



Cycle de formation des ingénieurs en Télécommunications

Rapport de P2M

Thème :

**« Etude comparative et implémentation
d'un modèle basé sur l'IA pour l'estimation
des empreintes carbone cas de l'UNFCCC »**

Réalisé par :

**LAKHDHAR Rayen
MAMI Mohamed Moez**

Encadrant (s) :

M. Mohamed AYADI

Travail proposé et réalisé en collaboration avec

GreenWay Maker

Année universitaire : 2021-2022

Ecole Supérieure des Communications de Tunis

SUP'COM

2083 Cité Technologique des Communications - Elghazala - Ariana - Tunisie

Tél. +216 70 240 900 – site web : www.supcom.mincom.tn

Remerciements

Avant tout développement sur cette expérience académique, il apparaît opportun de commencer ce rapport par des remerciements.

Nous remercions Mr Mohammed Ayadi notre encadrant qui nous accompagnés tout au long de cette expérience académique avec beaucoup de patience et de pédagogie.

Nous remercions aussi, tous les membres du jury pour avoir accepté de juger ce modeste travail.

Table des matières :

Introduction générale.....	6
Chapitre 1	9
1. Introduction.....	9
2. Définition UNFCCC	10
3. Définition de l'IPCC avec ses 5 répartitions.....	10
4. Problématique.....	12
5. Les outils utilisés	13
5.1. Python/Jupyter.....	13
5.2 Logiciel IPCC.....	13
Chapitre 2	14
Algorithmes utilisés et comparaison des résultats.....	14
1. La régression linéaire	14
1.1. Définition	14
1.2. Score calculé.....	14
2. SVM	15
2.1. Définition	15
3. DecisionTreeRegressor	15
3.1. Définition	15
La démarche suivie.....	16
1. Les données numériques	17
1.1. La méthode Mean	17

1.2. La méthode d'interpolation linéaire	18
1.3. La méthode dropping	19
2. Les données catégorielles	20
2.1. La méthode most frequent values	20
Chapitre 3	22
Comparaison des résultats	22
1. La régression linéaire	22
1.1 Solution proposée	22
1.2. Interprétation	23
2. SVM.....	23
2.1. Solution proposée	23
2.2. Interprétation	23
3. DecisionTreeRegressor	24
3.1. Solution proposée	24
3.2. Interprétation	25
Tableau récapitulatif	27
Conclusion générale.....	28
Bibliographie.....	29

Liste des figures :

<u>Figure 1</u> : Base de données UNFCC.....	9
<u>Figure 2</u> : Les 5 répartitions de l'IPCC	11
<u>Figure 3</u> : Estimation des émissions du CO2 selon La répartition	12
<u>Figure 4</u> : Capture qui montre le nombre de données manquantes.....	16
<u>Figure 5</u> : Remplir les données avec la méthode mean	17
<u>Figure 6</u> : Explication du principe de l'interpolation linéaire	18
<u>Figure 7</u> : Capture des données manquantes.....	18
<u>Figure 8</u> : Utiliser la méthode interpolation linéaire	19
<u>Figure 9</u> : Données remplies	19
<u>Figure 10</u> : La méthode dropping utilisée	20
<u>Figure 11</u> : Utilisation de la méthode most frequent values pour remplir les données	21
<u>Figure 12</u> – Solution proposée de la régression linéaire.....	22
<u>Figure 13</u> – Solution proposée de la SVM.....	23
<u>Figure 14</u> – Solution proposée de la DecisionTreeRegressor.....	24
<u>Figure 15</u> – Exemples de prédiction en utilisant la DecisionTreeRegressor	26

Introduction générale :

-Le réchauffement climatique est un phénomène global de changement climatique caractérisé par une augmentation générale de la température moyenne (notamment en relation avec l'activité humaine) qui altère durablement les équilibres météorologiques et les écosystèmes.

Quand on en parle aujourd'hui, c'est environ 100 à 150 ans d'augmentations de température qui se sont produites sur Terre. Ce phénomène est dû au effet de serre. Donnons ainsi une définition au effet de serre.

-L'effet de serre est un phénomène naturel. La surface de la Terre absorbe naturellement 70% du rayonnement solaire tandis que le reste est renvoyé vers l'espace sous l'effet de la réverbération. Le rayonnement solaire absorbé est alors transformé en rayonnement infrarouge et restitué dans l'atmosphère. Une partie de ce rayonnement infrarouge est alors à son tour renvoyée vers l'espace tandis que l'autre partie est retenue par les gaz à effet de serre présents dans l'atmosphère.

