

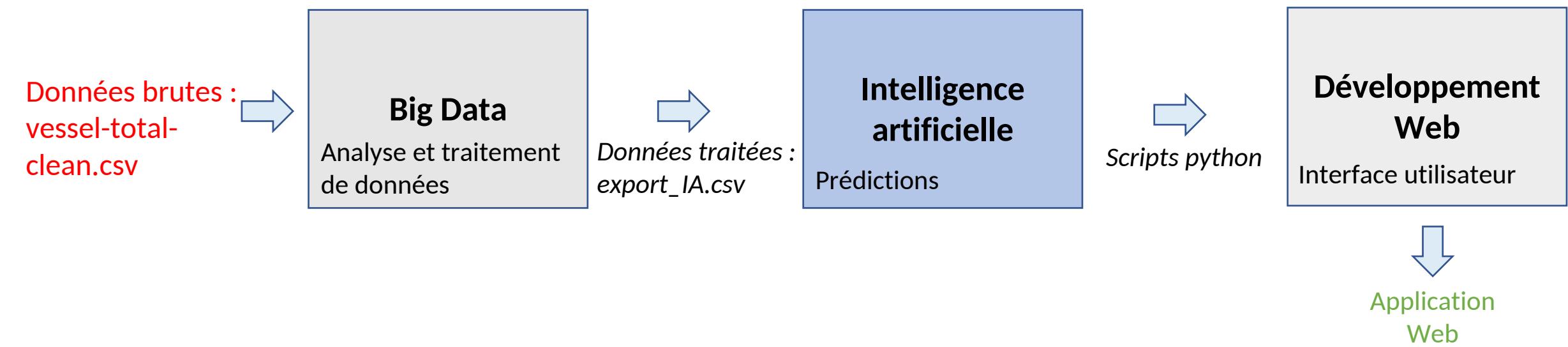


Projet Année 3 Big Data/IA/Web

Partie Intelligence Artificielle

Contexte du projet

Déroulement du projet



Objectif

Analyse et Modélisation des Comportements de Navigation des Navires à partir des Données AIS

Approfondir les compétences acquises dans les modules ***Big Data, Intelligence Artificielle, Développement Web et Base de Données*** à travers une application complète de traitements et de visualisation de données.

Objectifs de la partie Intelligence Artificielle :

- Comprendre les étapes d'un projet d'apprentissage automatique
- Apprendre à évaluer une méthode d'apprentissage automatique
- Manipuler les différentes techniques d'apprentissage automatique :
 - Apprentissage non supervisé : clustering
 - Apprentissage supervisé : classification, régression

Besoin client

1. Visualisation sur carte :

- Créer un visuel sur une carte qui regroupe automatiquement les navires selon des schémas de navigation similaires.

2. Modèle de prédiction du type du navire :

- Développer un modèle d'intelligence artificielle capable de prédire le type du navire.

3. Prédiction de la trajectoire des navires:

- Prédire les positions futures d'un navire à différents horizons temporels

Cahier des charges

Cahier des charges

Fonctionnalités

5 fonctionnalités principales sont attendues :

1. Préparation des données (pour les besoins client 1, 2 et 3)
2. Apprentissage non-supervisé (pour les besoins client 1)
3. Apprentissage supervisé (pour les besoins client 2 et 3)
4. Métriques (pour les besoins client 1, 2 et 3)
5. Création de scripts utilisables en ligne de commande (pour les besoins client 1, 2 et 3)

Attention

Les fonctionnalités additionnelles que vous développerez seront évaluées uniquement si les fonctionnalités principales sont en place.

Réalisation

Préparation de l'environnement de travail et gestion de projet :

- Outil de développement : VSCode
- Installation des packages : *pandas*, *sklearn*, *plotly*...
- *Gantt*
- *Git*

→ **Attendus : diagramme de Gantt (et dépôt Git si possible)**

Réalisation du Besoin client 1 : Visualisation sur carte

Besoin client 1 : Visualisation sur carte

Visualisation sur carte :

- Créer un visuel sur une carte qui regroupe automatiquement les navires selon des schémas de navigation similaires (vitesses, routes/directions, heading...) via des techniques de clustering.
- L'objectif est d'identifier des comportements typiques, détecter des anomalies ou optimiser les itinéraires.

