

Cahier des Charges Technique de l'application d'estimation et de Prédiction d'émission de GES* pour les Véhicules Électriques

1. Introduction

Le présent cahier des charges décrit les spécifications techniques pour le développement de l'application de prédiction d'émission de GES pour les véhicules électriques.

L'application sera basée sur les données de l'AEE* nettoyées à l'aide de KNIME, stockées dans une base de données PostgreSQL hébergée sur Azure, et les analyses exploratoires effectuées avec les bibliothèques de visualisation Pyplot (matplotlib version 3.7.2) et seaborn (version 0.12.2).

La conception du modèle prédictif utilisera ScikitLearn et Keras.

2. Objectifs

L'objectif principal de l'application est de fournir aux utilisateurs des prédictions fiables d'émission de GES pour leurs véhicules électriques. Les utilisateurs pourront ainsi prendre des décisions informées concernant leur utilisation des véhicules électriques et leur impact environnemental.

La comparaison avec les émissions de véhicules thermiques de même type permettra d'enrichir les données d'aide à la décision.

3. Spécifications Techniques

3.1 Plateforme et Langages de Programmation

L'analyse des données et la réalisation du modèle de ML* se feront en Python (version 3.11) sous Databricks (spark version 3.3.2).

Le développement de l'API et de l'application seront réalisés en Python dans un environnement de test Windows et de production Linux (utilisation de la librairie Flask pour l'application web).

L'application et l'API seront déployées sur une plateforme de cloud computing, telle qu'Azure, pour assurer leur disponibilité et leur mise à l'échelle.

Les données pourront être stockées de façon temporaire dans un serveur de stockage azure blob.

3.2 Traitement des données

Les données peuvent présenter des valeurs manquantes et des incohérences occasionnelles qui seront traitées comme suit dans la phase de préparation des données :

- Suppression des enregistrements avec des valeurs manquantes pour les features principales et conservation des valeurs manquantes pour les features secondaires.
- Suppression a priori des outliers, mais l'exploration pourra permettre d'identifier d'autres règles à mettre en œuvre.
- Renommage des colonnes pour que les noms soient plus explicites.
- Normalisation des types de carburant

3.3 Modèle de Prédiction

La prédiction sera effectuée par un modèle d'apprentissage supervisé. Etant donnée la nature des données (séries temporelles), nous choisissons de concevoir un modèle LSTM.

Les hyper paramètres du modèle seront ajustés en utilisant un benchmarking de performances. La métrique utilisée sera la RMSE (root mean squared error).

Un modèle de ML sera développé en utilisant les bibliothèques scikit-learn (version 1.3.2), keras et Tensor Flow (version 2.15). Le modèle sera entraîné sur les données nettoyées à l'aide de KNIME. Les caractéristiques pertinentes pour la prédiction d'émission de GES seront sélectionnées en fonction des résultats de l'analyse exploratoire.

3.4 Base de Données

Les données nettoyées seront stockées dans une base de données PostgreSQL v.14.8 hébergée sur un serveur flexible Azure Database B1ms, (1 vCore, RAM 2Go, stockage 32Go) du datacenter East US. Les informations spécifiques aux utilisateurs et leurs prédictions pourront être sauvegardées pour un suivi ultérieur.

La table principale, co2_vehicles_new_clean, stockera les données d'immatriculation et les caractéristiques des véhicules. Elle sera partitionnée en deux (véhicules électriques et véhicules thermiques) et indexée de façon à optimiser les requêtes.

D'autres tables seront mises en place pour stocker les données de mix énergétique et les données utilisateurs.

3.5 Sécurité

Des mesures de sécurité seront mises en place pour protéger les données des utilisateurs conformément à la RGPD. L'accès à la base de données et à l'application sera sécurisé, et les données sensibles seront cryptées.

Le développement se fera sur des machines privées, limitant ainsi les risques de sécurité.

L'accès à ces machines se fera via une double authentification lorsque cela est possible et, a minima, par des mots de passe forts. Le white listing d'adresses IP se fera au cas par cas.

3.6 Interface Utilisateur

L'interface utilisateur sera développée à l'aide du framework Flask. Les utilisateurs pourront entrer les caractéristiques de leur véhicule électrique, tels que le modèle, l'année de fabrication, etc. L'application affichera ensuite l'estimation des émissions de GES basée sur ces informations et d'éventuelles données permettant de comparer à d'autres véhicules.

4. Livrables Attendus

- Application fonctionnelle de prédiction d'émissions de GES pour les véhicules électriques.
- Code source complet et commenté.
- Documentation des endpoints de l'API pour une éventuelle intégration future.

5. Planning Prévisionnel

- Phase 1 (Mois 1-2) : Mise en place de la base de données PostgreSQL sur Azure et prétraitement des données.
- Phase 2 (Mois 3-4) : Analyse approfondie des données nettoyées, sélection des caractéristiques pertinentes et test de différents modèles de prédiction.
- Phase 3 (Mois 5) : Entraînement du modèle de prédiction choisi et développement de l'interface utilisateur avec Flask.
- Phase 4 (Mois 5) : Développement de l'API et intégration du modèle de prédiction.
- Phase 5 (Mois 6-7) : Mise en œuvre des mesures de sécurité et tests de l'ensemble de l'application. Documentation, finalisation du code source et préparation des livrables.

