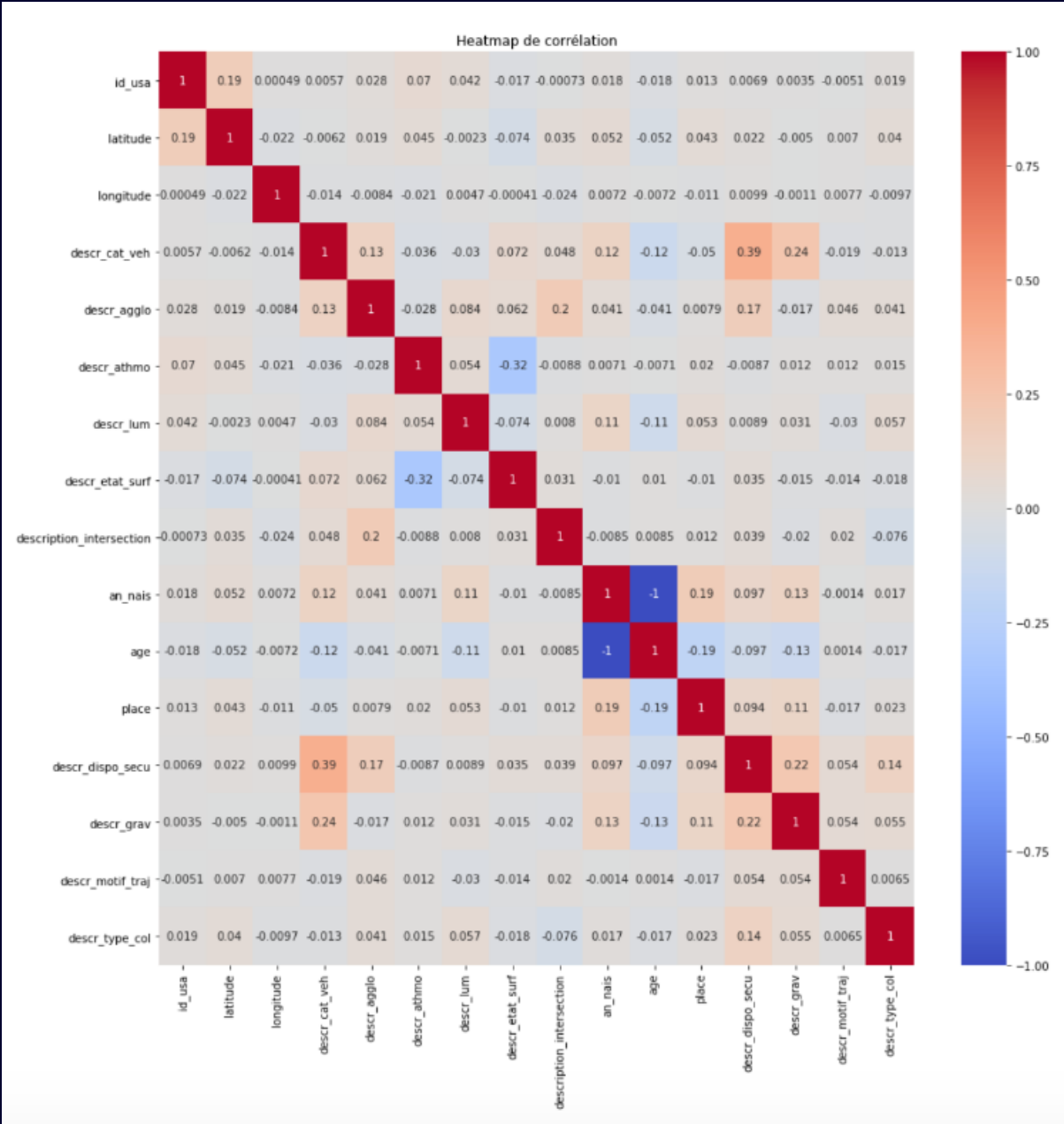


QUEST

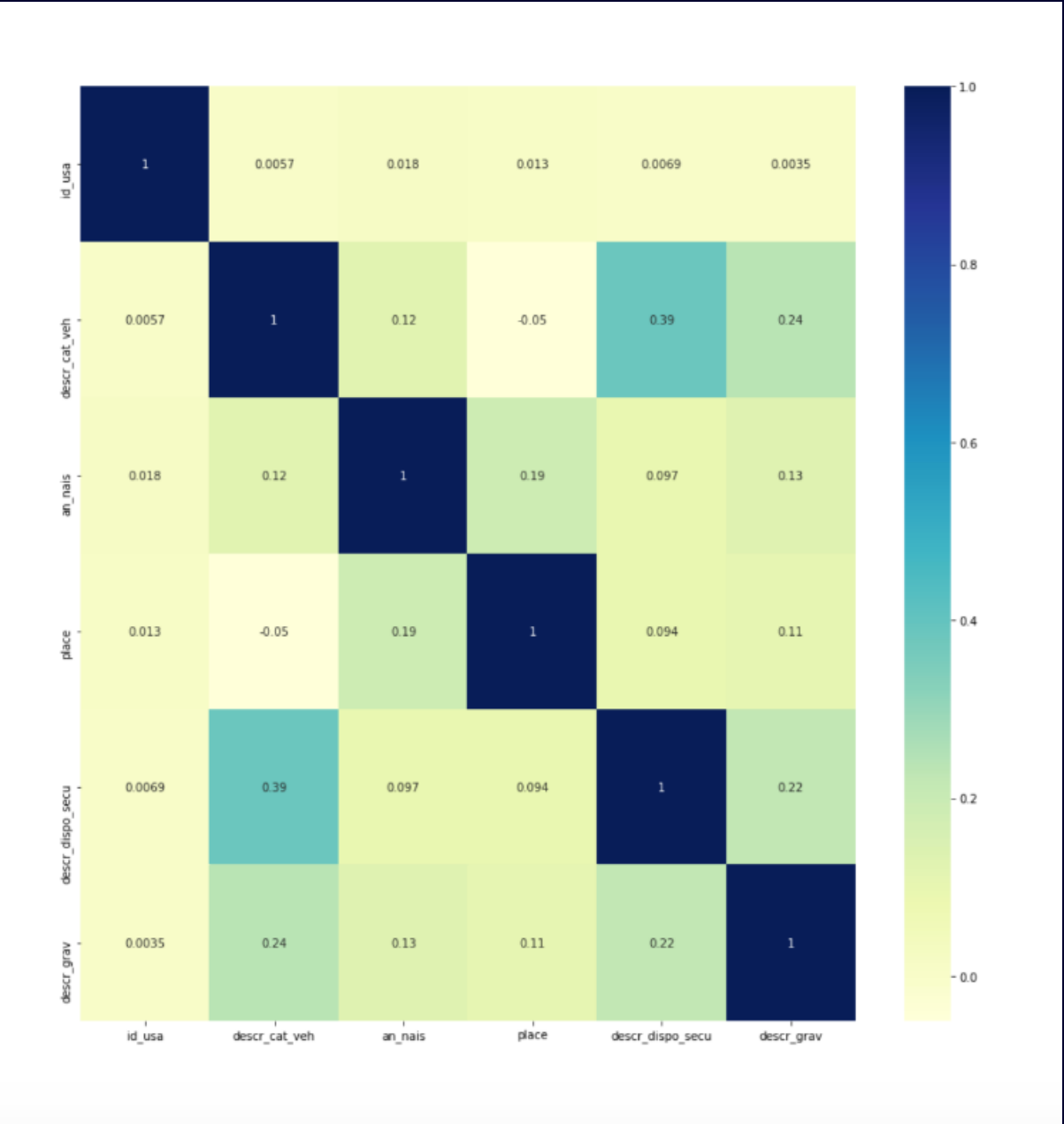


Nicolas
 Thomas
 Hermine

RÉDUCTION DE DIMENSION

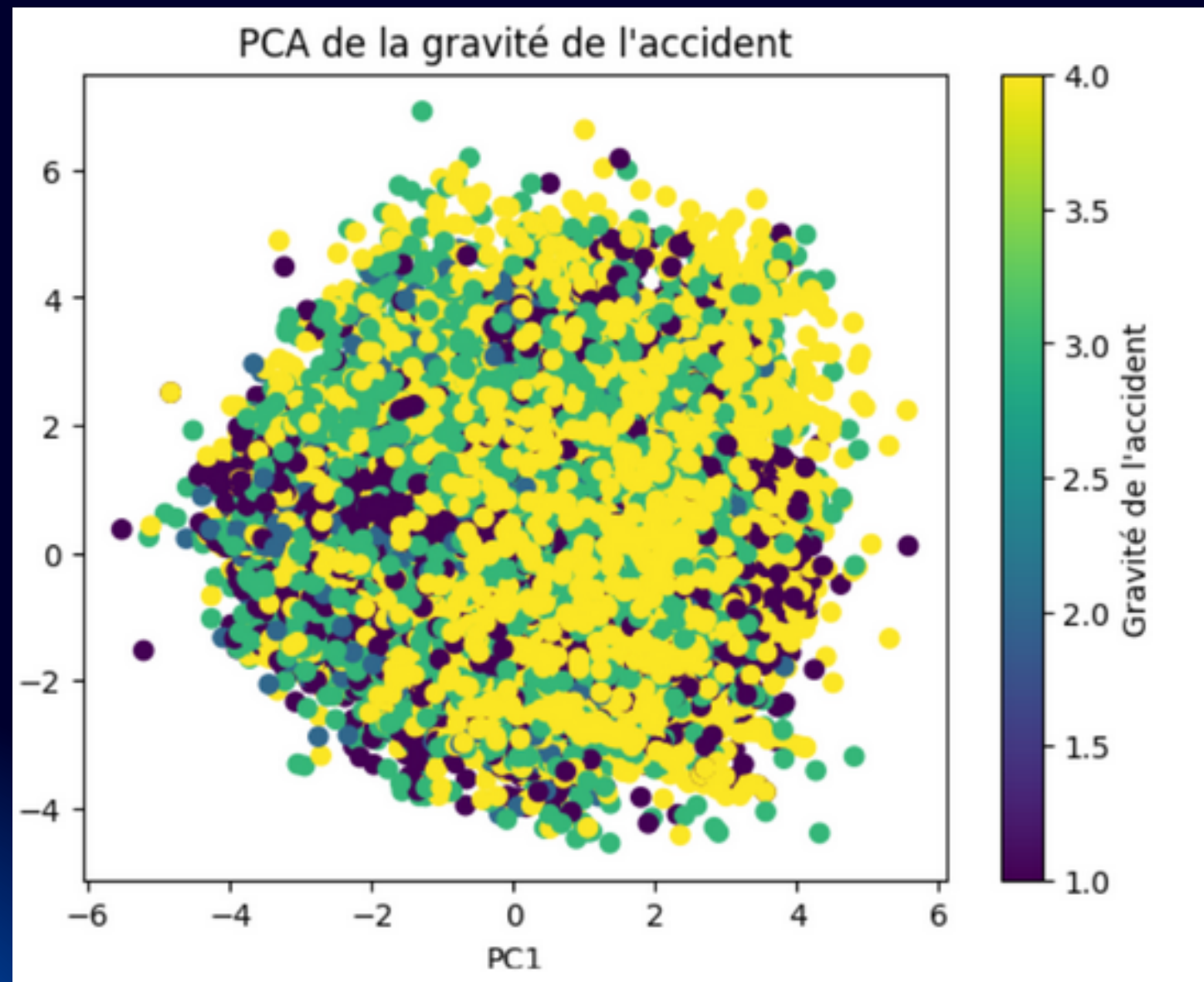


Avant



Après

PCA



- méthode de réduction non supervisée
- transformer un ensemble de de variables

CLUSTERING

- FROM SCRATCH

```
0      latitude longitude cluster_label
1      47.4833  -1.233330          0
2      45.6000   5.633330          2
3      45.6000   5.633330          2
4      42.7667   2.983330          1
