

Présentation du Projet P2M

Préparée par:
BEN AMOR Hadil
SLIM Baha Eddine

Encadré par:
Mr. AYADI Mohamed
Madame DRIDI Safa

Plan

- 1 Sujet
- 2 Etat de l'Art
- 3 Roadmap suivit
- 4 Visualisation des données
- 5 Préparation du dataset
- 6 Développement du Modèle

Sujet :

Notre sujet de P2M "Projet de Métier" consiste à élaborer **un modèle intelligent** pour le **calcul** des gaz à effet de serres en se basant sur les données ADEME.

Etat de l'Art

GreenWay-Maker

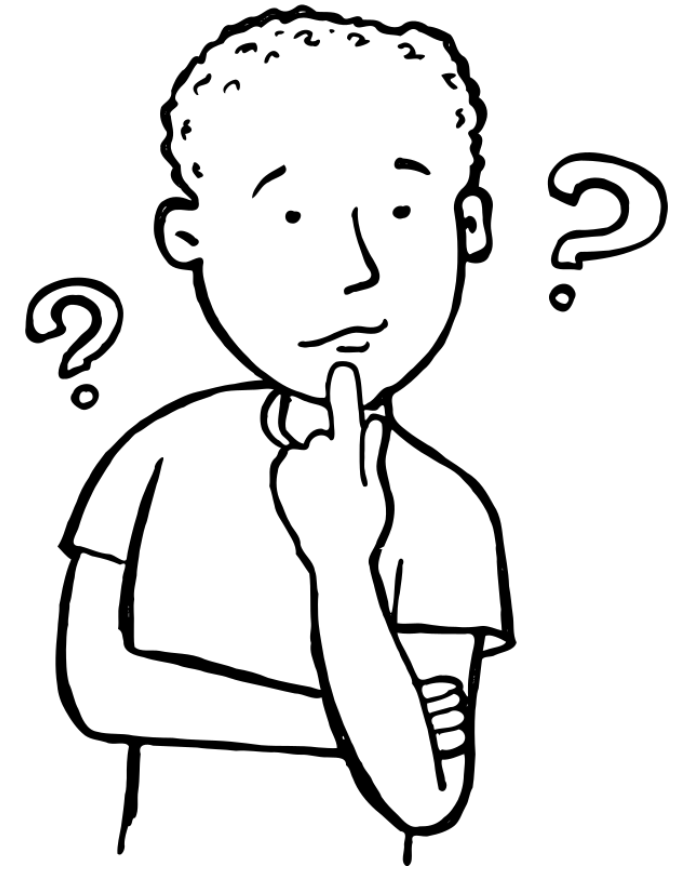


- Société par actions simplifiée.
 - Localisée à PARIS, elle est spécialisée dans le secteur d'activité de la programmation informatique.
-
- Le groupe Greenway-Maker International a développé des solutions logicielles innovantes qui permettent de réaliser les bilans Carbone et l'évaluation du coût de cycles de vie.

Problématique

Une petite quantité de gaz à effet de serre (GES) est produite naturellement, la majorité est émise par des activités humaines.

L'utilisation massive d'énergies fossiles (charbon, gaz, pétrole) pour les transports, les bâtiments, l'agriculture, l'industrie produisent de grandes quantités de GES qui se concentrent dans l'atmosphère.



La Solution de GreenWay-Maker

Nom

Le Logiciel Ozone Suite

Objectif 1

Réponds aux exigences des différentes méthodes et permet de passer du site au produit tout en utilisant les mêmes bases de données.

Objectif 2

Innovation logicielle qui offre sur une même plateforme les fonctionnalités des meilleurs outils ACV et des meilleurs outils dédiés aux bilan carbone.

--> Notre solution va être implémentée dans Ozone Suite

Les GES et ses sources

Gaz à effet de serre	Source naturelle	Source humaine
La vapeur d'eau (H ₂ O)	Évaporation de l'eau surtout au-dessus des océans	Centrales électriques - Irrigation
Le dioxyde de carbone (CO ₂)	Respiration des êtres vivants – Feux de forêt - Volcans...	Utilisation massive d'énergies fossiles pour les transports, les bâtiments et l'agriculture Déforestation
Le méthane (CH ₄)	Digestion des herbivores – Décomposition des végétaux- Volcans	Intensification des élevages (bovin) et des cultures (riz) - Décharge d'ordures
Le protoxyde d'azote (N ₂ O)	Marécages	Utilisation d'engrais azotés
Ozone de basse atmosphère (O ₃)	Foudre	Industrie - Circulation automobile
Les gaz fluorés (CFC, HFC, PFC)	N'existent pas dans la nature	Gaz des bombes aérosols et des climatiseurs

Les périmètres

✓ **01** Périmètre organisationnel

✓ **02** Périmètre opérationnel

✓ **03** Les postes d'émission

✓ **04** Les données à utiliser

Scopes : 3 catégories d'émissions

Scope 1

Émissions directes de GES

IL Y A 5 POSTES:

- SOURCES FIXES DE COMBUSTION
- SOURCES MOBILES À MOTEUR THERMIQUE
- PROCÉDÉS HORS ÉNERGIE
- ÉMISSIONS DIRECTES FUGITIVES
- ÉMISSIONS ISSUES DE LA BIOMASSE (SOLS ET FORÊTS)

Scope 2

Émissions à énergie indirectes

IL Y A 2 POSTES:

- ÉMISSIONS LIÉES À LA CONSOMMATION D'ÉLECTRICITÉ
- ÉMISSIONS LIÉES À LA CONSOMMATION DE VAPEUR, CHALEUR OU FROID

Scope 3

CONCERNE TOUS LES
AUTRES POSTES D'ÉMISSIONS.