Les étapes à suivre pour réaliser le besoin client 1 :

1. Préparation des données
2. Apprentissage non supervisé
3. Métrique pour apprentissage non supervisé
4. Visualisation sur carte
5. Préparation d'un script *Python*

Besoin client 1 : Visualisation sur carte

- **Préparation des données :**
 - Extraction des données d'intérêt : sélectionner les colonnes pertinentes de la base de données selon ce besoin.
 - Encodage des données catégorielles si nécessaire : utiliser des techniques de prétraitement pour convertir les données non numériques en données numériques si nécessaire. Référez-vous à <https://scikit-learn.org/stable/modules/preprocessing.html>
 - **Apprentissage non-supervisé :**
 - Choix de l'algorithme de clustering : sélectionner **un/des algorithme(s)** de clustering pour regrouper les navires selon des schémas de navigation similaires . Référez-vous à <https://scikit-learn.org/stable/modules/clustering.html>
 - Détermination du nombre de clusters : expérimenter avec différents nombres de clusters pour voir quel modèle offre le meilleur regroupement.
 - **Métriques pour l'apprentissage non-supervisé :**
 - Évaluation des clusters : utiliser des métriques pour évaluer la qualité du clustering. Référez-vous aux méthodes : *Silhouette Coefficient* , *Calinski-Harabasz Index* , *Davies-Bouldin Index*.
 - **Visualisation sur un carte :**
 - Création de la carte : utiliser une bibliothèque de visualisation pour représenter les bateaux sur une carte avec des couleurs différentes pour chaque cluster. Référez-vous à <https://plotly.com/python/scattermapbox/>
 - **Préparation d'un script :**
 - Écrire un script prenant en entré les spécificités d'un navire et qui renvoie le cluster associé. Un exemple de script est fourni dans les ressources du projet.
- Attention : le script ne doit pas relancer une recherche des clusters à chaque usage, il doit impérativement charger le modèle préalablement enregistré.**

Les attendus

- Justifier le choix des variables sélectionnées
- Justifier le choix du modèle de clustering + Expliquer le principe de son fonctionnement
- Justifier le choix des métriques et discuter les résultats obtenus (tableau + graphes)

Réalisation du Besoin client 2 : Modèle de prédiction du type de navire

Besoin client 2 : Modèle de prédiction du type de navire

Modèle de prédiction du type de navire :

- Développer un modèle d'intelligence artificielle capable de prédire le type du navire.

Les étapes à suivre pour réaliser le besoin client 2 :

1. Préparation des données
2. Apprentissage supervisé pour la classification
3. Métrique pour la classification
4. Préparation d'un script *Python*

Besoin client 2 : Modèle de prédiction du type de navire

- **Préparation des données :**

- Extraction des données d'intérêt : Sélectionner les colonnes pertinentes de la base de données selon ce besoin.
- Encodage des données catégorielles.
- Normalisation des données ou pas.
- Cible : **['VesselType']**
- Utiliser des techniques de prétraitement de données pour préparer les données d'apprentissage. Référez-vous à <https://scikit-learn.org/stable/modules/preprocessing.html>

Attention : Enregistrer vos modèles de prétraitement de données, c'est important pour le script final et le projet WEB.

- **Apprentissage supervisé pour la classification :**

- Choix de l'algorithme d'apprentissage : Sélectionner **des/un algorithme(s)** d'apprentissage pour prédire le type du navire. Référez-vous à https://scikit-learn.org/stable/supervised_learning.html#

- **Métriques pour la classification :**

- Évaluation de résultats de la classification : Utiliser des métriques pour évaluer la qualité du modèle de la classification. Référez-vous à https://scikit-learn.org/stable/modules/model_evaluation.html.
- Utilisation du GridSearchCV pour tester toutes les combinaisons possibles avec **plusieurs hyperparamètres du modèle** et trouver le meilleur modèle de prédiction. Référez-vous à https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

- **Préparation d'un script :**

- Écrire un script prenant en entrée les spécificités du navire et qui renvoie la prédiction de son type. Un exemple de script est fourni dans les ressources du projet.

