

Thèse professionnelle

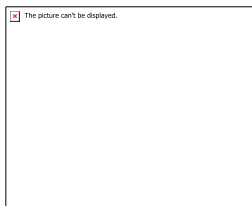
L'intelligence artificielle peut-elle nous aider à réduire notre empreinte carbone individuelle ?



Étudiant : Omar LAOUAN
Tuteur de thèse : Julien Robillard

MS Big Data
2021-2022

A renseigner par l'étudiant



☒ Non Confidentiel

☐ Confidentiel

Titre :

L'intelligence artificielle peut-elle nous aider à réduire notre empreinte carbone individuelle ?

Programme : MS - Big Data 2022 - Grenoble (2021 - 2022)

Année : 2021-2022

EA - Session de programme : Projet thèse Big Data 2021-2022

Nom de l'étudiant : Omar Farouk Laouan Yacouba

Nom du tuteur / évaluateur : Julien Robillard

A renseigner s'il s'agit d'un stage :

Nom de l'entreprise :

Ville :

Pays :

Fonction occupée durant le stage :

Résumé informatif de la mission : L'objectif principal est d'étudier, de challenger et de montrer les limites éventuelles de diverses solutions utilisant des techniques d'intelligence artificielle qui peuvent potentiellement aider à lutter contre le réchauffement climatique en permettant à des individus de réduire leur empreinte carbone. Pour ce faire, je m'intéresserai d'abord aux solutions qui peuvent permettre aux individus de mieux comprendre l'impact de leurs comportements, en rendant notamment l'information disponible ou en prédisant l'impact en fonctions de diverses caractéristiques liées à l'individu ou à son environnement. Ensuite je me focaliserai sur l'aide que peut apporter l'IA pour suggérer aux individus des comportements plus vertueux.

Mots-clés principaux décrivant la mission ([*cf. Thesaurus du Management*](#)) :

INTELLIGENCE ARTIFICIELLE, ANALYSE DES DONNEES, DONNEES CHIFFREES, RESISTANCE AU CHANGEMENT.

Table des matières

Abstract.....	4
Remerciements	5
1. Introduction	6
3. Contexte et objectif	7
3.1. Intérêt et objectif du projet	7
3.1.1. Intérêt de la recherche	7
3.1.2. Objectif de recherche	8
3.2. Périmètre du projet.....	8
3.3. Bilan carbone	8
3.3.1. Qu'est-ce que le bilan carbone ?	8
3.3.2. Quel est le bilan carbone des Français ?	9
3.4. Machine Learning	9
3.4.1. Qu'est-ce que le Machine Learning ?	9
3.4.2. Quand est-il pertinent de faire du Machine Learning ?	12
5. État de l'art des calculateurs carbone	15
5.1. Saisie entièrement manuelle	15
5.2. Automatisation partielle.....	17
6. La science des données au service du bilan carbone	19
6.1. Récupération automatique des données	19
6.1.1. Transport	20
6.1.2. Logement	23
6.1.3. Alimentation	27
6.1.4. Achats	30
6.1.5. Dépenses publiques.....	31
6.2. Communication plus efficace	31
6.3. Personnalisation dans la proposition d'alternative	33
1. Limites et challenges.....	34
1.1. Biais cognitifs empêchant les individus d'agir face au réchauffement climatique.....	34
1.2. Le Machine Learning a aussi son empreinte carbone	36
1.3. RGPD et utilisation des données	38
2. Cas pratique : calcul automatique de l'empreinte carbone du transport.....	39
2.1. Acquisition et préparation des données	39
2.2. Préparation des données (Feature engineering)	40
2.3. Modélisation.....	41
2.4. Calcul des émissions	42
Conclusion	43
Liste des figures et tableaux	44
Bibliographie	45

Abstract

Cette thèse professionnelle traite de l'application de techniques de Machine Learning pour aider les individus à réduire leur empreinte carbone.

L'objectif est d'étudier les challenges, limites et opportunités de plusieurs solutions visant à aider les individus à réduire leur empreinte carbone.

Remerciements

Je tiens à remercier mon tuteur Julien ROBILLARD pour son encadrement, ses conseils, ses enseignements tout au long de la rédaction de cette thèse.

Je remercie aussi tous les enseignants et intervenants du master Big Data pour la qualité des cours enseignés et pour toutes les compétences transmises.

1. Introduction

L'idée de ce projet vient avant tout d'une préoccupation personnelle que j'ai, comme beaucoup, face au réchauffement climatique. Les conséquences sont déjà observables aujourd'hui (épisodes caniculaires de plus en plus fréquents, élévation de la température, montée des eaux, chute de la biodiversité, etc.).

Si nous n'agissons pas collectivement, ces conséquences vont être de plus en plus désastreuses comme le soulignent les rapports successifs du Groupe d'Experts Intergouvernemental sur l'évolution du climat (GIEC)¹. En tant qu'ingénieur, j'ai envie d'utiliser la technique pour le bien commun. C'est pourquoi, armé des connaissances techniques et managériales que j'ai acquises durant le MS Big Data, j'ai fait le choix de traiter de ce sujet qui est à l'intersection de mes compétences et de mes centres d'intérêts et qui couvre des enjeux importants de notre siècle.

3. Contexte et objectif

3.1. Intérêt et objectif du projet

3.1.1. Intérêt de la recherche

Le changement climatique est l'un des plus grands défis de l'humanité pour les siècles à venir. Selon le dernier rapport du GIEC², il y aura déjà des impacts irréversibles à court et moyen terme du fait du surplus de CO₂ émis dans l'atmosphère. On peut citer par exemple :

- La température moyenne à la surface du globe s'élèvera de 1,5 °C d'ici à 2030, ce qui aura pour conséquences une augmentation de la sécheresse des sols, la raréfaction de l'eau, ainsi que des risques de famine à cause d'une baisse des rendements agricoles
- Le niveau de la mer continuera d'augmenter, ce qui entraînera des migrations massives, car beaucoup de grandes villes sont situées près des côtes.

Dans son dernier rapport, le GIEC juge que les actions mises en œuvre par les gouvernements à travers le monde sont insuffisantes et bien en dessous des objectifs fixés lors des conférences des parties (COP) successives.

Face à un problème de cet ordre, on peut se sentir dépassé individuellement et jugé que les moyens d'actions ne sont que dans la main des états. On développe dès lors des mécanismes de défense intellectuels qui vont du déni (climato-scepticisme) au découragement. On cède ainsi à des biais cognitifs qui nous permettent de justifier notre manque d'action. ³

Le changement climatique, même s'il est un des principaux des sujets de préoccupation des populations à travers le monde, ne génère pourtant pas de mobilisation citoyenne à grande échelle. Effet surprenant, alors même que les conséquences du changement climatique sont de plus en plus visibles, le nombre de climato-sceptiques augmentent⁴.

Une des causes de l'inaction est le manque d'information personnalisée et en temps réel sur les impacts de nos choix de consommation. En effet, les indicateurs d'émissions sont souvent globaux et les conséquences sont projetées sur le temps long. Par exemple, on parlera de moyenne d'émissions de CO₂ par habitants et de conséquences à horizon 2100. Ce type d'indicateurs peut avoir pour effet de diluer les responsabilités individuelles. Depuis 2020, l'agence Pour répondre à ce besoin d'information personnalisé, il existe de nombreux « calculateurs carbone » qui permettent d'estimer son bilan carbone personnel. Par exemple, depuis 2020, l'agence de l'environnement et de la maîtrise de l'énergie (ADEME) propose un calculateur carbone destiné aux particuliers⁵. Ces calculateurs ont le mérite de donner une première estimation des émissions de CO₂ en fonction des réponses de l'individu à une série de questions sur ses modes de consommation (alimentation, logement, numérique, etc.). Mais ces calculateurs ont deux grandes limites : ils reposent sur une saisie manuelle des données (avec toutes les imprécisions et erreurs potentielles), et ils ne sont pas en temps réel.

Il est possible, dans certaines conditions, de palier à ces limites en utilisant des solutions d'intelligence artificielle (IA).

Les algorithmes d'IA étant une révolution technologique qui nous permet de résoudre des problèmes complexes, on peut légitimement se demander dans quelle mesure ils peuvent nous assister pour d'une part mieux nous informer sur les impacts environnementaux de nos actions et d'autre part pour nous proposer des alternatives personnalisées auxquelles nous sommes le plus susceptibles d'adhérer. Il conviendra néanmoins de prendre en compte, l'impact inhérent à l'utilisation de ces technologies qui sont elles-mêmes consommatrices de ressources et émettrices de CO₂. C'est le sujet de cette recherche.

3.1.2. Objectif de recherche

L'objectif principal est d'étudier, de challenger et de montrer les limites éventuelles de diverses solutions utilisant des techniques d'intelligence artificielle qui peuvent aider des individus de réduire leur empreinte carbone en automatisant le calcul de l'empreinte carbone et en proposant des alternatives personnalisées.

3.2. Périmètre du projet

Seront exclus de cette étude :

- Les solutions qui se basent sur des technologies non matures (par exemple l'utilisation de l'IA dans l'optimisation des technologies de capture de CO₂, ou dans la découverte de nouveaux matériaux).
- Les solutions qui s'adressent aux entreprises ou aux industriels.

Le périmètre de l'étude se limitera donc aux solutions actionnables à l'état actuel de l'art technologique et à destination des individus.

3.3. Bilan carbone

3.3.1. Qu'est-ce que le bilan carbone ?

Le Bilan carbone est une méthodologie de calcul des émissions directes ou indirectes de gaz à effet de serre (GES) liées à l'activité humaine. Cette activité peut être celle d'une entreprise, celle de l'état, ou celle d'un individu. Dans ce travail de recherche, on se concentre uniquement sur la part due aux individus.

Sont pris en compte dans le bilan, les émissions de six gaz à effet de serre :

- Dioxyde de Carbone - CO₂
- Méthane - CH₄
- Hydrofluorocarbure - HFC
- Protoxyde d'Azote - N₂O
- Perfluorocarbure - PFC
- Hexafluorure de Soufre - SF₆

Parmi ces gaz, le carbone est le plus répandu et est utilisé comme base de calcul. Ainsi la quantité produite des autres gaz est convertie en équivalent carbone. (on parle alors de tonnes CO₂ équivalent). Le calcul des émissions se base sur les analyses de cycle de vie de produits et utilise la base de données d'émissions de l'ADEME comme source.

3.3.2. Quel est le bilan carbone des Français ?

Avant de voir dans quelle mesure l'IA peut nous aider à mieux connaître nos émissions carbone personnelle, faisons un tour d'horizon des différents postes d'émission des Français. Quels sont les principaux postes d'émission carbone pour un individu ?

En grosse maille, l'empreinte carbone moyenne en France s'élève à 9t CO₂ eq / personne en 2019. En rentrant dans le détail, on obtient les statistiques suivantes fournies par le cabinet carbone4 :

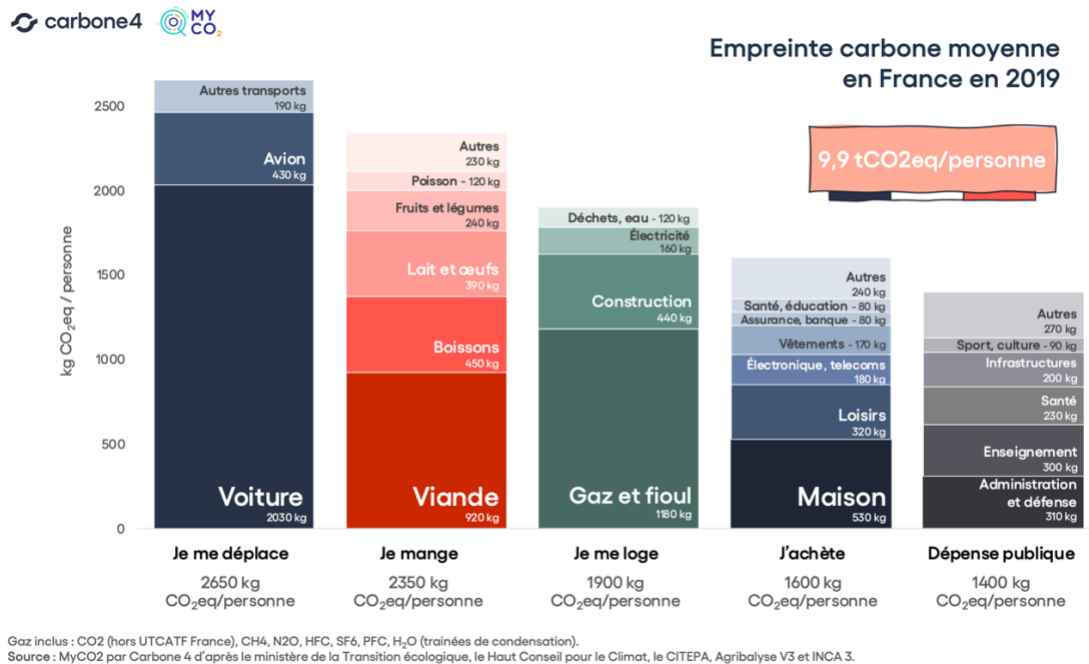


Figure 1 : Empreinte carbone moyenne en France en 2019.

(Source : carbone4.com)

Les émissions sont réparties selon plusieurs postes de dépense : *je me déplace, je mange, je me loge, j'achète et les dépenses publiques*.

Il est intéressant de noter que le principal poste de dépense est le transport dans lequel les déplacements en voiture constituent la plus large part avec 2030 kgCO₂eq/personne. À l'inverse, la catégorie la plus économe en carbone est la dépense publique.