Domaine du travail

**Secteur
Industrielle**

Poste 1

Emissions directes

Poste 6

Electricité

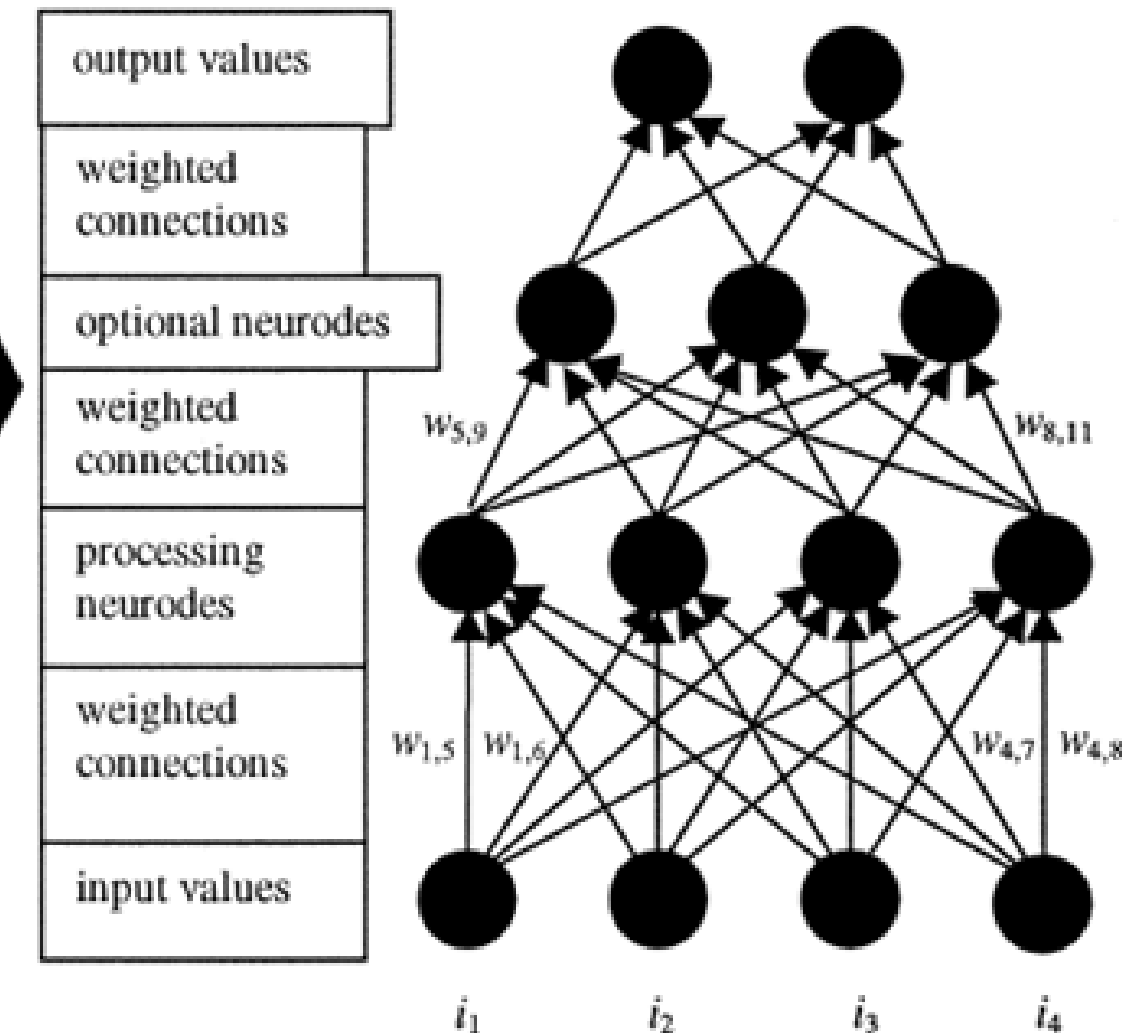
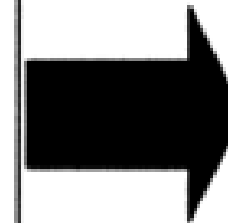
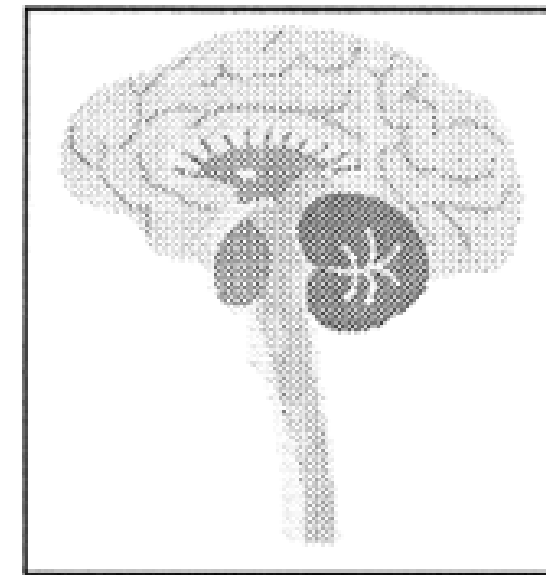
Poste 7

Chauffage

Les algorithmes Artificiels

ANN

- LES RÉSEAUX NEURONAUX ARTIFICIELS (RNA), GÉNÉRALEMENT APPELÉS RÉSEAUX NEURONAUX (RN), SONT DES SYSTÈMES INFORMATIQUES INSPIRÉS DES RÉSEAUX NEURONAUX BIOLOGIQUES QUI CONSTITUENT LE CERVEAU HUMAIN.
- LES ANN SONT DEVENUS UNE VASTE GAMME DE TECHNIQUES QUI ONT FAIT PROGRESSER L'ÉTAT DE L'ART DANS DE NOMBREUX DOMAINES.
- ANN EST CONSIDÉRÉE COMME MOINS PUISSANTE QUE CNN, RNN ET DNN.

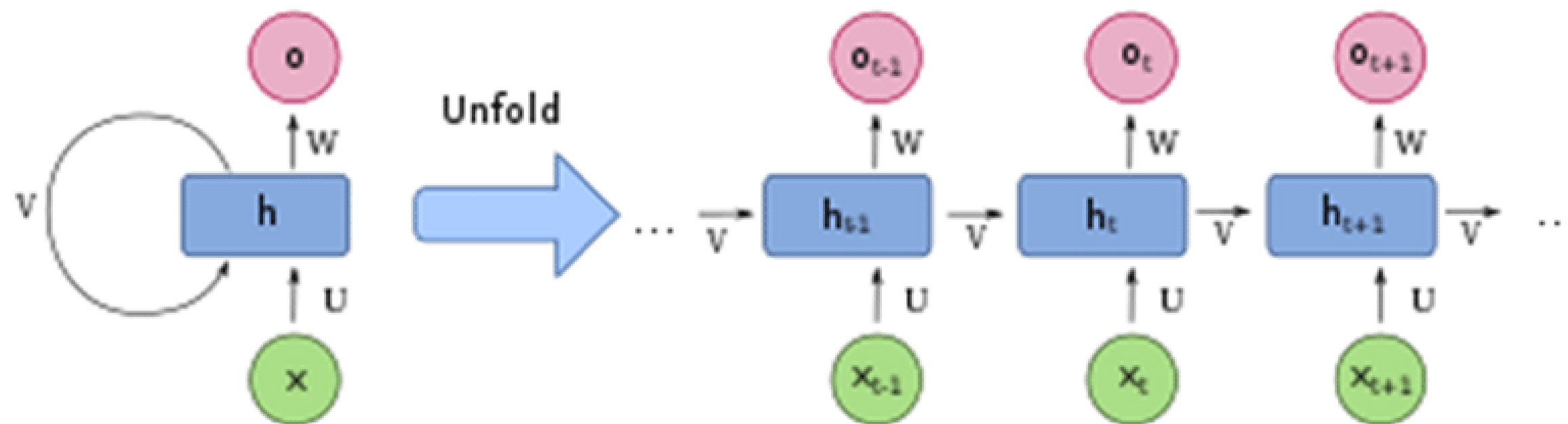


➡ **À REJETER PARCE QU'IL EST LE MOINS PERFORMANT**

RNN

- UN RÉSEAU NEURONAL RÉCURRENT (RNN) EST UTILISÉ POUR TRAITER LES SÉQUENCES
- PEUT ÊTRE APPELÉ DNN AU SENS « TEMPOREL » PLUTÔT QUE PUREMENT VERTICAL.
- LES ACTIVATIONS MASQUÉES DE L'ÉTAPE T SONT UTILISÉES COMME CONTEXTE POUR CALCULER LES ACTIVATIONS DE L'ÉTAPE $T+1$.

