

Projet IA

Corentin - Robin - Noé

Rapide

Synthétique

Visualisation sur carte

Clustering maritime

Regroupe les navires

Schémas de navigation similaires

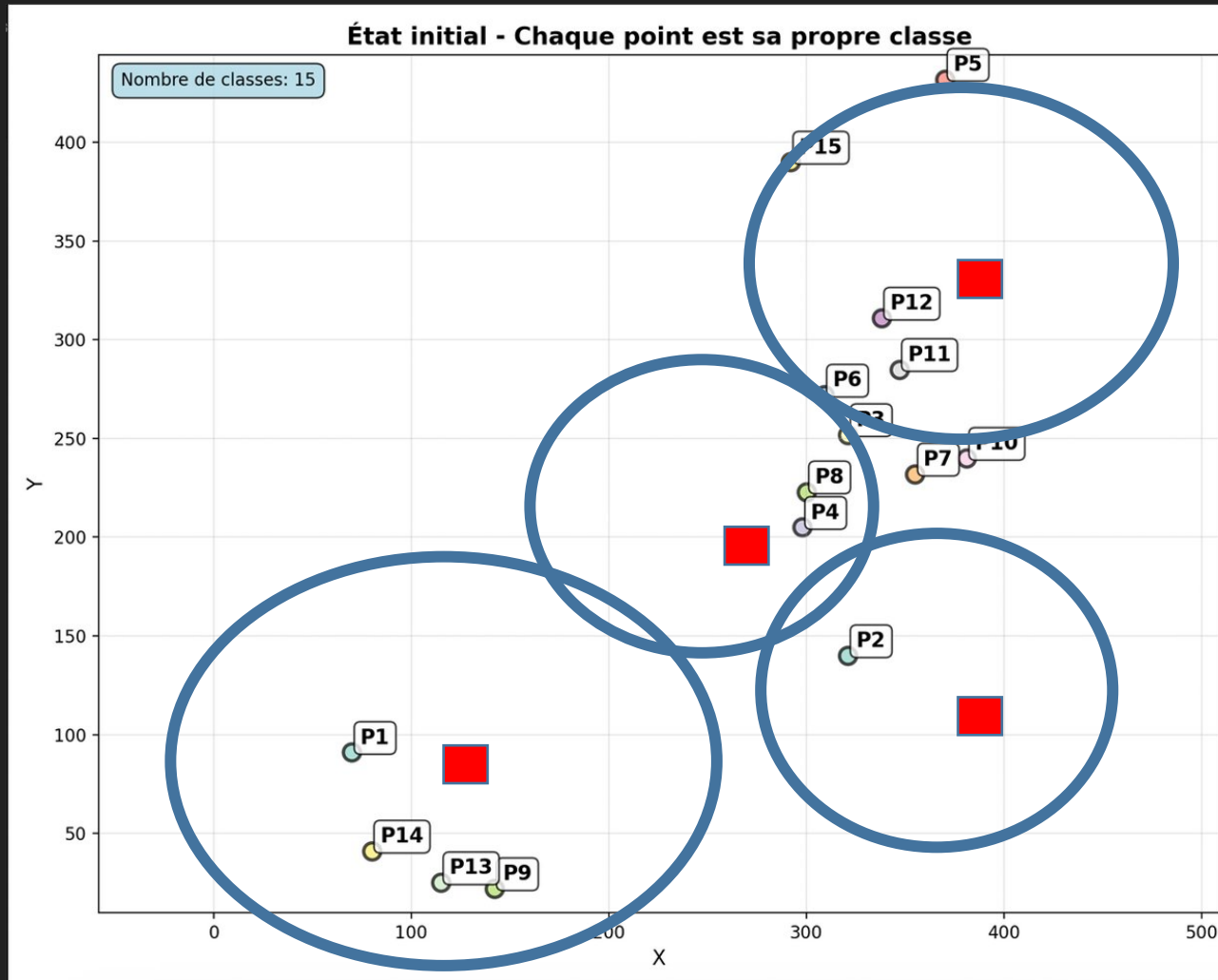


LAT, LON
SOG
COG
Heading

Visualisation sur carte

K-Means

Pour $k = 4$



Avantages

Rapide

Inconvénients

Connaitre à l'avance
le nombre de
cluster

Visualisation sur carte

Agglomerative

Est un algorithme de CAH

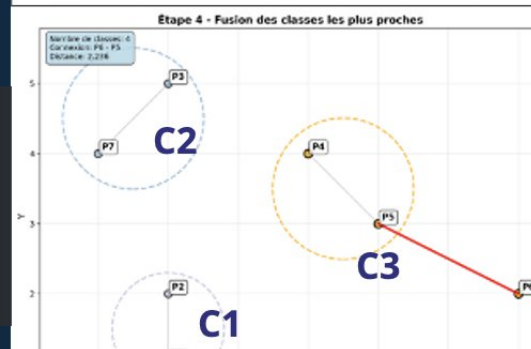
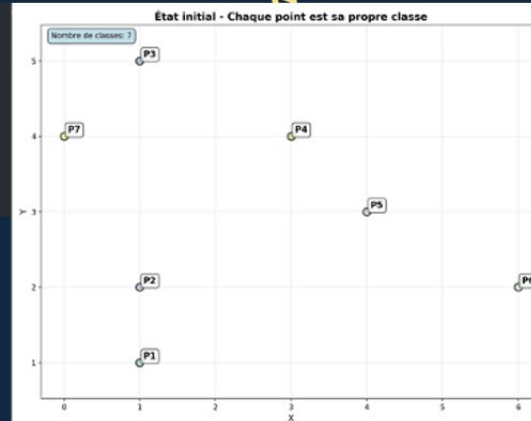
Partie 5

	C1	C2	C3	C4	C5	C6	C7
C1	0	1.00	4.00	3.61	3.61	5.10	3.16
C2	-	0	3.00	2.83	3.16	5.00	2.24
C3	-	-	0	2.24	3.61	5.83	1.41
C4	-	-	-	0	1.41	3.61	3.00
C5	-	-	-	-	0	2.24	4.12
C6	-	-	-	-	-	0	6.32
C7	-	-	-	-	-	-	0

Fusion: $C1\{P6\} \cup C4\{P4, P5\}$
Distance: 2.236

Matrice des distances (3 classes):

	C1	C2	C3
C1	0	2.24	2.83
C2	-	0	2.24
C3	-	-	0



ETAPE 1

Au début, chaque point forme son propre groupe (cluster)

