

**MODÉLISATION ET APPLICATION DE DATA MINING AUX  
ACTIVITÉS ÉMETTRICES DE CO<sub>2</sub> : « CAS DES INDUSTRIES DE  
FABRICATION DE BOISSONS ET DES PRODUITS ALIMENTAIRES  
DANS LA COMMUNE DE LIMETE/VILLE-PROVINCE DE KINSHASA,  
EN RÉPUBLIQUE DÉMOCRATIQUE DU CONGO »**

Par

**André MUSAS-A-MUSAS**

*Assistant à l'Université de Kinshasa/Faculté des Sciences et Technologies/Mention :  
Mathématiques, Statistique et Informatique*

**Eugène MBUYI MUKENDI**

*Professeur Ordinaire à l'Université de Kinshasa/Faculté des Sciences et  
Technologies/Mention : Mathématiques, Statistique et Informatique*

**Sylvestre FREY**

*Assistant à l'Université de Kinshasa/Faculté des Sciences et Technologies/Mention :  
Mathématiques, Statistique et Informatique*

**David WEMBELO MULENDA OSTHUDI**

*Assistant à l'Université de Kinshasa/Faculté des Sciences et Technologies/Mention :  
Mathématiques, Statistique et Informatique*

**RÉSUMÉ**

*Cet article explore les données des activités émettrices de CO<sub>2</sub> des industries de fabrication de boissons et de produits alimentaires situées dans la commune de Limete, à Kinshasa, RDC, en utilisant le data mining. Le data mining, ou exploration de données, est une discipline à l'intersection de la statistique, des mathématiques, et de l'informatique, qui permet d'extraire des informations pertinentes à partir de vastes ensembles de données, facilitant ainsi la prise de décision au sein des entreprises et organisations.*

*L'absence d'une base de données dédiée aux activités émettrices de CO<sub>2</sub> dans ces industries a conduit à la création d'une nouvelle base de données d'identification et de quantification des émissions, modélisée à l'aide de la méthode Merise. Cette modélisation a été enrichie par l'approche de premier niveau des lignes directrices du Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC), ainsi que par les référentiels de facteurs d'émission de l'Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME) et de l'Agence Internationale de l'Énergie (AIE).*

*Les résultats obtenus montrent une forte dépendance aux énergies non renouvelables, représentant 69 % des émissions, contre 31 % pour les énergies renouvelables. Le modèle prédictif, construit à partir d'un arbre de décision, révèle que, quel que soit le type d'énergie utilisé par ces industries et indépendamment de leurs niveaux d'émissions en KgCO<sub>2</sub>équiv/An, les énergies seront systématiquement classées comme renouvelables ou non renouvelables.*

**Mots-clés :** *Data mining, Modélisation de données, Merise, Base de données, Arbre de décision, Référentiel des facteurs d'émissions.*

## ABSTRACT

*This article explores the CO<sub>2</sub>-emitting activities of beverage and food manufacturing industries located in the Limete municipality of Kinshasa, DRC, using data mining. Data mining, or data exploration, is a discipline at the intersection of statistics, mathematics, and computer science, which allows for the extraction of relevant information from large datasets, thereby facilitating decision-making within companies and organizations.*

*The absence of a dedicated database for CO<sub>2</sub>-emitting activities in these industries led to the creation of a new database for identifying and quantifying emissions, modeled using the Merise method. This modeling was enhanced by the first-level approach of the Intergovernmental Panel on Climate Change (IPCC) guidelines, along with the emission factor benchmarks from the Environment and Energy Management Agency (ADEME) and the International Energy Agency (IEA).*

*The results indicate a strong reliance on non-renewable energies, accounting for 69% of emissions, compared to 31% for renewable energies. The predictive model, built using a decision tree, reveals that regardless of the type of energy used by these industries and their emission levels in KgCO<sub>2</sub>equiv/Year, the energies will consistently be classified as either renewable or non-renewable.*

**Keywords:** *Data mining, Data modeling, Merise, Database, Decision tree, Emission factor benchmarks.*

## INTRODUCTION

La recherche et la collecte de données au sein des entreprises sont devenues un défi majeur. Certes, ces données sont produites presque quotidiennement par certaines entreprises, mais dans la plupart des cas, elles ne sont pas organisées de manière optimale. Elles sont souvent consignées sur des supports non durables tels que des registres ou des cahiers, et il est parfois impossible de suivre leur traçabilité une fois générées. La perte de ces supports constitue une perte significative pour ceux qui souhaiteraient exploiter ces données pour acquérir des connaissances.<sup>1,2,3</sup>

À l'ère des Nouvelles Technologies de l'Information et de la Communication (NTIC), cette organisation de données basée sur des documents papier devient

---

<sup>1</sup> Benamar Houmadi, étude exploratoire d'outils pour le data mining, Thèse, Avril 2007. <http://depot-e.uqtr.ca/id/eprint/1423/1/030000495.pdf>

<sup>2</sup> Cigref, Valorisation des données dans les grandes entreprises, Maturité, pratiques et modèle, Novembre 2016. <https://www.cigref.fr/wp/wp-content/uploads/2016/11/CIGREF-Valorisation-des-donnees-Pratiques-Modele-2016.pdf>

<sup>3</sup> Chami Djazia, une plateforme orientée agent pour le data mining, Thèse, 2009-2010. <http://eprints.univ-batna2.dz/193/1/Djazia%20CHAMI.pdf>

de plus en plus rudimentaire, voire archaïque et obsolète. L'avènement des technologies de l'information et de la communication, grâce à l'informatique, nous permet aujourd'hui d'automatiser les systèmes d'information des entreprises.<sup>4</sup>