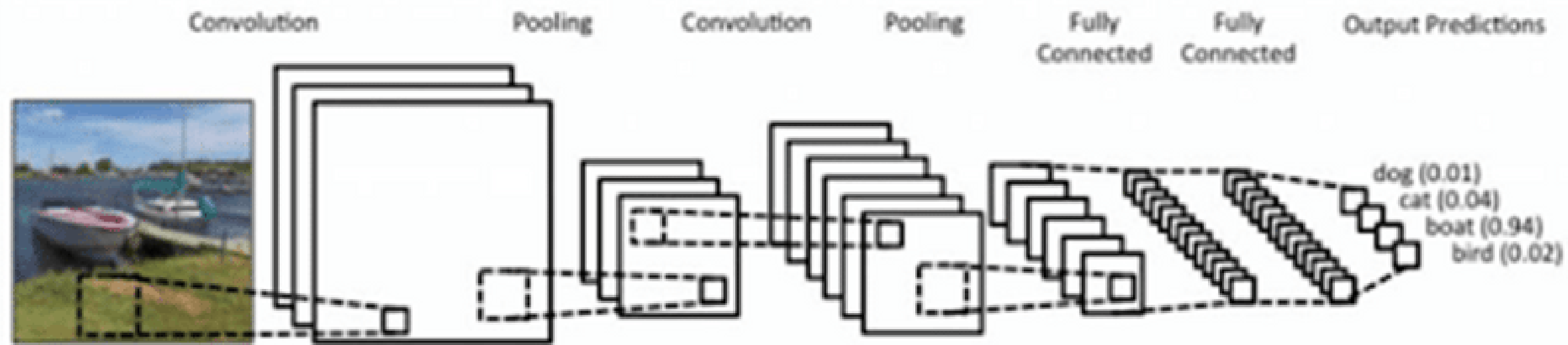
→ À REJETER PARCE QU'IL N'Y A PAS BESOIN D'UN RÉSEAU RÉCURRENT



CNN

- CONSIDÉRÉ COMME PLUS PUISSANT QU'ANN, RNN
- ALGORITHME DE RECONNAISSANCE EFFICACE QUI EST LARGEMENT UTILISÉ DANS LA RECONNAISSANCE DE MODÈLES ET LE TRAITEMENT D'IMAGE.
- A DE NOMBREUSES CARACTÉRISTIQUES TELLES QUE LA STRUCTURE SIMPLE,
- PLUS SEMBLABLE AUX RÉSEAUX DE NEURONES BIOLOGIQUES
- DÉTECTE AUTOMATIQUEMENT LES CARACTÉRISTIQUES IMPORTANTES SANS AUCUNE SUPERVISION HUMAINE
- PEUT APPRENDRE LES PRINCIPALES CARACTÉRISTIQUES DE CHAQUE CLASSE PAR LUI-MÊME.

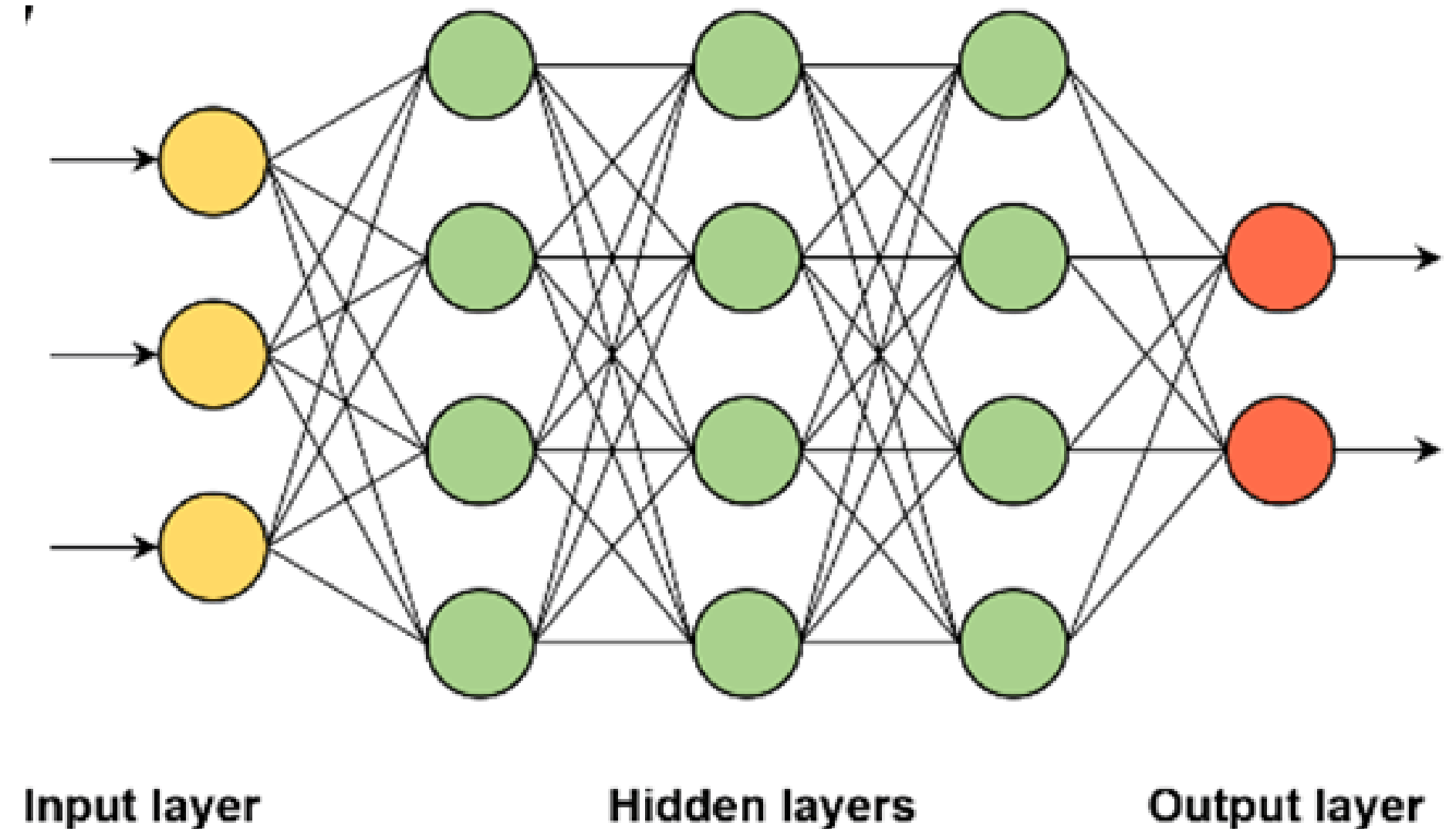
➔ **À REJETER PARCE QU'IL EST MIEUX ADAPTÉ AUX PROBLÈMES DE CLASSIFICATION**



DNN

- UN RÉSEAU NEURONAL PROFOND (DNN) EST UN ANN AVEC PLUSIEURS COUCHES ENTRE LES COUCHES D'ENTRÉE ET DE SORTIE.
- ILS PEUVENT MODÉLISER DES RELATIONS NON LINÉAIRES COMPLEXES
- SE CONCENTRE D'AVANTAGE SUR LA FAÇON DONT L'INFORMATION D'ENTRÉE EST REPRÉSENTÉE PAR UN GROUPE DE FONCTIONS NON LINÉAIRES (PAQUET DE COUCHES) AVANT D'ATTEINDRE LA COUCHE DE SORTIE.

➤ **À CONSIDÉRER PARCE QU'IL EST MIEUX ADAPTÉ AU PROBLÈMES D'EXTRACTION**



RoadMap suivit

Plan du travail



Approcher une base des données Tunisienne

- - - Croiser des différentes bases de données avec celle de ADEME



Préparer la base de données

- - - Eliminer les colonnes peut utiles
Traiter les valeurs manquantes et aberrantes



Développer un modèle de prédiction

- - - En se basant sur la BD

Visualisation des données

Outil:

- **Microsoft Power BI:**
Vise d'avantage un encodage de données rapide et flexible et la construction de modèles de données complexes.



Power BI



But:

- **Crée un rapport des données.**
- **Déterminer les zones de valeurs manquantes.**
- **Visualiser les distributions**

Scénario 1: "Par Scope"

Quick summary

Table

9865766

Sum of staff

216112349267

Sum of total_sco...

206680932413

Sum of total_sco...

384195604163

Sum of total_sco...