5      45.9667   0.066667          1
...
57797  46.8333   3.450000          2
57798  49.2667   2.383330          0
57799  46.9833   3.166670          0
57800  45.7833   3.083330          1
57801  48.7667   1.916670          0
```

```
[57802 rows x 3 columns]
```

- classification par étiquetage
- calculer la position de nouveaux centroïdes.
-

CLUSTERING

- FROM SCRATCH

avec $k = 10$

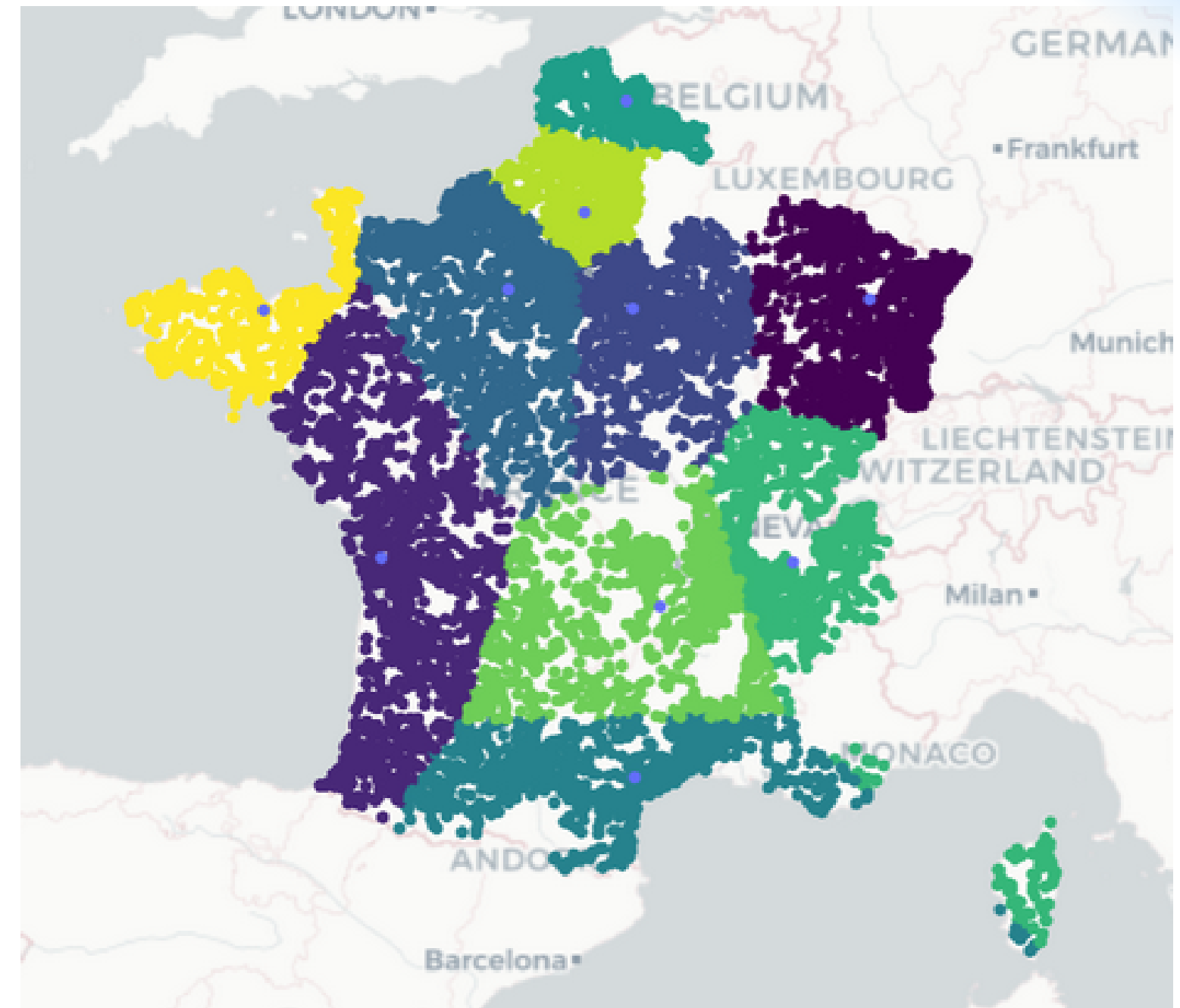
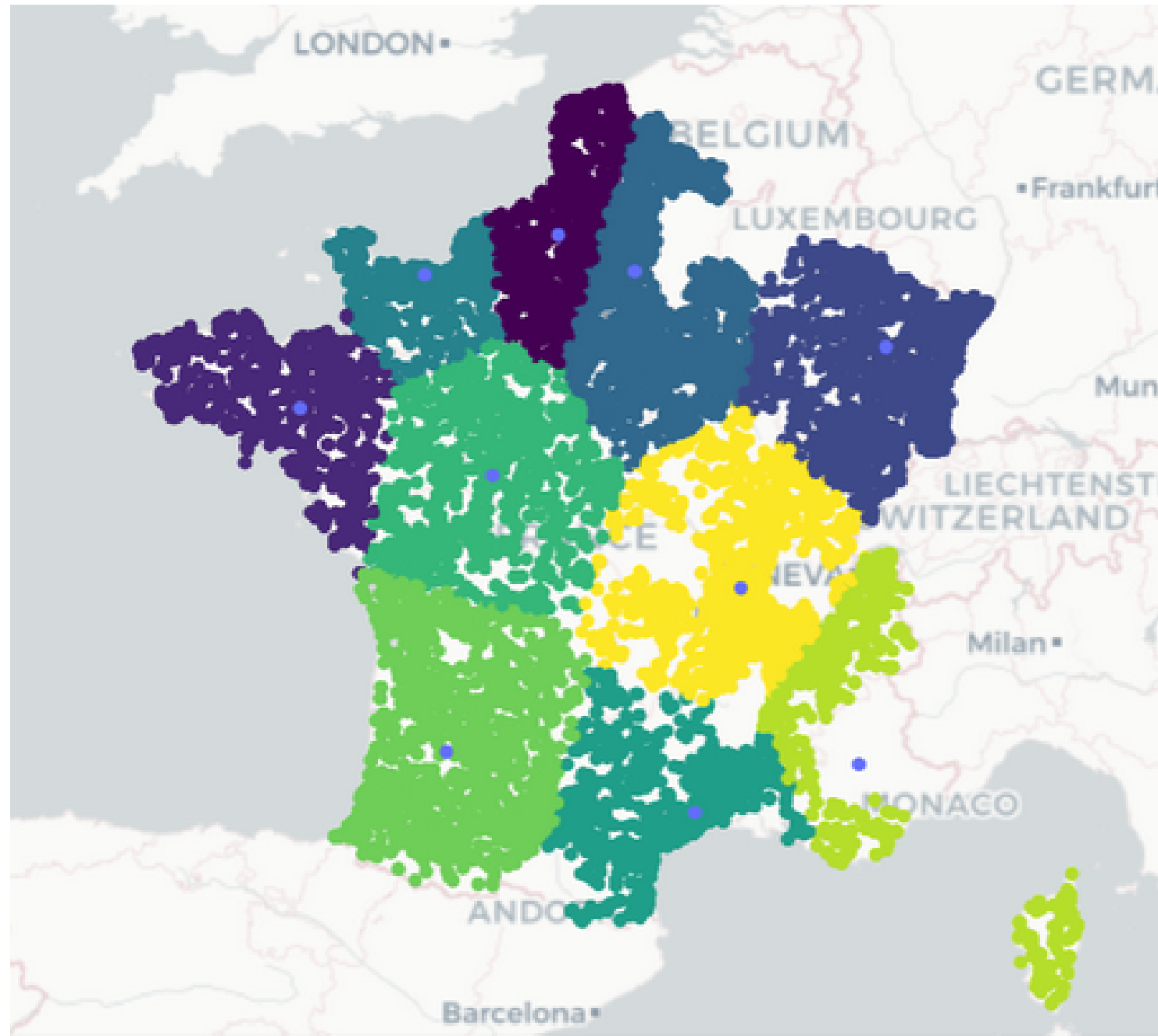
ISEN

