

Roadmap Technique : Préparer les 24 fichiers CSV pour l'Entraînement du Modèle

- **Vue d'ensemble du processus**

24 CSVs → (1. Ingestion & Nettoyage) → Objets Java Propres → Base de Données (SQL) → (3. Vectorisation) → Tableaux Numériques → (4. Split) → Training/Test

- **Étape 1 : Ingestion & Nettoyage (ETL)**

- Lire les 24 fichiers CSV
- Corriger les fourchettes du kilométrage en valeurs entières
- Normaliser le prix
- Déduire marque/modèle si absent
- Supprimer doublons

- **Étape 2 : Centralisation (Base de Données)**

- Stockage dans une table unique VehiculeAnnonce
- Nettoyage effectué avant insertion
- Utilisation Spring Data JPA
- Objectif : Master Dataset unique

- **Étape 3 : Vectorisation (Préparation Numérique)**

- Encodage One-Hot ou Label (Marque, Modèle, Carburant, Boîte)
- Normalisation des valeurs numériques
- Gestion des valeurs manquantes (NULL)
- Service : VectorisationService

- **Étape 4 : Train / Test Split**

- Mélanger les données
- 80% pour l'entraînement
- 20% pour le test

- **Checklist Finale**

- Une seule liste avec toutes les marques
- Kilométrage converti en entier unique
- Champs texte encodés en chiffres
- Données mélangées correctement