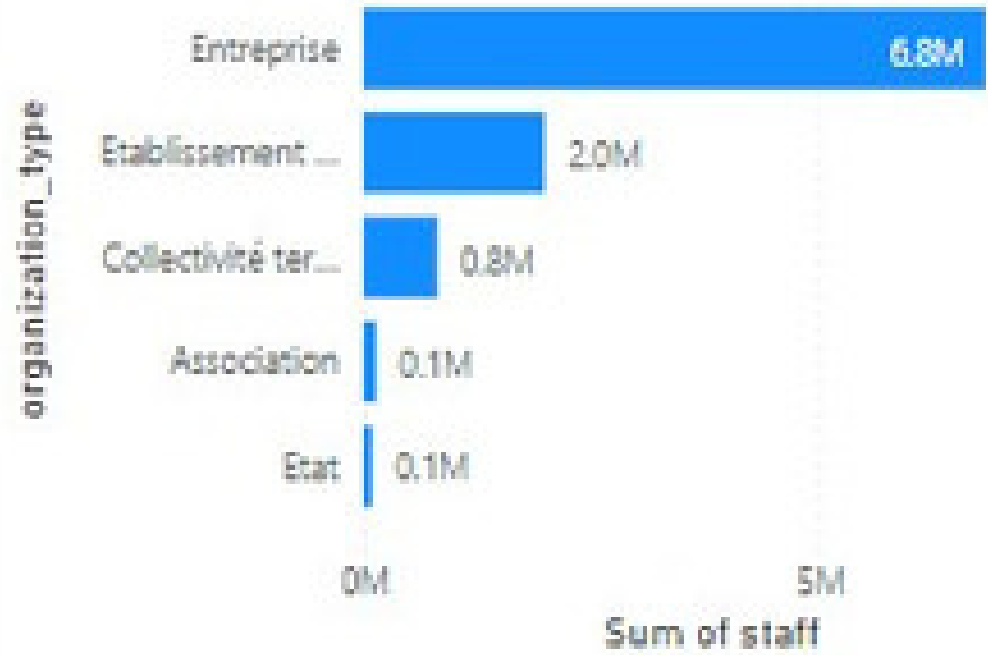
7817136

Sum of reporting...

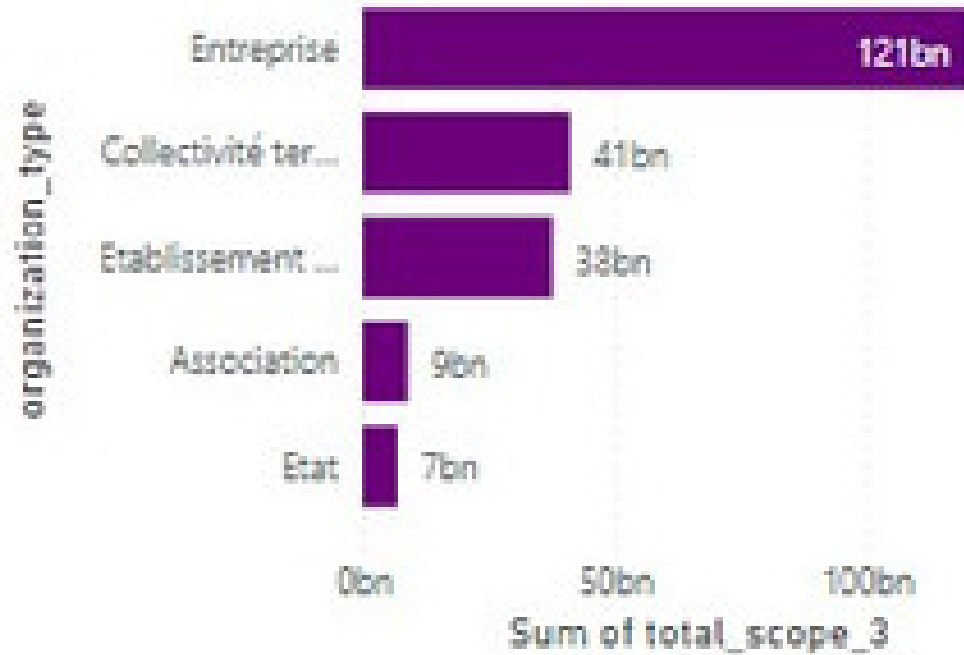
152519121

Sum of population

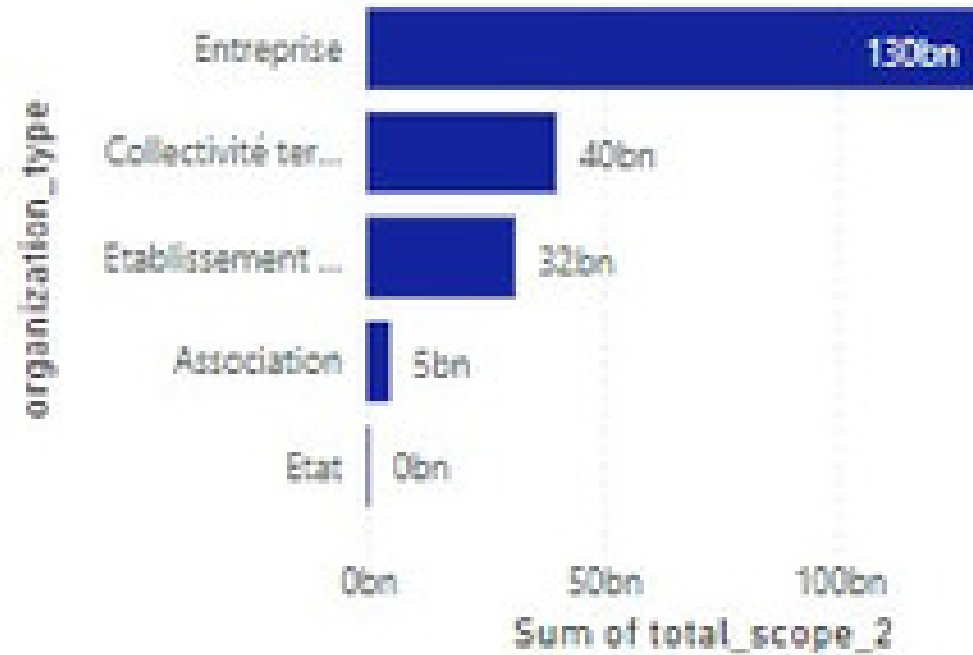
Sum of staff by organization_type



Sum of total_scope_3 by organization_type



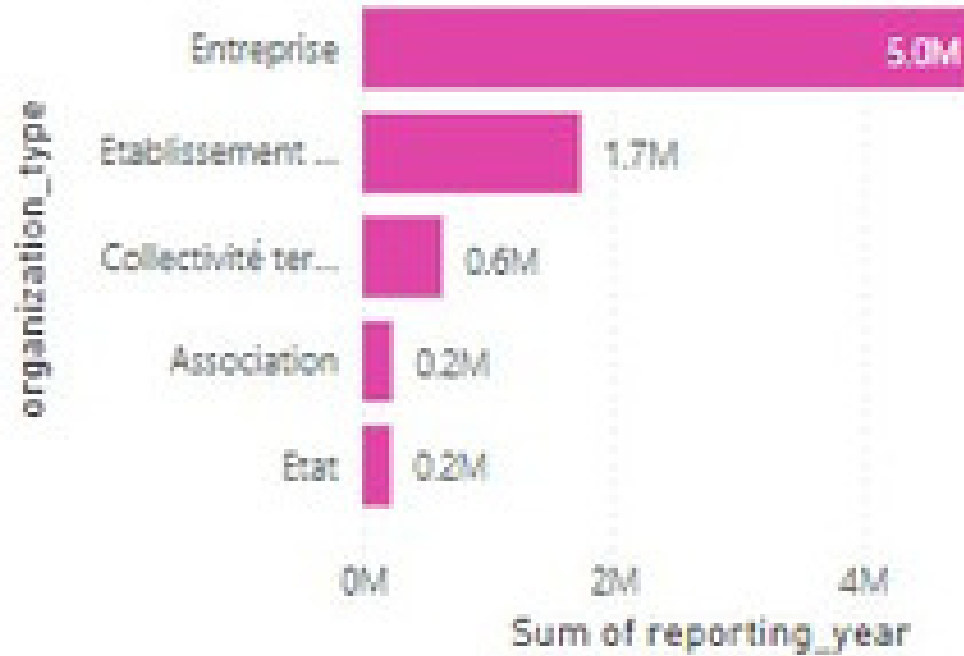
Sum of total_scope_2 by organization_type



