

Projet - Big Data / IA / Web

Présentation générale

A3 Brest / Caen / Nantes
CIPA4 Brest / Nantes

Objectif

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.

3 parties dans l'ordre suivant :

Big Data — Intelligence Artificielle — Développement Web

30 heures par bloc pour les A3

Précisions :

- **Projet fortement guidé** : le but est d'approfondir les compétences acquises
- **Une seule application globale** : mais les trois parties sont indépendantes

Sujet identique sur les trois sites impliqués

Compétences attendues :

- **Analyser un problème** et un cahier des charges
- Proposer des **solutions techniques** répondant aux fonctionnalités attendues
- Mettre en place les algorithmes proposés avec des **méthodes de programmation adaptées**
- Développer des **outils ergonomiques, réutilisables et évolutifs**
- **Analyser qualitativement et quantitativement** les résultats obtenus
- Mettre en place une **gestion de projet de type "Agile"**
- **Rédiger des documents** en lien avec les solutions techniques, les tests mis en place et les résultats obtenus

Gestion de projet :

- **Gestion de projet de type "Agile"** : trois sprints d'une semaine, chaque étudiant du trinôme aura ses proches tâches
- **Utilisation de Git** : organisation, partage du code source et gestion globale du projet

Évaluations (pour chaque bloc) :

compétences en séance + soutenance + rapport et/ou code

Les étudiants sont en trinômes imposés

Suivi des étudiants

10 à 15 trinômes par enseignants

	Brest	Caen	Nantes
Nombre d'étudiants	129	31	59
Nombre de groupes	= 43 trinômes	= 11 trinômes	~ 20 trinômes
Nombre d'enseignants	4	1	2

Périodes :

- **Caen et Nantes** : 10/06 au 27/06
- **Brest** :
 - session 1 : du 02/06 au 20/06
 - session 2 : du 10/06 au 27/06

Objectif

Étant donné un jeu de données initial, les étudiants devront effectuer l'ensemble des prétraitements nécessaire à la phase d'apprentissage et de prédiction de caractéristiques de la partie Intelligence Artificielle.

Notions abordées :

- **L'extraction des données** : à partir d'un fichier, d'un site web...
- **La visualisation d'un grand volume de données**
- **Le nettoyage des données** : suppression des données incomplètes, suppression des données erronées, agrégation de données similaires...
- **Le stockage des données dans une base de données non relationnelle adaptée**
- **L'application de modèles statistiques pour l'analyse des données recueillies**

Objectif

Étant donné le jeu de données nettoyé et stocké dans la partie précédente, les étudiants devront réaliser différents modèles de prédiction de données et fournir des résultats chiffrés des performances obtenues. Les modèles de prédiction, pré-entraînés, devront être utilisable par l'interface graphique mise en œuvre dans la partie Développement Web.

Notions abordées :