Avec l'émergence de nouvelles structures d'organisation des données, telles que les bases de données relationnelles et les bases de données orientées objets, ces structures sont devenues indispensables pour la gestion efficace des informations au sein des entreprises ou des organisations. Grâce aux méthodes de modélisation de données (Data modeling), il est désormais possible de formaliser, schématiser ou représenter les problématiques du monde réel.<sup>5,6</sup>

D'où la nécessité de recourir à des méthodes de modélisation de données comme MERISE ou UML, chacune apportant sa spécificité pour résoudre un problème donné. Par exemple, à travers le schéma conceptuel de données et le schéma logique de données proposés par la méthode MERISE, il est possible de modéliser le monde réel ou le problème auquel on est confronté.<sup>7</sup>

Les données produites au sein de ces entreprises augmentent chaque jour, dépassant ainsi les gigaoctets. Pour certaines organisations, ces données représentent de véritables mines d'or, renfermant des connaissances cruciales pour orienter et piloter la prise de décision. Le data mining, ou exploration et fouille de données, est une discipline à l'intersection de la statistique, des mathématiques et de l'informatique, qui permet aujourd'hui d'explorer, d'analyser et d'extraire des informations utiles à partir de vastes ensembles de données, facilitant ainsi la prise de décision au sein des entreprises ou des organisations.<sup>8,9</sup>

Notre étude vise à rechercher les données existantes sur les activités émettrices de CO<sub>2</sub> exploitées par les entreprises de la ville de Kinshasa, en particulier celles situées dans les industries de fabrication de boissons et de produits alimentaires de la commune de Limete.

---

<sup>4</sup> Dominique Crié, de l'extraction des connaissances au knowledge management, 2003. <https://www.cairn.info/revue-francaise-de-gestion-2003-5-page-59.htm>

<sup>5</sup> Merati Medjeded, bases de données avancées, polycopié de cours, 2019-2020. <http://fmi.univ-tiaret.dz/images/Polycopies/Bases-de-Donnes-Avanc.pdf>

<sup>6</sup> Nicolas Larrousse, Création de bases de données, Pearson Education France, 2009. <http://tony3d3.free.fr/files/Creation-de-bases-de-donnees.pdf>

<sup>7</sup> Georges Gardarin, bases de données, Eyrolles, 2003. [http://georges.gardarin.free.fr/Livre\\_BD\\_Contentu/XX-TotalBD.pdf](http://georges.gardarin.free.fr/Livre_BD_Contentu/XX-TotalBD.pdf)

<sup>8</sup> Benamar Houmadi, étude exploratoire d'outils pour le data mining, Thèse, Avril 2007. <http://depot-e.uqtr.ca/id/eprint/1423/1/030000495.pdf>

<sup>9</sup> Chami Djazia, une plateforme orientée agent pour le data mining, Thèse, 2009-2010. <http://eprints.univ-batna2.dz/193/1/Djazia%20CHAMI.pdf>

Les activités réalisées par ces entreprises sont à l'origine des émissions de gaz à effet de serre, dont l'accumulation dans l'atmosphère entraîne une hausse des températures à l'échelle planétaire. Ces activités anthropiques sont la cause principale du phénomène de changement climatique auquel nous assistons. Fonctionnant sans interruption, ces entreprises mènent des activités intensives et non régulées, contribuant ainsi au réchauffement climatique et à la dégradation de notre environnement.<sup>10,11</sup>

Ces entreprises, telles que les industries métallurgiques, pharmaceutiques, et manufacturières (notamment celles de fabrication de boissons et de produits alimentaires), contribuent grandement au développement de notre pays sur les plans politique, économique et social. Cependant, ce développement s'accompagne malheureusement d'une négligence de l'aspect environnemental.

À Kinshasa, ces entreprises ont recours aux énergies fossiles (gaz, pétrole, charbon) et à leurs dérivés (essence, gasoil, kérosène) pour mener à bien leurs activités. Ces énergies constituent les principales sources d'émissions de gaz à effet de serre dans le monde, représentant 80 % des émissions de CO<sub>2</sub> à l'échelle mondiale, et tendent progressivement à s'épuiser.<sup>12</sup>

Actuellement, la ville de Kinshasa fait face à de nombreux problèmes environnementaux, notamment l'accumulation des déchets non biodégradables tels que les sachets et les bouteilles en plastique, la pollution de l'air causée par les émissions de gaz à effet de serre provenant des moteurs de véhicules et des groupes électrogènes, ainsi que des inondations fréquentes dues à l'absence d'entretien des caniveaux. Ces problèmes ont des conséquences graves, comme la prolifération de maladies telles que le choléra et le paludisme.

Pour comprendre de manière précise comment les activités des entreprises contribuent aux émissions de CO<sub>2</sub>, nous avons d'abord soumis une requête au Secrétariat du ministère de l'Environnement et Développement Durable. Cette démarche a été accueillie favorablement, et nous avons été orientés vers la Direction du Développement Durable. De là, nous avons été dirigés à la Division du Changement Climatique, où nous avons pris conscience de la

---

<sup>10</sup> Changement climatique, genre et santé, Organisation mondiale de la Santé, 2016.  
[https://apps.who.int/iris/bitstream/handle/10665/204177/9789242508185\\_fre.pdf](https://apps.who.int/iris/bitstream/handle/10665/204177/9789242508185_fre.pdf)

<sup>11</sup> Harat Ahmed, Effet de gaz à effet de serre sur la santé humaine, Avril 2017.  
[https://www.researchgate.net/publication/316739125\\_EFFET\\_DE\\_GAZ\\_A\\_EFFET\\_DE\\_SERR\\_E\\_SUR\\_LA\\_SANTE\\_HUMAINE](https://www.researchgate.net/publication/316739125_EFFET_DE_GAZ_A_EFFET_DE_SERR_E_SUR_LA_SANTE_HUMAINE)

<sup>12</sup> Le changement climatique, ADEME, Juin 2012.  
<https://www.ademe.fr/sites/default/files/assets/documents/guide-changement-climatique.pdf>

réalité des inventaires nationaux des gaz à effet de serre attribuables aux activités industrielles à Kinshasa.