3.4. Machine Learning

3.4.1. Qu'est-ce que le Machine Learning ?

Le Machine Learning (ML) est un ensemble de technique qui consiste à programmer les ordinateurs pour leur apprendre à réaliser une ou plusieurs tâches. Selon Arthur Samuel, "Le Machine Learning est le domaine d'étude qui donne aux ordinateurs la capacité à apprendre sans être explicitement programmé" (Arthur Samuel, 1959). Autrement dit, le ML diffère de l'approche classique en ce sens qu'il n'y a pas besoin d'inscrire explicitement dans le programme toutes les étapes de résolution d'une tâche. Une définition plus formelle est donnée par Tom Mitchell : "Un ordinateur apprend d'une expérience E respectueusement à une tâche T et une mesure de performance P, si sa performance sur la tâche T, comme mesuré par P, augmente avec l'expérience E"

(Tom Mitchell, 1997). À l'ère du Big Data, on utilise au quotidien des services boosté par du Machine Learning. Par exemple, le détecteur de spams intégré à quasiment toutes les boîtes mails, les traducteurs de texte comme Google traduction, les détecteurs de personnes dans les photos, les agents conversationnelles (Siri, Alexa), etc. Le point commun entre tous ces services est qu'il n'y a pas dans le code de leurs programmes respectifs toutes les étapes nécessaires à la réalisation des tâches. Pour certaines tâches, il est même impossible d'explicitier ces tâches, car la définition du problème n'est pas aisée. Par exemple, écrire un programme de reconnaissance vocale nécessite d'identifier l'amplitude sonore et la fréquence de toutes les syllabes et de prendre en compte les variations dues aux individus comme le timbre de la voie et l'accent. Ainsi pour résoudre ce problème, on laisse la machine apprendre seul les règles d'association entre la version écrit et orale d'un mot et les caractéristiques les plus importantes pour faire dans cette association. Dans son livre, *Hands-on Machine Learning With Scikit-Learn, Keras and Tensorflow, 2ed.* Aurélien Géron donne une comparaison schématique entre les programmes basés sur du Machine Learning et les programmes classiques :

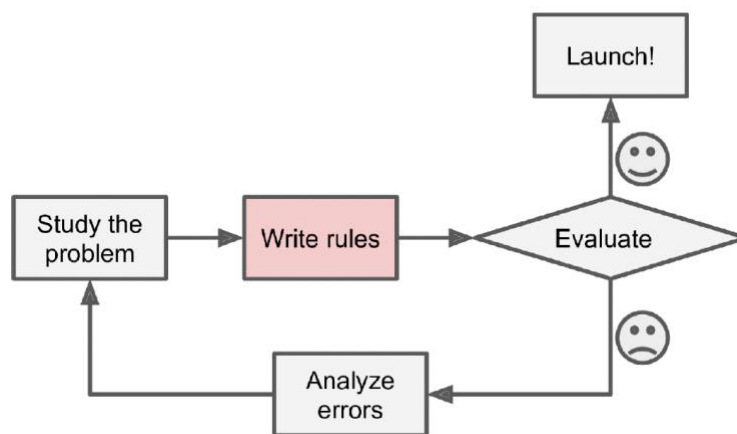


Figure 2 : Programmes traditionnels

(Source : Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow)

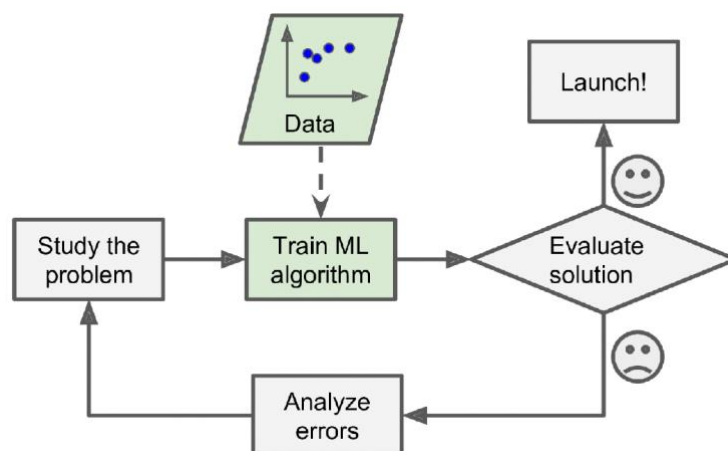


Figure 3 : l'approche Machine Learning

(Source : Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow)

On voit bien que la différence clé entre l'approche traditionnelle et l'approche Machine Learning que les programmes classiques reposent avant tout sur l'écriture de règles explicites sur la tâche à résoudre, alors que l'approche Machine Learning repose sur l'entraînement d'algorithme sur la base de données. De plus, un autre avantage des algorithmes de Machine Learning est qu'ils peuvent aider les êtres humains à découvrir des liens jusqu'ici inconnus dans la résolution d'un problème spécifique. Par exemple, si on entraîne un algorithme à détecter des fraudes bancaires avec succès, on peut dans un second temps prendre connaissance des caractéristiques sur les transactions qui ont eu le plus de poids dans la prédiction. On peut ainsi remarquer que si la transaction a lieu en dehors du pays de résidence du client, il est très probable qu'elle soit frauduleuse. Ce qui permet de mettre en place des actions de mitigations. ci-dessous une vue schématique du processus par lequel le Machine Learning nous aide à affiner notre compréhension d'un problème.

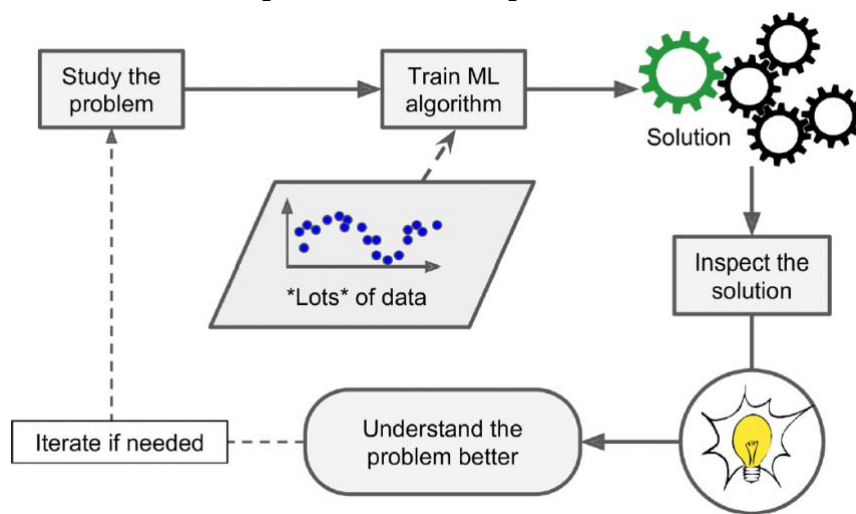


Figure 4 : le Machine Learning nous aide à apprendre

(Source : Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow)

On décompose les algorithmes de Machine Learning en deux grandes catégories : Les algorithmes d'apprentissage supervisé et les algorithmes d'apprentissage non supervisé.

On parle d'apprentissage supervisé quand on entraîne un algorithme pour réaliser une tâche sur la base de données étiquetées. Une étiquette est la « bonne réponse » associé à une observation. Par exemple, dans le cas d'un problème de prédiction de départ d'un client, la base de données sur laquelle sera entraîné un algorithme supervisé va contenir, en plus des données du client, une étiquette qui correspond au départ ou non du client (par exemple 1 si le client est parti, 0 sinon). Les algorithmes d'apprentissage supervisé sont eux même subdivisé en deux sous-catégories selon la nature de la variable à prédire. Si cette dernière est continue (par exemple : chiffre d'affaires, prix de vente) on parle d'algorithme de régression. Si la variable à prédire est discrète (par exemple : présence ou non d'une maladie, fraude ou pas fraude), on parlera alors d'algorithme de classification.

La deuxième grande catégorie est celle des algorithmes d'apprentissage non supervisé. À la différence du cas supervisé, ici, on n'a pas d'étiquette sur les données. On n'a que des observations sans le contexte associé. Reprenons comme exemple la détection de fraude bancaire, sans étiquettes nous indiquant quelles transactions historiques sont frauduleuses ou non, on ne pourra pas prédire la fraude avec un algorithme d'apprentissage supervisé. On pourra seulement faire de l'apprentissage non supervisé qui consistera par exemple à détecter des groupes homogènes (clusters) dans les données. Un autre exemple d'application de l'apprentissage non supervisé est la segmentation automatique d'une base de données client pour mieux comprendre leurs sociologies et mieux les cibler. Pour résumer, ci-dessous une vue schématique des principaux types d'algorithmes de Machine Learning.

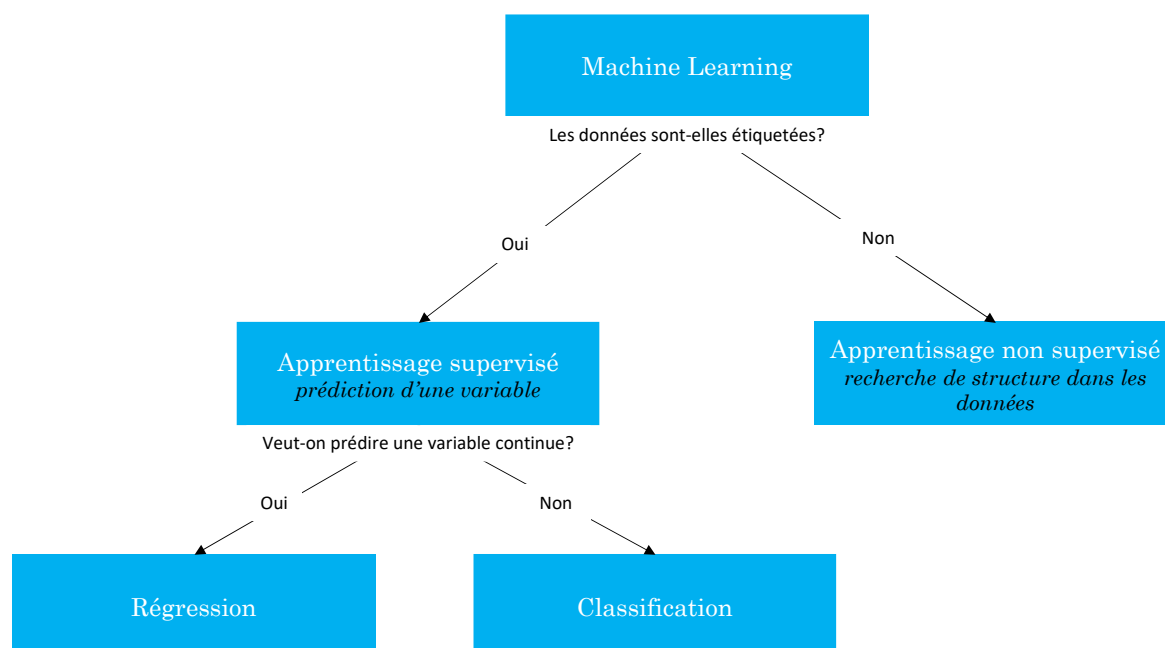


Tableau 1 : Principaux type d'algorithmes de Machine Learning

3.4.2. Quand est-il pertinent de faire du Machine Learning ?

Le Machine Learning est une technologie très à la mode. Il suffit de voir l'évolution de la recherche des mots clés « Machine Learning » sur Google Trends.



Dans tous les pays. 10/11/2010 – 01/08/2022. Recherche sur le Web.

Figure 5 : Tendence des mots clés "Machine Learning". Source : Google Trends.

Cette tendance s'explique aussi par le fait que le Machine Learning permet de résoudre des problèmes complexes jusqu'ici difficile à résoudre. Il est dès lors tentant de vouloir appliquer du Machine Learning à tout va, sans s'assurer au préalable de sa pertinence pour le problème qu'on cherche à résoudre. Il y a plusieurs cas pour lesquels le Machine Learning n'est pas pertinent. Pour répondre à la question des cas où il est pertinent de faire du Machine Learning, regardons les cas où il n'est pas pertinent d'en faire. Il n'est pas recommandé de faire du Machine Learning si :

- On manque de données : C'est le cas le plus trivial. Si on manque de données, il est alors impossible d'entraîner des algorithmes de Machine Learning, car il n'y aura aucune base d'apprentissage. Si on a des données, mais qu'elles ne sont pas labellisées, on ne pourra pas non plus faire de l'apprentissage supervisé.
- On manque de données en quantité suffisantes : Ce cas est moins trivial, car à l'état actuel de la recherche, on ne peut pas dire à l'avance et avec précision la quantité de données dont on a besoin pour résoudre un problème donné avec des algorithmes de Machine Learning. L'estimation de la quantité de données nécessaires est principalement basée sur l'expérimentation⁶ ⁷.
- On manque de données fiables et représentatives : avoir des données ne suffit pas, encore faut-il qu'elles soient fiables et représentatives du problème qu'on cherche à résoudre. Le non-respect de ces conditions introduira des biais importants dans les résultats de l'algorithme et peut avoir des conséquences graves selon le cas d'application.⁸
- On veut de reproductibilité parfaite : les algorithmes de Machine Learning ont du mal à être rigoureusement reproductible du fait de leur nature stochastique⁹. Ainsi, il est commun de voir des papiers de recherches dont les résultats publiés par les auteurs sont significativement différents des

résultats obtenus par des tiers. Sur des domaines d'application sensible nécessitant une reproductibilité parfaite, ce point peut justifier la non-utilisation de ces algorithmes.

- On peut résoudre le problème cible à bas coût grâce des systèmes experts : certains problèmes d'optimisation ou de classification à complexité faible peuvent être résolus par des systèmes experts (basés sur des règles explicites) ayant un coût de mise en œuvre et de maintenance plus faible que les algorithmes de Machine Learning¹⁰.

On peut par conséquent utiliser les algorithmes de Machine Learning dans les autres cas. C'est-à-dire quand on doit résoudre des problèmes complexes, qu'on a des données fiables, en quantité suffisantes et qu'on peut tolérer des variations mineures dans les prédictions. C'est le cadre dans lequel on se place pour cette recherche. Nous verrons au chapitre de ce document intitulé La science des données au service du bilan carbone que les conditions précédentes sont réunies et qu'on peut légitimement utiliser les algorithmes de Machine Learning pour la prédiction automatisée de l'empreinte carbone individuel.

5. État de l'art des calculateurs carbone

Les calculateurs de l'empreinte carbone individuelle existent depuis plusieurs années. Les calculateurs sur le marché ont plusieurs niveaux de complexité. Cela va du calculateur basé entièrement sur de la saisie manuelle de donnée (réponse à un questionnaire) au calculateur récupérant automatiquement les informations nécessaires pour faire le calcul de l'empreinte carbone lié à des modes de consommation. Nous allons ici détailler le fonctionnement de ces calculateurs et exposer certaines de leurs limites. Pour chaque type de calculateur, nous allons donner un exemple de solution qui existe déjà dans le commerce.

5.1. Saisie entièrement manuelle

Le calculateur carbone le plus élémentaire repose sur une saisie manuelle des données. L'utilisateur est guidé via une série de questions à communiquer des informations sur son mode de vie selon plusieurs catégories : le logement, le transport, l'alimentation et les divers achats.

Selon les réponses aux différentes questions, une empreinte carbone est associée à l'utilisateur, souvent exprimé en kilogrammes CO₂ équivalent, noté KgCO₂eq. Le calcul repose essentiellement sur des tables de correspondance entre différents modes de consommation et l'empreinte carbone moyenne des Français qui leur est associé. Ces tables de correspondances sont fournies par l'ADEME

Groupe d'aliments	Conso moyenne (g/j)	Facteur Émission (gCO ₂ /g)
Pain et panification sèche raffinés	108,1	1,52
Pain et panification sèche complets ou semi-complets	6,3	1,52
Céréales pour petit déjeuner et barres céréalières	4,8	0,34
Pâtes, riz, blé et autres céréales raffinées	62,7	0,71
Pâtes, riz, blé et autres céréales complètes et semi-complètes	1,6	0,71
Œufs et plats à base d'œufs	12,6	2,61
Viandes (hors volailles)	47,3	21,78
Volailles	26	5,91
Charcuterie	27,3	7,00
Poissons	23	6,18
Crustacés et mollusques	3,7	10,70
Abats	2,7	7,36
Légumes	130,7	1,29
Légumineuses	7,7	0,79
...

sur les principales catégories.¹¹ Ci-dessous un extrait qui donne les émissions de CO₂ liées à quelques aliments.