-L'effet de serre est un phénomène qui s'applique à toutes les planètes dotées d'une atmosphère comme la Terre ou Vénus.

Les activités humaines sont à l'origine de l'accroissement du phénomène que l'on nomme "effet de serre".

Ses conséquences, dont le réchauffement climatique, sont très préoccupantes pour la planète. Donc quels sont les gaz qui causent cet effet de serre nommés Gaz à effet de serre ?

-Les Gaz à Effet de Serre (GES) sont des gaz naturellement présents dans l'atmosphère. Ils absorbent une partie des rayons solaires et les redistribuent ensuite sous la forme de radiations. Leur concentration croissante dans l'atmosphère, due aux activités humaines, est à l'origine du réchauffement climatique. Comme leur nom l'indique, ces gaz sont la cause principale de l'effet de serre.

Les gaz responsables de l'effet de serre sont les suivants :

- La vapeur d'eau (H₂O) ;
- Le dioxyde de carbone (CO₂) ;
- Le méthane (CH₄) ;
- Le protoxyde d'azote (N₂O) ;
- L'ozone (O₃).

Les principales sources des gaz à effet de serre sont :

- le gaz carbonique ou dioxyde de carbone (CO₂), qui provient essentiellement de la combustion des énergies fossiles et de la déforestation

- le méthane (CH_4), qui a pour origine principale la digestion des ruminants, la culture du riz, les décharges d'ordures ménagères, les exploitations pétrolières et gazières
- le protoxyde d'azote ou oxyde nitreux (N_2O), qui provient de l'utilisation des engrais azotés et de certains procédés chimiques

C'est pourquoi il est essentiel d'agir contre les émissions de gaz à effet de serre liées à ce phénomène. En effet de nos jours et depuis plusieurs années plusieurs congrès ont été établis afin d'essayer de réduire au maximum le taux d'émission des Gaz à effet de serre et assurer la neutralité Carbone qui consiste à avoir le même taux de CO_2 consommé que émis. Prenons l'exemple du congrès COP26 qui s'est déroulé à Glasgow du 31 octobre au 12 novembre 2021 et qui avait pour objectif de maintenir le réchauffement climatique sous 2° - de préférence à 1.5° - d'ici 2100.

Tout cela n'est qu'une solution provisoire, donc que devons nous faire afin de réduire à maximum les émissions du CO_2 ?

Chapitre 1 :

Introduction :

Lors de nos recherches, on est tombé sur plusieurs base de données mais afin de réaliser notre projet on a décidé de travailler sur la base de données UNFCCC qui est liée aussi à la base données IPCC

Country_cod	Country	Format_nam	Pollutant_na	Sector_code	Sector_name	Parent_sectc	Unit	Year	emissions	PublicationD	DataSource
NL	Netherlands	IPCC Common f SF6 - (CO2 equi	2.E.2	2.E.2 - TFT Flat	2.E	Gg CO2 equival	2008			20210616	EEA
NL	Netherlands	IPCC Common f SF6 - (CO2 equi	2.E.3	2.E.3 - Photovo	2.E	Gg CO2 equival	1992			20210616	EEA
NL	Netherlands	IPCC Common f SF6 - (CO2 equi	2.E.3	2.E.3 - Photovo	2.E	Gg CO2 equival	2010			20210616	EEA
NL	Netherlands	IPCC Common f SF6 - (CO2 equi	2.E.3	2.E.3 - Photovo	2.E	Gg CO2 equival	2017			20210616	EEA
NL	Netherlands	IPCC Common f SF6 - (CO2 equi	2.E.5	2.E.5 - Other el	2.E	Gg CO2 equival	1993			20210616	EEA
NL	Netherlands	IPCC Common f SF6 - (CO2 equi	2.E.5	2.E.5 - Other el	2.E	Gg CO2 equival	2001			20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b	1.A.3.b - Road 1	1.A.3	Gg	2009	0.10775		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.i	1.A.3.b.i - Cars	1.A.3.b	Gg	2009	0.08124		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.i	1.A.3.b.i - Cars	1.A.3.b	Gg	2012	0.07069		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.i	1.A.3.b.i - Cars	1.A.3.b	Gg	2019	0.10371		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.iii	1.A.3.b.iii - Hea	1.A.3.b	Gg	2008	0.03074		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.iii	1.A.3.b.iii - Hea	1.A.3.b	Gg	2011	0.02634		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.iv	1.A.3.b.iv - Mot	1.A.3.b	Gg	1997	0.00029		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.iv	1.A.3.b.iv - Mot	1.A.3.b	Gg	2001	0.00019		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.v	1.A.3.b.v - Othe	1.A.3.b	Gg	1991			20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.v	1.A.3.b.v - Othe	1.A.3.b	Gg	1996			20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.v	1.A.3.b.v - Othe	1.A.3.b	Gg	1997			20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.v	1.A.3.b.v - Othe	1.A.3.b	Gg	2003			20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.v	1.A.3.b.v - Othe	1.A.3.b	Gg	2004			20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.b.v	1.A.3.b.v - Othe	1.A.3.b	Gg	2005			20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.c	1.A.3.c - Railwa	1.A.3	Gg	1996	0.09867		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.c	1.A.3.c - Railwa	1.A.3	Gg	2013	0.06277		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.c	1.A.3.c - Railwa	1.A.3	Gg	2017	0.06477		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.d	1.A.3.d - Dome	1.A.3	Gg	1990	0.00042		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.d	1.A.3.d - Dome	1.A.3	Gg	2003	0.00036		20210616	EEA
LT	Lithuania	IPCC Common f N2O	1.A.3.d	1.A.3.d - Dome	1.A.3	Gg	2007	0.00049		20210616	EEA