Depuis que la République Démocratique du Congo (RDC) a ratifié des accords internationaux majeurs, tels que la Convention-cadre des Nations unies sur les changements climatiques, le Protocole de Kyoto, et l'Accord de Paris, le pays s'est engagé à lutter contre le changement climatique. Une part importante des émissions de gaz à effet de serre provient de divers secteurs, notamment l'énergie, l'industrie, les transports, l'agriculture, la foresterie et d'autres usages des terres. Les données concernant ces secteurs sont issues de divers documents, tels que les rapports de la Banque Centrale du Congo (BCC) et l'annuaire de l'Institut National de la Statistique (INS).<sup>13</sup>

Après avoir extrait ces données, elles sont compilées dans des matrices de calcul, où toutes les méthodes proposées par les lignes directrices du Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC) sont appliquées pour produire les résultats des inventaires nationaux des gaz à effet de serre, comme exigé par les instances internationales. Cependant, il a été constaté qu'il n'existait pas de base de données spécifique pour les activités émettrices de CO<sub>2</sub> des industries de fabrication de boissons et de produits alimentaires, comme cela aurait été nécessaire.

Face à ce manque crucial de données pour répondre à notre requête, nous nous sommes interrogés : « *Comment identifier et quantifier efficacement les différentes sources d'énergie utilisées par les entreprises qui sont responsables des émissions de CO<sub>2</sub> ?* » Cette question est au cœur de notre étude, particulièrement dans le contexte des industries de fabrication de boissons et de produits alimentaires.

La modélisation du système d'identification et de quantification des activités émettrices de CO<sub>2</sub> par les industries de fabrication de boissons et de produits alimentaires pourrait nous permettre de concevoir une base de données dédiée à ces activités, sur laquelle les outils de data mining seraient appliqués pour analyser les données.

L'objectif principal de cette étude est de développer et de tester un outil informatique d'aide à la décision, visant à sensibiliser sur les enjeux environnementaux et à contribuer à la lutte contre le changement climatique dans la Ville-Province de Kinshasa. Pour atteindre cet objectif général, plusieurs actions spécifiques ont été définies :

---

<sup>13</sup> Programme - Pays de la RDC au Fonds Vert Climat (GCF), RDC, 20 Février 2019. <https://www.greenclimate.fund/sites/default/files/document/congo-country-programme.pdf>

- Modéliser le système d'identification et de quantification des activités émettrices de CO<sub>2</sub> des entreprises ;
- Concevoir une base de données dédiée à ces activités, permettant leur identification et leur quantification ;
- Analyser les données à l'aide d'outils de data mining pour identifier les tendances et relations entre les variables, et construire un modèle prédictif de classification automatique basé sur des arbres de décision.

L'objectif global de cette étude est de sensibiliser les entreprises et les acteurs concernés aux défis posés par le changement climatique. En utilisant les données fournies par ces entreprises, nos analyses les orienteront dans leurs efforts pour réduire les émissions de CO<sub>2</sub>, principalement causées par l'utilisation d'énergies non renouvelables (comme l'essence, le gasoil, et autres sources d'énergie). Ces analyses les aideront également à faire des choix éclairés quant aux équipements à adopter dans leurs installations.

Dans le cadre de cette recherche, nous avons identifié et quantifié les activités émettrices de CO<sub>2</sub> des entreprises de la Ville-Province de Kinshasa, en nous focalisant sur les industries de fabrication de boissons et de produits alimentaires situées dans la commune de Limete, pour la période allant de 2018 à 2020.

## MILIEU D'ÉTUDE

### 1.1. Description de la Commune de Limite

Limete est une commune de la ville province de Kinshasa en République démocratique du Congo. Elle se situe dans la partie Est de la ville, entre le Pool Malebo (dont elle occupe la façade Sud-Ouest) et le Boulevard Lumumba. C'est une commune accueillant de nombreuses industries et la majorité des installations portuaires de Kinshasa. Elle est traversée par la ligne de chemin de fer Matadi-Kinshasa. La commune héberge également la Foire internationale de Kinshasa ou FIKIN, dont l'entrée principale se trouve en face de la Commune de Lemba. C'est à partir de cette commune qu'on a développé le Quartier de la Cité du Fleuve Congo, gagné sur les eaux. Elle compte sur sa liste 14 quartiers ci-après : Mfumu Mvula, Kingabwa, Mombele, Mososo, Quartier industriel, Salongo, Quartier Résidentiel, Masiala, Mateba, Mayulu, Agricole, Ndanu, Mbamu, Nzadi.<sup>14</sup>

Avant la date de sa création 1958 la superficie de la commune était de 23,73 Km<sup>2</sup>, et après la promulgation de l'ordonnance-loi n°68-018 bis du 02 décembre 1968, elle passa à 67,60 Km<sup>2</sup>. La Densité de la population est le résultat du dernier recensement administratif de 2007, qui relève le nombre de la population à 242 531 habitants avec une densité de 3.57 habitants par Km<sup>2</sup>. Elle

---

<sup>14</sup> <https://fr.wikipedia.org/wiki/Limete>

est installée sur une plaine sablonneuse, d'un climat tropical, avec une pluviométrie étendue remarquable, c'est la raison pour laquelle quelques quartiers sont inondés à chaque fois qu'il y a des fortes pluies. La commune a une température de 23°C à 36°C pour la simple raison de sa proximité avec le Fleuve Congo. Et elle est baignée par trois rivières dont : la Rivière Funa et le Fleuve Congo au Nord ; la Rivière Yolo à l'Ouest et la Rivière N'djili à l'Est.<sup>15</sup>