ETAPE 2

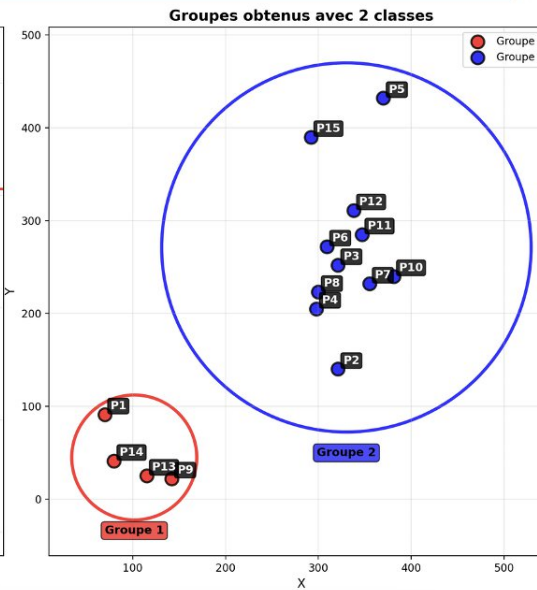
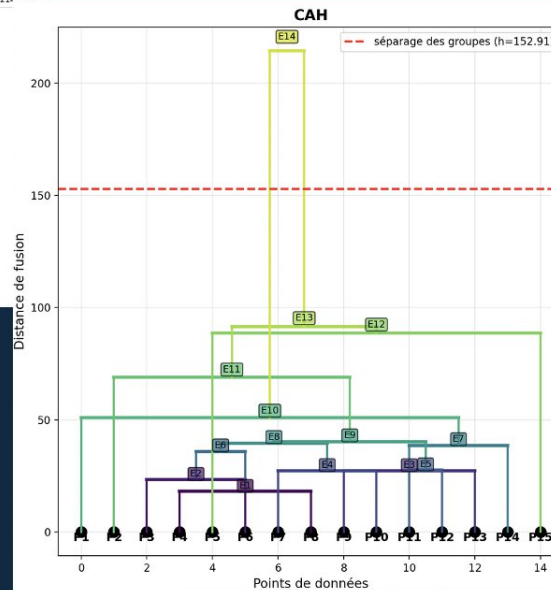
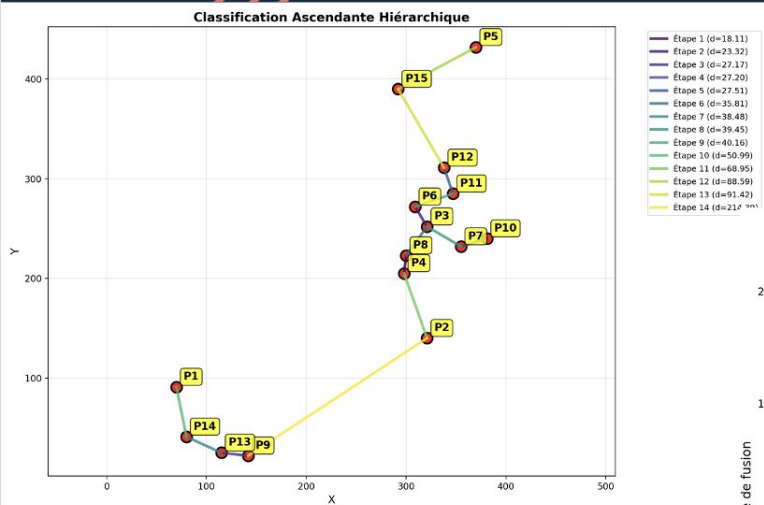
On calcul une matrice de distance de tous les cluster (on peut utiliser plusieurs algorithmes pour calculer les distances)

Visualisation sur carte

Agglomerative

Est un algorithme de CAH

Partie 5



Visualisation sur carte

Agglomerative

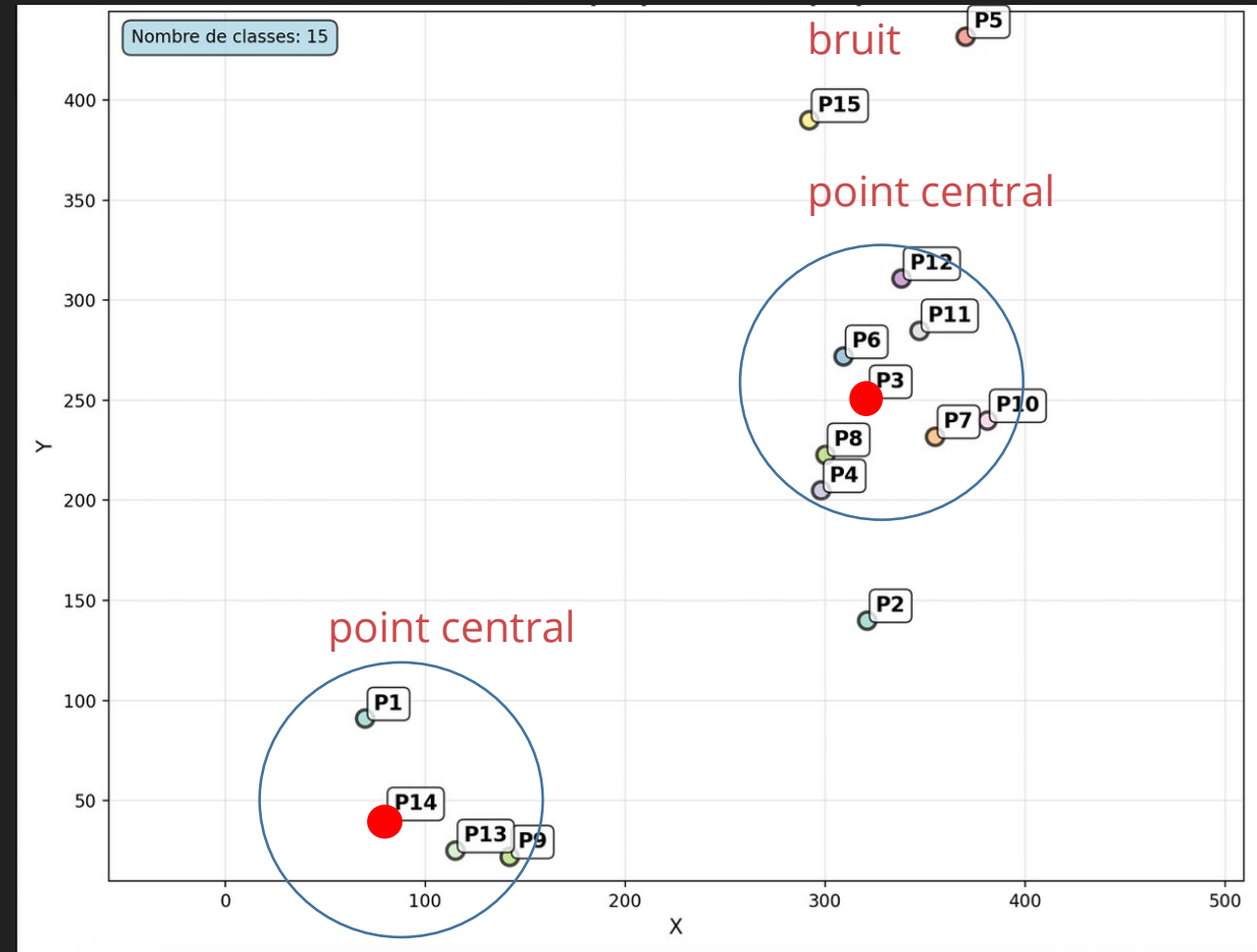
Avantages	Inconvénients
Facilement compréhensible	Lent si beaucoup de points

Visualisation sur carte DBSCAN

Density-Based Spatial Clustering

2 paramètres

- Epsilon (rayon)
- min_sample (nb dans le groupe minimum)