Tableau 3 : Émissions moyennes par aliments d'un Français.

(Source : bilans-ges.ademe.fr)

L'ADEME fournit aussi des tables de correspondance par repas type plutôt que par aliments unitaires pour faciliter l'estimation. On retrouve par exemple l'empreinte carbone d'un repas classique avec poulet, repas classique avec bœuf,

repas végétarien, etc. Ci-dessous l'exemple du bilan carbone associé à un repas classique avec bœuf.

Types de repas	Qte (g)	Kcal	Protéines (g)	Lipides (g)	gCO _{2e} /kg	gCO _{2e}
Entrée : légumes à la grecque						
légumes de saison	200	60	3	0	267	53,4
huile d'olive (1/2 c.s.)	7	63	0	7	2600	18,2
Plat principal : poulet au riz						
poulet	150	225	30	10,5	5160	774
riz	60	214	5	1	1410	84,6
beurre	10	76	1	8,4	9490	94,9
Plateau de fromages						
fromage à pâte molle	25	68	5	5	4280	107
fromage à pâte dure	25	100	7	7,5	5600	140
Pain	50	125	4	0,6	1520	76
Total		931	55	40		1350

Tableau 4 : Émissions moyennes d'un repas classique avec bœuf pour un Français.

(Source : bilans-ges.ademe.fr)

Vous trouverez toutes les listes publiées par l'ADEME sur le site dédié : bilans-ges.ademe.fr

Les calculateurs basés sur de la saisie manuelle de donnée ont plusieurs avantages :

- Ils sont faciles à mettre en place : comme vu précédent, le principal ingrédient de ce type de calculateur sont les données rentrées manuellement par un utilisateur et les tables de correspondances qui donnent les émissions moyennes par catégories. En comparaison avec des calculateurs basés sur du Machine Learning, la mise en place est plus simple et moins coûteuse puisqu'on n'a pas d'algorithme à entraîner et à maintenir en production.
- Ils sont faciles à maintenir : c'est un avantage non négligeable puisque la maintenance d'une infrastructure complexe peut être coûteuse, car plus sujette à des bugs. Pour les calculateurs à saisie manuelle, le principal point de vigilance est la mise à jour des tables de correspondance si celle-ci venaient à changer. Puisque la table est basée sur des estimations moyennes d'émissions, l'ordre de grandeur de la durée entre deux mises à jour se compte en années plutôt qu'en secondes. De ce fait, on n'a pas besoin de faire de la mise à jour proche du temps réel, ce qui simplifie l'infrastructure sous-jacente.

Ces qualités en font les calculateurs les plus répandus aujourd'hui sur internet. Deux exemples de calculateurs à saisie manuelle sont celui de l'ADEME (nosgestesclimat.fr) et celui du cabinet carbone4 fondé par un des créateurs du bilan carbone (www.carbone4.com/myco2).

Cependant, ces calculateurs présentent quelques limites qui empêchent leur utilisation récurrente et à grande échelle :

- La saisie manuelle est incompatible avec une utilisation récurrente et massive : Il est intéressant d'analyser les calculateurs carbone en les considérant comme des services et en considérant utilisateurs comme des clients. L'objectif serait alors de maximiser satisfaction client et la loyauté envers le service. En se plaçant dans ce cadre, on note que les calculateurs carbone à saisie manuelle, même s'ils ont un taux de satisfaction élevé, ne parviennent pas à créer de la loyauté. En effet, plus l'utilisateur d'un service fait d'effort pour l'utiliser, moins il sera loyal. Ainsi, la saisie manuelle peut être un frein à une utilisation récurrente puisqu'elle demande un certain effort et un certain temps de la part de l'utilisateur. Ce type de service aura tendance à être peu utilisé et à être utilisé très rarement. Pour maximiser la loyauté d'un client, il faut lui fournir un service qui lui demande le moins d'effort possible dans son utilisation. Ce phénomène est connu sous le nom d'expérience sans effort (en anglais, Effortless Experience) ¹². Il est difficile de trouver des chiffres précis sur la fréquence d'utilisation des calculateurs à saisie manuelle, car ces données ne sont pas publiées. Mais on peut en avoir une approximation de leur succès en comparant le taux de satisfaction, la taille de la population ciblée et le nombre d'utilisateurs l'ayant utilisé au moins une fois. Le service MyCO2 du cabinet carbone 4 est un des rares à publier ses chiffres de satisfaction et son nombre d'utilisateurs. Le site affiche ainsi une note de satisfaction grand public de 4,7/5 et 30.000 personnes ayant utilisé le service¹³. Sachant que ce service existe depuis 2 ans et qu'il est destiné à une large partie de la population, le nombre d'utilisateurs est relativement faible. Si on définit la population cible comme étant les personnes dont l'âge est compris entre 20 et 64 ans (51,7 millions, selon l'institut National d'Études Démographiques¹⁴), cela représente 0.05% de la population.
- Le bilan carbone est calculé sur une base annuelle : cela peut seulement conditionner l'utilisateur à ajuster la fréquence d'utilisation du service sur la même échelle de temps. De plus, être sur de grandes échelles de temps (1 an) peut aussi être moins appréhendable, surtout si l'on veut mettre en place et suivre l'impact de certaines de nos actions pour réduire l'empreinte carbone. Il est plus facile de faire un suivi sur de courtes échelles de temps plutôt que sur des temps longs.

5.2. Automatisation partielle

La deuxième catégorie de calculateur carbone concerne ceux basés sur une récupération automatique des données sur une partie des catégories. C'est donc une automatisation partielle puisque l'utilisateur n'aura qu'un bilan partiel de son impact global. Il existe par exemple des calculateurs qui dresse le bilan carbone de la catégorie achat, en catégorisant automatiquement les achats sur les comptes bancaires¹⁵. D'autres calculent automatiquement l'empreinte carbone liée à l'électricité, en récupérant les données du mix énergétique ainsi que les émissions associées à chaque mode de production d'électricité¹⁶.

Les avantages des calculateurs automatiques mono-catégorie est qu'ils permettent de connaître son empreinte carbone sur des catégories, quasiment sans intervention de l'utilisateur. Les catégories ciblées par ce type de calculateurs sont choisies soigneusement. Les catégories sont ciblées sur la base de 2 critères : un accès relativement simple aux données et la possibilité pour les utilisateurs de pouvoir agir pour réduire l'impact carbone calculé sur la catégorie. Les catégories achats et électricité citées plus haut remplissent ces deux conditions.

Les inconvénients de ce type de calculateurs sont inhérents à leur spécialisation. Ils donnent une partie seulement du bilan carbone global, car ils se limitent à une catégorie. L'utilisateur a donc un résultat partiel et selon la catégorie sur laquelle se spécialise le calculateur, ce bilan peut être mineur par rapport au bilan globale. Par exemple, comme vu au chapitre Quel est le bilan carbone des Français ? la catégorie achat est la quatrième la plus émissive avec 1,6 tCO₂eq sur un total de 9,9 tCO₂eq. Cela représente 16% du bilan carbone global et est donc peu représentatif. Les calculateurs mono-catégories peuvent donc biaiser la connaissance de l'empreinte carbone individuelle en présentant une part minoritaire de ce bilan. Ces calculateurs peuvent compléter le bilan en le rendant semi-automatiser : calcul automatique sur certaines catégories et saisie manuelle sur les autres. Mais cela peut nuire à leur adoption pour les mêmes raisons que les calculateurs à saisie manuelle.

6. La science des données au service du bilan carbone

6.1. Récupération automatique des données

La force des algorithmes de Machine Learning, c'est qu'on peut les entraîner pour qu'ils puissent récupérer automatiquement les bonnes informations permettant dans un second temps de faire le calcul de l'empreinte carbone. L'avantage d'une telle démarche, c'est qu'elle est plus précise qu'une rentrée manuelle de données, car elle ne repose pas sur la capacité d'un individu à se rappeler ses modes de consommation, mais sur ses consommations réelles. Un autre avantage est que c'est plus facile de faire adopter sur le long terme un système automatisé qu'un système de saisie manuelle.

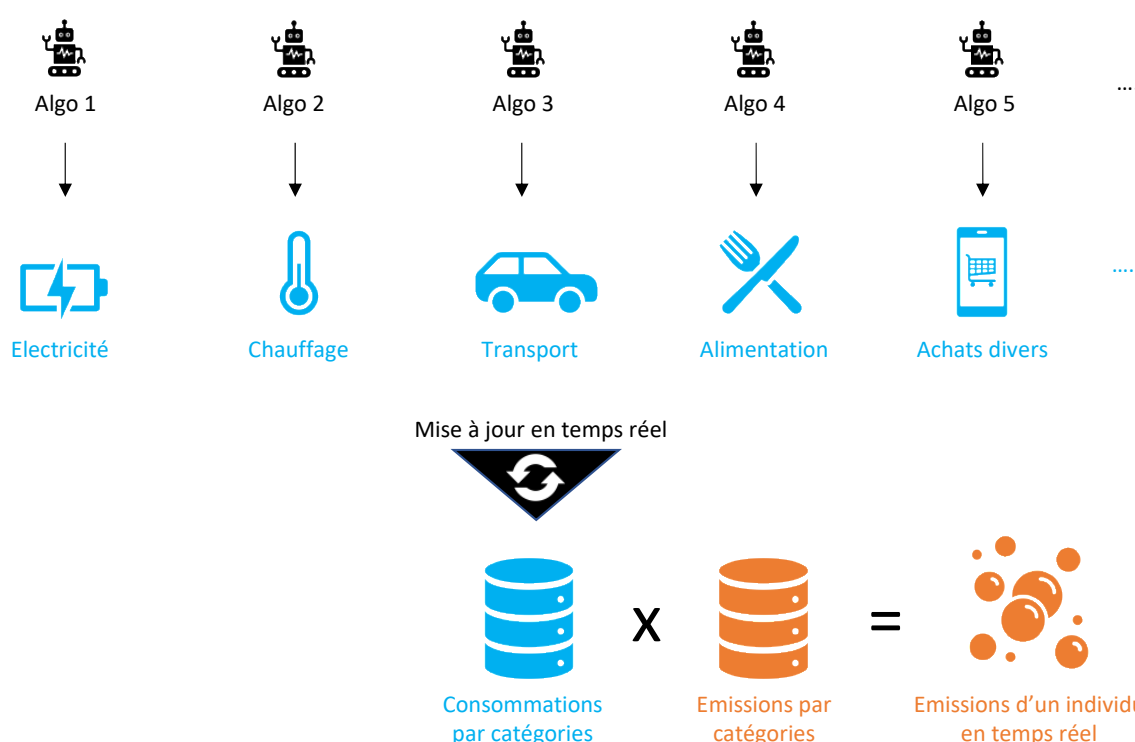


Figure 6 : récupération automatique de données

Pour récupérer les données automatiquement, il faut développer des algorithmes pour chaque catégorie. Les données cibles étant de nature et de structures différentes selon les catégories, il faut calibrer les algorithmes à chacune des catégories. Par exemple, pour la catégorie électricité, il s'agira de récupérer la consommation électrique et le mix énergétique associé alors que dans la catégorie transport, il s'agira de détecter automatiquement le mode de transport grâce aux informations enregistrées par des capteurs présent dans les smartphones (gyroscope, accéléromètre, etc.).

La multiplicité des modèles est une contrainte importante, car c'est autant d'algorithmes qu'il faudra suivre et maintenir en production. Ce suivi peut être coûteux, car il peut nécessiter des infrastructures de monitoring dédiées et une

expertise. Ce cout peut être réduit en privilégiant les algorithmes les plus simples et en mettant en place des standards de contrôle qui peuvent réduire le niveau d'expertise nécessaire.

Explorons maintenant comment, sur les différentes catégories, on peut entraîner des algorithmes de Machine Learning pour récupérer automatiquement la donnée nécessaire au calcul du bilan carbone. Nous aborderons ici les aspects méthodologiques, les aspects pratiques seront couverts dans le chapitre Cas pratique

6.1.1. Transport

Le transport est un des postes d'émission les plus importants en France avec 2650 kgCO₂eq/par personne. Comme vu précédemment, la massification de l'usage de la voiture (2030 kgCO₂eq/par personne) et le cout faible du voyage en avion (430 kgCO₂eq/par personne) contribuent à ce bilan Carbone élevé. Parmi les usagers des transports, il y a une conscience de la pollution engendrée par l'avion ou la voiture par exemple. Mais l'absence de visibilité en temps réel sur cet impact peut être un frein au changement de comportement. Il est donc important trouver des solutions qui peuvent donner une estimation en temps réel des émissions carbone liée au mode de transport. Des calculateurs carbone existent, mais ils demandent un effort de saisie à l'utilisateur, ce qui ne favorisent pas leur adoption.

À l'ère du Big Data et de l'intelligence artificielle, il est possible d'automatiser le calcul de l'empreinte carbone lié au transport.

Pour ce faire, il faut dans un premier temps trouver un moyen de connaître le mode de transport d'un individu en temps réel. Cela est possible en exploitant les données présentes dans les capteurs de smartphones comme le gyroscope et l'accéléromètre. Mais ces seules données ne suffiraient pas, car il faudrait aussi que ces données soient étiquetées pour qu'elles puissent être exploitables. Il est plus facile de traiter ce problème par de l'apprentissage supervisé. En effet, si on associe les données des capteurs et le contexte dans lequel elles ont été collectées (immobile, vélo, voiture, train, etc.), on pourra dans un second temps entraîner des algorithmes qui pourront apprendre la règle d'association et prédire en temps réel les modes de transport dans les futurs. Une fois collectées, les données auront comme structure :

Capteur 1	Capteur 2	Capteur 3...	...	+	Étiquette
x	x	x	...		Voiture
x	x	x	...		Vélo
x	x	x	...		Train
...

Il y a donc deux parties dans la collecte des données :

- La récupération des données de capteurs des smartphones
- L'étiquetage des données

Pour la récupération des données de capteurs, cela dépend du système d'exploitation du smartphone. Une des contraintes qu'on peut rencontrer à ce

niveau est que la multiplicité des systèmes d'exploitation exige d'avoir des applications pour chacun de ces systèmes. Si cette variété de systèmes pouvait être problématique il y a quelques années, ce n'est plus le cas aujourd'hui puisqu'il n'y a plus de grande diversité dans les systèmes d'exploitation, Android et iOS ayant complètement dominé le marché. La part de marché mondiale de ces deux systèmes combinés était 38,9 % en 2010 contre quasiment 100% aujourd'hui (figure 6).