## 1.2. Carte administrative de la Commune de Limete

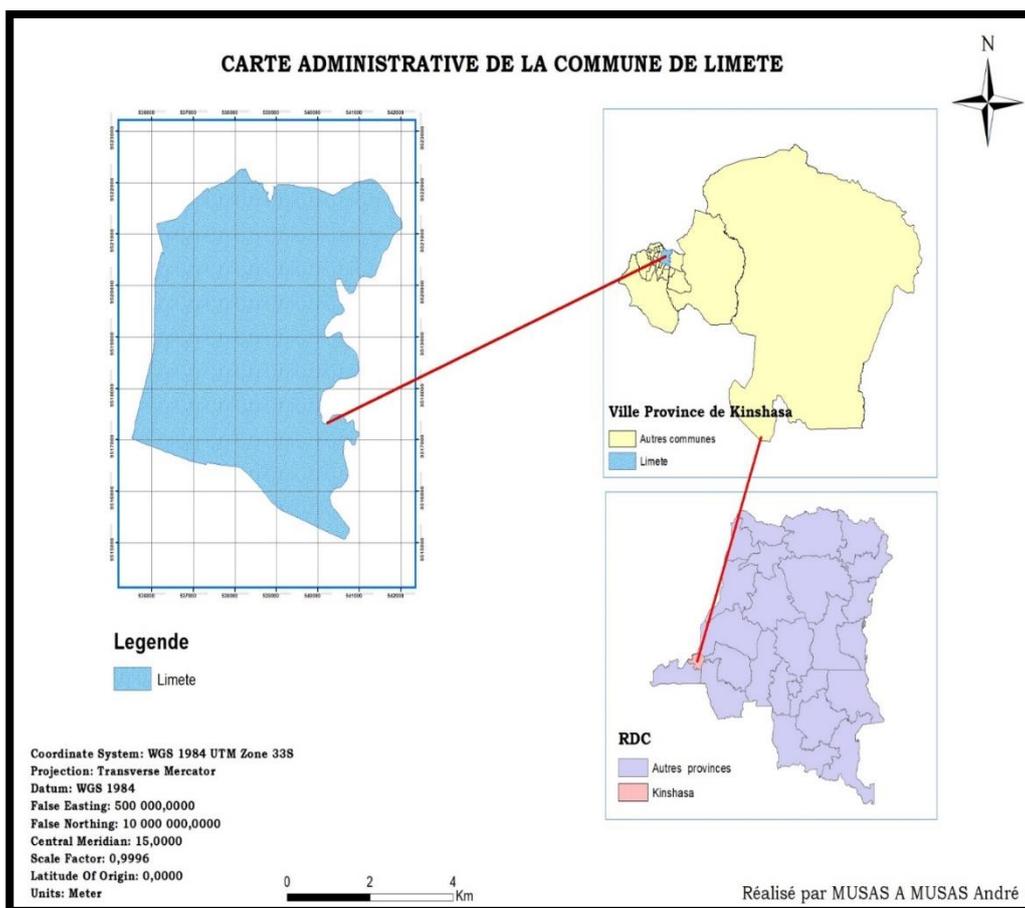


Figure 1.1 : Carte administrative de la Commune de Limete (Source : MUSAS-A-MUSAS André)

<sup>15</sup> [https://www.congovirtuel.com/page\\_rapport\\_travaux/rapport\\_unikin/rapport\\_ndjodji.pdf](https://www.congovirtuel.com/page_rapport_travaux/rapport_unikin/rapport_ndjodji.pdf)

### 1.3. Cas d'étude : Industries de fabrication de boissons et de produits alimentaires

#### a) Situation industrielle dans la ville province de Kinshasa

D'après le rapport annuel de l'Institut National de Statistique de 2015, la ville province de Kinshasa renferme un nombre important d'industries dans son actif, on peut en dénombrer plus d'une centaine. Les tableaux 1 et 2 montrent respectivement la répartition de la situation industrielle de la ville de Kinshasa et de la commune de Limete. <sup>16</sup>

**Tableau 1 : Répartition de la situation industrielle dans la ville de Kinshasa**

Code Citi	Branches d'activités	Effectifs
10	Fabrication de produits alimentaires	46
11	Fabrication de boissons	23
12	Fabrication de produits à base de tabac	1
13	Fabrication de textiles	1
15	Fabrication de cuir et articles en cuir	1
16	Production bois et articles en bois	9
17	Fabrication de papier et articles en papier	3
18	Imprimerie et reproduction supports enregistrés	12
19	Cokéfaction et fabrication de produits pétroliers raffinés	2
20	Fabrication de produits chimiques	28
21	Fabrication de préparation pharmaceutique	8
22	Fabrication d'articles en caoutchouc et matières en plastiques	27
23	Fabrication de produits minéraux non métalliques	15
24	Fabrication de produits métallurgiques de base	9
25	Fabrication d'ouvrages en métaux	11
27	Fabrication de matériels électriques	1
31	Fabrication de meubles	3
	<b>Total</b>	<b>200</b>

Source : INS, 2015

<sup>16</sup> Annuaire statistique 2015, Mars 2017

<https://www.ins.cd/wp-content/uploads/2021/04/Annuaire-statistique-2015-Web.pdf>

**Tableaux 2 : Répartition de la situation industrielle dans la commune de Limete**

Commune	2013	2014	Part en % 2013	Part en % 2014
Limete	134	132	67	69,1
Gombe	37	35	18,5	18,3
Autres commune	29	24	14,5	12,6
<b>Total</b>	<b>200</b>	<b>191</b>	<b>100</b>	<b>100,0</b>

Source : INS, 2015

#### 1.4. Choix de la zone d'étude et du cas d'étude

La communication nationale sur l'inventaire de gaz à effet de serre de 1994 jusqu'à 2003 révèle, les principales activités contribuant à l'émission de CO<sub>2</sub> dans le secteur d'utilisation de combustibles fossiles à savoir les industries de production d'énergie, les industries de manufacture et de construction, le transport terrestre, les activités commerciales, institutionnelles, agriculture, la foresterie et la pêche ainsi que le transport ferroviaire, fluvial et maritime.<sup>17</sup>

En effet, le choix de la Commune de Limete se justifie par le nombre important des industries qu'occupent son territoire et par leurs secteurs d'activités. Notre étude étant limité dans le temps et dans l'espace, nous avons collecté les données auprès des industries de fabrication de boissons et de produits alimentaires (tous, des industries manufacturières).