Visualisation sur carte

Algorithme de score

Silhouette

$$-1 < x < 1$$

Si 1 = loin des autres clusters, proche du centre du sien

0 = pile entre 2 clusters

Calinski-Harabasz

Compare la dispersion entre les clusters et à l'intérieur des clusters

Il faut le score le plus **élevé** possible

Davies-Bouldin

Regarde la similarité entre chaque cluster et son voisin le plus proche

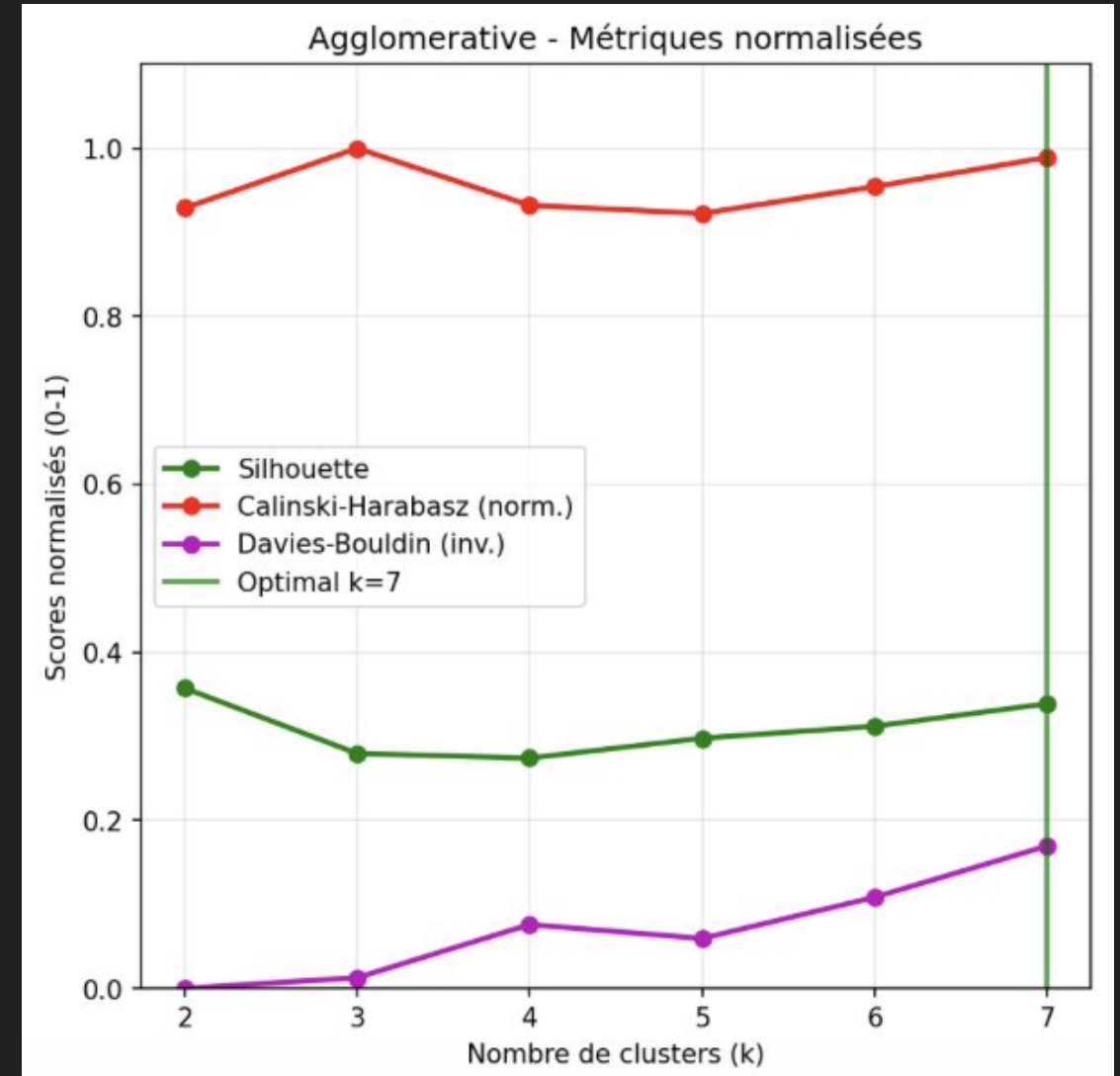
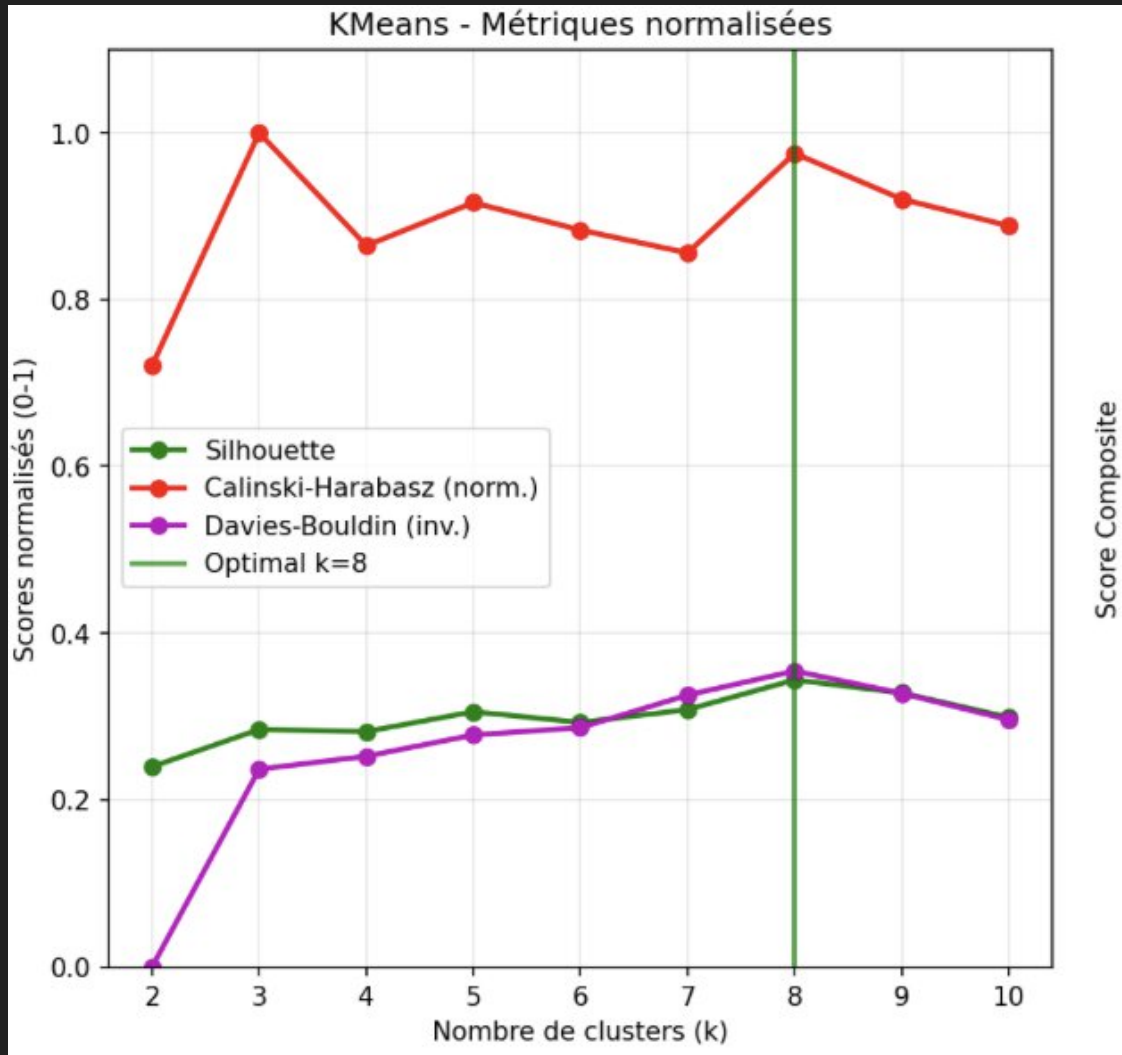
Il faut le score le plus **BAS** possible