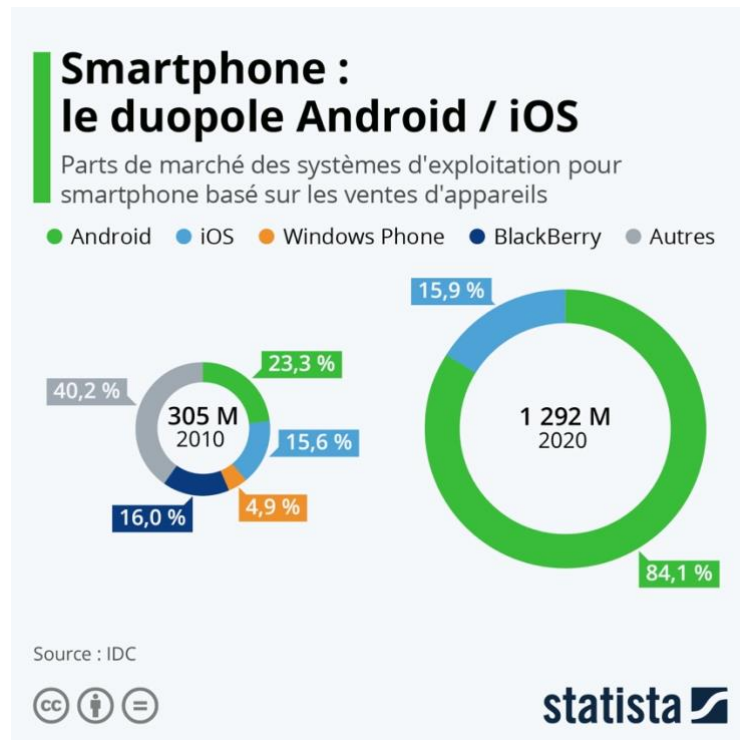


Figure 7 : Part de marché des systèmes d'exploitation de smartphone

(Source : statista.com)

La récupération des données de capteurs se limitera donc à Android et iOS. Ces deux systèmes permettent aux développeurs d'applications d'avoir accès à ces données via des Api dédiées^{17 18}, à condition bien sûr d'avoir l'approbation de l'utilisateur. Un exemple de protocole de collecte serait de sélectionner un groupe de volontaire et de collecter les données des capteurs de leurs smartphones sur plusieurs heures. C'est ce protocole qu'a suivi des chercheurs de l'Université de Bologne en Italie pour constituer la plus grande base de données de capteurs sur smartphone, nommée Transportation Mode Detection dataset¹⁹. Donnons quelques chiffres descriptifs clés (figure 7) : ces données contiennent 31 heures d'enregistrement d'activités de capteurs de smartphone, de 13 volontaires (10 hommes et 3 femmes), reparti quasiment équitablement (24% immobile, 26% marche, 25% voiture, 20% train, 5% bus).

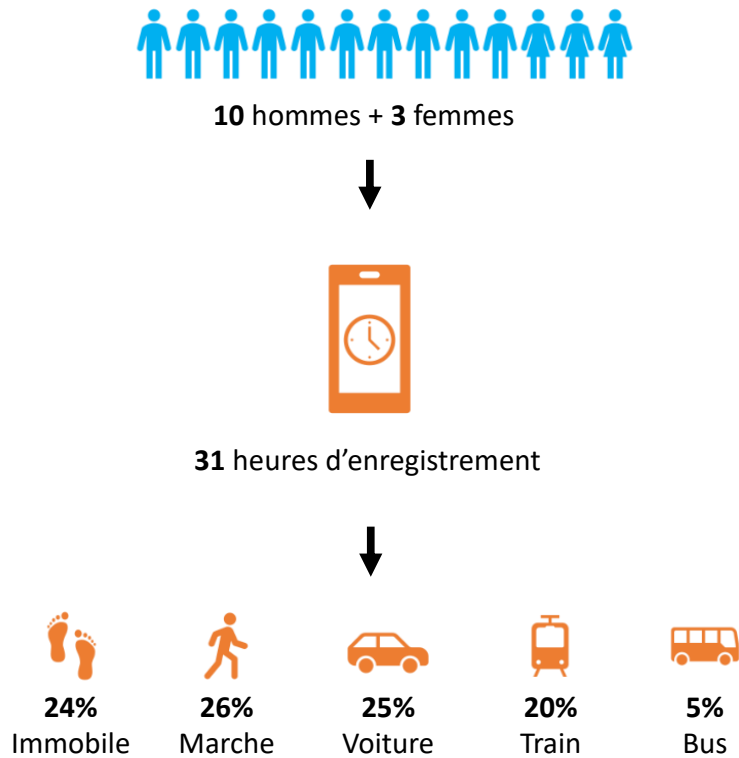


Figure 8 : Vue schématique du TMD dataset

Vous remarquerez dans la figure précédente qu'il y a une répartition par mode de transport. Cette répartition a nécessité un étiquetage à la main pendant le processus de collecte. C'est la deuxième partie du processus de collecte de données cité plus haut. Les chercheurs ayant collecté ces données ne donnent pas de détails sur leur protocole de labellisation, mais on peut identifier au moins 3 méthodes pour de labelliser ces données :

- Labellisation manuelle indirecte : les étiquettes sont saisies à la main sur un support indépendant du flux de collecte des données de capteurs. Peuvent être indiqués l'heure de début et l'heure de fin associé aux différents modes de transport. Ensuite, une jointure est faite manuellement entre les données capteurs et les modes de transports relevés via la variable temps communs aux deux tables. Ce protocole, étant manuelle et couvrant beaucoup de données venant de plusieurs volontaires, peut être sujet à des erreurs.
- Labellisation manuelle directe : même processus que précédemment, à la différence que les étiquettes sont saisies via un support connecté au flux de collecte des données de capteurs. Ce support peut être une application sur laquelle l'utilisateur indique en un clic le mode de transport qu'il utilise sur le moment. La jointure est alors faite automatiquement puisque les données sont collectées sur le même flux. Ce protocole peut également introduire des erreurs, mais dans une moindre mesure par rapport au protocole précédent puisqu'une partie de la collecte est automatisé.
- Labellisation semi-manuelle : cette technique fait appel à l'utilisation d'algorithme de clustering (apprentissage supervisé). L'astuce consiste à

labelliserseulement une petite partie des données couvrant tous les modes de transport. Ensuite, on entrainera un algorithme de clustering (comme le K-Means, le DBSCAN ou le HAC²⁰) qui détectera des groupes dans les données. Enfin, on labellisera chaque groupe selon l'étiquette la plus représentée parmi les données déjà étiquetées au sein du groupe. [Schéma] ?

Pour plus de détails concrets sur ce chapitre, voir le chapitre Cas pratique : calcul automatique de l'empreinte carbone du transport.

6.1.2. Logement

Comme vu précédemment, la catégorie logement est la troisième source d'émission individuelle en France avec 1900 kgCO₂eq/personne dont la majorité est due à la consommation de gaz et fioul (1180 kgCO₂eq), suivi des émissions liées à la construction (440 kgCO₂eq), la consommation d'électricité (160 kgCO₂eq) et le traitement des déchets et la gestion de l'eau (120 kgCO₂eq).



Figure 9 : Émission carbone moyennes par personne, pour la catégorie logement.

(Source : carbon4.com)

Il y a une prise de conscience générale quant à l'impact de la consommation énergétique des logements sur le climat. Selon un sondage Opinionway sur un échantillon de 1039 personnes, 73% des Français juge qu'il est nécessaire d'organiser une société plus sobre énergétiquement pour préserver les ressources et réduire l'impact environnemental de leurs modes de vie (figure 9).

« Nous allons parler de sobriété, c'est-à-dire la modification des comportements qui vise à réduire la consommation d'énergie, à préserver les ressources et à réduire l'impact environnemental de nos modes de vie »

Q. Pour vous, l'objectif d'organiser une société plus sobre énergétiquement en France ... ?

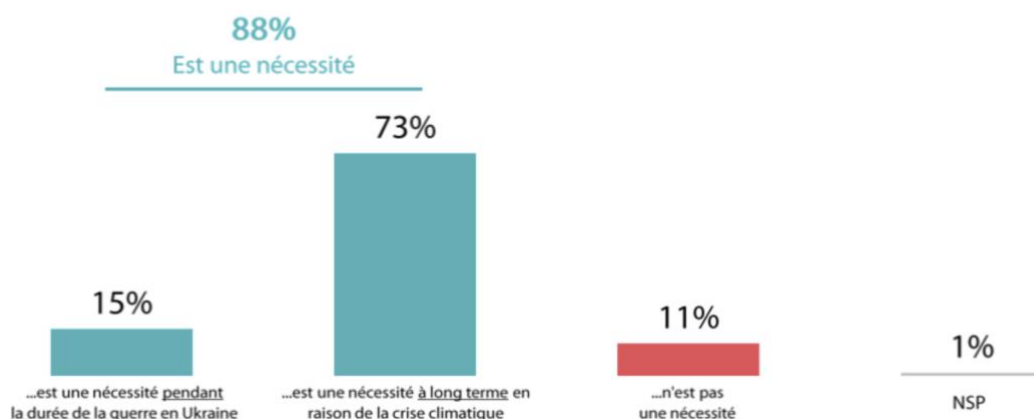


Figure 10 : Résultats de sondage sur la sobriété énergétique.

(source : www.opinion-way.com)

Toujours d'après le même sondage, la diminution de la consommation énergétique est perçue comme ayant un impact positif relativement important par rapport à d'autres actions. C'est la troisième catégorie d'action jugée la plus utile (Figure 11).

Q. Diriez-vous que chacune des actions suivantes a des effets positifs importants ou pas importants sur la protection de l'environnement ?

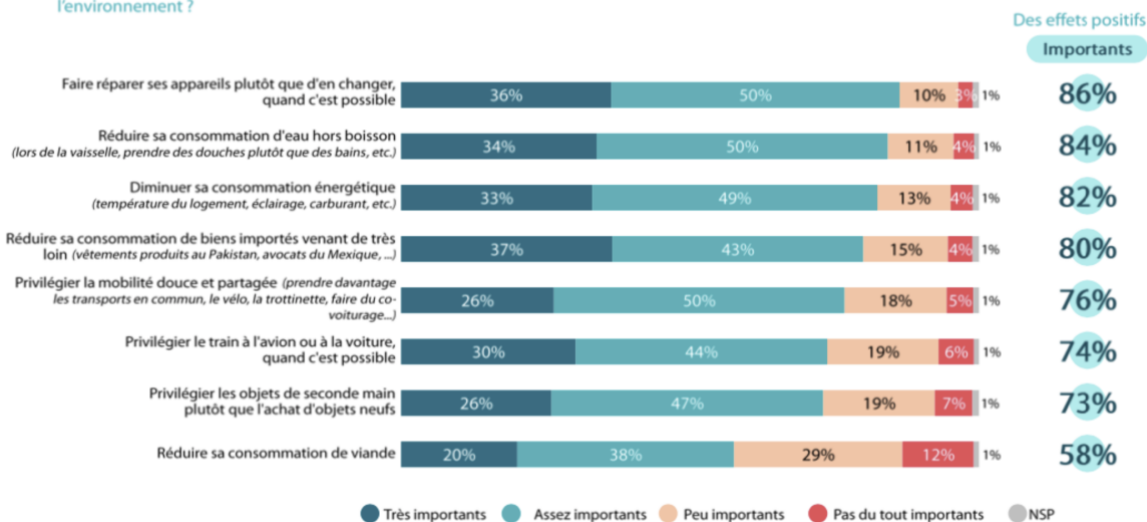


Figure 11 : Sondage sur l'impact perçu de différentes actions pour la protection de l'environnement

(source : www.opinion-way.com)

La consommation énergétique des logements est donc une catégorie sur laquelle les consciences sont plutôt favorables au changement. C'est aussi l'une des catégories sur laquelle on dispose de données pouvant être récupéré automatiquement pour calculer un bilan carbone en temps réel.

Le processus de collecte de données va dépendre de la sous-catégorie ciblée : gaz et fioul, électricité, construction, déchets et eau (figure 8).

Détaillons-le pour chaque cas :

- Gaz et fioul : les compteurs communicants Gazpar sont en déploiement partout en France avec l'ambition de couvrir la quasi-totalité des foyers équipés d'un compteur gaz d'ici à la fin de l'année 2022²¹. Le taux de couverture est aujourd'hui de 95% du territoire desservi en gaz. Ce compteur nouvelle génération permet d'avoir un suivi de sa consommation de gaz en kWh jusqu'à la maille horaire (Figure 12).

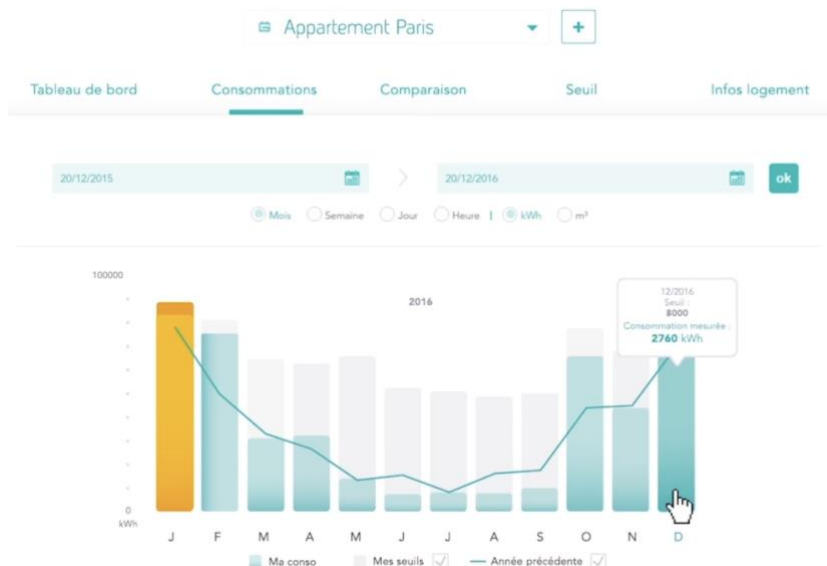


Figure 12 : Tableau de bord de suivi de consommation de gaz.

(Source : grdf.fr)

Les données sont accessibles via l'API GRDF ADICT (Accès aux Données Individuelles des Clients par des Tiers)²². Le site de GRDF indique que ce service est destiné aux collectivités ou entreprises voulant suivre la consommation de leurs sites, mais aussi aux « fournisseurs de services de Maîtrise De l'Énergie.

Les données sont également accessibles pour les particuliers via leurs tableaux de bord de suivi sur le site de GRDF. De ce fait, le processus de collecte automatique en vue du calcul du bilan carbone individuel peut se faire de deux manières : soit via l'API en se qualifiant comme fournisseur de service, soit en téléchargeant les données à une certaine fréquence sur l'espace client de l'utilisateur (en créant un canal de connexion par exemple). Dans les deux cas, il faudra un accord explicite de l'utilisateur du service.