Ces données ne seront pas seulement stockées dans une base de données préalablement créée à partir d'un système d'identification de sources énergétiques émettrices de CO<sub>2</sub>, mais aussi explorées par les outils de data mining pour voir les différents liens ou tendances qui existent réellement entre les variables ou attributs. Ainsi, nous avons eu à collecter les données dans quatre entreprises réparties de la manière suivante :

- Une entreprise de fabrication de Boisson ;
- Trois entreprises de fabrication de produits alimentaires.

## MATÉRIELS ET MÉTHODES

#### 1.5. Matériels ou outils utilisés

La réalisation d'une étude complexe, comme celle visant à identifier et quantifier les activités émettrices de CO<sub>2</sub> dans les industries, exige une infrastructure technique appropriée. La réussite de la collecte, du traitement et de l'analyse des données dépend autant des méthodes employées que des outils technologiques utilisés. Le choix des matériels et des logiciels est essentiel pour assurer la précision des modèles, la gestion efficace des données, et la production de résultats fiables.

<sup>17</sup> Programme - Pays de la RDC au Fonds Vert Climat (GCF), RDC, 20 Février 2019. <https://www.greenclimate.fund/sites/default/files/document/congo-country-programme.pdf>

L'intégration de l'informatique dans la recherche environnementale permet d'automatiser les processus, d'améliorer la gestion des informations, et d'appliquer des techniques avancées de modélisation et de data mining. Les logiciels, tels que les systèmes d'information géographique (SIG) et les bases de données relationnelles, offrent des capacités accrues pour manipuler et visualiser de grandes quantités de données, tandis que les outils de data mining permettent d'extraire des informations clés à partir de ces données.

Dans ce contexte, il est indispensable de disposer de matériels robustes et de logiciels spécialisés pour répondre aux exigences techniques de l'étude. Les équipements informatiques sélectionnés doivent avoir une puissance de calcul suffisante pour exécuter des tâches complexes, comme le traitement de grands volumes de données et l'application de modèles prédictifs.

Voici les principaux matériels et outils utilisés dans le cadre de cette étude :

- **Ordinateur portable**

**Tableau 3 : Caractéristiques hardware et software de l'ordinateur utilisé**

<i>Marque</i>	<i>Dell Latitude E6410</i>
<i>Système d'exploitation</i>	<i>Windows 10 Professionnel 64-bit</i>
<i>Processeur</i>	<i>Intel(R) Core (TM) i5 CPU M 560 @ 2.67GHz (4 CPUs), ~2.7GHz</i>
<i>Mémoire</i>	<i>4096MB RAM</i>

- **ArcMap :**

Ce composant clé de la suite de logiciels ArcGIS d'Esri nous a permis de créer la carte administrative de la commune de Limete, qui constitue notre zone d'étude, comme présenté dans le chapitre précédent.

- **Microsoft Access 2016**

Après avoir modélisé le système d'identification et de quantification des sources d'énergie utilisées par les entreprises à l'aide de la méthode MERISE, et appliqué l'équation d'estimation des émissions de CO<sub>2</sub> proposée par les lignes directrices du GIEC, nous avons créé une base de données. Cette base a été mise en place en utilisant Microsoft Access 2016, un système de gestion de bases de données développé par Microsoft. Cet outil nous a permis de créer, stocker, organiser, et gérer efficacement les données recueillies, facilitant ainsi leur manipulation et visualisation.

- **Tanagra**

Nous avons utilisé Tanagra pour appliquer le data mining et explorer notre base de données. Tanagra est un logiciel gratuit de data mining conçu pour l'enseignement et la recherche, qui nous a permis d'analyser en profondeur les données collectées.<sup>18</sup>

<sup>18</sup> [https://eric.univ-lyon2.fr/ricco/tanagra/fichiers/le\\_logiciel\\_tanagra\\_egc\\_2005.pdf](https://eric.univ-lyon2.fr/ricco/tanagra/fichiers/le_logiciel_tanagra_egc_2005.pdf)

- **Sipina**

Sipina, un autre logiciel libre de data mining destiné à l'enseignement et à la recherche, a été utilisé pour construire un modèle prédictif. Ce logiciel offre une large gamme d'algorithmes pour la construction d'arbres de décision. Pour notre jeu de données, nous avons utilisé l'algorithme Improved CHAID (CHI-squared Automatic Interaction Detector), une version améliorée de l'algorithme CHAID.<sup>19</sup>

## 1.6. Méthodes

Pour atteindre les objectifs de cette étude, nous avons mis en place une méthodologie combinant plusieurs méthodes de recherche rigoureuses, adaptées à la complexité des données à analyser. Cette approche repose sur l'intégration de méthodes quantitatives et de modélisation, permettant de structurer, quantifier et analyser les émissions de CO<sub>2</sub> des industries concernées.