ALL IS DIGITAL!

OUEST



yncrea



L1 – distance de Manhattan

L2 – distance euclidienne

CLUSTERING

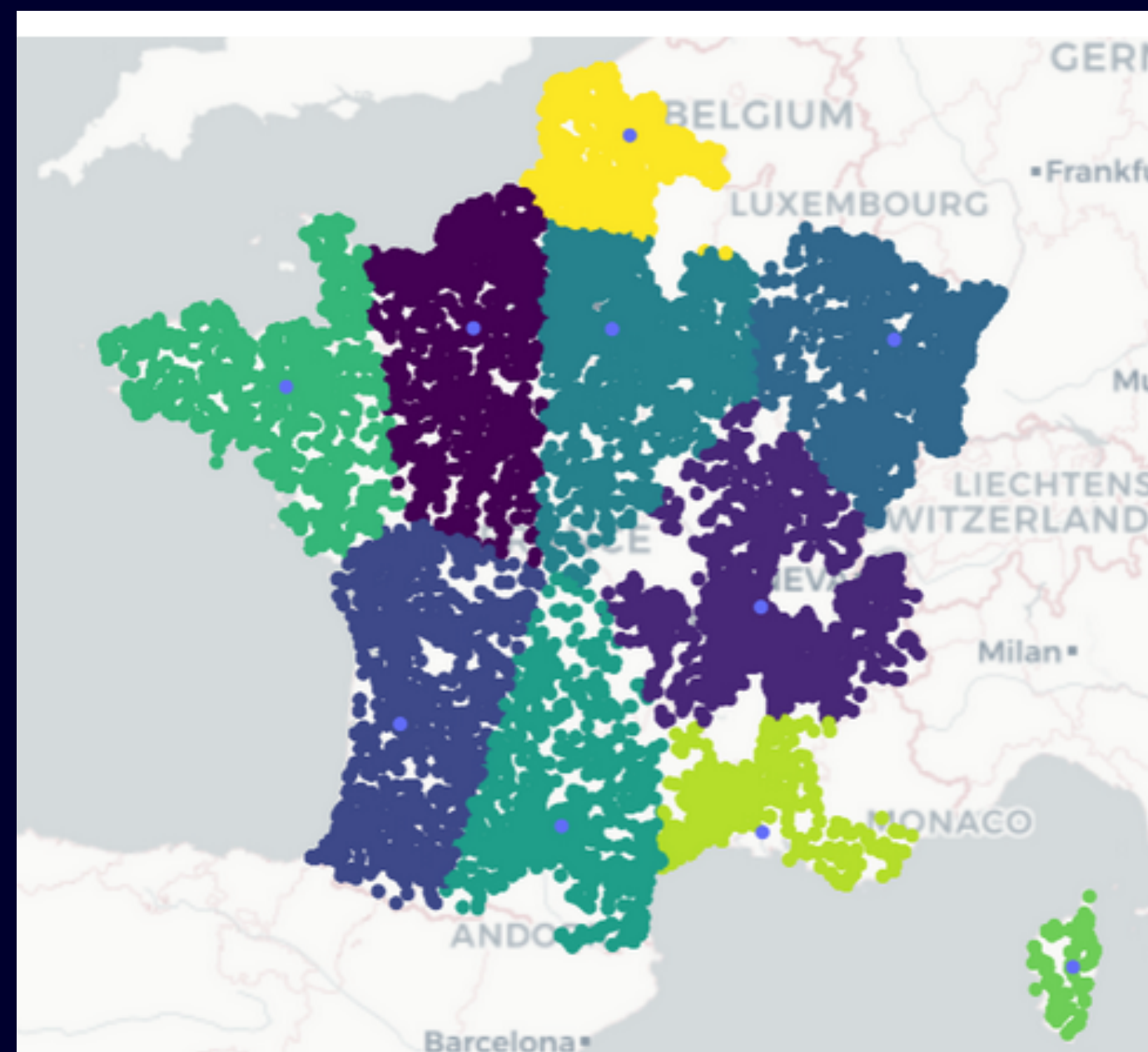
• FROM SCRATCH

ISEN
ALL IS DIGITAL!
OUEST



avec $k = 10$

distance de Harversine



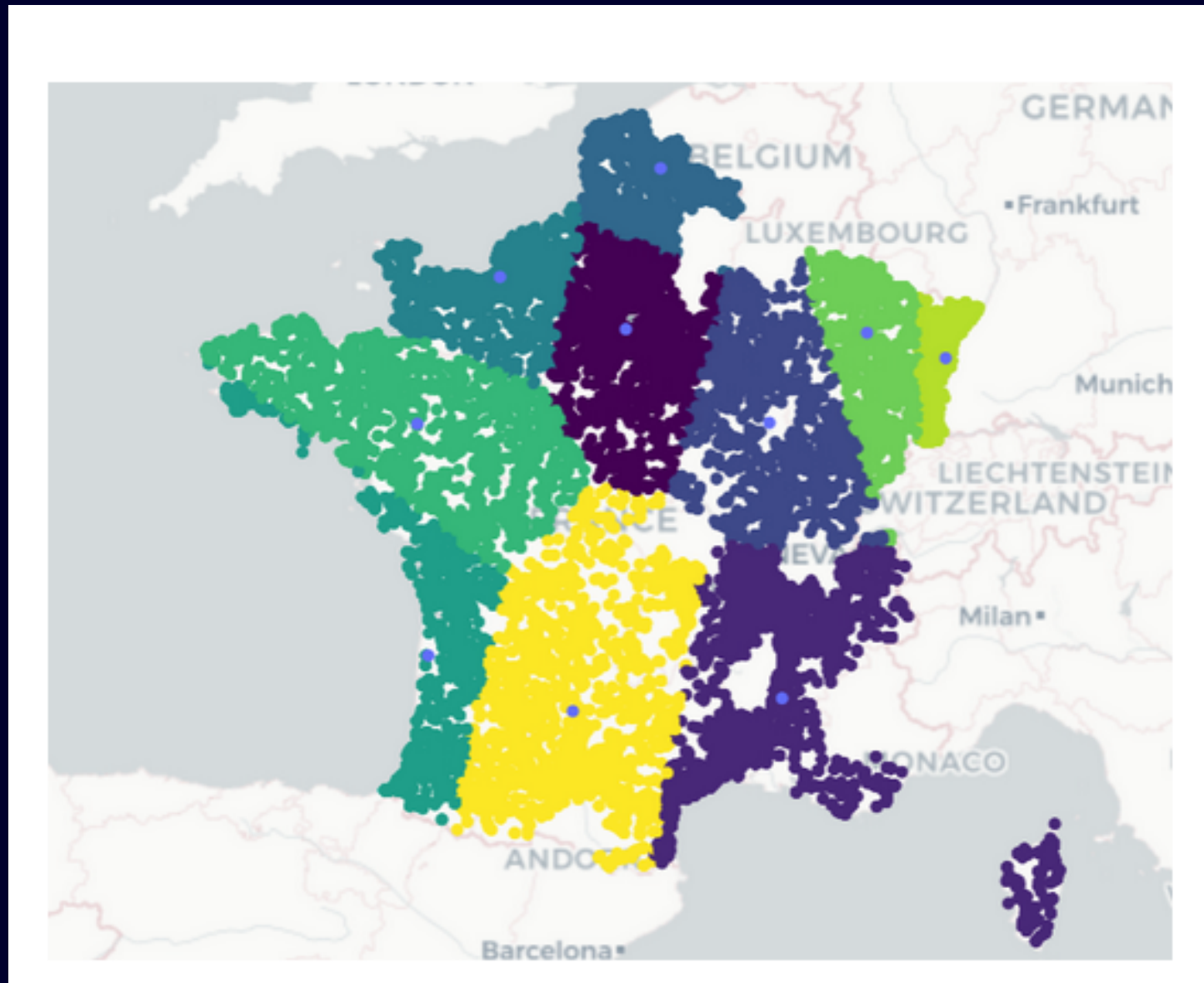
CLUSTERING

- SCIKIT-LEARN

ISEN

ALL IS DIGITAL

OUEST



Calcul des distances avec
les indices de performances
les meilleurs possibles :

- Calinski-Harbasz
- Davies - Bouldin,
- Silhouette Coefficient

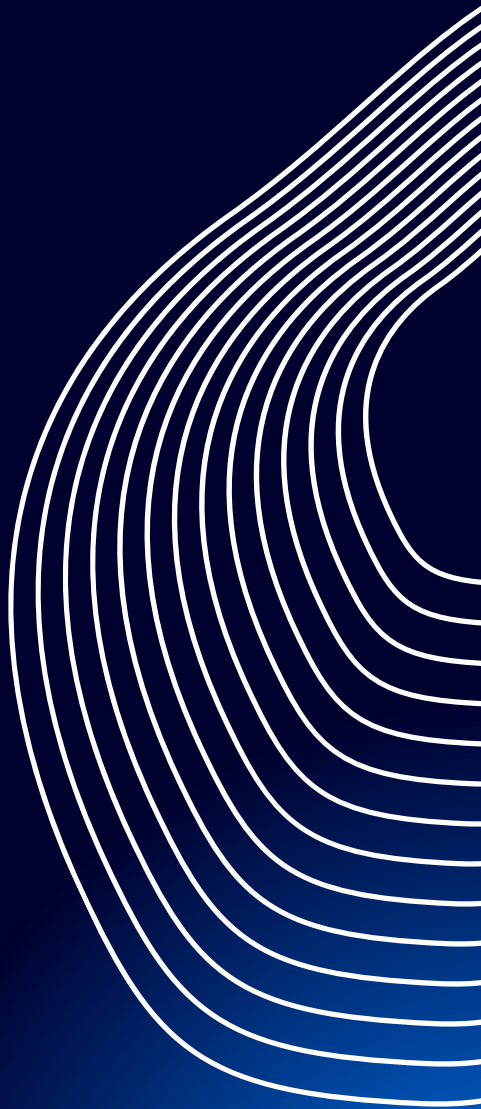
RÉPARTITION DES DONNÉES

EN BASE D'APPRENTISSAGE AVEC L'ALGORITHME
HOLDOUT

Précision (Holdout) : 0.4861898664715566

EN BASE D'APPRENTISSAGE AVEC
L'ALGORITHME LEAVE ONE OUT:

- validation croisée
- évaluer les performances d'un modèle d'apprentissage automatique
- coûteuse en temps de calcul



CLASSIFICATION KNN (AVEC SKLEARN)

Accuracy: 39.4%

Rappel: 29.9%

F1-score: 30.2%

Matrice de confusion: (KNN avec Sklearn)

```
[[83 37 43 35]  
 [72 70 42 23]  
 [65 60 44 23]  
 [92 33 27 51]]
```

CLASSIFICATION RNN (SANS SKLEARN)

les 5 plus proches instances géographiques de l'accident sont:

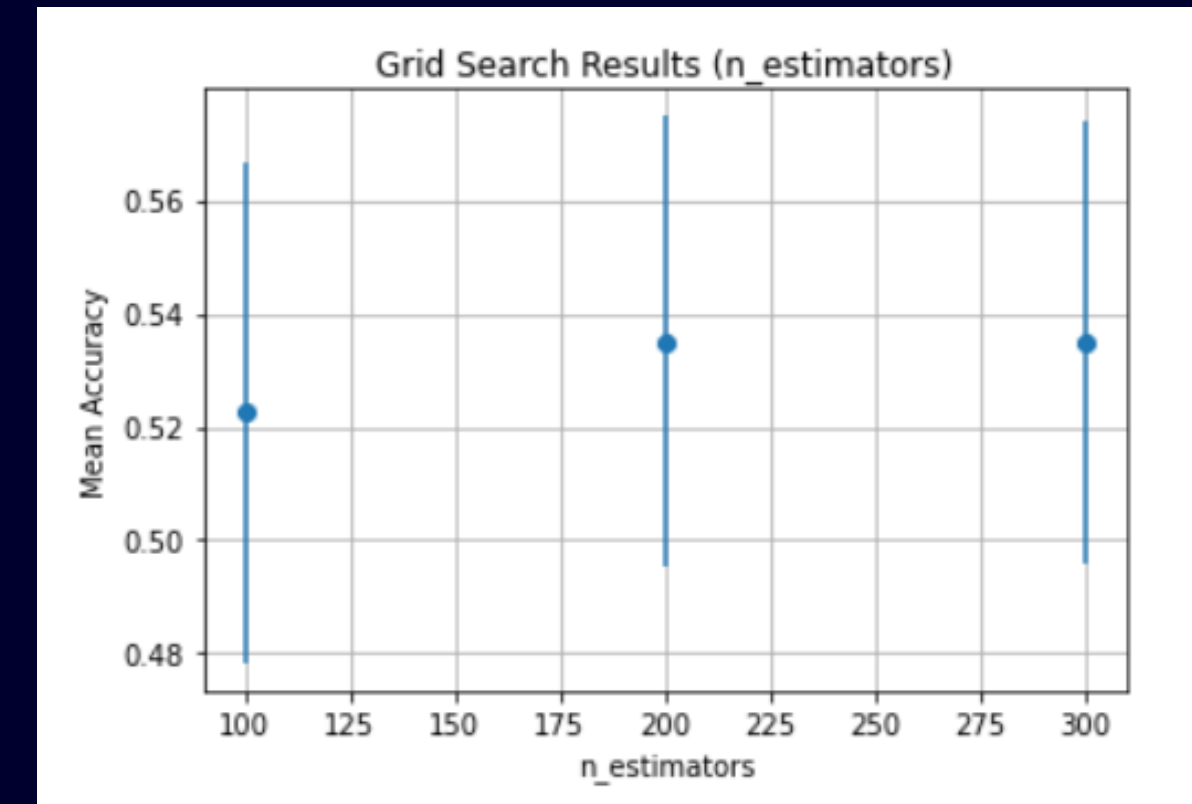
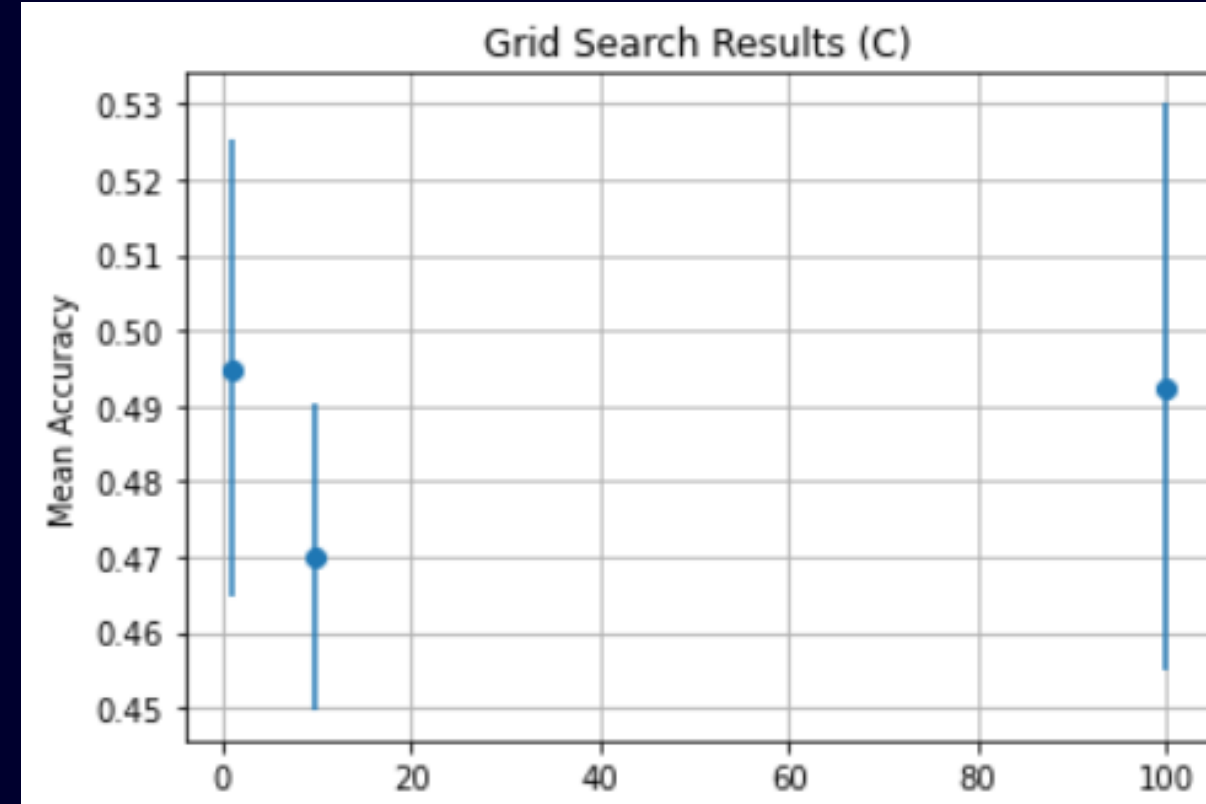
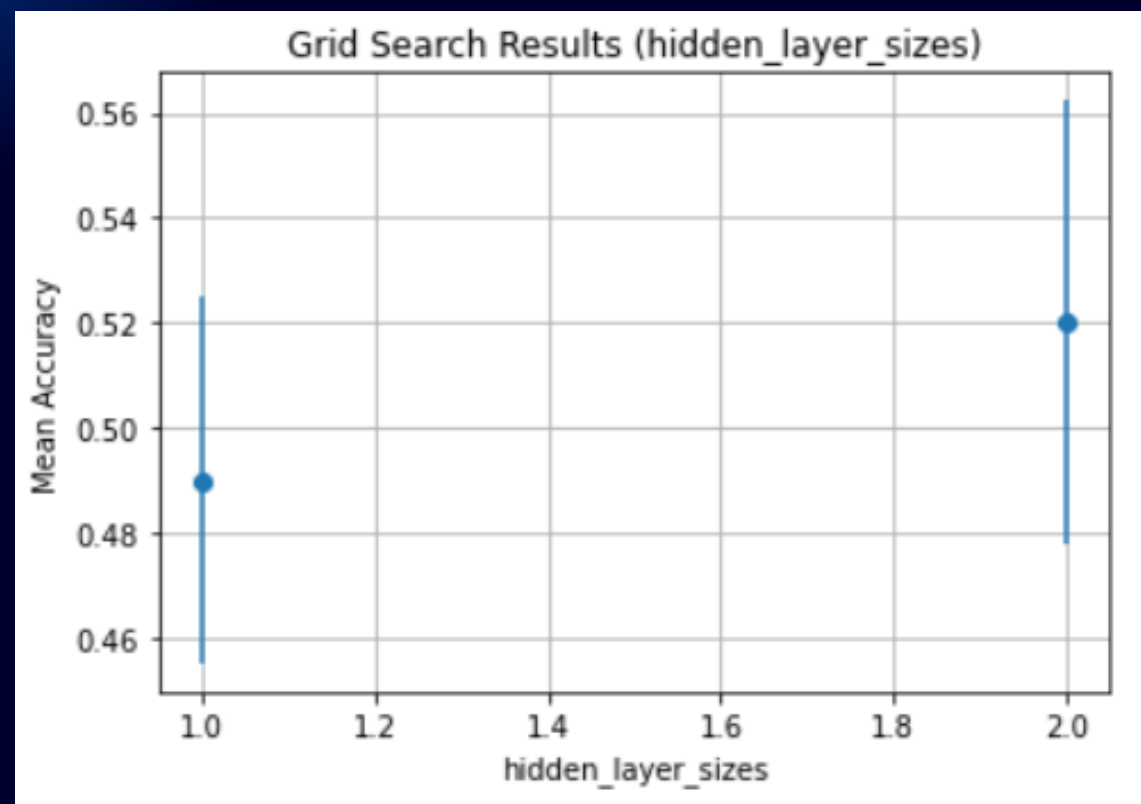
	id_usa	latitude	longitude	descr_cat_veh	an_nais	place
18201	784972.0	47.9167	-3.33333	3	1988.0	1.0
6103	852046.0	47.9167	-3.33333	7	1988.0	1.0
5792	852045.0	47.9167	-3.33333	3	1947.0	1.0
55157	831819.0	47.9167	-3.33333	3	1980.0	1.0
18369	784969.0	47.9167	-3.33333	2	1966.0	1.0

	descr_dispo_secu	descr_grav	distance
18201	1	1	0.060654
6103	2	3	0.060654
5792	1	1	0.060654
55157	3	2	0.060654
18369	1	3	0.060654

ALGORITHMES DE 'HAUT NIVEAU'

	precision	recall	f1-score	support
1	0.48	0.89	0.62	45
3	0.62	0.19	0.29	52
accuracy			0.50	100
macro avg	0.37	0.36	0.30	100
weighted avg	0.54	0.50	0.43	100