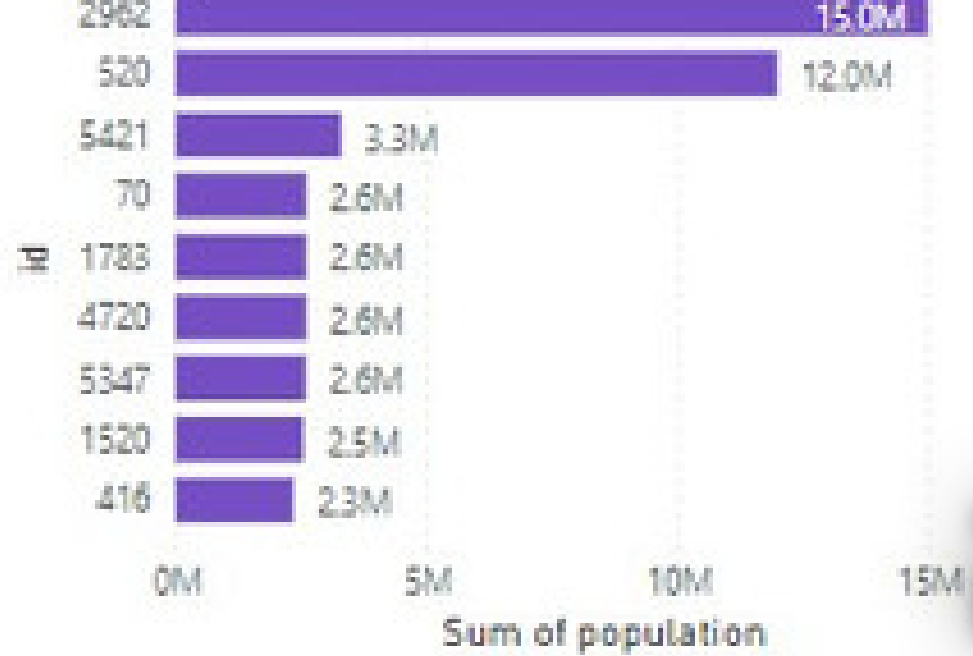
Sum of total_scope_1 by organization_type



Sum of reporting_year by organization_type



Sum of population by id



Remarques:

- Le type d'organization a une influence importante sur la somme des total_scope_1, 2, 3
(Les entreprises ont la plus grande somme)
 - Une Entreprise a le plus grand nombre de "Staff" par rapport aux autres types
(Etablissement, Collectivité territoriale, Association, Etat)
- > Le stuff influe la somme de total_scope**

Scénario 2: "Par Poste"

Quick summary
Table

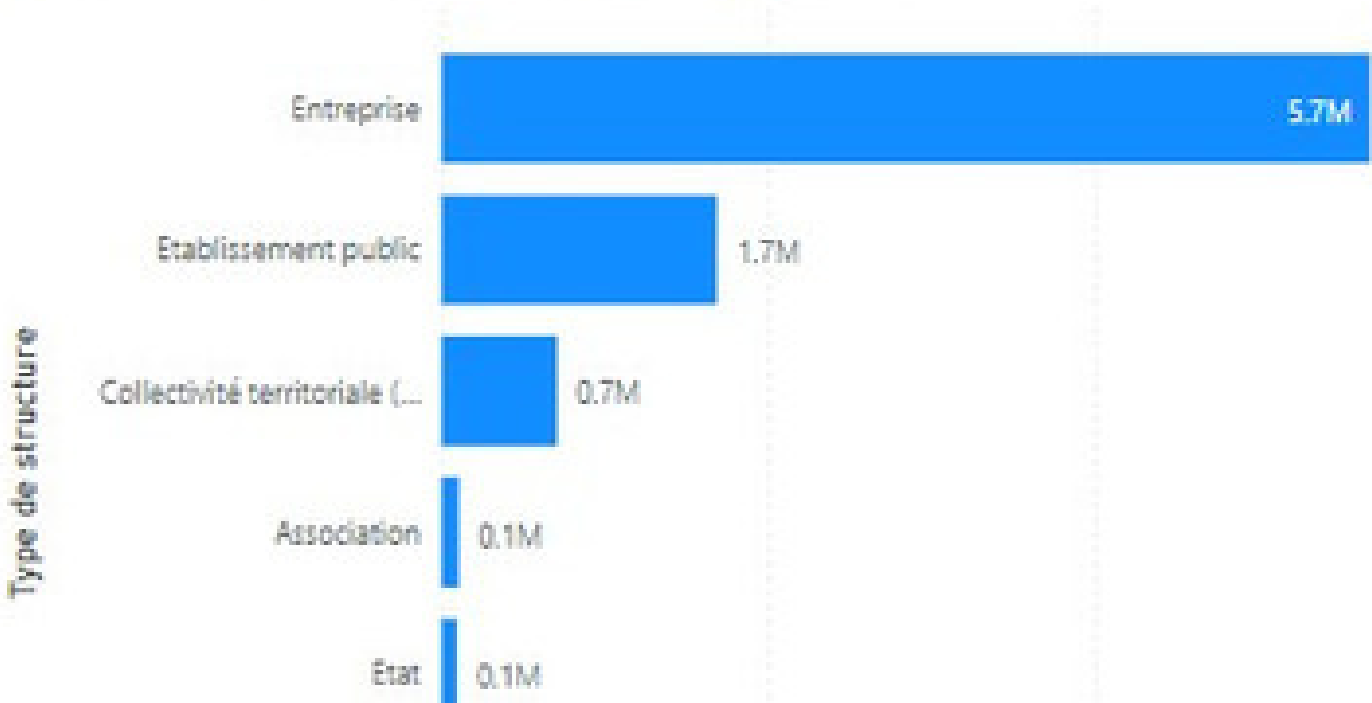
8279625
Sum of Nombre de salari...