Attention : le script ne doit pas relancer un apprentissage à chaque usage, il doit impérativement charger les modèles préalablement enregistrés.

Les attendus

- Justifier le choix des variables sélectionnées (tableau + graphes)
- Justifier le choix du modèle de classification + Expliquer le principe de son fonctionnement
- Justifier le choix des métriques et discuter les résultats obtenus (tableau + graphes)

Réalisation du Besoin client 3 : Prédiction de la trajectoire des navires

Besoin client 3 : Prédiction de la trajectoire des navires

Problème de séries temporelles

Objectif:

Prédire la position future (LAT, LON) d'un navire à partir de ses données historiques :

- BaseDateTime (timestamp)
- LAT, LON (position)
- SOG (vitesse), COG (cap), Heading (orientation)
- Potentiellement enrichi par : VesselType, Length, Draft, etc.

Variables dépendantes (à prédire) :

- LAT, LON (prédiction de trajectoire)

Variables explicatives :

- SOG, COG, Heading, VesselType, Length, etc.
- Index temporel : BaseDateTime

Output:

- Entainer un modèle pour prédire les positions futures (LAT, LON) à différents horizons temporels (ex. : 5, 10, 15 min).

Déroulement

Déroulement

Organisation

Travail en trinôme :

- Chaque étudiant dans le trinôme connaît l'ensemble du projet : **utiliser le gestionnaire de code Git si possible**
- Attention à bien se répartir le travail en prévoyant les tâches de chacun avec un **diagramme de Gantt**

Point d'avancement :

- Des points d'avancement sont prévus avec votre client (qui est votre enseignant)

Ressources externes :

- Tous les documents sont autorisés
- Attention à utiliser avec une grande précaution tout document extérieur : site de vulgarisation, forum, code d'autrui

Documentation du projet :

- Au fur et à mesure
- Standardisée
- Livraison de code ou de documents :
 - Ne pas attendre la dernière minute pour poster un livrable
 - Préparer des livrables intermédiaires (surtout pour les sources)
 - Sauvegarder régulièrement vos scripts et résultats

Livrables et évaluations

Format de l'archive :

Archive ZIP, TGZ, 7ZIP, pas de RAR : projetia_groupeX.zip (remplacer X par votre numéro de trinôme)

Le rendu final doit contenir :

- Trois dossiers : *Besoin_Client_1*, *Besoin_Client_2* et *Besoin_Client_3*
- Dans chaque dossier on doit trouver :
 - La base de données utilisée (.csv)
 - Le Notebook de la partie expérimentale (.ipynb) commenté et exécutable sans erreur
 - Le script final (.py)
 - Les ressources nécessaires (comme les modèles enregistrés (.pkl))
 - Le Readme.txt pour décrire comment utiliser le script final avec un script d'exemple
- Un rapport (.pdf) de 11 pages (dont 1 page de diagramme de Gantt) dans lequel sont formalisés les différents attendus

Remarques :

- Malus possible sur l'un des membres du groupe si l'investissement est jugé trop faible
- Possibilité d'être interrogé durant le projet de façon individuelle
- Plagiat sévèrement sanctionné pour TOUS les membres du / des groupe(s)

Attention

Les livrables seront à poster sur l'intranet. Tout retard sera sanctionné (l'heure du réseau faisant foi).

Les fichiers au mauvais format ou avec un mauvais nommage seront pénalisés.

Présentation orale vendredi 20 juin à partir de 13h30 :

- Soutenance de 10 minutes (strict) + 5 minutes de questions
- Présentation en trinôme (pensez à vous répartir la parole)
- Présentez l'essentiel de votre projet
- Chaque membre de l'équipe présente sa contribution au projet
- Réservez 2 minutes pour exécuter la démonstration des 3 scripts finaux

Code et rapport :

- Rendu de l'intégralité de vos codes sources
- Rapport du trinôme présentant les attendus du projet
- **Aucun code ne doit apparaître dans le rapport et la présentation**

Evaluations des compétences :

- Des questions seront posées tout au long du projet pour vérifier les acquis

Barème indicatif : Soutenance 40% - Compétences 40% - Évaluation du rapport + code 20%

ISEN
ALL IS DIGITAL!



MERCI
Des questions ?