REPRÉSENTATION GRAPHIQUE



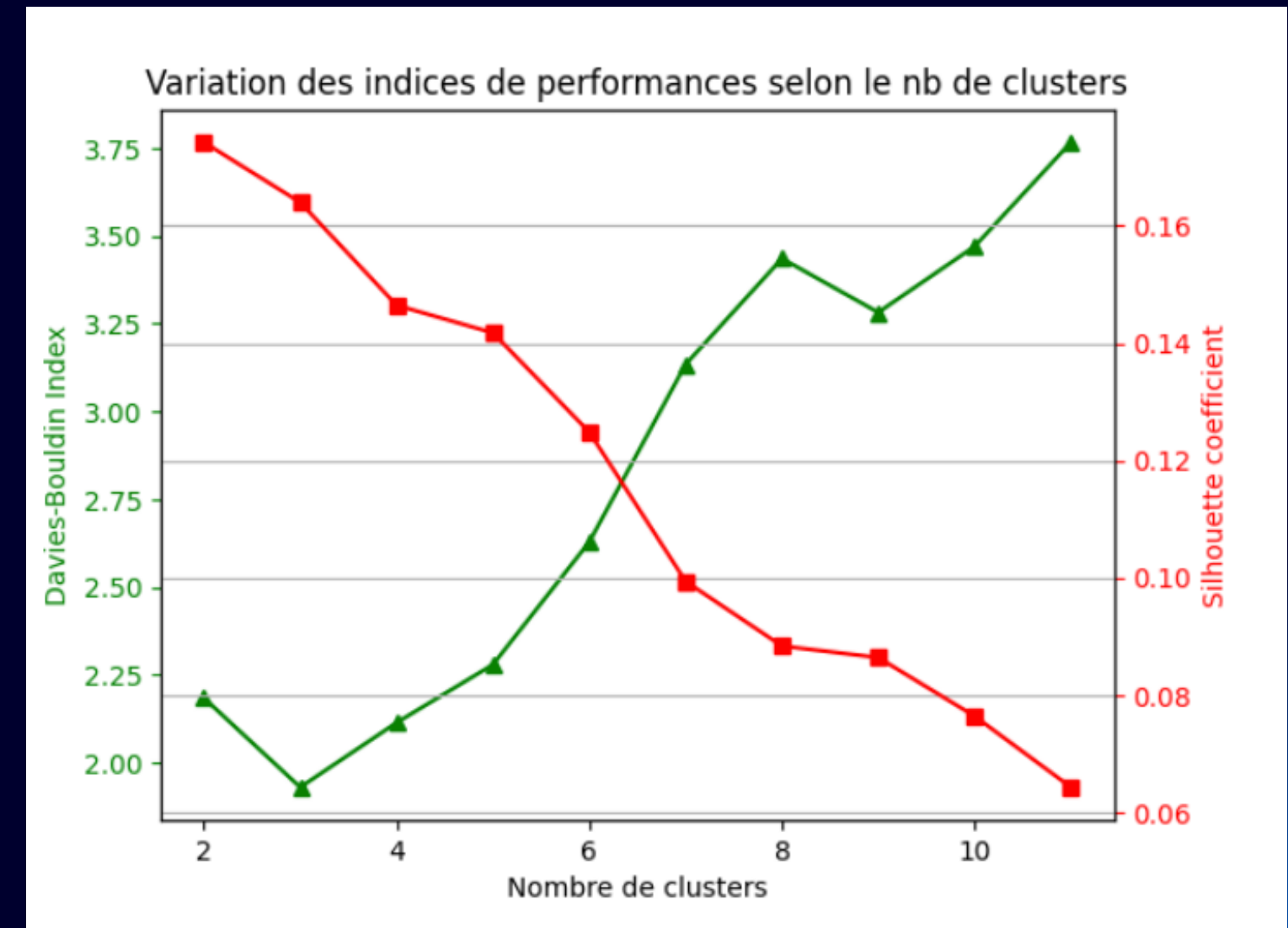
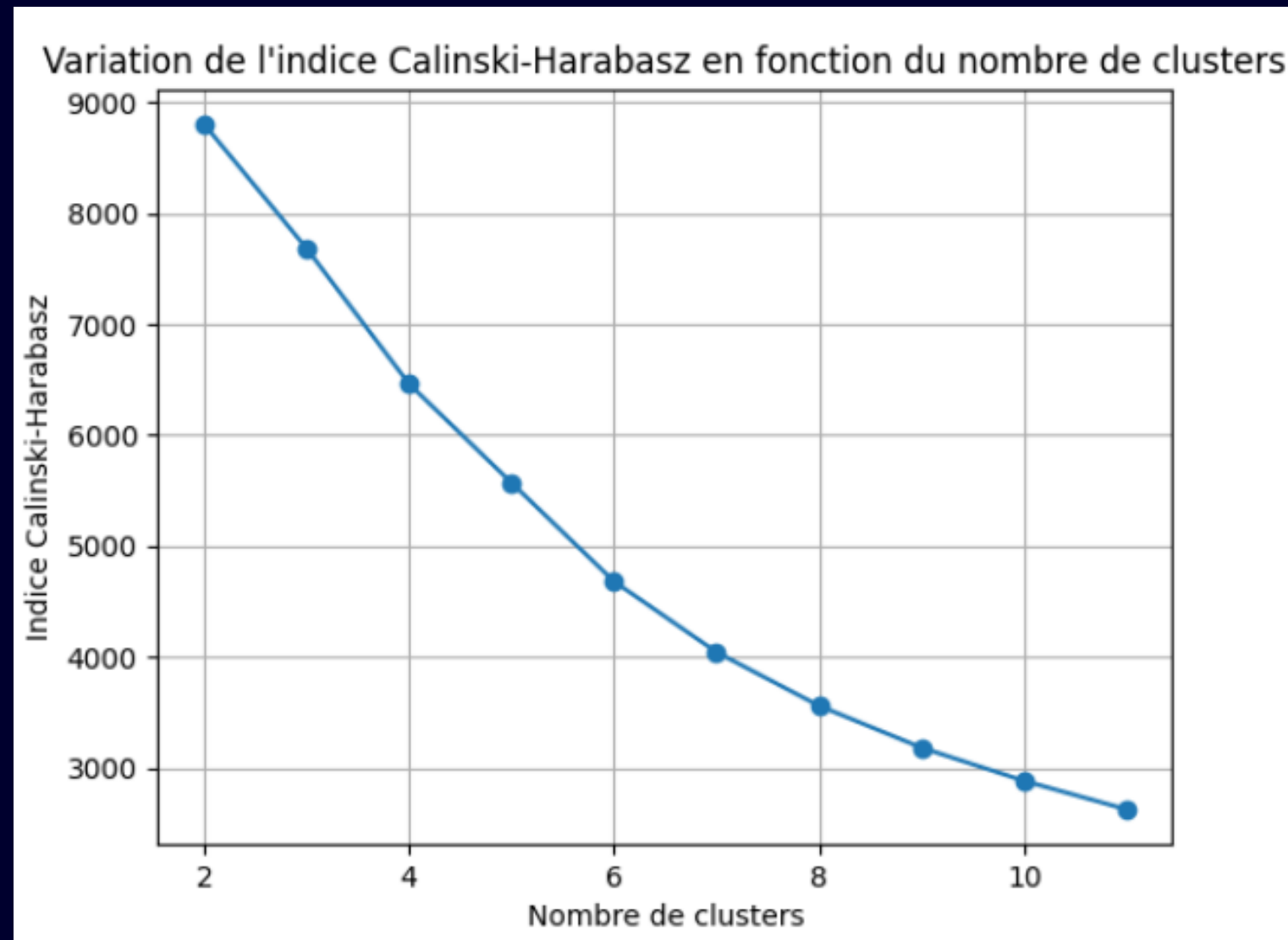
L'ÉVOLUTION DE LA
PRÉCISION MOYENNE EN
FONCTION DES VALEURS DE
MLP

L'ÉVOLUTION DE LA
PRÉCISION MOYENNE, EN
FONCTION DES VALEURS DE
C

L'ÉVOLUTION DE LA
PRÉCISION POUR LA
MÉTHODE RANDOM FOREST
EN FONCTION DU NOMBRE
D'ESTIMATEURS DE RANDOM
FOREST.

ÉVALUATION QUANTITATIVE DES RESULTATS

«NON SUPERVISÉ»



ÉVALUATION QUANTITATIVE RÉSULTATS SUPERVISES

Précision (SVM) : 39.0%
Rappel (SVM) : 39.0%
Taux d'apprentissage(SVM) : 90.0%
Matrice de confusion: (SVM)
[[181 9 8 0]
[92 85 30 0]
[86 62 44 0]
[121 43 38 1]]