- Électricité : la récupération automatique des données d'électricité fonctionnera en grande partie comme pour les données de gaz. Le compteur communicant Linky offre les mêmes services que Gazpar, à savoir la connexion à une API et la possibilité d'avoir un suivi de consommation quasi instantané (30 min)^{23 24}, ce qui n'est pas idéal pour notre besoin de suivi temps réel. Une autre façon de collecter les données

est de le faire via l'espace client, mais la maille la plus fine proposée est journalière et il faut compter entre 2 à 4 jours entre le moment de consommation et la disponibilité des données. Des fournisseurs comme TotalEnergies donnent la possibilité d'avoir une consommation en temps réel (taux de rafraichissement ~ 1 seconde), en souscrivant au service payant Conso Live(2 euros/mois). Ce service s'active par l'insertion d'une clé qui est fournie, directement sur le compteur Linky²⁵. La clé communiquera ensuite les informations de consommation en temps réel sur l'application smartphone via une connexion wifi.



Figure 13 : Principe de fonctionnement du service conso live.

(Source : www.totalenergies.fr)

La principale contrainte liée à ce service est qu'il est payant. L'alternative la plus intéressante est la connexion à l'API Enedis (maille de 30 min). Dans les deux cas il faudra un accord explicite de l'utilisateur.

- Construction, traitement des Déchets et gestion eau : dans ces sous-catégories, les infrastructures sont partagées ou mutualisées, il est donc difficile d'avoir une répartition individuelle. La manière la plus simple de traiter ce problème est de considérer les émissions associées à ces sous-catégories comme une constante et de diviser cette constante par le pas de temps sur lequel est calculé le bilan carbone pour avoir une constante par unité de temps. Cette constante par unité de temps sera automatique, rajouté au calcul final de l'empreinte carbone sur le pas de temps choisi. Par exemple, si on veut calculer l'émission carbone de ces 2 sous-catégories sur une minute, pour une personne vivant en France, on peut partir des données d'émissions moyennes et faire le calcul suivant :

$$\begin{aligned} \text{Émissions (Construct° + Déchets/Eau)} &= (440 + 120) \text{ kgCO}_2\text{eq/personne/an} \\ &= 560 \text{ kgCO}_2\text{eq/personne/an} \end{aligned}$$

$$\begin{aligned}
&= \frac{560}{525600} \text{ kgCO}_2\text{eq/personne/minute (car 1 an = 525600 minutes)} \\
&= 1,065 \cdot 10^{-3} \text{ kgCO}_2\text{eq/personne/minute} \\
&= 1,065 \text{ gCO}_2\text{eq/personne/minute}
\end{aligned}$$

Ce mode de calcul n'est pas forcément idéal, mais c'est un bon compromis quand on manque de données par individu. Notons aussi que bénéficier d'un calcul plus individualisé n'apportera pas forcément de la valeur puisque les individus ont très peu de moyen d'action directe sur ces sous-catégories qui sont dans le périmètre d'action des collectivités. On peut tout de même admettre qu'il y a des leviers d'actions indirects comme la réduction de la consommation d'eau et la réduction de la production des déchets, qui peuvent entraîner une baisse des émissions liées à la gestion de ces matières s'il y a suffisamment de personnes impliquées. Mais attention à ne pas compter les économies en double puisque nous aborderons à la section suivante les catégories alimentation et achats dont la réduction de la consommation entraîne, toute chose étant égale par ailleurs, une réduction des déchets.

6.1.3. Alimentation

L'alimentation a un poids non négligeable dans le bilan carbone des Français. Avec une émission moyenne de 2350 kgCO₂eq/an, c'est la deuxième catégorie la plus émissive après les transports. L'alimentation représente 23,7% du total du bilan carbone. Les aliments les plus émissifs sont les protéines d'origine animale avec en tête la viande rouge (Figure 14).

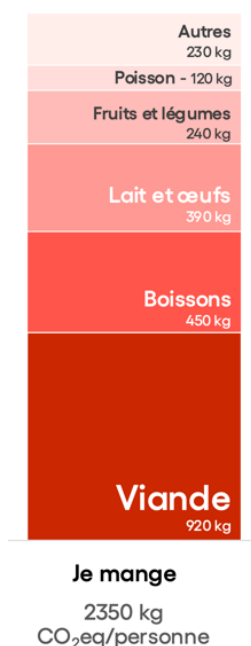


Figure 14 : Émissions moyennes liées à l'alimentation. (Source : carbone4.com)

En plus d'être plus émissives dans leur production (besoin important en eau et en céréales²⁶), les protéines d'origine animales sont les plus consommées en France. Un Français mange deux fois de protéine animale que la recommander par l'OMS²⁷ :

	Protéines (total)	Protéines d'origine animale	Protéines d'origine végétale
Consommation moyenne d'un Français (INRAE, 2019b)	90 grammes / jour	60 grammes / jour	30 grammes / jour
Recommandations nutritionnelles de l'OMS pour un adulte en bonne santé (2011)	50 à 70 grammes / jour	25 à 35 grammes / jour	25 à 35 grammes / jour

Tableau 5 : Consommation de protéine des Français et recommandations de l'OMS

(Source : ecotoxicologie.fr)

En plus du facteur carbone, il y a des enjeux de santé publique, car selon l'OMS, une consommation excessive de viande rouge et transformé est associée à un risque plus élevé de cancer²⁸.

Même si le thème de la santé n'est pas dans le scope de la thèse, il est important d'associer les indicateurs d'émission carbone avec ceux de la santé dans le cadre de pour mieux convaincre les individus à initier des changements des comportements.

La récupération automatique des données pertinentes pouvant permettre de calculer le bilan carbone d'un aliment est une tâche ardue. En effet, ce bilan va dépendre de plusieurs données qui sont difficiles à récupérer, soit parce qu'elles ne sont pas communiquées par les acteurs du domaine, soit par qu'elles n'existent pas. Parmi ces facteurs d'influence, on peut citer :

- Le pays d'origine
- Les moyens de transports utilisés
- La supplychain (les entrepôts intermédiaires, les centres de distribution)
- Le type d'emballage du produit
- Le mode de production (bio, label rouge, sans label)

Ces facteurs influent énormément sur le bilan carbone de l'aliment. Pour donner un exemple, il y a un facteur 12 entre la viande de bœuf qui pollue le plus (105 kgCO₂eq) et celle qui pollue le moins (9,1 kgCO₂eq)²⁹.

Un moyen de récupérer ces informations est de créer une base de données communautaire, sur le même modèle que l'application Yuka³⁰. Notons cependant que même avec ce mode de collecte, il sera difficulté de collecter avec précision certaines données comme le mode d'acheminement. On peut toujours tenter de contourner ce problème en faisant des hypothèses sur le moyen de transport en fonction de la provenance de l'aliment. Par exemple, si un aliment vient d'un autre pays européen, on peut supposer qu'il a été transporté uniquement dans

des camions frigorifiques, de même si cet aliment vient d'Afrique du Sud, on peut faire l'hypothèse qu'il y a eu du transport en bateau puis en camion selon la destination finale.

L'éco-score est une de ces initiatives communautaires ayant pour but de créer l'équivalent du nutri-score pour l'impact carbone. C'est l'indicateur le plus connu et celui qui est la plus adoptée par les plus grands distributeurs. Cet indicateur a été créé par un ensemble d'acteurs engagés comme Yuka ou Scan'Up. On peut lire sur le site que Scan'Up que : « l'éco-Score est calculé sur la base de l'Analyse du Cycle de Vie (ACV) du produit issu de la base de données Agribalyse de l'ADEME. Ce score est ensuite pondéré selon des indicateurs qualitatifs complémentaires (mode de production, origine des matières premières, circularité de l'emballage, politique environnemental du pays producteur, etc.) ». La documentation officielle donne tous les détails du calcul³¹. Ci-dessous une vue imagée du processus de calcul.

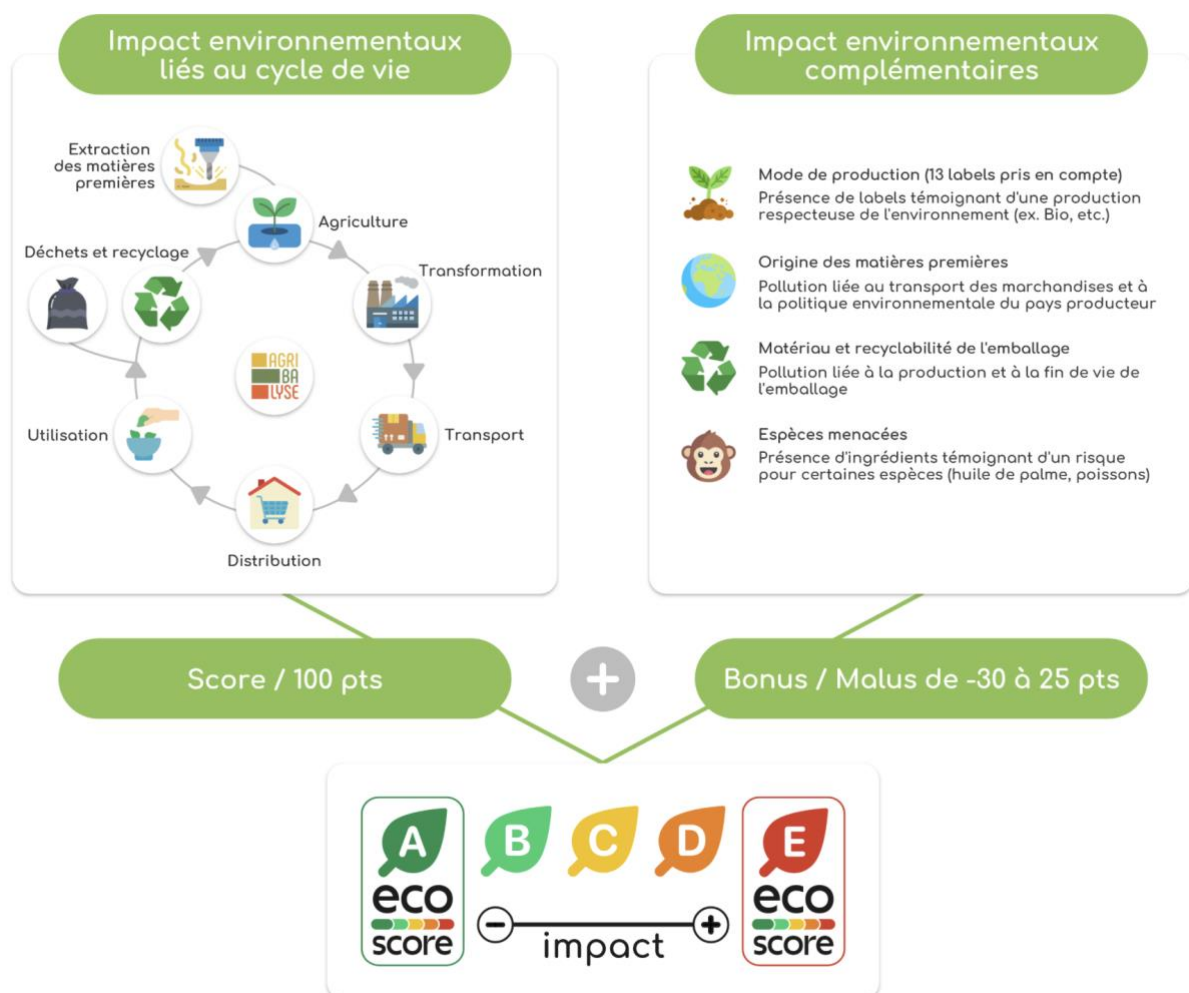


Figure 15 : Aperçu de méthode de calcul de l'éco-score

(Source : docs.score-environnemental.com)

L'éco-score, comme le nutri-score a vocation à s'étendre à la majorité des aliments qu'on trouve en rayon. En janvier 2022, 1 an après son lancement, déjà 400.000 produits et 100.000 recettes affichaient un score. C'est donc un indicateur sur lequel on peut se projeter et se baser pour faire l'estimation de l'impact carbone d'un aliment.

Le protocole de collecte dépendra de l'endroit où l'utilisateur fait ses courses (grandes surfaces ou petits magasins) et de la possession ou non d'une carte de fidélité.

Si le client fait ses courses dans une grande surface et a une carte fidélité, il est possible d'avoir accès aux tickets de caisse via l'espace client. On pourra alors entraîner un algorithme qui récupèrera automatiquement les références produit et la quantité associée. Enfin, on fait du scrapping sur le site du commerçant pour retrouver l'article via la référence. Si l'article a un éco-score, on le récupère et on le multiplie par la quantité. S'il n'y a pas d'éco-score, on peut se baser sur les tables de correspondance d'émission moyenne de l'ADEME.

Si le client fait ses courses en dehors des grandes surfaces et/ou pour les rares clients n'ayant pas de carte fidélité (moins de 5% par enseigne³²), on peut mettre en place un système de reconnaissance optique de caractère (Optical Character Recognition – OCR) pour digitaliser un ticket de caisse papiers et extraire dans un second temps les noms des aliments. Ensuite, on pourra utiliser les tables de correspondance fournies par l'ADEME pour avoir l'empreinte carbone moyenne associée à chaque aliment identifié avec éventuellement une confirmation manuelle de l'utilisateur en cas de lecture difficile sur certaines lignes.

Notons que tous les points précédents ne concernent que des aliments unitaires. Que faire pour les repas achetés au restaurant ? le plus simple dans le cas est d'utiliser des algorithmes de computer vision pour automatiquement détecter le plat puis récupérer dans la base de données de l'ADEME les données d'émission moyenne pour chacun des principaux ingrédients identifiés. Il existe déjà une multitude d'applications qui permettent de détecter avec une bonne précision les ingrédients d'un plat sur une photo. La technologie existe et est suffisamment mature.

6.1.4. Achats

La catégorie j'achète comprends toutes les activités liées à des achats de biens ou de services. Par exemple l'achat de meubles ou objets de décoration pour la maison, les travaux, l'achat de vêtements, etc. En France, c'est l'avant-dernière catégorie en termes d'émissions juste avant les dépenses publiques. Pourtant, elle mérite qu'on y prête une attention particulière, car elle est sur-représentée chez les jeunes. Les achats de vêtement sur le site de vente en ligne Shein représenteraient 22% du bilan carbone des adolescentes françaises³³.

Les problématiques liées à la catégorie achats sont les mêmes que celle de la catégorie alimentation. Comme pour l'alimentation, des initiatives sont en cours pour mettre en place un éco-score pour les vêtements, les chaussures³⁴ et même les meubles³⁵.