Méthode Composite

$$0.4 * silhouette + 0.3 * calinski + 0.3 * davies$$

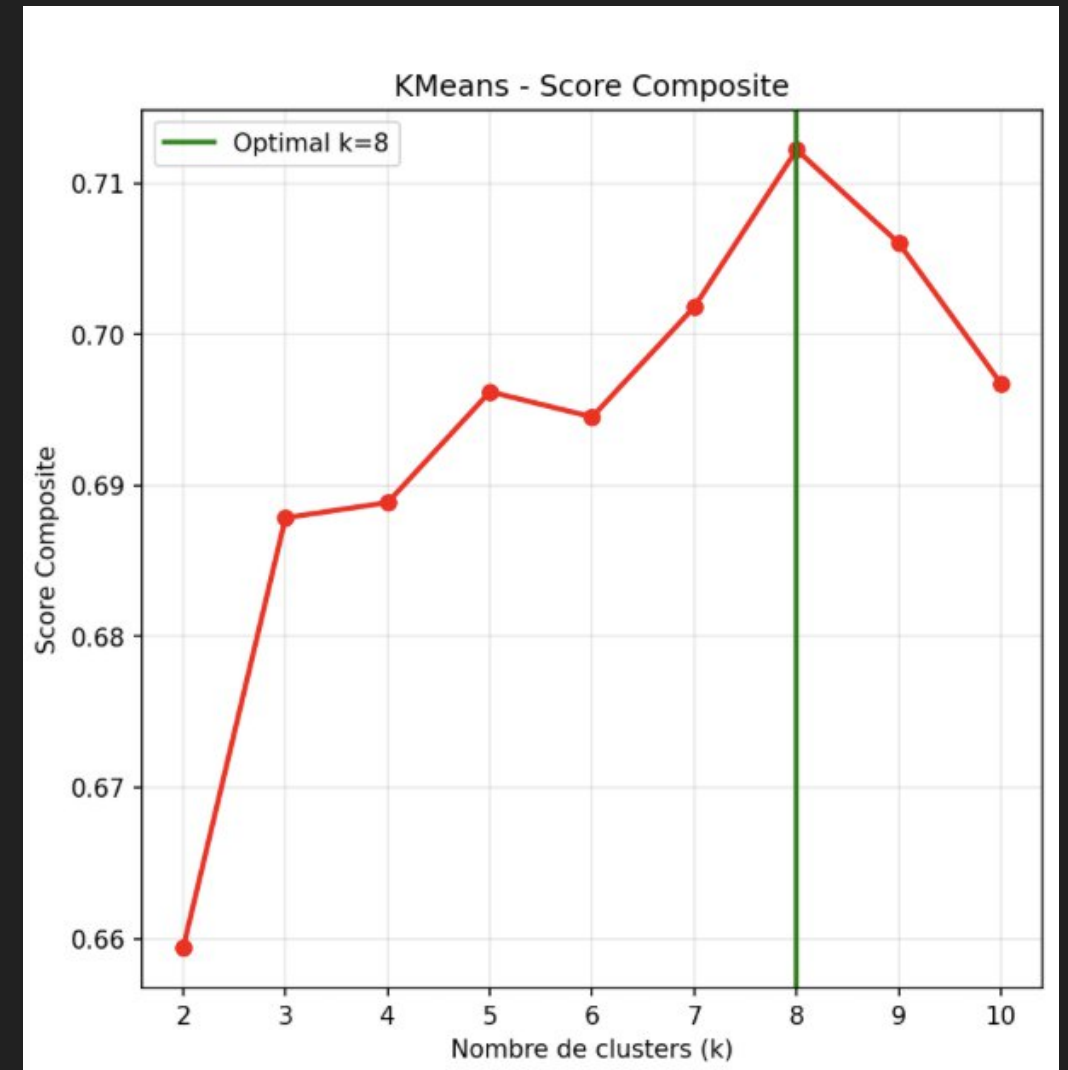
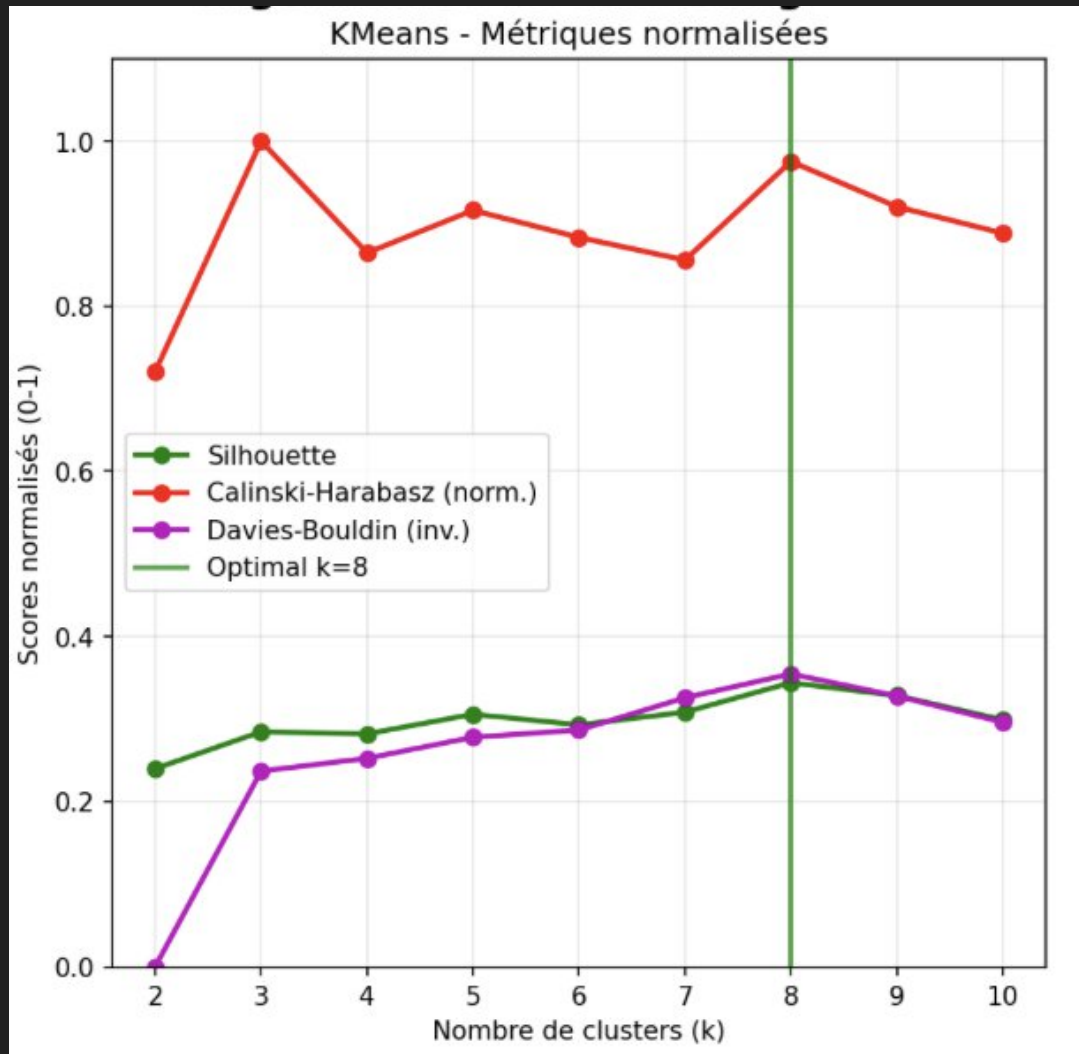
Visualisation sur carte