Tâches	Mai	Juin	Juillet	Août	Sept	Oct	Nov	Dec
Pré-traitement des données et stockage en production								
Analyse exploratoire								
Analyse descriptive								
Test de différents modèles de prédiction								
Entrainement du modèle de prédiction choisi								
Développement de l'interface utilisateur								
Développement de l'API								
Sécurité et tests de l'ensemble de l'application								
Documentation et préparation des livrables								
Livraison de l'application								

6. Validation et Tests

L'application sera soumise à des tests approfondis pour vérifier sa fiabilité, sa précision et sa sécurité.

7. Références

Les références aux bibliothèques, aux technologies et aux méthodologies utilisées pour le développement de l'application seront documentées.

9. Architecture de projet RNCP :

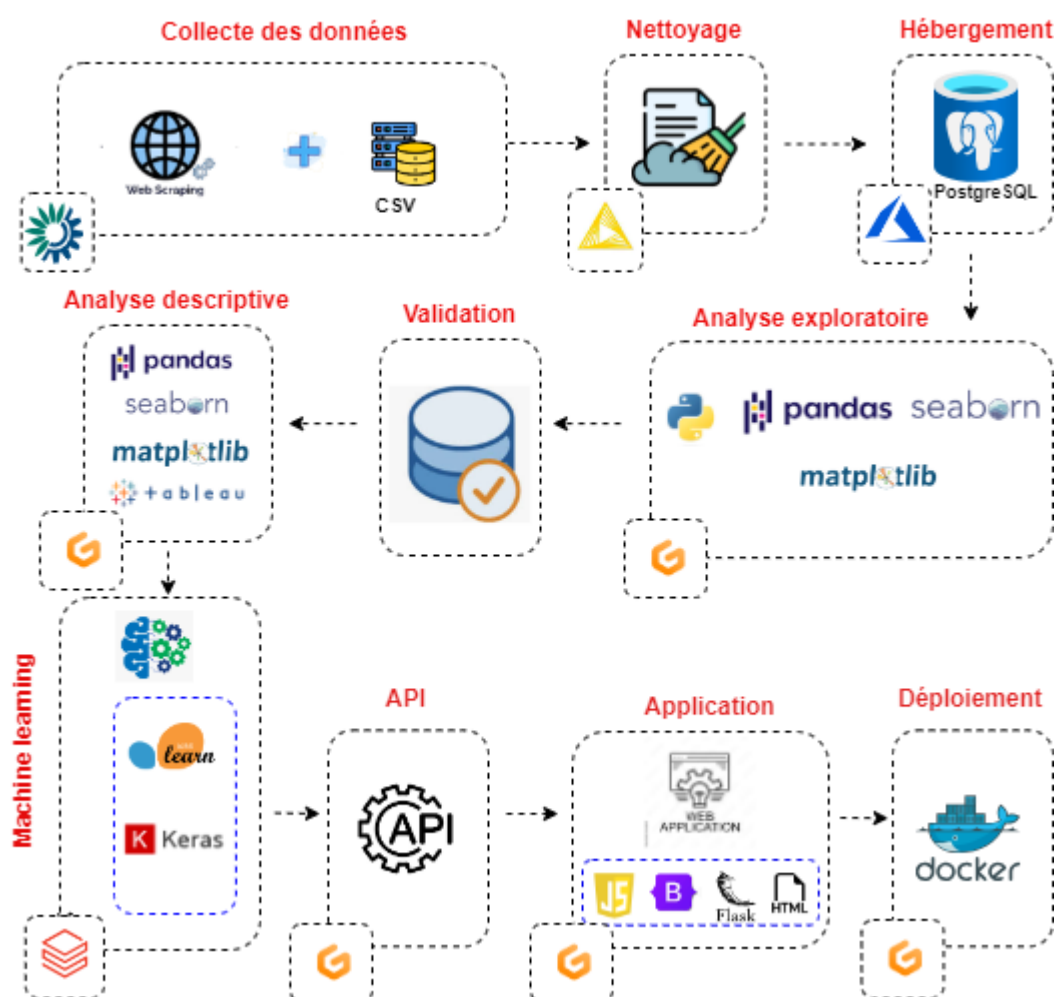


Figure 2: Architecture du projet RNCP

Lexique :

AEE : Agence Européenne de l'Environnement. Elle fournit de nombreuses données à l'échelle européenne, notamment concernant les véhicules immatriculés en Europe et leurs émissions de GES associées.



**European
Environment
Agency**

GES : Gaz à Effet de Serre. Les émissions de GES sont exprimées en équivalent CO2 (gCO2_{eq})

Mix de production de l'énergie électrique : il s'agit de la part de chaque type d'énergie primaire (Hydraulique, pétrole, gaz, solaire...) dans la production d'énergie électrique. Ce mix permet de déterminer la quantité de GES générés pour chaque MWh d'énergie électrique

ML : machine learning