6.1.5. Dépenses publiques

La catégorie dépense publique regroupe toutes les émissions concernant les dépenses de l'état, ramené à chaque Français sur une année civile, comme l'administration (les bâtiments publics et leurs usages), les services publics, les infrastructures. Ces données sont mutualisées et il est difficile de faire mieux qu'une moyenne. Donc, similairement aux sous-catégories Construction, traitement des Déchets et gestion de l'eau dans la catégorie logement, nous allons considérer ces émissions comme des constantes et les ramener à l'échelle de temps qui nous intéressera au moment du calcul de l'impact carbone.

6.2. Communication plus efficace

À ce stade, il est clair que les changements climatiques constituent une « tempête cognitive parfaite » défiant leur résolution. Il n'existe donc pas de raccourci vers l'action climatique. La communication sur le changement climatique a un rôle majeur à jouer dans la préparation à une telle action.

Fondamentalement, la communication sur le climat consiste à informer, éduquer, guider les gens à initier des actions pour sortir de la crise climatique. Néanmoins, à un niveau plus profond, la communication sur le climat est façonnée par nos différentes expériences, nos modèles mentaux et culturels, nos valeurs sous-jacentes et nos visions du monde. La façon dont nous parlons du changement climatique, que ce soit en tant que défenseurs, scientifiques ou citoyens ordinaires, affecte la façon dont les gens comprennent et réfléchissent à la fois au problème et aux solutions.^{36 37}

Par conséquent, le même discours sur l'urgence climatique qu'utilisent les grands acteurs et décideurs ne suffit pas pour impliquer suffisamment le grand public. Cette formulation singulière peut même provoquer un sentiment d'impuissance ou de déconnexion de la part de ces derniers. Il ne suffit plus de donner raison aux scénarios les plus pessimistes possibles afin de susciter l'inquiétude du public, mais aussi à inciter les gens à agir de leur bon vouloir.

Le problème pointé, ici, est la pauvreté des discours - le changement climatique est abordé presque toujours de la même manière. L'objectif de la communication sur le climat est de passer de la communication de nature menaçante du changement climatique à des histoires d'action.

Les histoires sur le changement climatique proviendront de conversations que nous avons avec nos amis, notre famille et nos collègues - nous devons tous devenir des multiplicateurs de ce message. Et pour ce faire, nous devons réfléchir profondément à cette question et comprendre comment cela nous affectera. Nous devons donc réfléchir de manière critique au type d'histoire que nous racontons. Racontons-nous une histoire de désespoir et de soi-disant inévitabilité ? Racontons-nous une histoire sur les émissions négatives, les boucles de rétroaction, les points de basculement ou d'autres choses que la plupart des gens ne comprennent pas du tout ? Où racontons-nous une histoire d'espoir, de possibilité et d'action humaine ? Plus important encore, indiquons-nous clairement lorsque nous parlons de cela

qu'il y a des choix que nous pouvons faire, qui détermineront le genre d'avenir que nous allons avoir.

C'est le combat qu'a entrepris le *Climate Action Unit* (CAU) qui est un groupe d'experts en neurosciences, psychologie, communication scientifique et facilitation d'experts. L'objectif du CAU est d'accélérer l'adoption d'actions climatiques et la transition vers une consommation nette zéro dans la société. Le CAU est né de la Commission de politique de communication de la science climatique de l'UCL, dont les membres ont mis en œuvre des projets visant à éliminer les obstacles à une communication efficace.³⁸

**Entendre parler du
réchauffement climatique dans
les médias au moins une fois
par semaine**

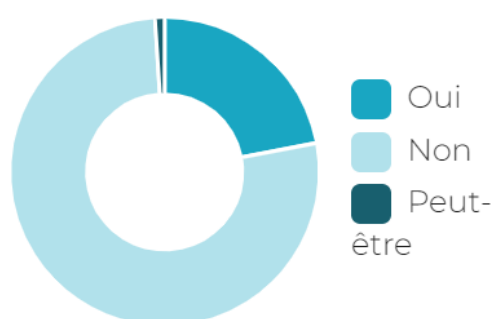


Figure 16 : Sondage d'opinions sur le changement climatique.

(Source : climate-xchange.org/communicating-the-climate-crisis)

Une bonne pratique en matière de communication sur le climat est certes d'utiliser des informations scientifiques faisant autorité. Cependant, les communicants doivent arriver à porter leur message au public via un langage fluide. Employer des termes et des codes scientifiques peut créer de la confusion et conduit souvent à des retards d'action ou même à des actions nuisibles. Il est peu utile, à titre d'exemple, de parler de g/CO₂ puisque cela ne parle à tout le monde. À l'inverse, transmettre un message d'espoir axé sur les solutions, montrer l'impact en des termes qui parlent à la personne en fonction de ses caractéristiques (âge, localisation, etc.) est un bon moyen de contourner la « fatigue de crise » en aidant les gens à se sentir concernés, responsabilisés et motivés à s'engager. Par exemple, en racontant une histoire locale et personnelle.

Les histoires individuelles peuvent forger un lien émotionnel, amener le public à s'intéresser et rendre les défis mondiaux partagés moins intimidants : « j'habite en France, près de la cote. Le niveau de la mer a augmenté de 30 cm ces dernières années. Face à la montée des eaux, il faut déjà se préparer à quitter le bord de mer. Et si tout le monde continuait à vivre sans se soucier de la planète, vivait comme vous, le niveau de la mer monterait de 2,5 m en 2100 et nos descendants ne connaîtront pas la Bretagne. » /

De plus, les individus utilisent leurs points de vue culturels et politiques pour filtrer les informations sur le changement climatique, ce qui leur permet de peser les risques du changement climatique et les solutions disponibles. Il y a une dépendance excessive à un flux linéaire d'informations où une approche « problème-solution » est adoptée et où l'on suppose que fournir des informations sur le changement climatique n'est donc pas suffisant pour aider à réduire les émissions par le changement de comportement. De ce fait, les experts s'intéressent de plus en plus à la segmentation du public en tant que stratégie possible pour développer des communications plus efficaces, adaptées et ciblées sur des sous-groupes du public qui partagent des valeurs, des croyances, des comportements et/ou des préférences politiques similaires en matière de changement climatique.³⁹

6.3. Personnalisation dans la proposition d'alternative

Un autre avantage du bilan carbone automatisé est qu'il offre la possibilité de faire des propositions d'alternatives personnalisées par rapport aux informations que l'utilisateur aura accepté de partager dans l'application. En fonction des informations que l'utilisateur partagera avec le service, des alternatives hautement personnalisées. Les informations comme la localisation, les caractéristiques socio-démographiques, les historiques d'achats sont autant de sources sur lesquels sont peu s'appuyer pour proposer des bonnes alternatives pertinentes au bon moment. Passons en revue quelques alternatives en fonction de la catégorie dans laquelle des émissions sont détectées.

Illustrons ce point sur la catégorie transport : mettons-nous dans le cadre d'un utilisateur dont le service détecte un trajet en voiture d'un point A à un point B. Li proposer systématiquement comme alternative de prendre les transports en commun n'est pas la solution optimale puisque l'utilisateur risque de trouver l'application trop intrusive et trop mécanique dans ses propositions. Une solution plus nuancée serait de comparer le trajet en voiture à un trajet en transports en commun/vélo/marche et de ne proposer d'alternatives que si la différence de temps de trajet entre la voiture et le mode alternatif est inférieur à un seuil soigneusement choisi (par exemple 10%).

Il ne suffit pas de passer le bon message, encore faut-il le délivrer au bon moment. Dans ce cas, le moment idéal pour déclencher le message de proposition d'alternative doit se situer suffisamment tôt pour que l'usage puisse passer à l'alternative proposée au cas où il la trouverait pertinente. Dans cet exemple, on pourra par exemple identifier des trajets récurrents fait à des moment précis et proposer une alternative 30 min avant le départ. On peut aussi créer de tunnels de connexion avec l'application Waze, Maps ou Plans installés sur le téléphone de l'utilisateur et anticiper des départs. Par exemple si l'utilisateur regarde un trajet sur Waze et sort de son appartement, on peut en déduire avec l'aide de son historique de données qu'il va prendre la voiture. Ensuite on comparera le trajet en voiture avec un mode alternatif qui ne sera proposé que si la différence est inférieure au seuil fixé.

1. Limites et challenges

1.1. Biais cognitifs empêchant les individus d'agir face au réchauffement climatique.

Les gens se soucient du changement climatique. Une série d'enquêtes récentes a montré que le changement climatique est le problème le plus important que la plupart des personnes pensent que la lutte contre le changement climatique nécessite un niveau d'urgence élevé ou extrêmement élevé (enquête commandée par le UK Centre for Climate Change and Social Transformation en août 2019). Les données montrent que les gens sont plus inquiets maintenant qu'ils ne l'étaient il y a 3 ans, et la plupart admettent que leur inquiétude a considérablement augmenté au cours de la dernière année en raison de l'augmentation des conditions météorologiques extrêmes, des reportages dans les médias et de l'activisme climatique

Pourtant, bien qu'il ait un consentement scientifique sur la réalité et la gravité du changement climatique, les principaux acteurs - grand public, politiques et entreprises - ne montrent pas d'intérêt particulier pour passer aux actes, du moins peu d'entre eux s'engagent effectivement dans la lutte pour le climat.

Les barrières structurelles constituent une partie de la raison de l'indifférence des acteurs, mais leurs choix comportementaux sont aussi entravés par des biais cognitifs. Par conséquent, de tous les nombreux individus engagés dans des actions d'amélioration, la plupart d'entre eux pourraient le faire davantage, mais sont freinés par des biais cognitifs. C'est ce que Gifford R. (2011) appelle *les dragons de l'inaction* dans son livre *Psychological barriers that limit climate change mitigation and adaptation*.

Qu'est-ce donc un biais cognitif ? Essentiellement, Un biais cognitif est un schéma de pensée trompeur et faussement logique. Ce schéma de pensée permet à l'individu d'émettre un jugement, ou de rapidement décider. Les biais cognitifs ont une grande influence dans nos choix, en particulier lorsque la quantité d'information à gérer est importante ou que le temps est limité. Ainsi, une sorte de dysfonctionnement se produit dans le raisonnement ⁴⁰. Les *dragons de l'inaction* ou biais cognitifs se répartissent, d'après Gifford R., en sept catégories : la cognition limitée, l'idéologie, la comparaison aux autres, les coûts irrécupérables, l'incrédulité, la perception du risque et les changements dont l'impact est limité.

Concrètement, le biais proximité (la difficulté à anticiper à très long terme), qui est l'essence même de la lutte pour le climat, pourrait nous aider à comprendre des biais. Il est difficile pour nous, humains, de se projeter sur un très long terme, car le cerveau est fait pour prendre en compte l'immédiat et non le futur lointain. L'expérience de la grenouille qui bout lentement montre que, lorsqu'un changement s'effectue d'une manière suffisamment lente, il échappe à la conscience et ne suscite la plupart de temps aucune réaction, aucune opposition, aucune révolte. C'est pourquoi agir sur le changement climatique implique de compromettre les avantages à court terme au profit des avantages à long terme, et

ce compromis est parmi les plus difficiles à faire pour les gens. Nous aurons donc plus tendance à réagir face aux problèmes directs provoqués par l'activité humaine telle que la hausse des températures, la hausse du niveau de la mer ou encore le recul et la fonte des glaciers. Parallèlement, le scénario élaboré par les scientifiques au-delà de l'horizon 2100 ne provoqueront pas ou peu d'engouement. Nous sommes donc face à un problème dont le caractère transgénérationnel - les générations futures sont celles qui subiront le plus le changement climatique si aucune action n'est prise par les générations actuelles - vient renforcer l'inaction, voire le déni climatique auquel nous faisons face aujourd'hui.

Un autre défi cognitif qui explique l'inaction face à l'urgence climatique est le biais de confirmation qui désigne la tendance à sélectionner uniquement les informations qui confirment des croyances ou des idées préexistantes. C'est le Parrain de tous les biais cognitifs. Il sera encore plus prononcé dans des contextes idéologiques, politiques ou les contextes sociaux chargés d'émotions.

Gilbert D., auteur de *Et si le bonheur vous tombait dessus*, nous fait cette illustration du biais : « quand la balance de notre salle de bains nous indique la mauvaise nouvelle, nous descendons et remontons immédiatement, juste pour être certain que nous n'avons pas mal lu l'écran ou mis trop de pression sur un pied. Mais quand elle indique une perte de poids (le Saint Graal !), ça nous met de bonne humeur pour le reste de la journée. En acceptant d'office les preuves quand cela nous arrange, et en insistant plus dans le cas contraire, on fait subtilement pencher la balance en notre faveur. »

Bien que le changement climatique soit scientifiquement documenté, il existe encore bon nombre de personnes qui n'y croient tout simplement pas. Ces personnes s'efforceront de chercher toutes les informations qui prouvent que les scientifiques se trompent et que l'urgence climatique est un vaste complot. En persévérant dans leurs recherches, ils renforceront l'idée selon laquelle le réchauffement de la planète n'est pas réel, ce qui est chose facile via GOOGLE, car il est relativement facile de rechercher et vérifier même les arguments les plus absurdes possibles.

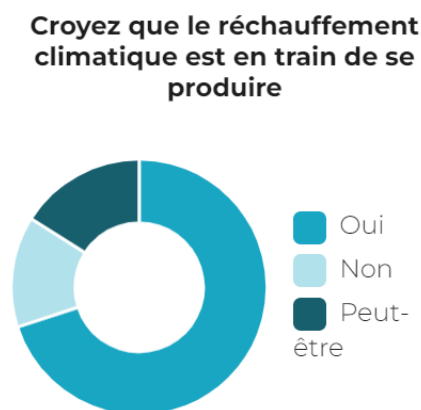


Figure 17 : Sondage du Yale Program on Climate Change Communication

(Source : climate-xchange.org/communicating-the-climate-crisis)

1.2. Le Machine Learning a aussi son empreinte carbone

En tenant compte du fait que l'IA a un bilan carbone, il va de soi qu'il faut chercher à optimiser les systèmes afin de se cantonner à des modèles peu énergivores. Plusieurs aspects sont à prendre en compte pour estimer la quantité de carbone que l'IA émet. Ces facteurs peuvent inclure : l'emplacement du serveur utilisé pour former le modèle et le réseau électrique qu'il utilise ; la durée de la procédure d'apprentissage ; et même la marque et le modèle du matériel sur lequel l'apprentissage a lieu.

*Quantifying the Carbon Emissions of Machine Learning*⁴¹ d'Alexandre Lacoste et al. est un calculateur d'émission de l'IA permettant aux communautés de mieux comprendre l'impact environnemental des modèles d'apprentissage automatique. Cet outil est accompagné d'une explication des facteurs cités ci-dessus, ainsi que des actions concrètes que les praticiens individuels et les organisations peuvent prendre pour atténuer leurs émissions de carbone. Une bonne base, certes, pour voir une idée de l'empreinte de la technologie et, plus important encore, de garder à l'esprit que ces modèles peuvent être énergivores afin de viser l'utilisation de modèles plus respectueux de l'environnement en termes d'empreinte carbone.