769963655
Sum of Emissions publica...

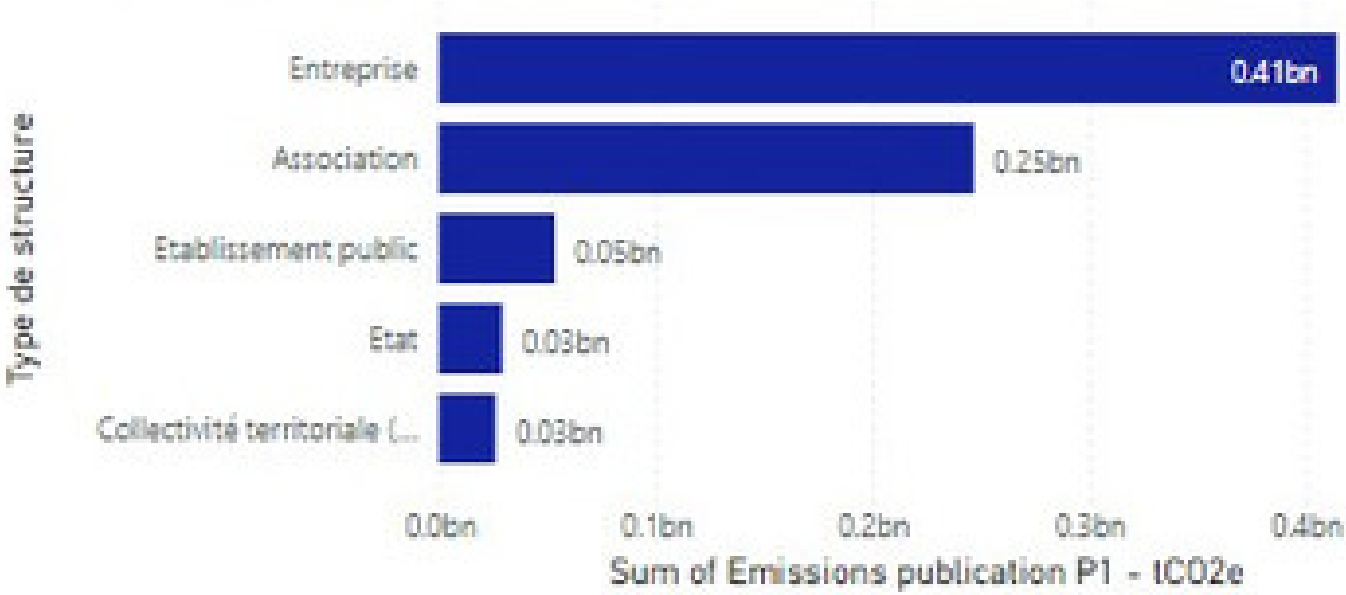
234016000
Sum of Emissions publica...

102703506
Sum of Emissions publica...

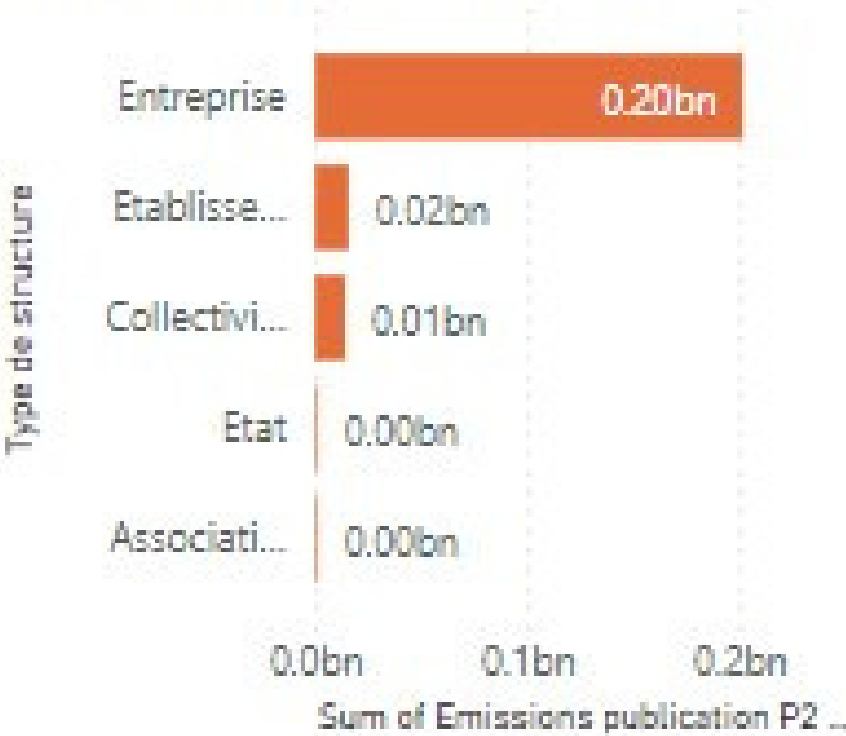
Sum of Nombre de salariés / d'agents by Type de structure



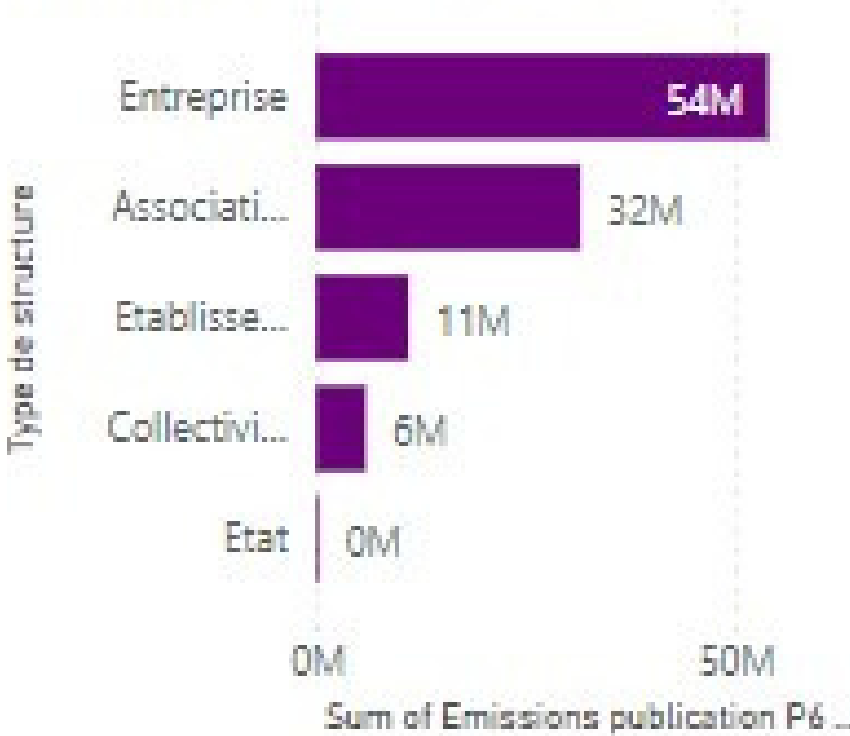
Sum of Emissions publication P1 - tCO2e by Type de structure



Sum of Emissions publication P2 - tCO2e by Type de structure



Sum of Emissions publication P6 - tCO2e by Type de structure



Remarques:

- La somme des salariés/agents dans les entreprises de la base de données est approximativement 5.7M.
- Les entreprises ont une somme des publications plus importante que les autres structures.
- La somme des publications P1 et P2 pour les structures dont la mode de consolidation est opérationnelle est plus grande que celles dont la mode de consolidation est financière.

--> On ne peut pas négliger la structure de l'entité et sa mode de consolidation.