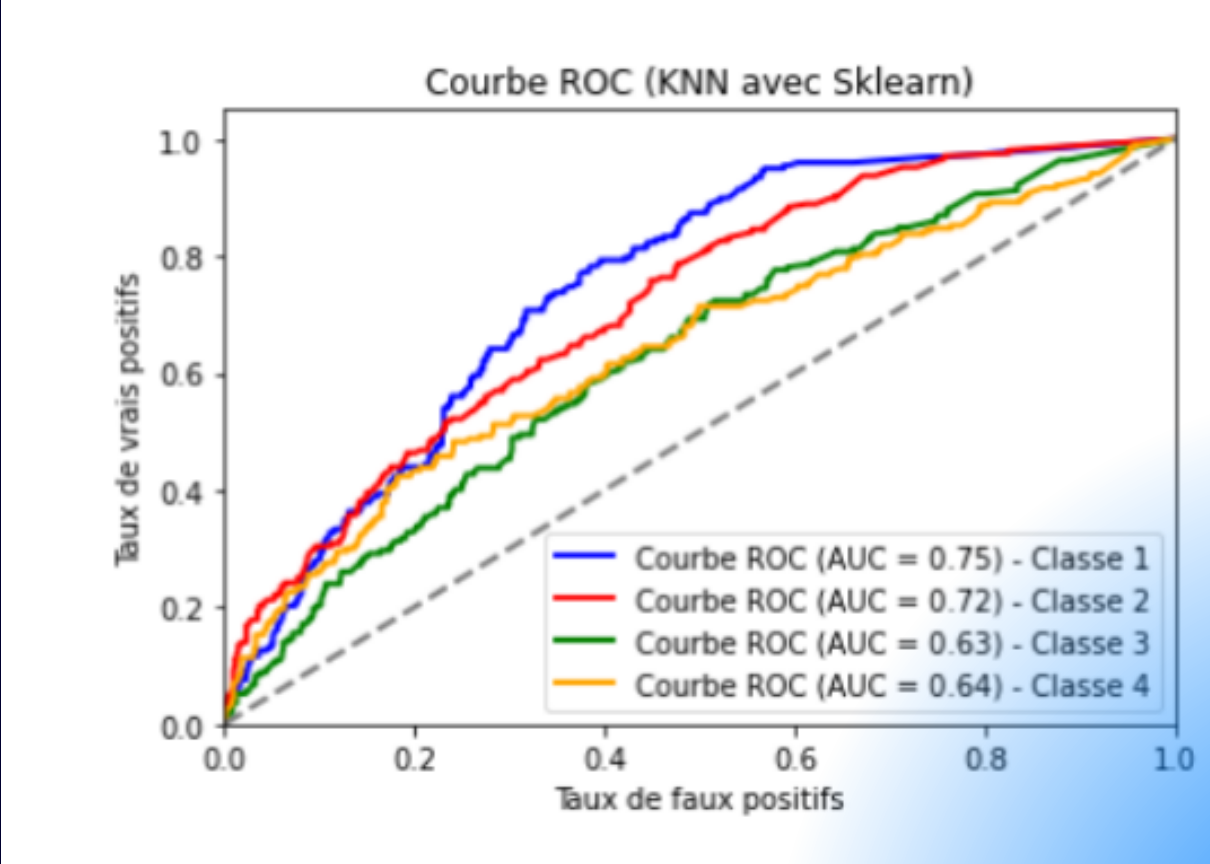
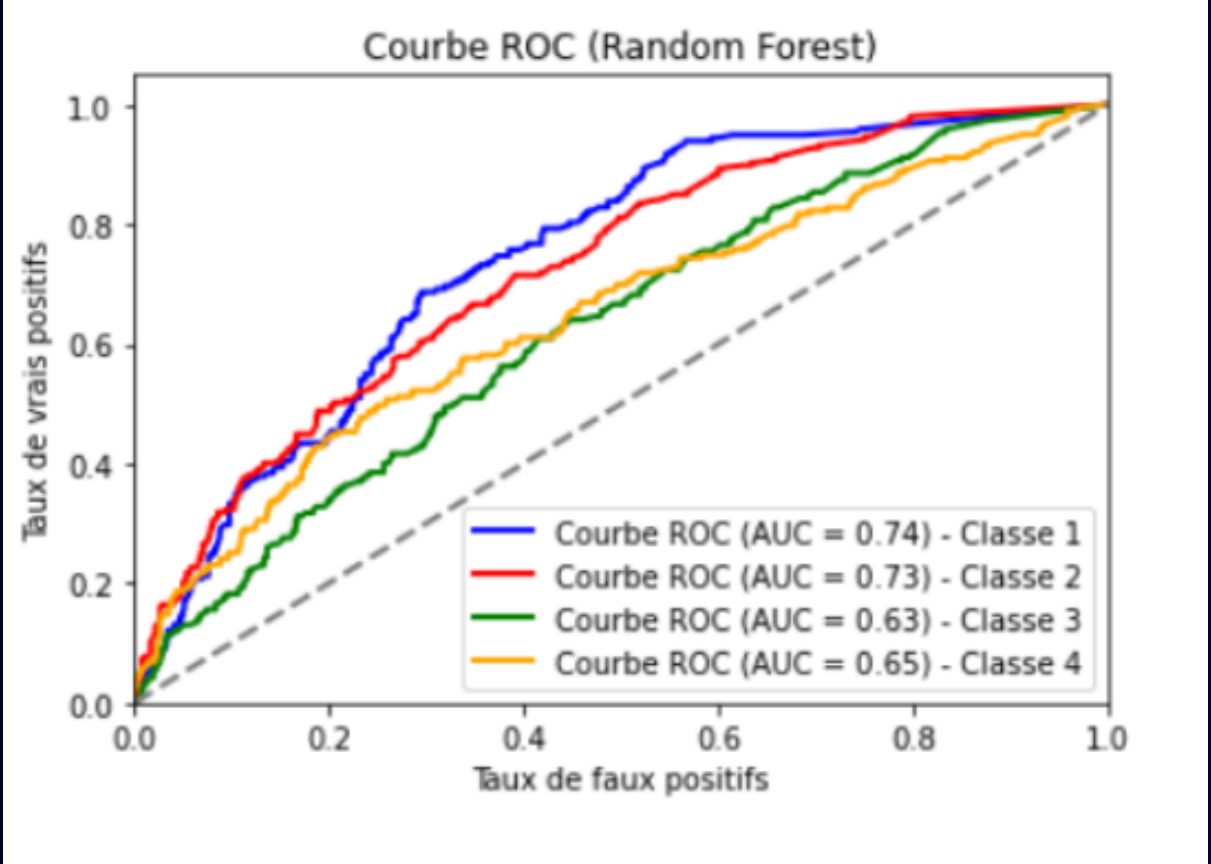
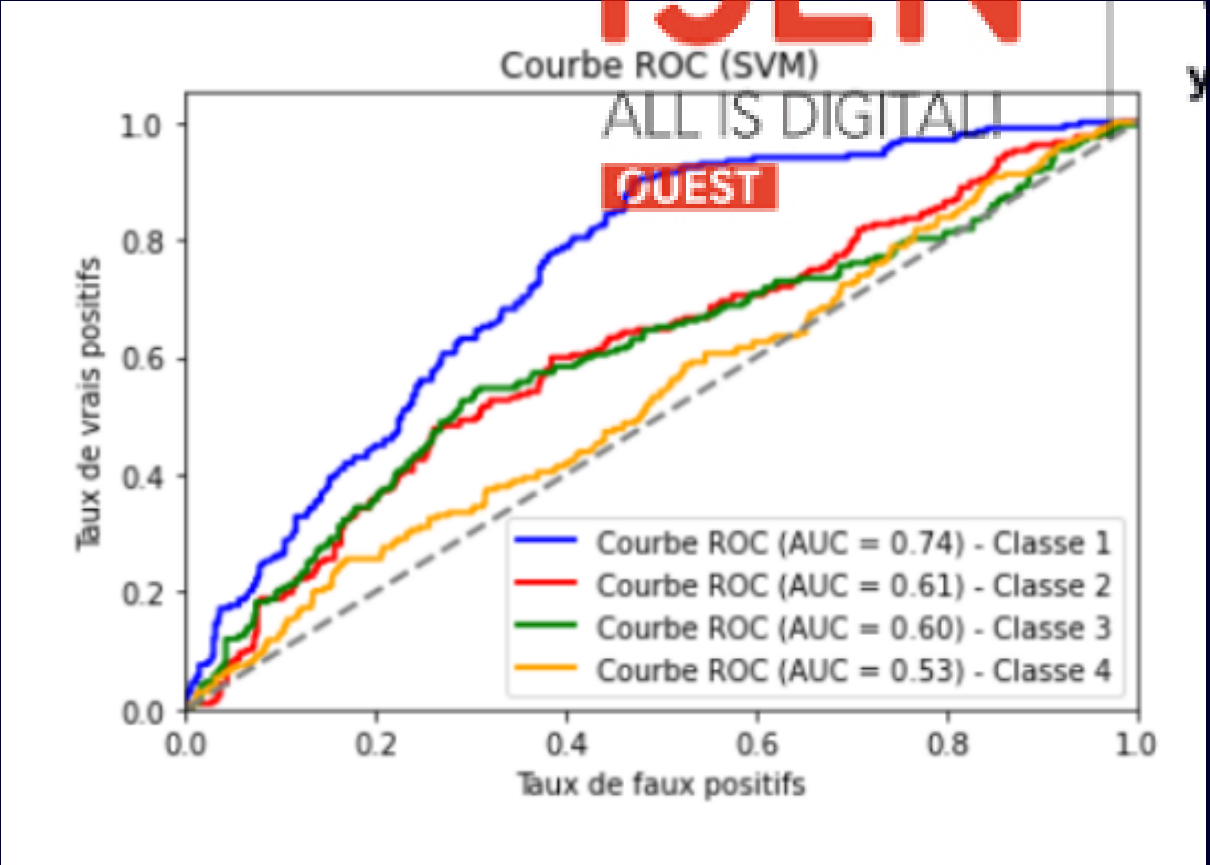
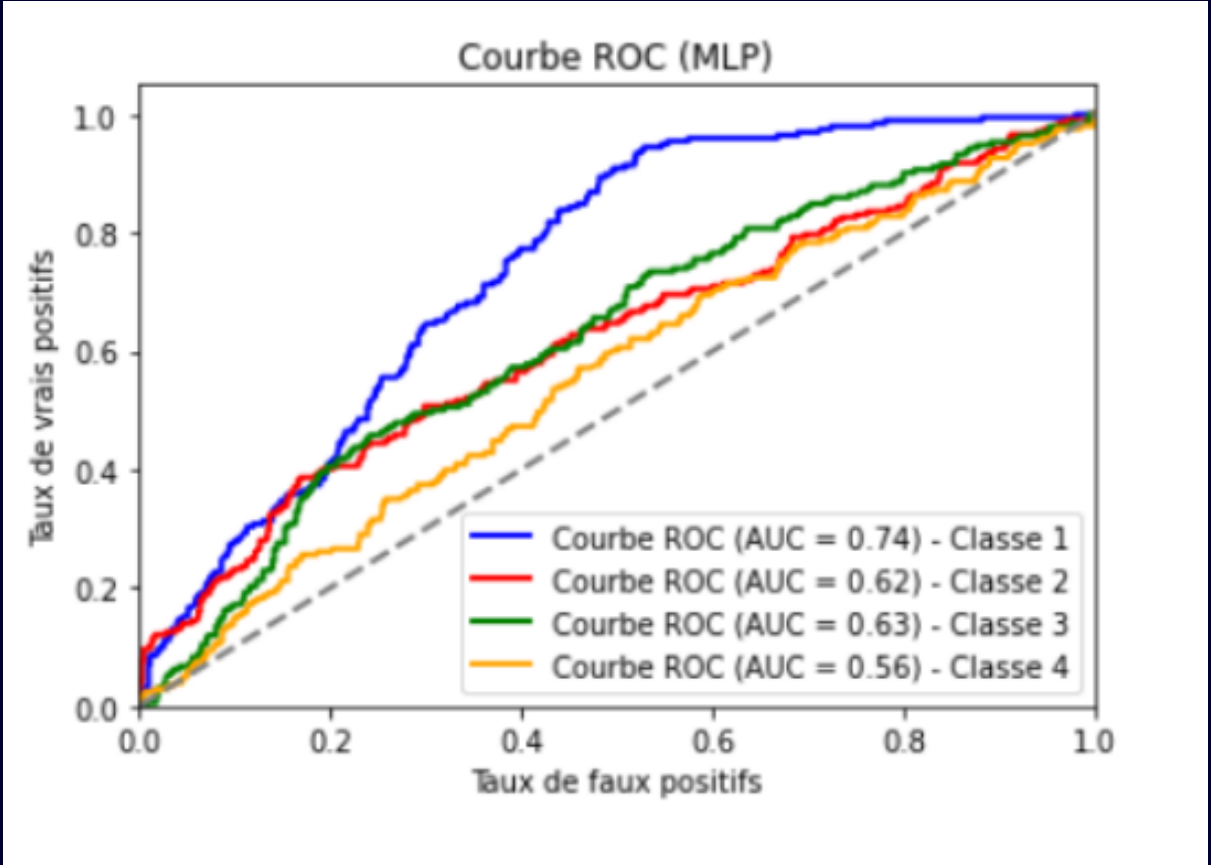
Précision (Random Forest) : 40.0%
Rappel (Random Forest) : 40.0%
Taux d'apprentissage(Random Forest) : 90.0%
Matrice de confusion: (Random Forest)
[[98 22 29 49]
[33 82 56 36]
[37 50 61 44]
[65 22 37 79]]