ML CO2 Impact

Compute Learn Act About Authors

Machine Learning Emissions Calculator

Choose your hardware, runtime and cloud provider to estimate the carbon impact of your research.

This calculator will give you 2 numbers: the **raw** carbon emissions produced and the approximate **offset** carbon emissions. The latter number depends on the grid used by the cloud provider and we are open to update our estimates if anything looks inaccurate or outdated.

Hardware type: Titan V

Hours Used: 48

Cloud Provider: Google Cloud Platform

Region of Compute: asia-east1

COMPUTE

CARBON EMITTED: 6.72 kg CO₂ eq.

CARBON ALREADY OFFSET BY PROVIDER: 6.72 kg CO₂ eq.

Power consumption x Time x Carbon Produced Based on the Local Power Grid:
250W x 48h = 12 kWh x 0.56 kg eq. CO₂/kWh = 6.72 kg eq. CO₂

Had this model been run in Google Cloud Platform's europe-west6 region, the carbon emitted would have been of 0.19 kg eq. CO₂

WHAT CAN I DO?

Figure 18 : D  mo de l'outil de calcul des   missions par Alexandre Lacoste et al.

(Source : github.com/mlco2/impact)

Toutefois, il est important de souligner qu'une estimation hautement précise de l'empreinte carbone de l'IA peut s'avérer être une tâche plus ou moins difficile en fonction de la complexité des différents facteurs inhérents à son fonctionnement.

Rappelons que les facteurs à prendre en compte pour estimer la quantité de carbone que l'IA émet sont : l'emplacement du serveur utilisé pour former le modèle et le réseau électrique qu'il utilise ; la durée de la procédure d'apprentissage ; et même la marque et le modèle du matériel sur lequel l'apprentissage a lieu. Certains paramètres, à l'instar des deux premiers cités, sont problématiques et ne peuvent tout simplement pas être précisément estimés à ce jour.

- L'emplacement des serveurs : Il est difficile, voire impossible à ce jour d'estimer avec une bonne précision ce facteur. En effet, la plupart des entreprises utilisent le cloud pour leur stockage. Il existe néanmoins quelques-unes qui le font localement, mais elle se compte du bout des doigts. La complexité de cette tâche réside donc dans le fait que la localisation des serveurs n'est pas connue. Ce qui fausse la précision de l'estimation, car la dépense énergétique ne sera pas la même en fonction que les serveurs soient situés en France, aux USA ou à Singapour.
- Le réseau électrique : Là encore, plusieurs paramètres sont à considérer comme la plage horaire d'utilisation (le moment de la journée pendant lequel le modèle a été entraîné), la position géographique (le pays /région/localité dans lequel le système se trouve physiquement) ou encore la saison (certains setups consomment plus en période de chaleur à cause de leur système de refroidissement). Par exemple, en France, le mix énergétique fait que les différentes sources d'énergie sont réparties entre énergies fossiles (pétrole, gaz naturel, charbon), énergies nucléaires et énergies renouvelables⁴². En raison de ce mix énergétique, la source d'énergie et donc l'empreinte écologique varient. Et cela, sans compter le fait que l'électricité produite et consommée est assurée, elle-même, par différentes sources d'énergie.

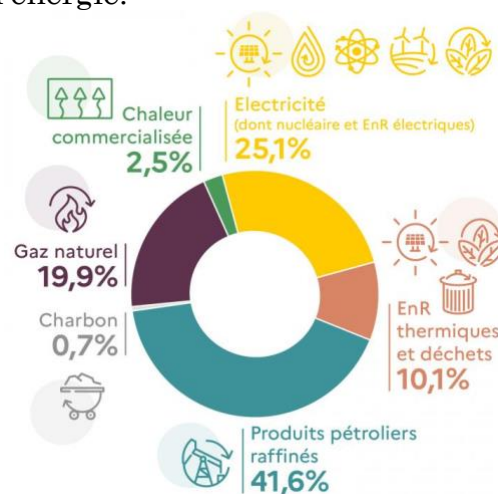


Figure 19 : Mix énergétique en France en 2022

(Source : ecologie.gouv.fr/transition-energetique-en-France)

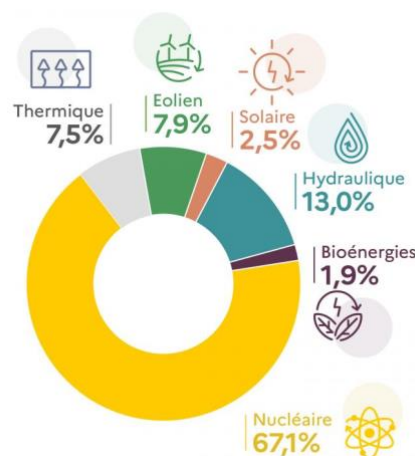


Figure 20 : Mix électrique en France en 2022

(Source : ecologie.gouv.fr/transition-energetique-en-France)

1.3. RGPD et utilisation des données

Face à l'enjeu que représente l'IA et la lutte pour le climat pour notre futur, la réaffirmation des libertés individuelles et en particulier la mise en valeur du RGPD au service de ces valeurs doit être au cœur de tout. Ainsi dit, la dimension collective des risques engendrés par l'IA, l'absence de transparence des systèmes d'apprentissage automatique, les difficultés de leur évaluation sont toutes des questions qui méritent d'être soulevées et résolues, car oui, l'IA peut nous aider à réduire notre impact, mais seulement 'si et seulement si' cela est conforme au RGPD.

En règle générale, le RGPD régleme déjà largement l'IA. Le cadre juridique actuel aborde bon nombre des risques et défis potentiels associés au traitement des données à caractère personnel par des algorithmes"

Tout comme le calculateur pour les émissions des modèles d'apprentissage automatique, la CNIL⁴³ a mis en place une grille d'auto-évaluation pour les systèmes d'intelligence artificielle ayant pour objectif de permettre aux organismes sondés par eux-mêmes la maturité de leurs dispositifs d'intelligence artificielle au regard du RGPD. Ce guide est également accompagné de bonnes pratiques en matière de protection de données.

Notons que même après avoir respecté tous les standards en matière de réglementation, la décision finale de partage des données reviendra à l'individu. Sarigol et al. suggèrent que la motivation d'un utilisateur de service à partager ses données sur ce service dépend non seulement de la valeur apportée par le service rendu, mais aussi des interactions avec les données des utilisateurs⁴⁴. On peut alors inclure dans l'application des indicateurs de comparaison entre un utilisateur et les utilisateurs qui lui sont proche (par exemple en termes de tranche d'âge ou de localisation).

2. Cas pratique : calcul automatique de l'empreinte carbone du transport

Dans cette section, on mettra en pratique certains concepts discuté dans les chapitres précédents. L'objectif ici est de montrer qu'il est bien possible d'automatiser le calcul de l'empreinte carbone en recueillant les bonnes données et en utilisant des algorithmes de Machine Learning. Nous nous focaliserons sur la catégorie transport. Les codes python sont disponibles sur mon github via le lien : www.github.com/Omarlaouan/Transportation-automatic-carbon-footprint-calculator

La méthodologie globale pour mener ce cas pratique va être la suivante :

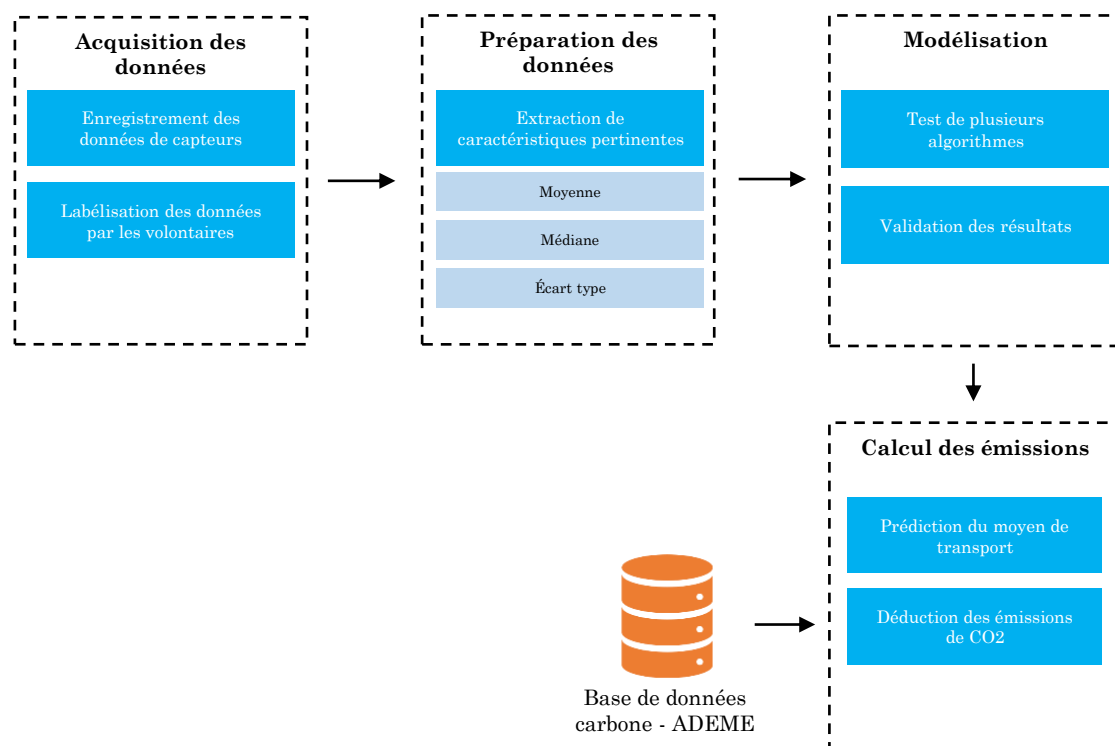


Figure 21 : Méthodologie globale du cas pratique

2.1.Acquisition et préparation des données

Les smartphones sont aujourd'hui quasiment tous équipés de capteurs qui en font des IOTs pouvant fournir des données précises pour la détection automatique du mode de transport. Les données que nous allons utiliser pour ce cas pratique ont été mises à disposition sur la plateforme Kaggle⁴⁵. Ces données ont été collectées par Di Felice et al. de l'université de Bologne via une application mobile qui récupère les informations de capteurs des smartphones de volontaires. Ensuite, l'utilisateur renseigne son mode de transport et une jointure est faite pour avoir une seule table contenant les données de capteurs et les étiquettes associées (immobile, marche, voiture, train). On obtient ainsi des données labellisées sur lesquelles on peut entraîner des algorithmes d'apprentissage supervisé. Ci-dessous une vue schématique du processus de collecte. Durant la collecte, une attention particulière a été portée sur la répartition des différents modes de

transport : Les données sont équilibrées du point de vue de la variable à prédire pour réduire le biais d'apprentissage qui peut être induit par la sur-représentation de certaines classes au détriment d'autres.

Still	Walking	Car	Bus	Train	Total
1179	1177	1180	1178	1179	5893

2.2.Préparation des données (Feature engineering)

Les données brutes à notre disposition sont les amplitudes des signaux de plusieurs capteurs. Par soucis de simplification, on va se restreindre à 3 capteurs (accéléromètre, gyroscope et le son en fonction du temps). Pour aider les algorithmes à trouver la règle d'association entre les features et les labels, il peut être intéressant de créer des variables à partir des observations à notre disposition. Les chercheurs proposent comme transformations :

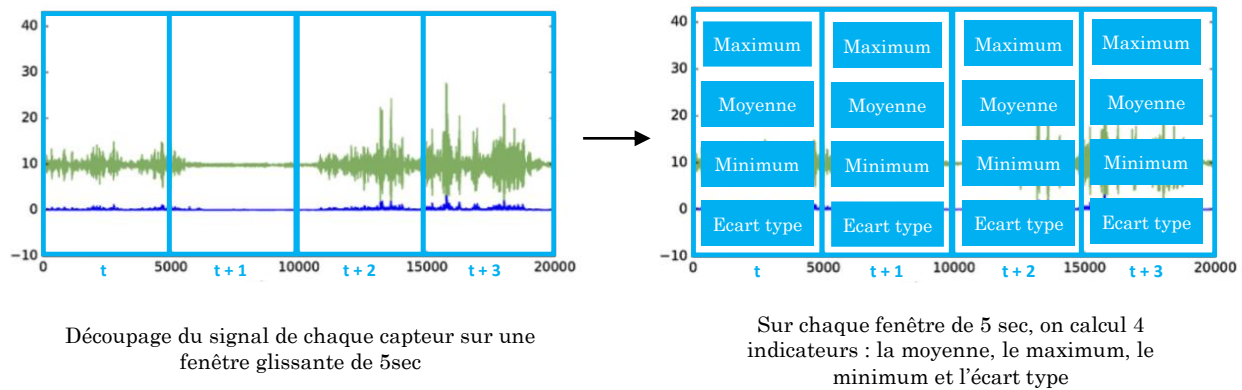


Figure 22 : création de nouvelles caractéristiques (features)

Après transformation et jointure avec les étiquettes encodées, les données auront la forme suivante :

timestamp	acc#mean	acc#min	acc#max	acc#std	gyro#mean	gyro#min	gyro#max	...	target
78.0	9.8114	9.7588	9.8494	0.0146	0.0016	0.0	0.0035	...	1
145.0	9.9392	7.7074	17.14663	1.7759	0.0363	0.0116	0.0593	...	2
...

Tableau 6 : Extrait des données de capteurs après transformation

Règle d'encodage de la target : 1=immobile, 2=marche, 3=voiture, 4=bus, 5=train

2.3. Modélisation

- Choix des algorithmes :

Comme tenu de la taille des données (5893 exemples) et de leur natures (données structurés), on se privera d'utiliser des algorithmes complexes du type Deep Learning. Nous allons tester quelques algorithmes classiques de classification supervisée : Les arbres de décision (Decision Trees), les forêts aléatoires (Random Forest) et la régression logistique (logistic regression).

- Résultats de la modélisation

Avant d'entraîner les algorithmes, il convient de faire une séparation des données en deux parties : un ensemble d'entraînement et un ensemble de test. Ainsi on pourra tester nos algorithmes sur des données qui leur sont inconnues.

Après entraînement des algorithmes sur l'ensemble d'entraînement, le Random Forest est l'algorithme ayant les meilleures performances sur l'ensemble de test :

	precision	recall	f1-score	support
1	0.9256198347107438	0.9218106995884774	0.9237113402061856	243.0
2	0.9227467811158798	0.9641255605381166	0.9429824561403508	223.0
3	0.8436018957345972	0.7447698744769874	0.7911111111111111	239.0
4	0.8513513513513513	0.8289473684210527	0.8400000000000001	228.0
5	0.8228782287822878	0.9065040650406504	0.8626692456479691	246.0
accuracy	0.8727735368956743	0.8727735368956743	0.8727735368956743	0.872
macro avg	0.8732396183389721	0.873231513613057	0.8720948306211233	1179.0
weighted avg	0.872650684896244	0.8727735368956743	0.8715514277984369	1179.0

Tableau 7 : Résultats de test - Random Forest

On peut noter que toute classe confondue et quel que soit la métrique choisie, le Random Forest a une performance supérieure à 87%. La classe 1 est celle ayant la meilleure performance, c'est-à-dire que le modèle a le plus de facilité à prédire qu'une personne est immobile. Ce n'est pas étonnant puisqu'en immobilité les valeurs sont proches de zéro, il est donc plus facile pour l'algorithme de trouver la règle d'association de cette classe.