Préparation du Dataset

Outils



Préparation du Dataset

1

Elimination des colonnes peut utiles

- Les colonnes avec faible valeur ajouté :
ex. « organization name »
- Les colonnes difficilement exploitables :
ex. « organization description »
- Les colonnes avec une grande pourcentage des valeurs manquantes :
ex. « collectivity type » (0.918)

2

Elimination des lignes peut utiles

Eliminer les lignes avec une grande pourcentage des valeurs manquantes

3

Quantification des données qualitatives

Convertir les données catégorielles qualitatives en données numériques :
ex.«organization type»

4

Remplissage des données manquantes

Procéder à remplacer les valeurs manquantes :
-Réelles : par la moyenne.
-entiers : Par un entier aléatoire d'un ensemble bien définis

Dataset avant préparation

Colonnes peut utiles
Valeurs manquantes
Formats non normalisées

ID du Bilan GE	Type de struct	Type de collec	Raison sociale	Année de rep	Date de publi	Nombre de sa	Population	Mode de cor	SIREN	PE(NAF)	Libellé	Région	Ville	Une année d
1	Entreprise		Mutuelle d'As	01/01/2011	04/03/2015	6848		Opérationne	775709702	512Z	Autres assurances		NIORT	
2	Entreprise		ADIDAS Franc	01/01/2011	05/03/2015	768		Opérationne	85480069	642Z	Commerce de gros (commer		SAVERNE	
3	Etablissement public		CROUS Nice-1	01/01/2011	05/03/2015	26		Opérationne	180600041	899B	Activités sociales sans héberger		Nice	
4	Collectivité te	Communes	Fort de France	01/01/2010	25/03/2015	300	94000						Martinique	
5	Collectivité te	Communes	concarneau C	01/01/2011	05/03/2015	10	49000						Bretagne	
6	Collectivité te	Communes	Communauté	01/01/2012	05/03/2015	73	57490						Normandie	
7	Etablissement public		OPAC Saône e	01/01/2011	05/03/2015	55		Opérationne	778596502	820A	Location de logements		MACON	
8	Entreprise		Système U Ou	01/01/2012	06/03/2015	2711		Opérationne	867800427	617A	Centrales d'achat alimentair		CARQUEFO	
9	Entreprise		Caisse d'Éparg	01/01/2011	06/03/2015	2732		Opérationne	353821028	419Z	Autres intermédiations mon		BORDEAUX	
10	Collectivité te	Communauté	Communauté	01/01/2012	06/03/2015	210	123476						Martinique	
11	Etablissement public		ADEME (Agen	01/01/2011	09/03/2015	963		Opérationne	385290309	413Z	Administration publique (tut		ANGERS	
12	Etablissement public		Chambre des	01/01/2011	09/03/2015	357		Opérationne	180600033	411Z	Activités des organisations p		TOULON	
13	Etablissement public		CENTRE HOSF	01/01/2018	09/03/2015	250		Opérationne	262405871	610Z	Activités hospitalières		NONTRON	1
14	Entreprise		NICOLLIN SAS	01/01/2011	10/03/2015	1113		Opérationne	775644149	811Z	Collecte des déchets non da		Saint Fons	
15	Etablissement public		CENTRE HOSF	01/01/2011	10/03/2015	250		Opérationne	266800192	610Z	Activités hospitalières		ROUFFACH	
16	Entreprise		Société Médit	01/01/2011	10/03/2015	1113		Opérationne	326180544	811Z	Collecte des déchets non da		Montpellier	
17	Collectivité te	Communauté	Valenciennes	01/01/2011	09/03/2015	149	192000						Hauts de France	
18	Collectivité te	Régions	Conseil Régio	01/01/2011	09/03/2015	1887	1350683						Auvergne-Rhône-Alpes	
19	Collectivité te	Département	Conseil Génér	01/01/2011	09/03/2015	1819	343000						Auvergne-Rhône-Alpes	
20	Collectivité te	Département	Conseil Dépar	01/01/2013	09/03/2015	2227	555999						Bourgogne-Franche-Comté	1
23	Collectivité te	Département	Conseil dépar	01/01/2013	10/03/2015	2262	571675						Hauts de France	
24	Etablissement public		Centre Hospit	01/01/2011	10/03/2015	450		Opérationne	260304936	610Z	Activités hospitalières		Ainay le château	
25	Entreprise		Coop Atlantiq	01/01/2011	10/03/2015	4300		Opérationne	525580130	711D	Supermarchés		Saintes	
27	Etablissement public		Centre Hospit	01/01/2016	10/03/2015	1099		Opérationne	264702689	610Z	Activités hospitalières		AGEN	1
29	Entreprise		Nice Matin	01/01/2012	12/03/2015	1000		Opérationne	955801204	813Z	Édition de journaux		Nice	
30	Etablissement public		Centre hospit	01/01/2011	10/03/2015	1554		Opérationne	267901213	610Z	Activités hospitalières		PARTHENAY	
31	Entreprise		MAAF Assurar	01/01/2012	10/03/2015	3299		Opérationne	781423280	512Z	Autres assurances		CHAURAY	
32	Entreprise		GIE EUROPAC	01/01/2012	12/03/2015	1823		Opérationne	389272295	299Z	Autres activités de soutien a		CHAURAY	

Dataset après préparation

index	Type de structure	Année de reporting	Date de publication	Nombre de salariés / d'agents	Mode de consolidation	Libellé	Emissions publication P1 - tCO2e
0	5	2015	2016	1270	1.0	1.0	2450258.34
1	5	2015	2017	1033	1.0	1.0	10586.0
2	5	2019	2019	900	1.0	2.0	5810.0
3	5	2015	2016	3000	1.0	3.0	5611.0
4	5	2018	2019	3100	1.0	1.0	4911.0
5	5	2018	2020	3029	1.0	3.0	4823.0
6	5	2016	2017	2356	1.0	4.0	4367.0
7	5	2015	2016	1000	1.0	5.0	3749.0
8	5	2015	2017	2051	1.0	6.0	3646.0
9	5	2015	2016	1487	1.0	7.0	3580.0
10	5	2015	2016	1800	1.0	8.0	3342.0
11	5	2017	2018	5500	1.0	9.0	3268.0
12	5	2016	2017	2400	1.0	1.0	3053.0
13	5	2014	2016	3058	2.0	10.0	3049.0
14	5	2016	2017	1148	1.0	7.0	3033.0
15	5	2019	2020	3200	1.0	2.0	2614.0
16	5	2014	2016	1700	1.0	1.0	2451.0
17	5	2014	2016	1700	2.0	6.0	2427.0
18	5	2014	2016	1200	1.0	11.0	2385.0
19	5	2014	2016	2192	1.0	7.0	2323.0
20	5	2015	2018	2700	1.0	12.0	2226.04
21	5	2014	2017	2800	1.0	13.0	2193.63
22	5	2011	2017	963	1.0	14.0	2004.0
23	5	2019	2020	604	1.0	7.0	1934.0

Développement du modèle

Architecture du modèle

```
model = Sequential()
model.add(Dense(l1_out, input_dim=first_input, kernel_initializer='he_uniform', activation='elu'))
model.add(Dense(l2_out, activation='elu'))
model.add(Dense(l3_out, activation='elu'))
model.add(Dense(l4_out, activation='elu'))
model.add(Dense(last_output))

model.compile(loss='mae', optimizer='adam', metrics=['accuracy'])

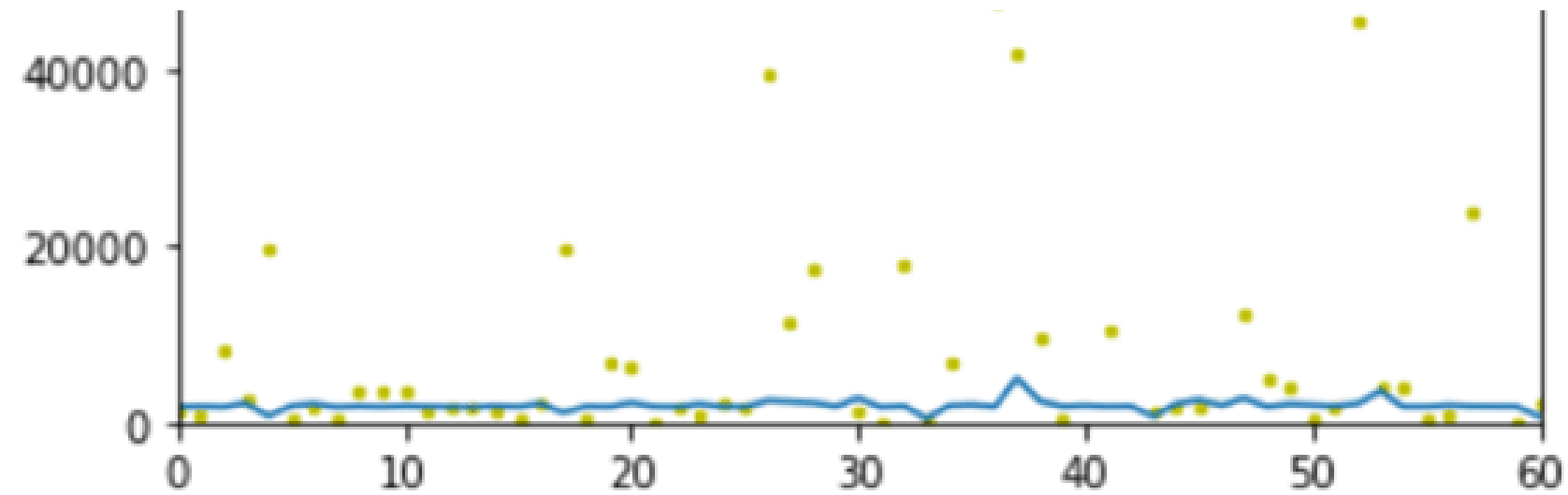
model.summary()
```