Algorithme de score



Visualisation sur carte

Algorithme de score



Visualisation sur carte

Algorithme de score

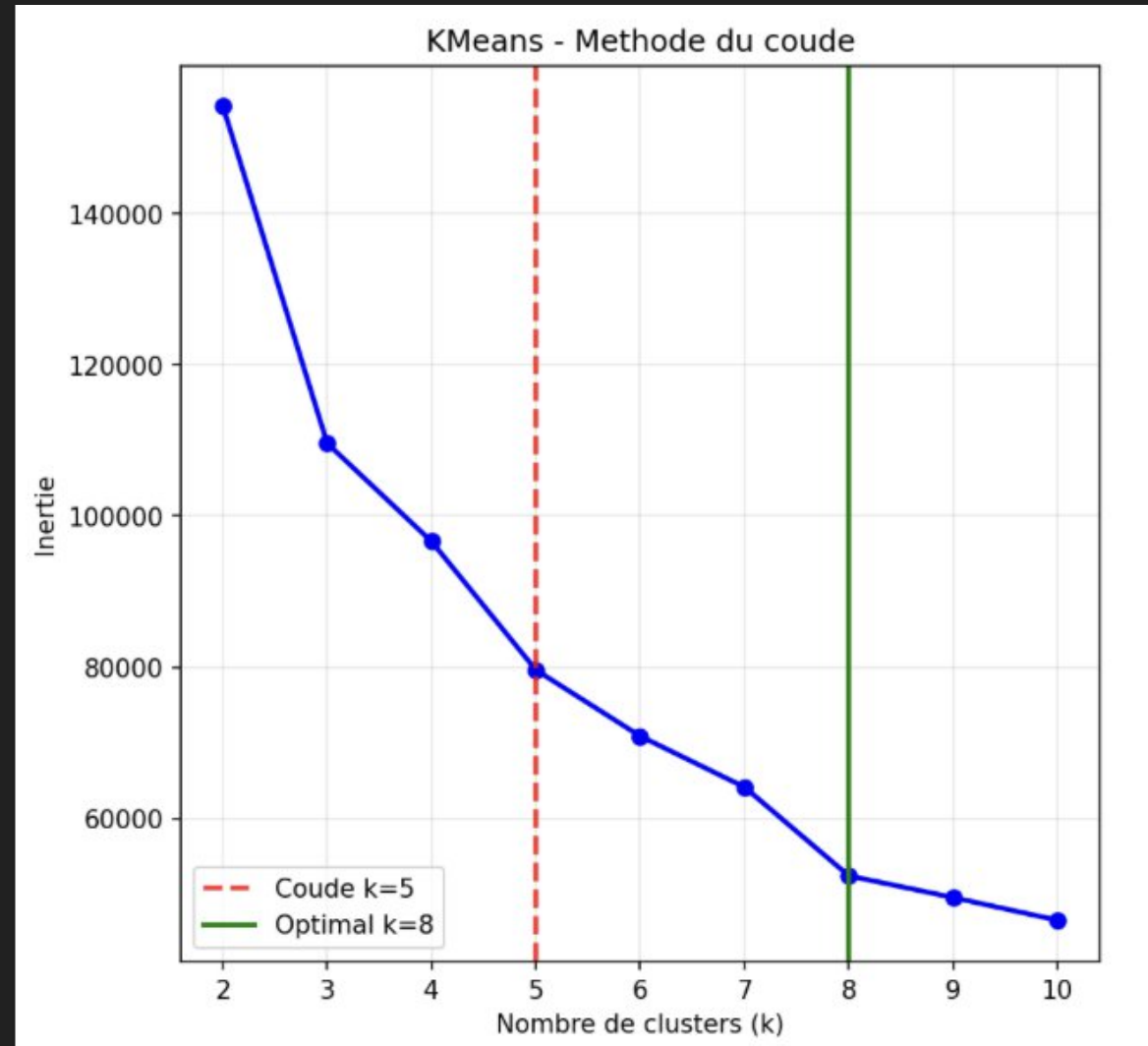
Coude pour K-mean

Pourquoi y a-t-il un coude ?

Quand il y a peu de clusters, l'inertie diminue rapidement,

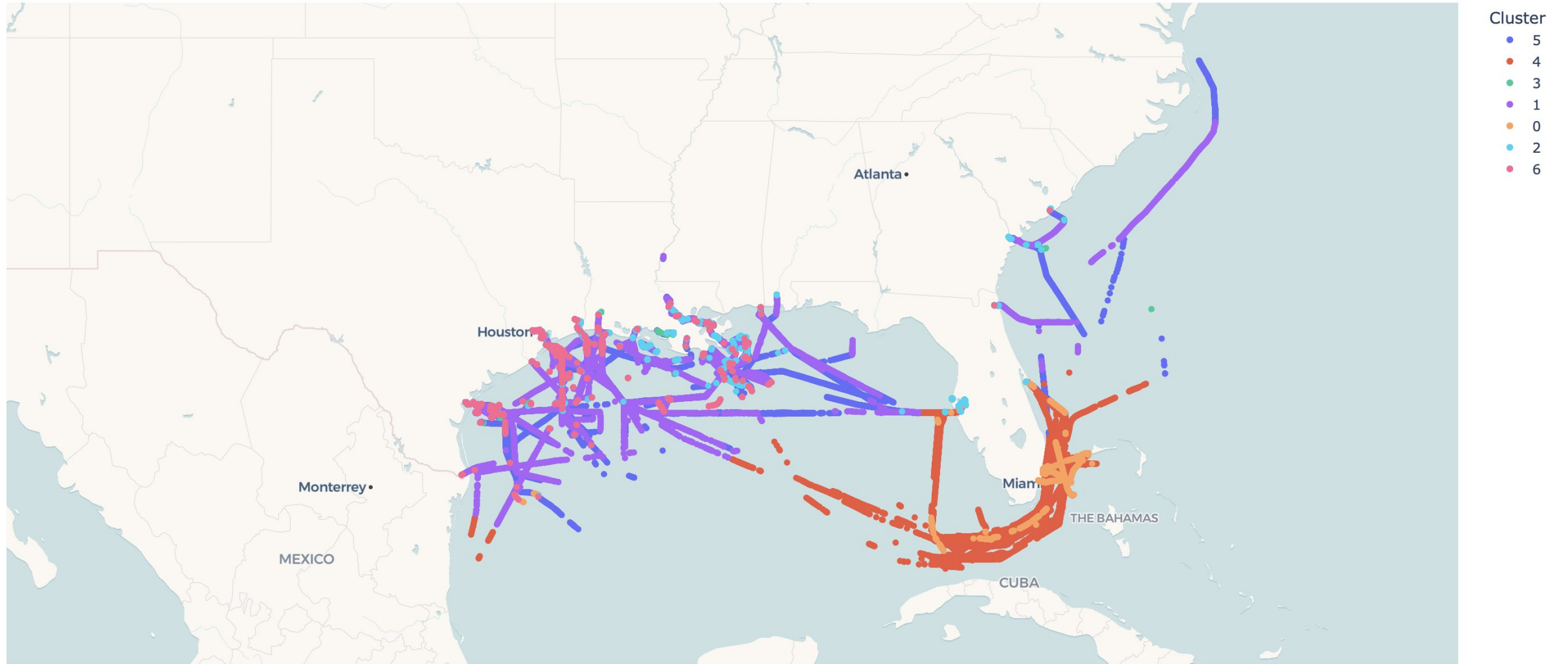
Quand il y a beaucoup de clusters, l'inertie diminue lentement.

Regarde la variance inter-cluster en fonction du nombre de clusters



Visualisation sur carte

Résultat



Prédiction du type d'un navire

Première version de la prédiction

1. Avec les données brutes
2. Sélection des variables explicatives pertinentes
3. Normalisation et encodage
4. On split train/test en veillant à ce que les bateaux soient différents
5. Supervisé : Random Forest, SVM, Nearest Neighbor

Prédiction du type d'un navire

Résultat de la première version

RandomForest: 86.5% de recall

Sur la grande BDD (1200 navires au lieu de 110)

Prédiction du type d'un navire

Super-paramétrage

Données brutes nettoyées \rightsquigarrow Segmentation + quelques pré-calculs

Computing des features additionnelles:



Prédiction du VesselType avec les nouvelles variables explicatives

- mean_dist_travel
- stderr_dist_travel
- mean_duration_stop
- stderr_duration_stop
- mean_cruise_speed
- stderr_cruise_speed
- mean_dist_coast_travel
- stderr_dist_coast_travel
- mean_draft
- stderr_draft
- number_occurence_travel
- number_occurence_significative_draft_variation_onshore
- number_occurence_significative_draft_variation_offshore
- duration_onshore_ratio
- ratio_length_width|
- ratio_directionnal_coherence

Prédiction du type d'un navire

Seconde version de la prédiction VesselType

Méthodes durant l'entraînement ?