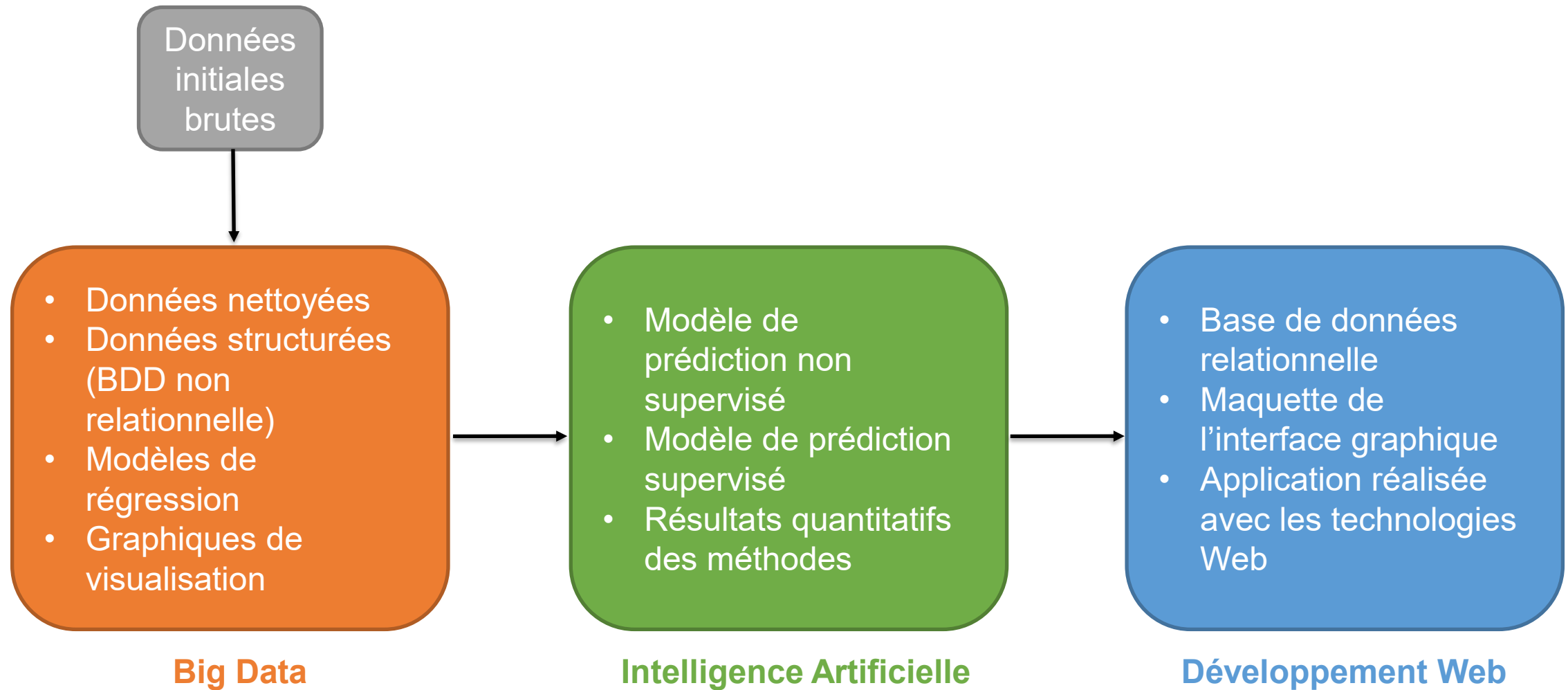
- **La séparation des jeux de données** : entraînement, validation et test
- **La création de modèles d'apprentissage non supervisés et supervisés**
- **L'apprentissage et l'inférence des modèles** : validation croisée, taux d'apprentissage, nombre d'époques, sur-apprentissage
- **L'évaluation quantitative des résultats** : courbe ROC, courbe précision/rappel, matrice de confusion

Objectif

Étant donné les données nettoyées et les modèles de prédiction développés, les étudiants devront mettre en place une interface web permettant l'utilisation de ces modèles par un utilisateur lambda.

Notions abordées :

- **La réalisation d'un modèle conceptuel de données et d'un modèle physique de données**
- **La création d'une maquette** : définition de l'interface graphique et des interactions utilisateurs
- **La définition des interfaces client-serveur**
- **Les développements de la partie statique de l'interface web** : technologies HTML et CSS
- **Le développement de la partie dynamique de l'interface web** : technologies JavaScript et PHP



Données de trafic maritime, fournies au travers des données du Système d'Identification Automatique (AIS) collectées par la Garde Côtière des États-Unis. Les données contiennent des informations telles que :

- la localisation
- l'heure
- le type de navire
- la vitesse
- la longueur
- la largeur
- le tirant d'eau

Lien : <https://hub.marinecadastre.gov/pages/vesseltraffic>

Pour chaque partie du projet :

- **Évaluation des compétences en séance**
- **Soutenance**
- **Rapport et/ou code**

Compétences en séance :

- Évaluation individuelle basée sur des compétences fixées (voir référentiel)
- Conversion en note à postériori
- [Liste des compétences de la formation](#)

Soutenance :

- Évaluation en groupe basée sur des compétences fixées (voir référentiel)
- Les compétences acquises peuvent être différentes entre les membres du groupe
- Conversion en note à postériori

Rapport et/ou code :

- Évaluation en groupe basée sur une grille de notation
- Conversion en compétence à postériori

Chaque compétence évaluée entre : acquis / partiellement acquis / en cours d'acquisition / non acquis

Des questions ?