Figure 1 – Base de données UNFCCC

Cette base de données comporte plusieurs paramètres, on peut en citer le pays concerné, le type de gaz émis , les différents secteurs selon l'IPCC , l'unité et la date de publication des données.

Mais ici la question de se pose :

Qu'est ce que ça veut dire UNFCCC et qu'est ce que l'IPCC ?

❖ **Définition UNFCCC :**

The **United Nations Framework Convention on Climate Change (UNFCCC)** : a été adoptée au cours du sommet de la Terre de Rio de Janeiro en 1992 par 154 États auxquels il faut ajouter la totalité des membres de la Communauté européenne. Elle est entrée en vigueur le 21 mars 1994, et le siège est à Bonn, en Allemagne. En 2004, elle était ratifiée par 189 pays, en 2015 on recense 195 pays² et en 2018, par 197 pays. La CCNUCC est la première tentative, dans le cadre de l'ONU, de mieux cerner ce qu'est le changement climatique et comment y remédier

❖ **Définition IPCC avec ses 5 répartitions :**

The Intergovernmental Panel on Climate Change (IPCC) :

C'est l'organisme des Nations Unies chargé d'évaluer la science liée au changement climatique.

IPCC fournit des évaluations régulières de la base scientifique du changement climatique, de ses impacts et des risques futurs, et des options d'adaptation et d'atténuation.

IPCC est composé de 5 répartitions :