- Tout d'abord, nous avons utilisé la méthode Merise, une méthode de modélisation systémique bien établie dans la conception et la gestion des systèmes d'information. Cette méthode a été cruciale pour créer un modèle d'identification et de quantification des activités émettrices de CO<sub>2</sub> des entreprises. Grâce à Merise, nous avons pu élaborer une base de données structurée, permettant de stocker et de gérer efficacement les informations nécessaires à notre étude.
- Ensuite, pour estimer les émissions de gaz à effet de serre, nous avons appliqué l'approche de Niveau 1 proposée par les lignes directrices du Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC). Cette méthode quantitative nous a permis de calculer les émissions de CO<sub>2</sub> résultant de l'utilisation de combustibles stationnaires dans les entreprises, à l'aide d'une équation spécifique. Cette équation a été enrichie avec les référentiels de facteurs d'émission de l'Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME) et de l'Agence Internationale de l'Énergie (AIE), garantissant ainsi la précision des résultats.
- Enfin, nous avons utilisé une méthode prédictive pour développer un arbre de classification, un outil puissant pour analyser les tendances et les relations présentes dans les données. Cette technique de modélisation prédictive, employée dans le data mining, nous a permis de concevoir un modèle capable de classer automatiquement les données selon des critères prédéterminés, offrant ainsi des perspectives précieuses pour la prise de décision.

---

<sup>19</sup> <https://www.rocq.inria.fr/axis/modulad/archives/numero-33/tutorial-rakotomalala-33/rakotomalala-33-tutorial.pdf>

Le graphique suivant illustre la méthodologie adoptée pour cette étude, résumant les principales étapes de manière séquentielle. Le diagramme commence par l'initialisation de l'étude, suivie de l'application de la méthode Merise pour modéliser le système d'identification et de quantification des activités émettrices de CO<sub>2</sub>. Il continue avec l'application de l'approche de Niveau 1 du GIEC pour le calcul des émissions, en utilisant les référentiels de l'ADEME et de l'AIE. Enfin, la méthodologie se termine par l'utilisation d'une méthode prédictive pour analyser les données et construire un modèle d'arbre de classification, avant de conclure l'étude.

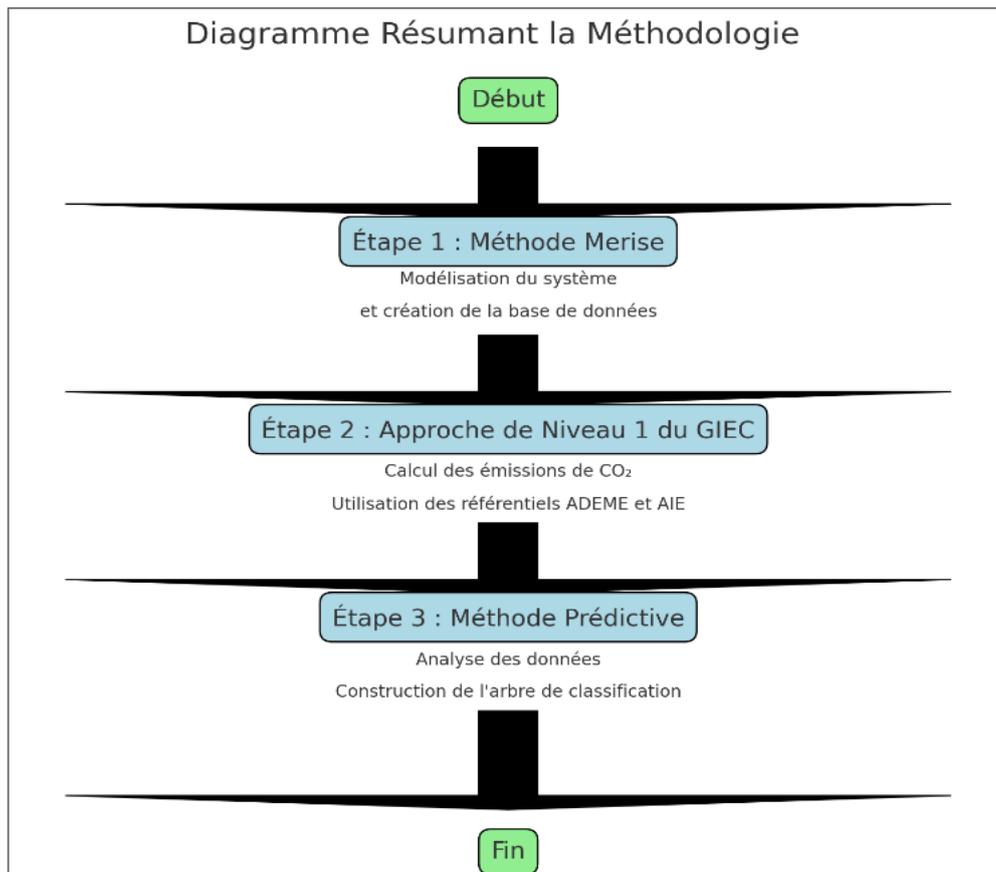


Figure 2.2 : Diagramme résumant la méthodologie

## RÉSULTATS

### 1.7. Implémentation de l'application

Les figures ci-dessous illustrent quelques formulaires de l'implémentation de notre base de données :

#### a) Formulaire graphique

La mise en relation des tables issues du modèle logique données sous Access 2016 :

