

## Résumé du projet

Dans le contexte actuel de prise de conscience environnementale, l'analyse des émissions de CO2 des véhicules commercialisés dans le monde revêt d'une importance capitale. Dans ce projet, il est question de prédire les émissions de CO2 générées par un véhicule à partir de ses principales caractéristiques (physiques, mécaniques, chimiques, marque, gamme,.....). Plusieurs techniques d'intelligence artificielle (Machine Learning et Deep Learning) ont été utilisées et appliquées aux données disponibles dans les liens ci-dessous pour répondre à la cible de ce projet.

[Emissions de CO2 et de polluants des véhicules commercialisés en France](#)

[Monitoring of CO2 emissions from passenger cars Regulation \(EU\) 2019/631](#)

La démarche proposée dans ce projet peut être scindée en 3 parties clés :

- Une première partie sur l'analyse statistique et exploratoire des données. Dans cette partie, plusieurs fichiers sous format CSV ont été récupérés dans les liens cités plus haut. Ces fichiers ont été examinés, analysés puis certains ont été concaténés pour avoir un fichier final qui a conduit à une analyse de forme et de fond. Ces analyses ont permis d'orienter le sujet sur l'utilisation de machine Learning et de Deep Learning. Le code utilisé pour cette première partie est tracé dans le notebook nommé : **Projet CO2 EDA & Preprocessing**
- Dans un projet de machine ou Deep Learning, l'étape en aval de l'analyse exploratoire consiste à mettre les données dans un format propice au développement des modèles. Cela consiste à avoir des données les plus propres possible afin d'améliorer la performance des modèles. Pour cela, il existe une tonne d'opérations possibles. Les données peuvent être encodées, normalisées et nettoyées. Dans certains cas, des méthodes de sélection et de création de variables ou d'extraction de caractéristiques de variables sont utilisées pour assainir les données afin qu'elles soient prises facilement en compte dans des modèles. Cette étape a été l'objet de la deuxième partie du projet, et est documentée dans le notebook : **Projet CO2 EDA & Preprocessing**
- La troisième partie du projet a été, d'embarquer les données traitées dans les parties 1 et 2, dans des modèles de Machine Learning et de Deep Learning. Pour ce faire, des méthodes de régression et de classification ont été utilisées dans les deux rubriques (Machine Learning et Deep Learning). Dans ce cadre, plusieurs modèles ont été entraînés, évalués, puis optimisés à l'aide des méthodes RandomSearchCV et d'ensemble Learning basé sur les techniques de Bagging et de Boosting. Les modèles RandomForest, DecisionTree et KNN présentent de meilleurs scores par rapport aux autres modèles testés dans la rubrique Machine Learning. Les codes des modèles de régression et de classification de la rubrique machine Learning sont documentés dans les notebooks **Projet CO2 Regression Modelisation ML** et **Projet CO2 Classification Modelisation ML**. Au niveau du Deep Learning, Nous avons utilisé un modèle séquentiel de la bibliothèque Keras qui permet d'empiler des couches de neurones. Les scores obtenus au niveau de la régression et de la classification sont proches de ceux obtenus par les modèles utilisés dans la rubrique classification. Les données sont tracées dans les notebooks **Projet CO2 Regression Modelisation DL** et **Projet CO2 Classification Modelisation DL**.