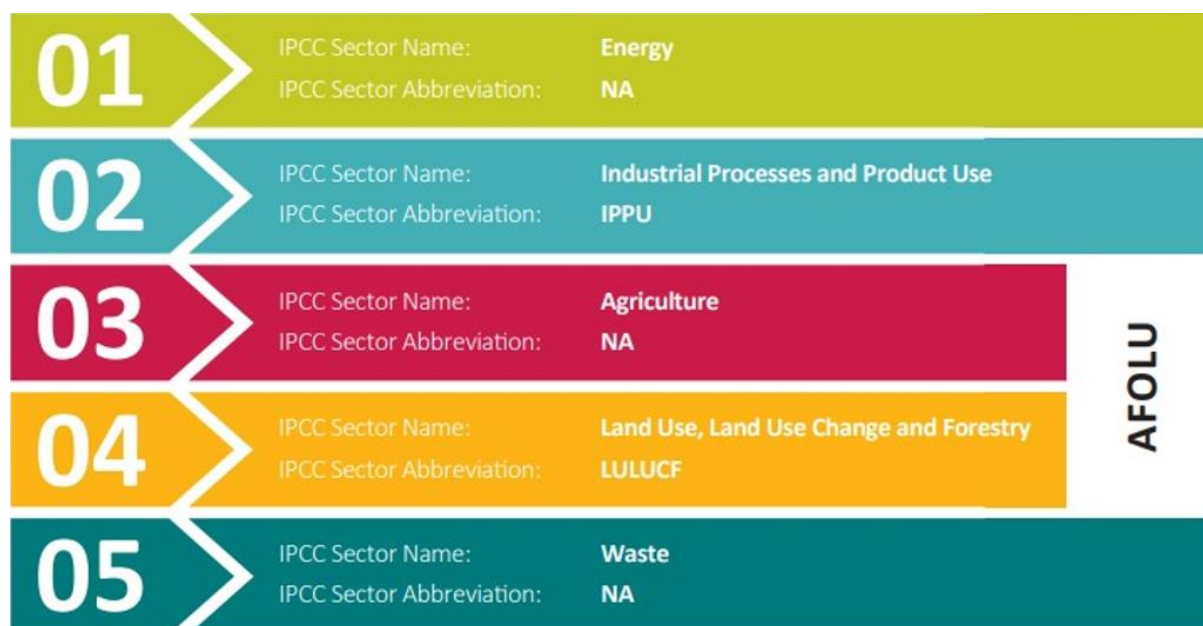


Figure 2 – Les 5 répartitions de l'IPCC

Maintenant on va présenter une statistique faite qui montre les différentes émissions du CO₂ selon les différents secteurs de l'IPCC :

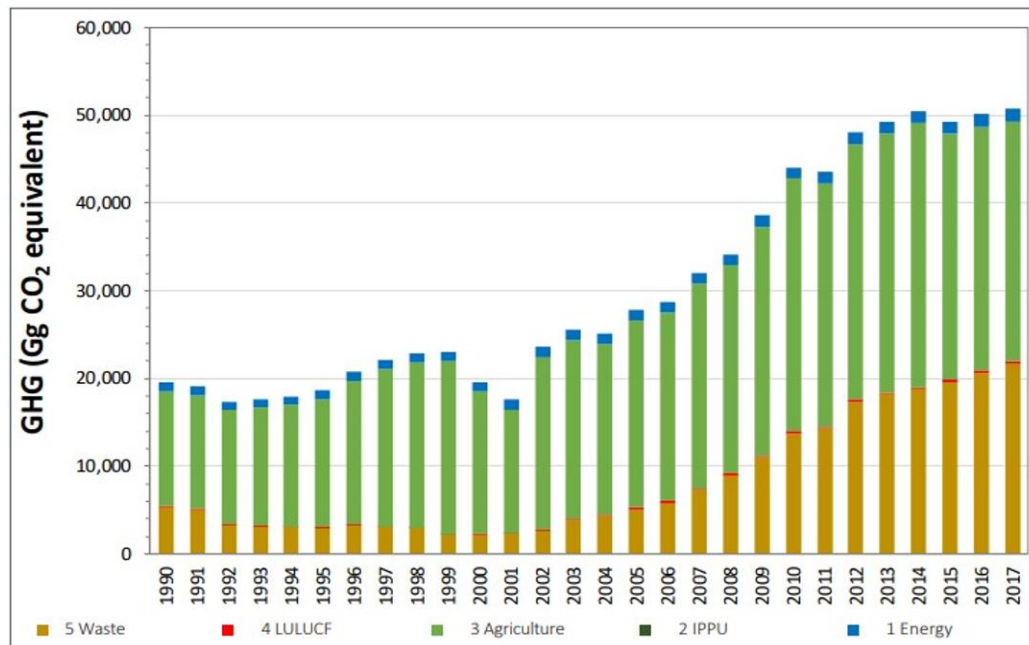


Figure 3 – Estimation des émissions du CO₂ selon La répartition

On remarque l'augmentation énorme des émissions du CO₂ selon les années, ce qui nous incite à trouver des solutions efficaces pour régler ce problème. Que devons nous faire alors ?

Problématique :

Ce projet consiste à prédire les quantités des gaz à effet de serre en développant un modèle intelligent capable d'effectuer cette tâche.

Ainsi pour réaliser ce modèle intelligent, on va procéder à différentes méthodes intelligentes basés sur des mathématiques.

Les outils utilisés :

❖ Python/Jupyter :

Un langage de programmation qui comprend de nombreux outils open-source dans le Domaine de la science des données.

❖ Logiciel IPCC :

Un logiciel disponible à télécharger qui contient les différentes classifications des secteurs dans une base de données.

Chapitre 2 :

Algorithmes utilisés :

Pour travailler notre projet on a choisi d'essayer 3 différents algorithmes pour pouvoir comparer les résultats après.