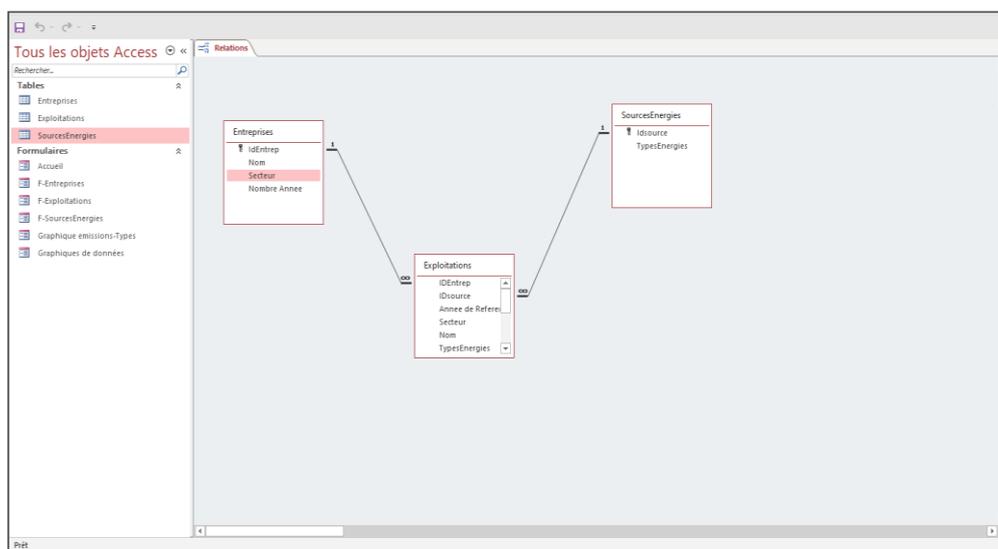


Figure 3.1: Formulaire de mise relation des tables (Entreprises, Sources d'énergie)

#### b) Formulaire d'accueil



Figure 3.2 : Formulaire d'accueil (Source image arrière-plan <sup>20</sup>)

<sup>20</sup> <https://www.espace-aubade.fr/blog/chauffage/se-chauffer-energies-renouvelables.html>

### c) Formulaire de saisie de données

IDsource	Annee de R.	Secteur	Nom	TypesEnerg.	Fournisseur	Equipement_Utilise
EL1	2018	ALIMENTAIRE	GFCD	Electricité	CABINE PRIVEE	RESIDENCE
GA1	2018	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC CATERPILAR/1C
GA1	2018	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC VOLVO/15
GA1	2018	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC PERKINS/1
ES1	2018	ALIMENTAIRE	GFCD	Essence	STATION	VEHICULE
EL1	2019	ALIMENTAIRE	GFCD	Electricité	CABINE PRIVEE	RESIDENCE
GA1	2019	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC CATERPILA
GA1	2019	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC VOLVO/15

*Figure 3.3: Formulaire de saisie de données*

### d) Formulaire de la table d'analyse

IDEntrep	IDsource	Annee de R.	Secteur	Nom	TypesEnerg.	Fournisseur	Equipement_Utilise	Quantite-Cc	Unite	FacteurEmis
GF1	EL1	2018	ALIMENTAIRE	GFCD	Electricité	CABINE PRIVEE	RESIDENCE	1	Kwh/An	0,00
GF1	GA1	2018	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC CATERPILAR/1C	1	Litres/An	3,1
GF1	GA1	2018	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC VOLVO/150KV	1	Litres/An	3,1
GF1	GA1	2018	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC PERKINS/100K\	1	Litres/An	3,1
GF1	ES1	2018	ALIMENTAIRE	GFCD	Essence	STATION	VEHICULE	1	Litres/An	2,7
GF1	EL1	2019	ALIMENTAIRE	GFCD	Electricité	CABINE PRIVEE	RESIDENCE	1	Kwh/An	0,00
GF1	GA1	2019	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC CATERPILAR/1C	1	Litres/An	3,1
GF1	GA1	2019	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC VOLVO/150KV	1	Litres/An	3,1
GF1	GA1	2019	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC PERKINS/100K\	1	Litres/An	3,1
GF1	ES1	2019	ALIMENTAIRE	GFCD	Essence	STATION	VEHICULE	1	Litres/An	2,7
GF1	EL1	2020	ALIMENTAIRE	GFCD	Electricité	CABINE PRIVEE	RESIDENCE	1	Kwh/An	0,00
GF1	GA1	2020	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC CATERPILAR/1C	1	Litres/An	3,1
GF1	GA1	2020	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC VOLVO/150KV	1	Litres/An	3,1
GF1	GA1	2020	ALIMENTAIRE	GFCD	Gasoil	STATION	GRUPE ELEC PERKINS/100K\	1	Litres/An	3,1
GF1	ES1	2020	ALIMENTAIRE	GFCD	Essence	STATION	VEHICULE	1	Litres/An	2,7
MB1	EL1	2018	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	17510	Kwh/An	0,00
MB1	GA1	2018	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	GRUPE ELEC PERKINS/105K\	113150	Litres/An	3,1
MB1	GA1	2018	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	VEHICULE	15330	Litres/An	3,1
MB1	EL1	2019	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	22794	Kwh/An	0,00
MB1	GA1	2019	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	GRUPE ELEC PERKINS/105K\	113150	Litres/An	3,1
MB1	GA1	2019	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	VEHICULE	15330	Litres/An	3,1
MB1	EL1	2020	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	25027	Kwh/An	0,00
MB1	GA1	2020	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	GRUPE ELEC PERKINS/105K\	113150	Litres/An	3,1
MB1	GA1	2020	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	VEHICULE	15330	Litres/An	3,1
ME1	EL1	2018	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	40609	Kwh/An	0,00
ME1	GA1	2018	BOISSON	MEGA CONGO	Gasoil	STATION TOTAL	VEHICULE	219000	Litres/An	3,1
ME1	GA1	2018	BOISSON	MEGA CONGO	Gasoil	SEP	GRUPE ELEC POWER TECH	1	Litres/An	3,1
ME1	EL1	2019	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	104036	Kwh/An	0,00
ME1	GA1	2019	BOISSON	MEGA CONGO	Gasoil	STATION TOTAL	VEHICULE	219000	Litres/An	3,1
ME1	GA1	2019	BOISSON	MEGA CONGO	Gasoil	SEP	GRUPE ELEC POWER TECH	1	Litres/An	3,1
ME1	EL1	2020	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	196363	Kwh/An	0,00

*Figure 3.4: Formulaire de la table d'analyse*

### e) Formulaire de graphiques de données

Il affiche les diagrammes de répartitions des effectifs des variables Types d'énergie, Secteur et Classe ci-après :