Précision (MLP) : 37.0%
Rappel (MLP) : 37.0%
Taux d'apprentissage(MLP) : 90.0%
Matrice de confusion: (MLP)
[[162 9 3 24]
[82 26 42 57]
[77 11 35 69]
[104 12 17 70]]

- La précision représente le pourcentage de prédictions correctes
- la matrice de confusion mettre en évidence les résultats de prédiction du modèle par rapport aux classes réelles
- Le rappel permet d'indiquer à quel point le modèle est capable de prédire les instances positives parmi toutes les instances réellement positives.

COURBES ROC

ISEN



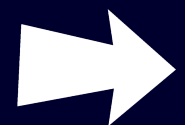
SCRIPTS

LES 3 TYPES DE SCRIPTS:

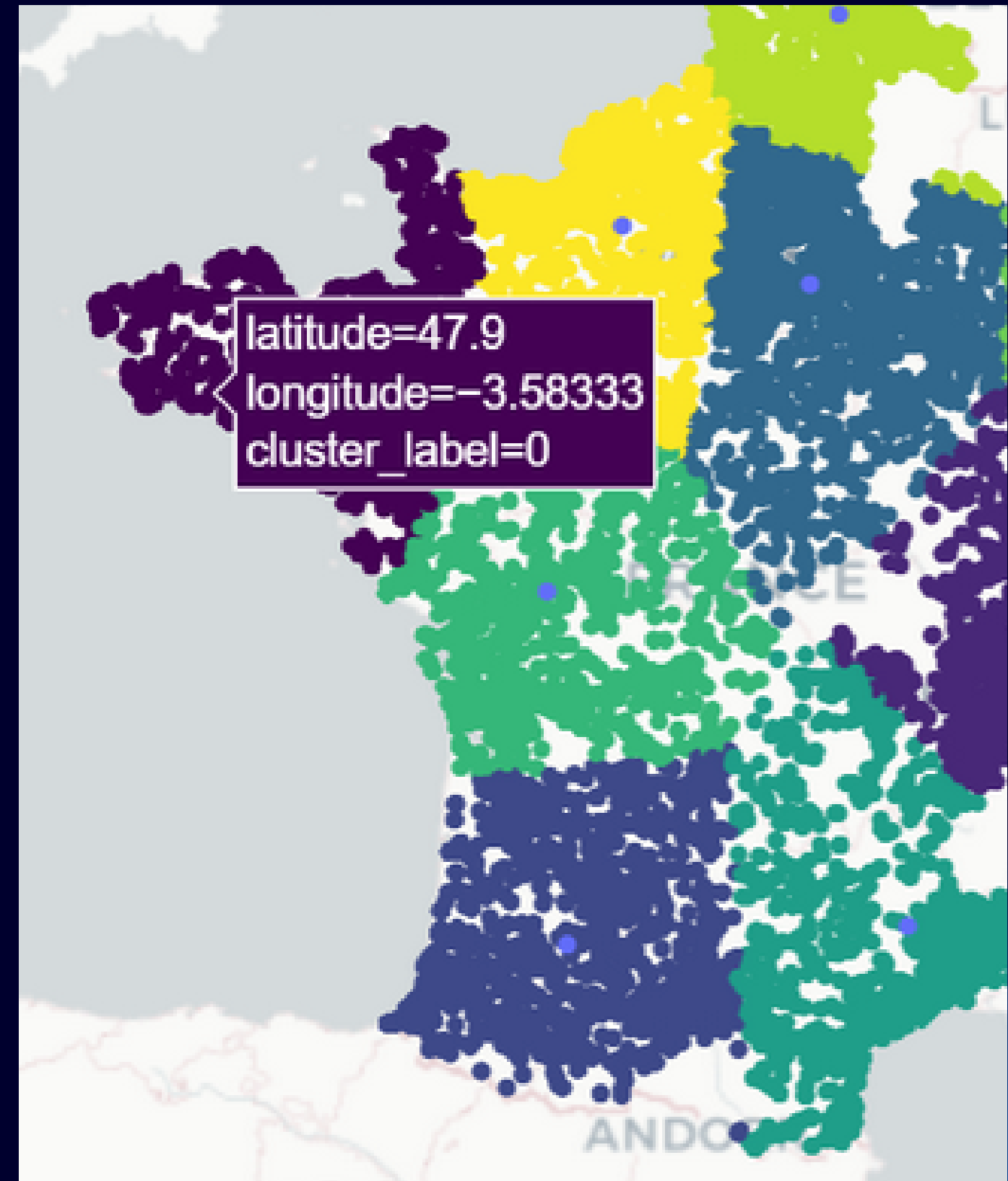
- K-MEANS
- KNN
- CLASSIFICATION DE «
HAUT NIVEAU»

k-means

```
# Préparez les données de l'accident à prédire
latitude_accident = 47.8566
longitude_accident = -3.3622
descr_cat_vehicule = 1
an_naissance= 2000.0
place = 1.0
descr_dispo_secu = 1
```



```
{"cluster_label": 0}
```



SCRIPTS

Script pour la classification
de « haut niveau »

```
{"Gravite de l'accident selon SVC": 3}  
{"Gravite de l'accident selon RF": 4}  
{"Gravite de l'accident selon MLP": 4}
```

Script pour KNN

```
'Gravite de l'accident': 3
```

Merci de
votre
attention!