➤ La régression linéaire multiple :

-Définition :

La régression linéaire multiple utilise un algorithme d'apprentissage supervisé en vue de déterminer la valeur de variables sur la base de précédentes données. La relation linéaire multiple permet de dégager différentes tendances et prédictions

-Score calculé :

C'est un indicateur utilisé en statistiques pour juger de la qualité d'une régression linéaire. Mathématiquement, il s'agit de la proportion de la variance d'une variable dépendante qui s'explique par une ou plusieurs variables indépendantes dans le modèle de régression.

➤ **SVM (Support Vector Machine)**

-Définition :

Les machines à vecteurs de support ou séparateurs à vaste marge sont un ensemble de techniques d'apprentissage supervisé destinées à résoudre des problèmes de discrimination et de régression. Les SVM sont une généralisation des classifieurs linéaires.

➤ **DecisionTreeRegressor :**

-Définition :

Un arbre de décision est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre. Les différentes décisions possibles sont situées aux extrémités des branches (les « feuilles » de l'arbre), et sont atteintes en fonction de décisions prises à chaque étape

Mais avec la présence de ces 3 algorithmes, on a rencontré des problèmes. C'est que la base de données UNFCCC admet beaucoup de valeurs manquantes. Donc comment remédier à cela ?

La démarche suivie :

```
[25] # checking for missing values
data.isnull().sum()
```

Python

```
... Country_code      0
Country              0
Format_name          0
Pollutant_name       0
Sector_code          0
Sector_name          0
Parent_sector_code   57582
Unit                 0
Year                 0
emissions            259850
PublicationDate       0
DataSource            0
dtype: int64
```

On a 37,85% de données manquantes pour les quatités d'émission et 8,38% de données manquantes pour les Parent_sector_code

```
[26] data.describe()
```

Python

```
...      emissions  PublicationDate
count  4.264980e+05      686348.0
mean    2.549875e+04      20210616.0
std     2.517595e+05         0.0
min    -4.673036e+05      20210616.0
25%     4.483300e-01      20210616.0
```

Figure 4 – Capture qui montre le nombre de données manquantes

La solution est de procéder à une ingénierie de données pour compléter ces données manquantes.

Tout d'abord, on doit faire la différence dans notre base de données entre les données numériques et les données catégorielles.

Les méthodes varient selon le type des données.

Voici les différentes méthodes possibles pour effectuer ceci :

○ Les données numériques :

➤ La méthode Mean :

Cela fonctionne en calculant la moyenne/médiane des valeurs non manquantes dans une colonne, puis en remplaçant les valeurs manquantes dans chaque colonne séparément et indépendamment des autres. Il ne peut être utilisé qu'avec des données numériques.

- Les Aspects positifs :

1. Facile et rapide
2. Fonctionne bien avec de petits ensembles de données numériques

- Les Aspects négatifs :

1. Pas très précis
2. Ne tient pas compte des corrélations entre les Inputs



```
Dealing with missing values

data['emissions']=data['emissions'].interpolate(method="linear")
```

[14] ✓ 0.3s Python

Figure 5 – Remplir les données avec la méthode mean

➤ La méthode d'Interpolation linéaire :

Cette méthode essaie d'estimer les valeurs d'autres observations dans la plage d'un ensemble discret de points de données connus

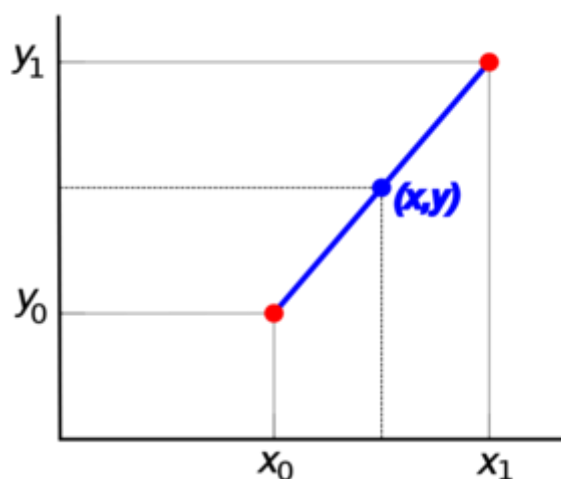


Figure 6 – Explication du principe de l'interpolation linéaire

```
data.head()
```