Model: "sequential_48"

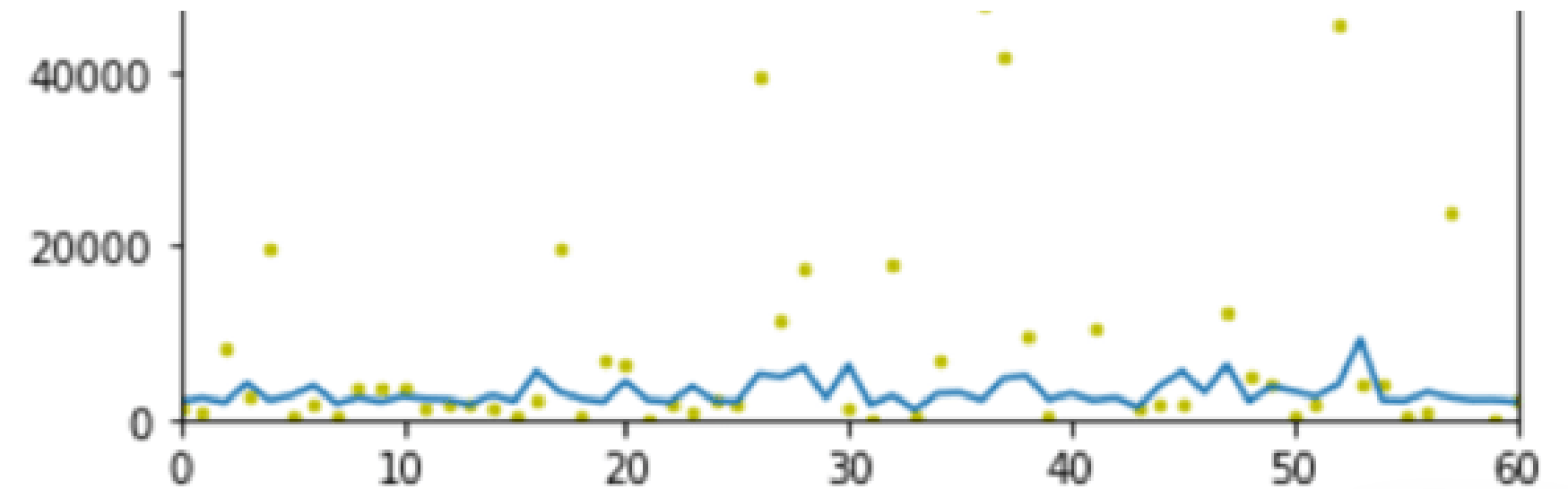
Layer (type)	Output Shape	Param #
dense_240 (Dense)	(None, 8)	56
dense_241 (Dense)	(None, 32)	288
dense_242 (Dense)	(None, 128)	4224
dense_243 (Dense)	(None, 8)	1032
dense_244 (Dense)	(None, 4)	36

```
=====
Total params: 5,636
Trainable params: 5,636
Non-trainable params: 0
```

Avant optimisation



Après optimisation



Performance de la modèle

```
Epoch 244/250
80/80 - 0s - loss: 15719.0986 - accuracy: 0.4067 - val_loss: 5002.6709 - val_accuracy: 0.3407 - 171ms/epoch - 2ms/step
Epoch 245/250
80/80 - 0s - loss: 15726.4199 - accuracy: 0.4087 - val_loss: 5017.0361 - val_accuracy: 0.3481 - 164ms/epoch - 2ms/step
Epoch 246/250
80/80 - 0s - loss: 15720.8086 - accuracy: 0.4112 - val_loss: 4992.8975 - val_accuracy: 0.3556 - 166ms/epoch - 2ms/step
Epoch 247/250
80/80 - 0s - loss: 15720.5117 - accuracy: 0.4087 - val_loss: 5003.0234 - val_accuracy: 0.3407 - 173ms/epoch - 2ms/step
Epoch 248/250
80/80 - 0s - loss: 15723.8867 - accuracy: 0.4092 - val_loss: 5014.0972 - val_accuracy: 0.3481 - 154ms/epoch - 2ms/step
Epoch 249/250
80/80 - 0s - loss: 15719.7930 - accuracy: 0.4067 - val_loss: 5003.2720 - val_accuracy: 0.3259 - 167ms/epoch - 2ms/step
Epoch 250/250
80/80 - 0s - loss: 15725.2266 - accuracy: 0.4062 - val_loss: 5003.9907 - val_accuracy: 0.3407 - 180ms/epoch - 2ms/step
```

- Accuracy 40.62%
- Le val_Loss est important

Raisons:

- Nombre limité des entrées
- Plusieurs valeurs manquantes qui ont été remplacé par la moyenne
- Taille relativement faible du dataset

Conclusion générale

Le modèle intelligent réalisé peut permettre un gain de temps et d'argent énorme pour les entreprises en permettant aux clients d'estimer leurs production en GES et prendre les mesures nécessaires afin de respecter les normes imposés par l'état et protéger l'environnement.

Une future étape, c'est trouver des méthodes et outils plus sophistiqués pour améliorer les performances du modèle

Word cloud featuring various expressions of gratitude in multiple languages, including:

- merci
- thank you
- gracias
- obrigado
- danke
- grazie
- arigatō
- ngiyabonga
- bedankt
- terima kasih
- moachakkeram
- dakujem
- chokrane
- maith agat
- go raibh maith agat
- спасибо
- murakoze
- manana
- dankon
- sobodi
- mauruuru
- asante
- taiku
- nanni
- trugarez
- paldies
- kam sah hamnida
- faafetai lava
- najis tuke
- kop khun
- mamnun
- ভোমকে ধন্যবাদ
- moachakkeram
- tau
- akun
- aciō
- barka
- huala
- hvala
- koszonom
- blagodaram
- xvala
- teşekkür ederim
- teşekkür ederim
- dhanyavad
- dhanyavadagalu
- matondo
- misaotra
- gratias ago
- welalin
- enkosi
- spas
- teşekkü ederim
- gracies
- diolch
- mersi
- tanemirt
- rahmat
- dekuji
- bayanalaa
- kia ora
- ありがとう
- merci
- grazzi
- merci
- leat
- nandri
- mahalo
- merce
- krap
- kop khun
- didi madloba
- rahmet
- tapadh
- gracie
- krap
- mési
- takk
- tenki
- 謝謝
- dziekuję