1. Encodage - Preprocessing
2. SMOTE : Équilibrage
3. VIF - Multicolinéarité
4. GridSearchCV
5. Cross-validation stratifiée
6. Optimisation des hyperparamètres

Prédiction du type d'un navire

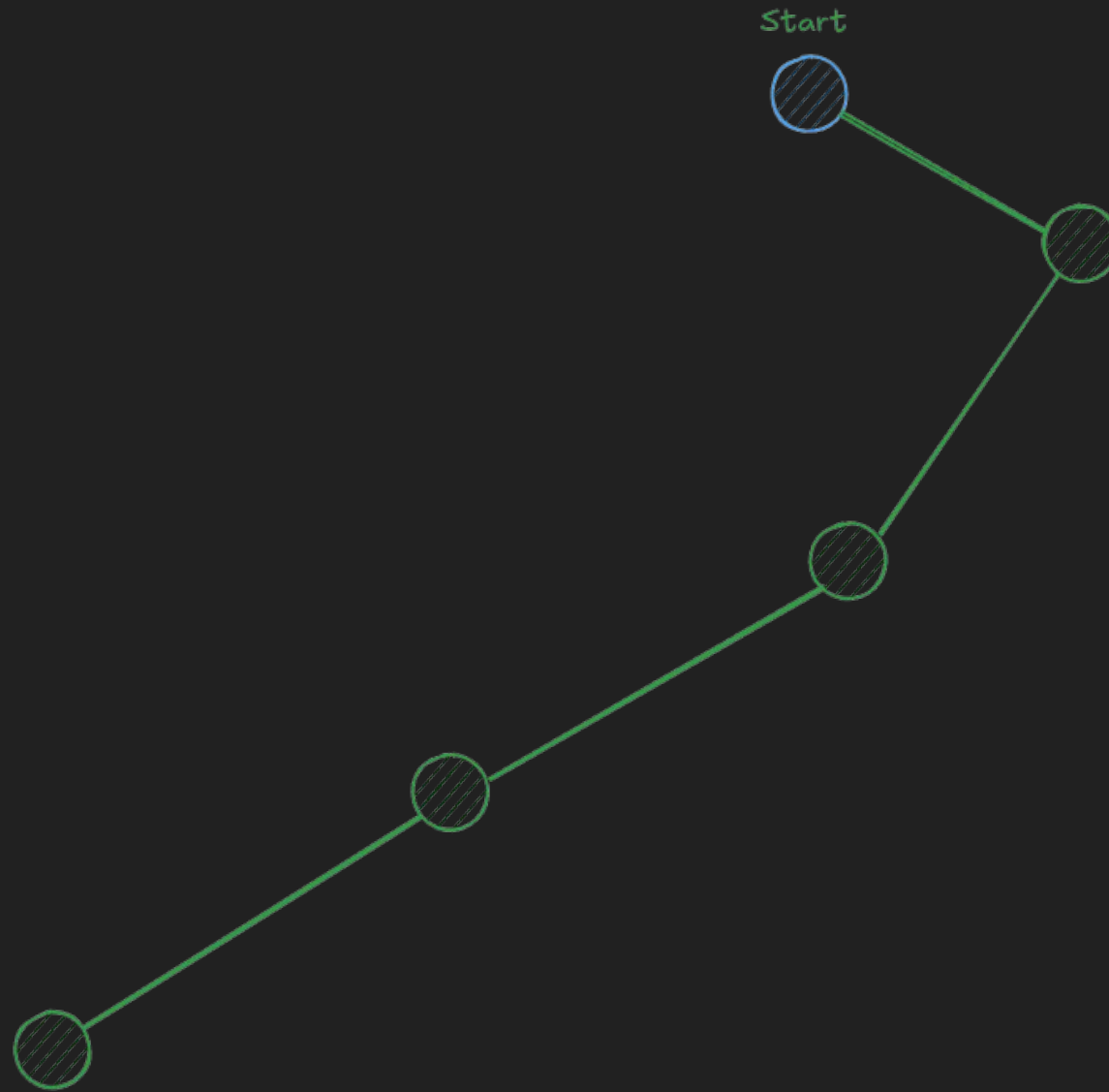
Résultat de la seconde version

LightGBM : 87.5% de recall

Sur la grande BDD (1200 navires au lieu de 110)