✓ 0.4s Python

	Country_code	Country	Format_name	Pollutant_name	Sector_code	Sector_name	Parent_sector_code	Unit	Year	emissions	PublicationDate
0	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.2	2.E.2 - TFT Flat Panel Display	2.E	Gg CO2 equivalent	2008	NaN	20210616
1	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.3	2.E.3 - Photovoltaics	2.E	Gg CO2 equivalent	1992	NaN	20210616
2	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.3	2.E.3 - Photovoltaics	2.E	Gg CO2 equivalent	2010	NaN	20210616
3	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.3	2.E.3 - Photovoltaics	2.E	Gg CO2 equivalent	2017	NaN	20210616
4	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.5	2.E.5 - Other electronics industry	2.E	Gg CO2 equivalent	1993	NaN	20210616

Figure 7 – Capture des données manquantes

Dealing with missing values

```
data['emissions']=data['emissions'].interpolate(method="linear")
```

[14] ✓ 0.3s

Python

Figure 8 – Utiliser la méthode interpolation linéaire

data.head(7)

✓ 0.4s Python

	Country_code	Country	Format_name	Pollutant_name	Sector_code	Sector_name	Parent_sector_code	Unit	Year	emissions	PublicationDate
0	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.2	2.E.2 - TFT Flat Panel Display	2.E	Gg CO2 equivalent	2008	2	20210616
1	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.3	2.E.3 - Photovoltaics	2.E	Gg CO2 equivalent	1992	2	20210616
2	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.3	2.E.3 - Photovoltaics	2.E	Gg CO2 equivalent	2010	2	20210616
3	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.3	2.E.3 - Photovoltaics	2.E	Gg CO2 equivalent	2017	2	20210616
4	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.5	2.E.5 - Other electronics industry	2.E	Gg CO2 equivalent	1993	2	20210616
5	NL	Netherlands	IPCC Common Reporting Format	SF6 - (CO2 equivalent)	2.E.5	2.E.5 - Other electronics industry	2.E	Gg CO2 equivalent	2001	2	20210616
6	LT	Lithuania	IPCC Common Reporting Format	N2O	1.A.3.b	1.A.3.b - Road Transportation	1.A.3	Gg	2009	0.10775	20210616

Figure 9 – Données remplies

➤ La méthode dropping :

C'est une méthode qui consiste à effacer la colonne ou il y a une donnée manquante

```
#dropping the observations where there are missing values
data.dropna(inplace=True)
```

Python

Figure 10 – La méthode dropping utilisée

○ Les données catégorielles :

➤ La méthode Most frequent values :

Most frequent est une autre stratégie statistique pour implémenter les valeurs manquantes. Il fonctionne en remplaçant les données manquantes par les valeurs les plus fréquentes dans chaque colonne.

Cette méthode a été utilisée parce que les données catégorielles ne peuvent pas être remplacées en utilisant des algorithmes.

- Les Aspects positifs :

Fonctionne bien avec les fonctionnalités catégorielles (Strings ou Numerical representations)

- Les Aspects négatifs :

Cela peut introduire un biais dans les données

Ne tient pas compte des corrélations entre les Inputs

```
#replacing the missing values of the column "Parent_sector_code" with the most frequent value  
data = data.fillna(data['Parent_sector_code'].value_counts().index[0])  
✓ 0.7s
```

Figure 11 – Utilisation de la méthode most frequent values pour remplir les données

Chapitre 3 :

Comparaison des résultats :

➤ La régression linéaire multiple :

-Solution proposée :

```
Model Training

1. Linear Regression

#loading linear regression model
lin_reg_model = LinearRegression()

lin_reg_model.fit(X_train, Y_train)

... LinearRegression()

#Prediction on Training data
training_data_prediction = lin_reg_model.predict(X_train)

# R squared Error
error_score = metrics.r2_score(Y_train, training_data_prediction) #should be as low as possible
print('R squared Error:', error_score)

... R squared Error: 0.013896776978032599
```

Figure 12 – Solution proposée de la régression linéaire

-Interprétation :

En utilisant cette méthode, on a obtenu un score négligeable devant 1 (0.013)_ce qui montre que cette méthode n'est pas efficace dans notre cas.

Avec cet algorithme on n'a pas pu remplir toutes les colonnes de la base de données.

➤ **SVM (Support Vector Machine)**

-Solution proposée :

```
from sklearn.svm import SVR
svr = SVR()
svr.fit(X_train, Y_train)
```

Python

Figure 13 – Solution proposée de la SVM

-Interprétation :

En utilisant cette méthode, on n'a pas pu malheureusement obtenir des résultats.