2.4. Calcul des émissions

Pour faire le calcul des émissions, on peut se placer dans la cadre où l'algorithme est déployé et utilisé par un individu. Supposons que l'algorithme Random Forest détecte (avec une précision théorique de 100%) que l'utilisateur est en train et ce pendant 37 min avec une vitesse moyenne de 87 km/h. La prochaine étape consiste à récupérer dans la base de données de l'ADEME les données d'émissions moyennes du train, soit 7kgCO₂eq/km parcourue. On pourra alors calculer l'empreinte carbone comme suit :

$$\begin{aligned}\text{Émissions [trajet train]} &= (\text{distance}) \times (\text{émissions par km}) \\ &= (\text{vitesse} \times \text{temps}) \times (\text{émissions par km}) \\ &= (87\text{km/h} \times 0,62\text{h}) \times (7 \text{ kgCO}_2\text{eq/km}) \\ &= 377,6 \text{ kgCO}_2\text{eq pour ce trajet}\end{aligned}$$

Conclusion

On a pu explorer diverses pistes de réflexion sur l'apport de l'IA dans le développement d'un calculateur carbone automatisé et complet. Nous avons montré que des sources de données sont disponibles sur les principaux postes d'émissions d'un individu vivant en France et même quand ces données n'étaient pas disponibles ou pas assez granulaires, on a exposé des méthodes qui permettent de contourner le problème pour calculer des émissions sur de courtes échelles de temps. Les principaux challenges de ce projet résident dans le monitoring des modèles de Machine Learning une fois en production. Plus il y a de modèles, plus il faudra de temps, plus il y a de risque d'avoir des dérives de performances et plus il faut de la ressource humaine pour les monitorer. Une perspective de développement de ce projet serait d'estimer le coût de mise en production (en temps de gestion, en matériel nécessaire et en coût carbone) d'une telle application et de le mettre en perspective avec la valeur ajoutée du service en termes de réduction d'émissions. Pour aller plus loin, on peut aussi développer un premier modèle d'application basé sur une architecture très simple, pour tester la pertinence du service et collecter les avis des utilisateurs en vue d'une amélioration de la proposition de valeur.

Liste des figures et tableaux

Figure 1 : Empreinte carbone moyenne en France en 2019	9
Figure 2 : Programmes traditionnels	10
Figure 3 : l'approche Machine Learning	10
Figure 4 : le Machine Learning nous aide à apprendre	11
Figure 5 : Tendance des mots clés "Machine Learning". Source : Google Trends	13
Figure 6 : récupération automatique de données	19
Figure 7 : Part de marché des systèmes d'exploitation de smartphone	21
Figure 8 : Vue schématique du TMD dataset	22
Figure 9 : Émission carbone moyennes par personne, pour la catégorie logement	23
Figure 10 : Résultats de sondage sur la sobriété énergétique	24
Figure 11 : Sondage sur l'impact perçu de différentes actions pour la protection de l'environnement	24
Figure 12 : Tableau de bord de suivi de consommation de gaz	25
Figure 13 : Principe de fonctionnement du service conso live.	26
Figure 14 : Émissions moyennes liées à l'alimentation. (Source : carbone4.com)	27
Figure 15 : Aperçu de méthode de calcul de l'éco-score	29
Figure 16 : Sondage d'opinions sur le changement climatique.	32
Figure 17 : Sondage du Yale Program on Climate Change Communication	35
Figure 18 : Démo de l'outil de calcul des émissions par Alexandre Lacoste et al.	36
Figure 19 : Mix énergétique en France en 2022	37
Figure 20 : Mix électrique en France en 2022	38
Figure 21 : Méthodologie globale du cas pratique	39
Figure 22 : création de nouvelles caractéristiques (features)	40

Tableau 1 : Principaux type d'algorithmes de Machine Learning	12
Tableau 2 : Principaux type d'algorithmes de Machine Learning	12
Tableau 3 : Émissions moyennes par aliments d'un Français	15
Tableau 4 : Émissions moyennes d'un repas classique avec bœuf pour un Français.	16
Tableau 5 : Consommation de protéine des Français et recommandations de l'OMS	28
Tableau 6 : Extrait des données de capteurs après transformation	40
Tableau 7 : Résultats de test - Random Forest	41

Bibliographie

- ¹ *Reports — IPCC*. (n.d.). Retrieved September 14, 2022, from <https://www.ipcc.ch/reports/>
- ² *AR6 Synthesis Report: Climate Change 2022 — IPCC*. (n.d.). Retrieved November 12, 2022, from <https://www.ipcc.ch/report/sixth-assessment-report-cycle/>
- ³ Gifford, R. (2011). The Dragons of Inaction: Psychological Barriers That Limit Climate Change Mitigation and Adaptation. *American Psychologist*, 66(4), 290–302. <https://doi.org/10.1037/A0023566>
- ⁴ *Changement climatique : 72% de la population mondiale se déclare inquiète, pourtant la mobilisation des citoyens ne progresse pas / Ipsos*. (n.d.). Retrieved November 12, 2022, from <https://www.ipsos.com/fr-fr/obscop-2021>
- ⁵ *Connaissez-vous votre empreinte sur le climat ? - Nos Gestes Climat*. (n.d.). Retrieved November 12, 2022, from <https://nosgestesclimat.fr/>
- ⁶ Nowruzi, F. E., Kapoor, P., Kolhatkar, D., Hassanat, F. al, Laganier, R., & Rebut, J. (2019). *How much real data do we actually need: Analyzing object detection performance using synthetic and real data*. <https://doi.org/10.48550/arxiv.1907.07061>
- ⁷ Cho, J., Lee, K., Shin, E., Choy, G., & Do, S. (2015). *How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?* <https://doi.org/10.48550/arxiv.1511.06348>
- ⁸ Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (n.d.). *A Survey on Bias and Fairness in Machine Learning*. Retrieved November 13, 2022, from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- ⁹ Kapoor, S., & Narayanan, A. (2022). *Leakage and the Reproducibility Crisis in ML-based Science*. <https://doi.org/10.48550/arxiv.2207.07048>
- ¹⁰ Ben-David, A., & Frank, E. (2009). Accuracy of machine learning models versus “hand crafted” expert systems – A credit scoring case study. *Expert Systems with Applications*, 36(3), 5264–5271. <https://doi.org/10.1016/J.ESWA.2008.06.071>
- ¹¹ *Documentation Base Carbone*. (n.d.). Retrieved November 14, 2022, from https://bilans-ges.ademe.fr/documentation/UPLOAD_DOC_FR/index.htm?repas.htm
- ¹² *The Effortless Experience Conquering the New Battleground for Customer Loyalty*. (2013).
- ¹³ *Qui sommes-nous ? - MyCO2 - Comprendre ensemble son empreinte carbone, agir dès demain !* (n.d.). Retrieved November 14, 2022, from <https://www.myco2.fr/qui-sommes-nous>
- ¹⁴ *Population par groupe d'âges - Structure de la population - France - Les chiffres - Ined - Institut national d'études démographiques*. (n.d.). Retrieved November 14, 2022, from <https://www.ined.fr/fr/tout-savoir-population/chiffres/france/structure-population/population-ages/>
- ¹⁵ *Net Zero adds up*. (n.d.). Retrieved November 14, 2022, from <https://www.joro.app/>

¹⁶ *Electricity Maps / Émissions CO₂ de la consommation électrique en temps réel.* (n.d.). Retrieved November 14, 2022, from <https://app.electricitymaps.com/map>

¹⁷ *SensorKit / Apple Developer Documentation.* (n.d.). Retrieved November 15, 2022, from <https://developer.apple.com/documentation/sensorkit>

¹⁸ *Sensors Overview / Android Developers.* (n.d.). Retrieved November 15, 2022, from https://developer.android.com/guide/topics/sensors/sensors_overview

¹⁹ *TMD Dataset - 5 seconds sliding window / Kaggle.* (n.d.). Retrieved November 6, 2022, from <https://www.kaggle.com/datasets/fschwartz/tmd-dataset-5-seconds-sliding-window>

²⁰ *The 5 Clustering Algorithms Data Scientists Need to Know / by George Seif / Towards Data Science.* (n.d.). Retrieved November 15, 2022, from <https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>

²¹ *Découvrez le calendrier du déploiement du compteur Gazpar.* (n.d.). Retrieved November 15, 2022, from <https://www.totalenergies.fr/particuliers/gaz/compteur-gazpar/calendrier-compteur-gazpar>

²² *Accédez à vos données de consommation gaz - GRDF.FR.* (n.d.). Retrieved November 15, 2022, from <https://www.grdf.fr/entreprises/grdf-et-vous-au-quotidien/consommations-gaz/donnees-consommation-gaz-entreprises>

²³ *Data Connect – Enedis DataHub.* (n.d.). Retrieved November 15, 2022, from <https://datahub-enedis.fr/data-connect/>

²⁴ *consometers/data-connect: Notre expérience avec le service Data Connect d'Enedis.* (n.d.). Retrieved November 15, 2022, from <https://github.com/consometers/data-connect>

²⁵ *Connaître ma consommation en temps réel avec Conso Live - TotalEnergies.* (n.d.). Retrieved November 15, 2022, from <https://www.totalenergies.fr/particuliers/parlons-energie/actualites/totalenergies/connaître-ma-consommation-en-temps-reel-avec-conso-live>

²⁶ *Pourquoi la viande est-elle si nocive pour la planète ?* (n.d.). Retrieved November 16, 2022, from https://www.lemonde.fr/les-decodeurs/article/2018/12/11/pourquoi-la-viande-est-elle-si-nocive-pour-la-planete_5395914_4355770.html

²⁷ *Comment réduire l'empreinte carbone de mon alimentation - Ecotoxicologie.fr.* (n.d.). Retrieved November 16, 2022, from <https://ecotoxicologie.fr/empreinte-carbone-alimentation>

²⁸ *Cancer : cancérrogénicité de la consommation de viande rouge et de viande transformée.* (n.d.). Retrieved November 16, 2022, from <https://www.who.int/fr/news-room/questions-and-answers/item/cancer-carcinogenicity-of-the-consumption-of-red-meat-and-processed-meat>

²⁹ Poore, J., & Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. *Science*, 360(6392), 987–992. https://doi.org/10.1126/SCIENCE.AAQ0216/SUPPL_FILE/AAQ0216_DATAS2.XLS

³⁰ *Yuka - L'application mobile qui scanne votre alimentation*. (n.d.). Retrieved November 16, 2022, from <https://yuka.io/>

³¹ *Présentation - Eco-score*. (n.d.). Retrieved November 16, 2022, from <https://docs.score-environnemental.com/>

³² *96% des Clients Principaux Des Grandes Surfaces Ont La Carte de Fidélité | Marketing PGC*. (n.d.). Retrieved November 16, 2022, from <https://www.marketing-pgc.com/2012/11/09/96-des-clients-principaux-des-grandes-surfaces-ont-la-carte-de-fidelite/>

³³ *Shein responsable de 22% du total des émissions CO2 des adolescentes françaises*. (n.d.). Retrieved from <https://www.presse-citron.net/shein-pret-a-porter-ravage-le-bilan-carbone-des-jeunes-francaises/>

³⁴ *Un éco-score pour évaluer l'empreinte environnementale de l'habillement - Environnement Magazine*. (n.d.). Retrieved November 16, 2022, from <https://www.environnement-magazine.fr/cleantech/article/2022/01/06/137611/ecoscore-pour-evaluer-empreinte-environnementale-habillement>

³⁵ *Eco Impact, le premier éco-score du mobilier – TRÈS. écodesign*. (n.d.). Retrieved November 16, 2022, from <https://tres-ecodesign.com/pages/l-co-score>

³⁶ O'Neill, S., & Nicholson-Cole, S. (2009). "fear won't do it": Promoting positive engagement with climate change through visual and iconic representations. *Science Communication*, 30(3), 355–379. <https://doi.org/10.1177/1075547008329201>

³⁷ Barnosky, A. D., Matlock, T., Christensen, J., Han, H., Miles, J., Rice, R. E., Westerling, L., & White, L. (2016). Chapter 9. Establishing Common Ground: Finding Better Ways to Communicate About Climate Disruption. *Collabra*, 2(1). <https://doi.org/10.1525/COLLABRA.68>

³⁸ *Climate Action Unit | Climate Action Unit - UCL – University College London*. (n.d.). Retrieved November 15, 2022, from https://www.ucl.ac.uk/climate-action-unit/sites/climate_action_unit/files/lets_talk_about_how_we_talk_about_climate_change_-_environment_magazine_-_march_2022_issue.pdf

³⁹ Howarth, C., Parsons, L., & Thew, H. (2020). Effectively Communicating Climate Science beyond Academia: Harnessing the Heterogeneity of Climate Knowledge. *One Earth*, 2(4), 320–324. <https://doi.org/10.1016/J.ONEEAR.2020.04.001>

⁴⁰ Haselton, M. G., Nettle, D., & Andrews, P. W. (2015). The Evolution of Cognitive Bias. *Si nous n'agissons pas collectivement*, 724–746. <https://doi.org/10.1002/9780470939376.CH25>

⁴¹ comme le soulignent les Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). *Quantifying the Carbon Emissions of Machine Learning*. <https://doi.org/10.48550/arxiv.1910.09700>

⁴² *Comment se compose le mix énergétique primaire de la France ? Brèves - notre-environnement.gouv.fr.* (n.d.). Retrieved November 15, 2022, from <https://www.notre-environnement.gouv.fr/actualites/breves/article/comment-se-compose-le-mix-energetique-primaire-de-la-france>

⁴³ *Guide d'auto-évaluation pour les systèmes d'intelligence artificielle (IA) | CNIL.* (n.d.). Retrieved November 15, 2022, from <https://www.cnil.fr/fr/intelligence-artificielle/guide>

⁴⁴ Sarigol, E., Garcia, D., & Schweitzer, F. (2014). Online Privacy as a Collective Phenomenon. *COSN 2014 - Proceedings of the 2014 ACM Conference on Online Social Networks*, 95–105. <https://doi.org/10.48550/arxiv.1409.6197>

⁴⁵ *TMD Dataset - 5 seconds sliding window / Kaggle.* (n.d.). Retrieved November 16, 2022, from <https://www.kaggle.com/datasets/fschwartz/tmd-dataset-5-seconds-sliding-window>