Prédiction de la trajectoire des navires

Techniques utilisées

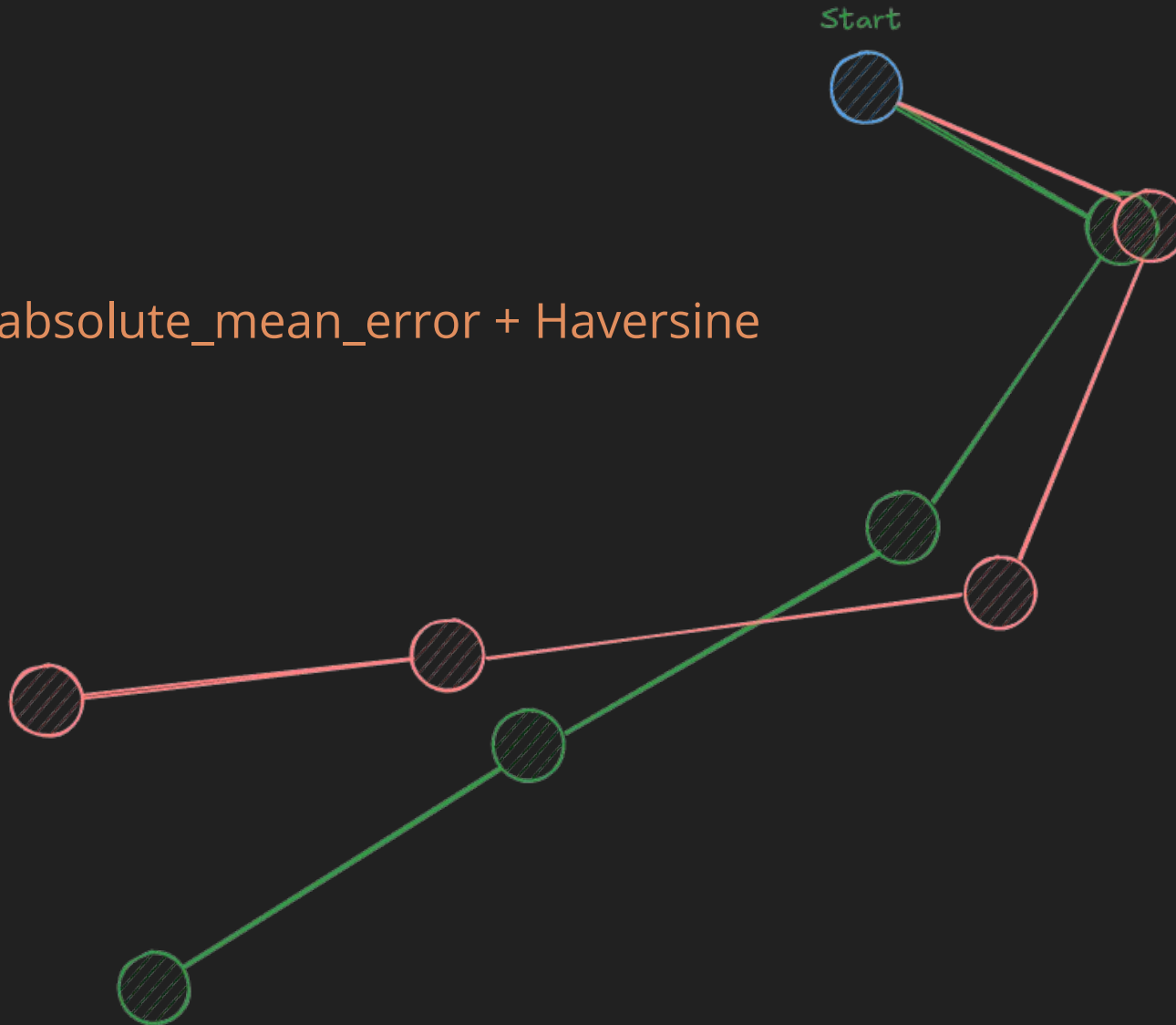


Prédiction de la trajectoire des navires

Techniques utilisées

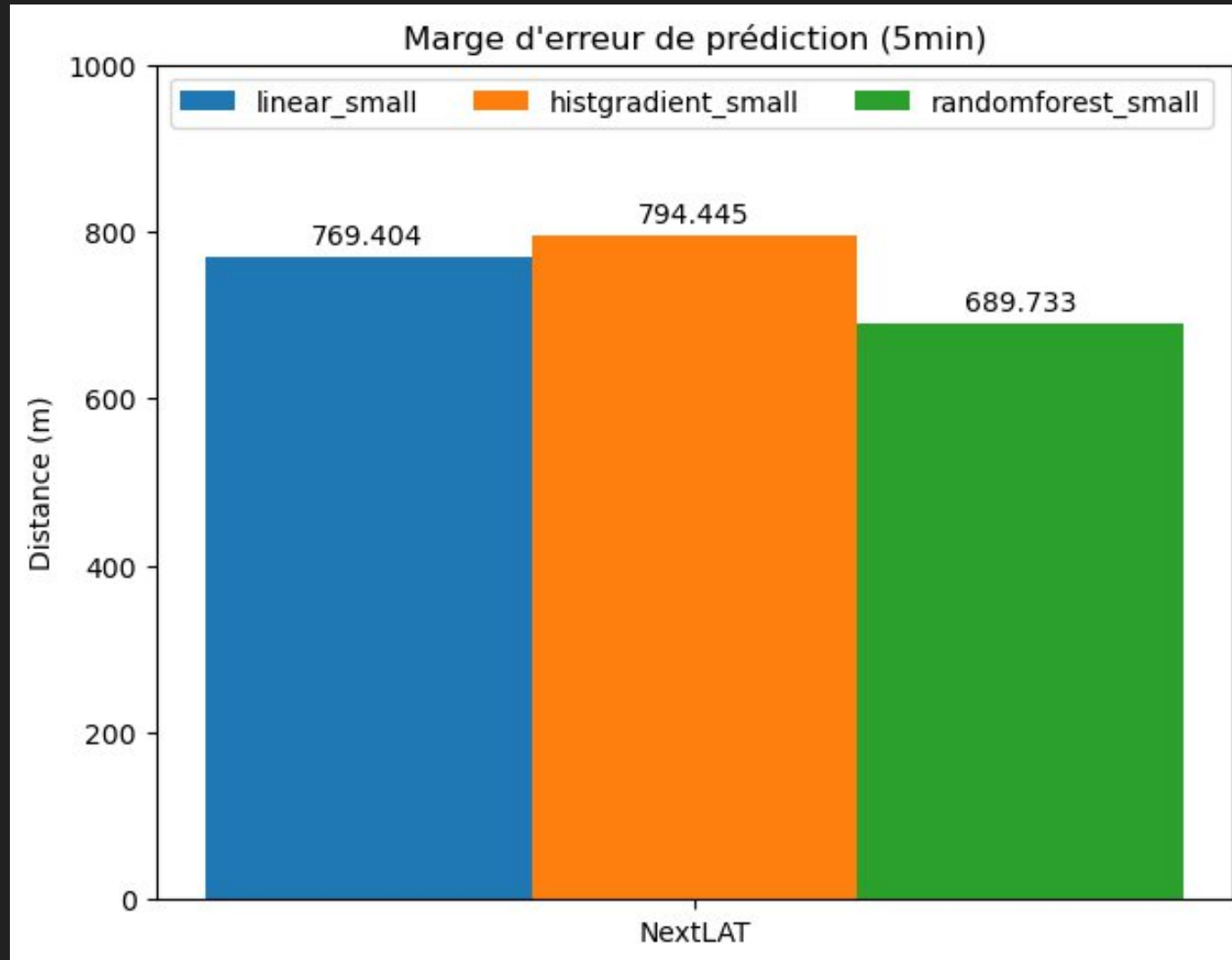
Évaluation → $\text{absolute_mean_error} + \text{Haversine}$

Valeurs → deltas
Angles → Sin/Cos



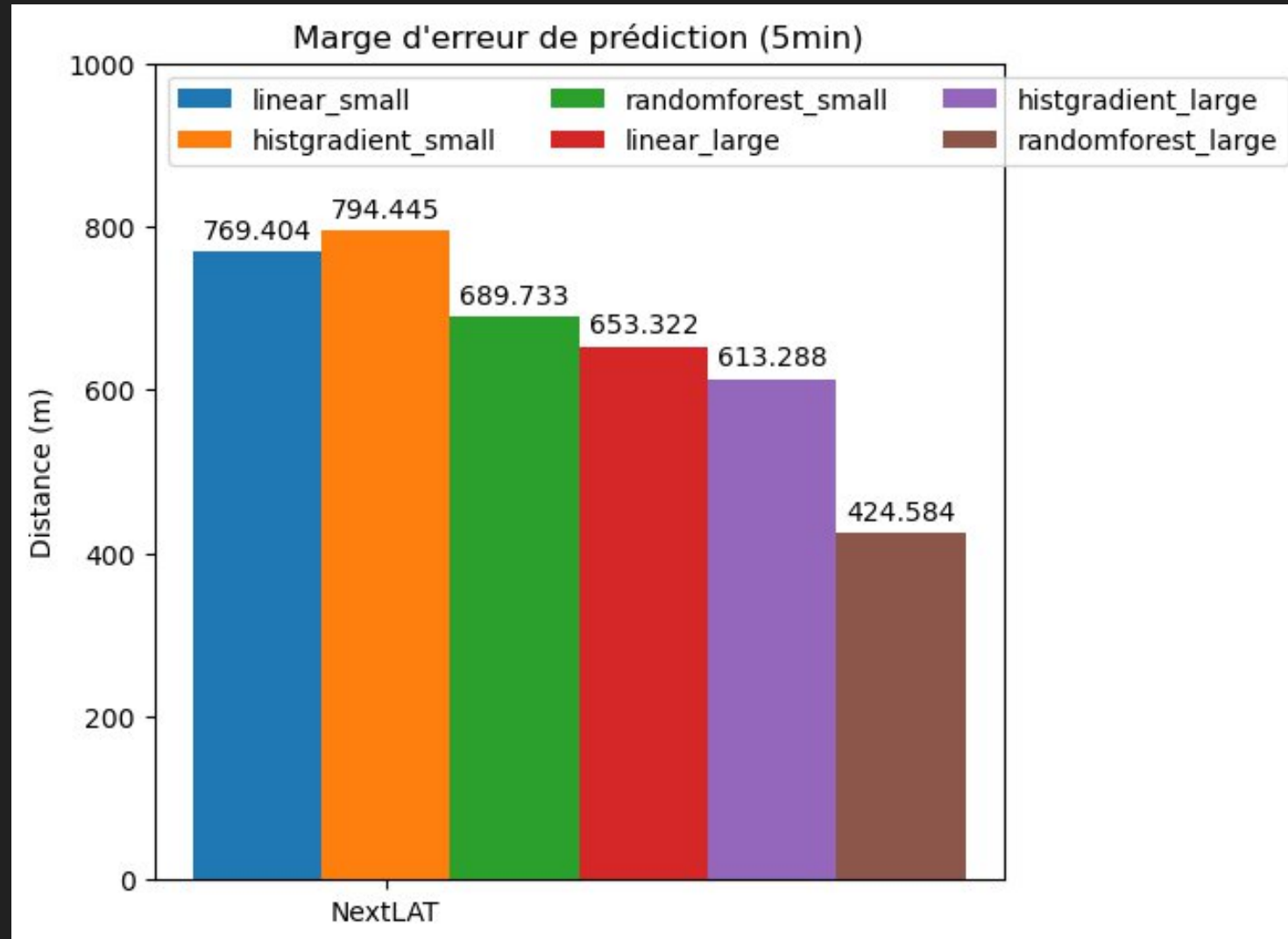
Prédiction de la trajectoire des navires