En effet, l'algorithme n'a fait que tourner en boucle pendant des heures jusqu'à notre machine surchauffe.

➤ **DecisionTreeRegressor :**

-Solution proposée :

```
2.DecisionTreeRegressor
Markdown

from sklearn.tree import DecisionTreeRegressor
model = DecisionTreeRegressor()
[31] Python

model.fit(X_train, Y_train)
[32] Python

... DecisionTreeRegressor()

predictions = model.predict(X_test)
[33] Python

from sklearn.metrics import mean_absolute_error
model.score(X_test, predictions)
[34] Python

... 1.0
```

Figure 14 – solution proposée pour la decisionTreeRegressor

-Interprétation :

En utilisant cette méthode, on a obtenu des bons résultats. En effet, la prédiction a été précise.

Lorsqu'on a voulu tester notre modèle, on a utilisé des colonnes déjà connues dans la base de données et l'output a été presque identique à celui qu'on a déjà.

Aussi on a remarqué qu'avec cette méthode l'interpolation linéaire a donné de bons résultats

Contrairement aux autres méthodes.

On conclut que le choix de la DecisionTreeRegressor est le meilleur pour la prédiction et aussi le plus compatible avec l'interpolation linéaire pour compléter les données.

Voici quelques exemples :

```

features = np.array([[21, 20,0, 4,23,23,3,0,2014,0,0]])
print(model.predict(features))
Python

[0.0004]

C:\Users\LENOVO\AppData\Roaming\Python\Python310\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but
DecisionTreeRegressor was fitted with feature names
warnings.warn(

features = np.array([[21, 20,0, 4,32,32,0,0,2008,0,0]])
print(model.predict(features))
Python

[0.51773]

C:\Users\LENOVO\AppData\Roaming\Python\Python310\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names, but
DecisionTreeRegressor was fitted with feature names
warnings.warn(

features = np.array([[22, 21,0, 1,137,137,30,0,1992,0,0]])
print(model.predict(features))
Python

[13.2699535]

```

Figure 15 – Exemples de prédiction en utilisant la DecisionTreeRegressor

Tableau récapitulatif:

	DecisionTree	SVM	Linear Regression
Score	1	NA	0.013
Rendement avec l'ingénierie de données	Bon pour utiliser	NA	A donner de mauvais résultats avec les méthodes d'implémentation de data
Appréciation	A utiliser	NA	A éviter

Conclusion générale :

Afin d'atteindre l'objectif de notre projet, qui consiste à prédire les émissions des gaz à effet de serre dans le futur, on a eu recours à plusieurs méthodes pour résoudre cette problématique.

On a commencé avec les algorithmes de la régression linéaire multiple et SVM mais ils n'ont pas été efficaces d'où on a opté pour l'algorithme de decisionTreeRegressor qui s'est avéré qu'elle est la bonne solution grâce aux résultats fournis et qu'elle est le meilleur algorithme pour avoir une base de données complète avec les différentes méthodes d'implémentations de données manquantes.

Cependant si on avait opté pour d'autre type de méthodes comme les réseaux de neurones ou la logique floue, est ce que le résultat aurait été le même ? On ne peut le savoir qu'en essayant ces méthodes.

Bibliographie :

- [1] <https://climate.selectra.com/fr/comprendre/effet-de-serre#:~:text=L%27effet%20de%20serre%20est%20un%20ph%C3%A9nom%C3%A8ne%20naturel,et%20restitu%C3%A9%20dans%20l%27atmosph%C3%A8re.>
- [2] <https://youmatter.world/fr/definition/scope-1-2-3-definition/>
- [3] <https://www.gouvernement.fr/actualite/les-decisions-cles-de-la-cop-26-contre-le-rechauffement-climatique#:~:text=Maintenir%20le%20r%C3%A9chauffement%20climatique%20sous,Un%20accord%20juridiquement%20contraignant.>
- [4] <https://www.climate-chance.org/agenda/cop26/>
- [5] <https://towardsdatascience.com/6-different-ways-to-compensate-for-missing-values-data-imputation-with-examples-6022d9ca0779>
- [6] IPCC (2012a). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of IPCC Intergovernmental Panel on Climate Change. , eds. C. B. Field, V. Barros, T. F. Stocker, Q. Dahe, D. J. Dokken, K. L. Ebi, et al. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press doi:9781107607804.