Modèles de régression



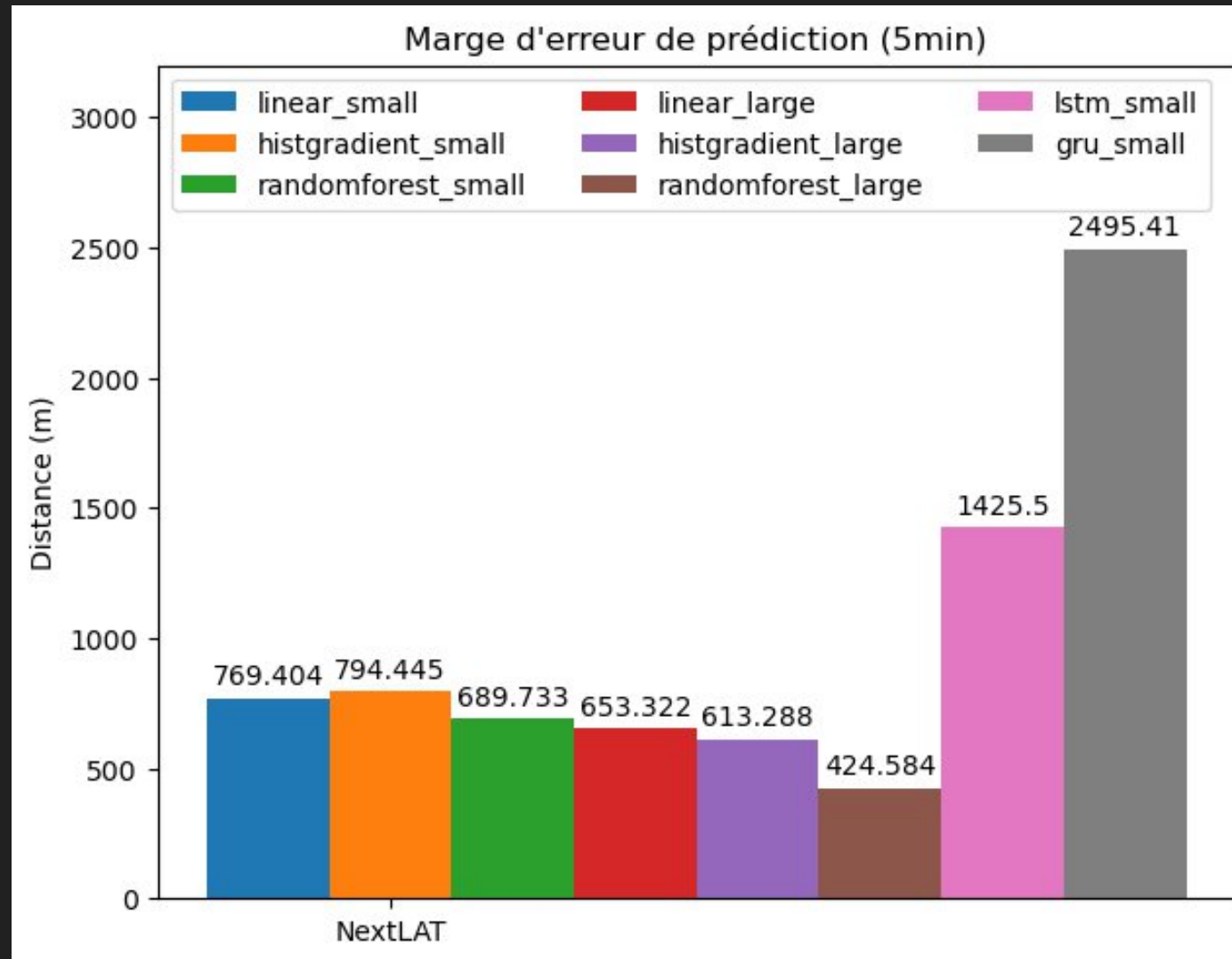
Prédiction de la trajectoire des navires

Modèles de régression

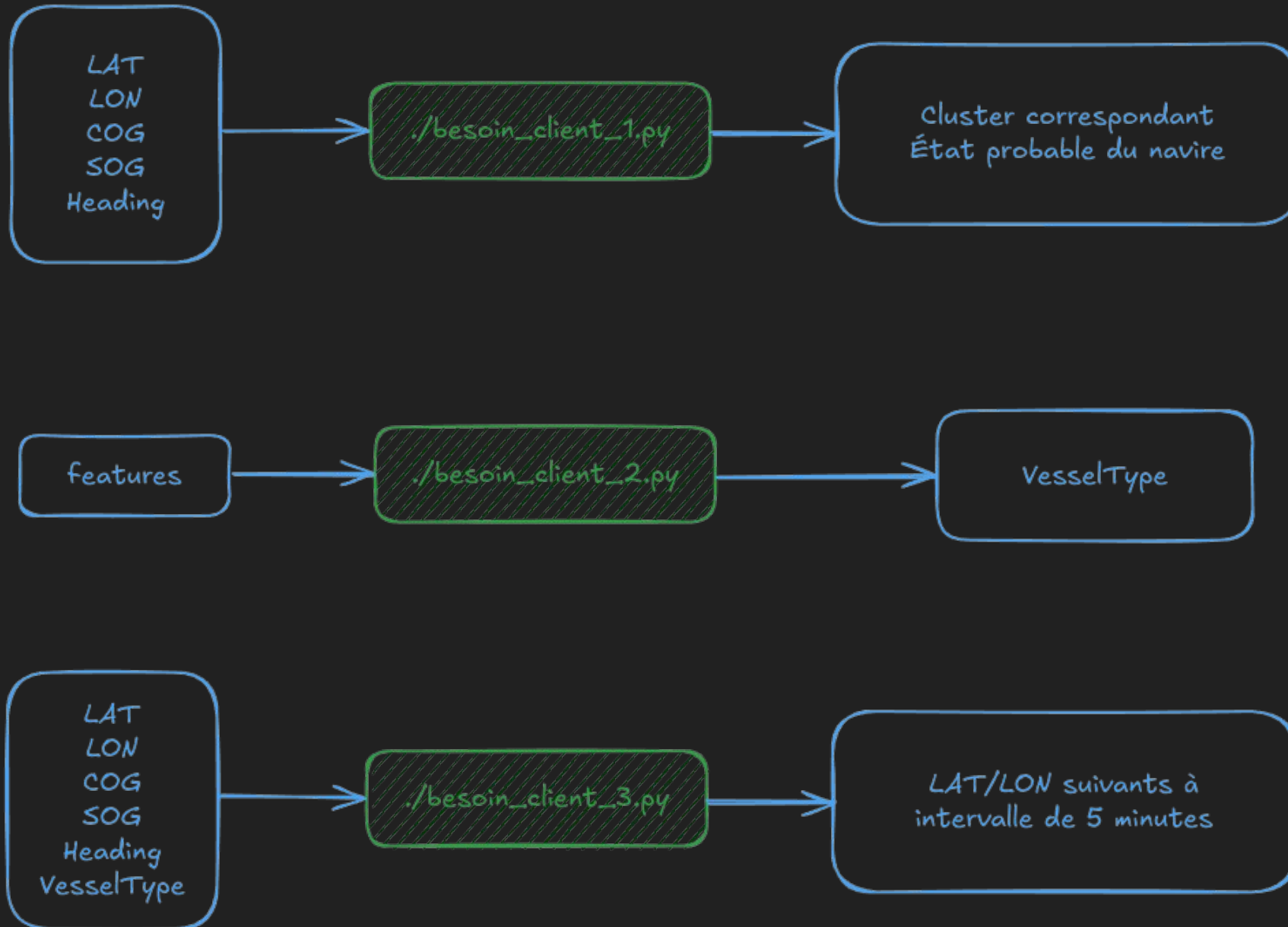


Prédiction de la trajectoire des navires

Neural networks (LSTM & GRU)



Diagrammes blocs



Merci de votre écoute

Démonstration des scripts !