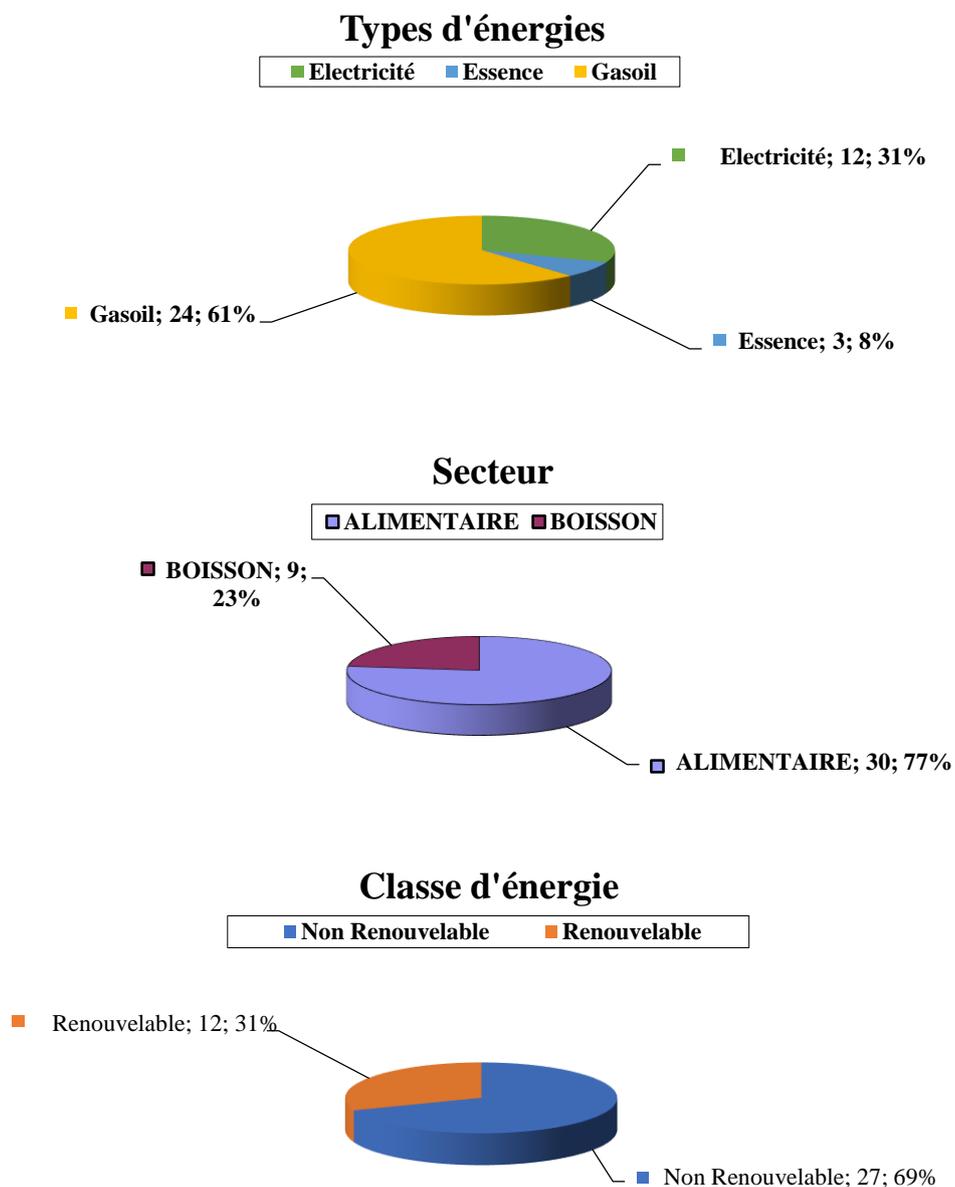


Figure 3.5: Diagrammes de répartitions des effectifs des variables Types d'énergie, Secteur et Classe d'énergie

1.8. Résultats obtenus avec TANAGRA

Les figures ci-dessous nous illustrent comment nous avons exporté la table d'analyse et générer notre premier jeu de données (Dataset) avec Tanagra.

a) Formulaire d'exportation de la Table d'analyse via Access 2016

IDsource	Année de Ré	Secteur	Nom	TypesEnerg	Fournisseur	Equipement_Utilise	Quantité-Cr
GF1	2018	ALIMENTAIRE	GFCC	Electricité	CABINE PRIVEE	RESIDENCE	1
GF1	2018	ALIMENTAIRE	GFCC	Gasoil	STATION	GROUPE ELEC CATERPILAR/1C	1
GF1	2018	ALIMENTAIRE	GFCC	Gasoil	STATION	GROUPE ELEC VOLVO/150KV	1
GF1	2018	ALIMENTAIRE	GFCC	Gasoil	STATION	GROUPE ELEC PERKINS/100KI	1
GF1	2018	ALIMENTAIRE	GFCC	Essence	STATION	VEHICULE	1
GF1	2019	ALIMENTAIRE	GFCC	Electricité	CABINE PRIVEE	RESIDENCE	1
GF1	2019	ALIMENTAIRE	GFCC	Gasoil	STATION	GROUPE ELEC CATERPILAR/1C	1
GF1	2019	ALIMENTAIRE	GFCC	Gasoil	STATION	GROUPE ELEC VOLVO/150KV	1
GF1	2019	ALIMENTAIRE	GFCC	Gasoil	STATION	GROUPE ELEC PERKINS/100KI	1
GF1	2019	ALIMENTAIRE	GFCC	Essence	STATION	VEHICULE	1
GF1	2020	ALIMENTAIRE	GFCC	Electricité	CABINE PRIVEE	RESIDENCE	1
GF1	2020	ALIMENTAIRE	GFCC	Gasoil	STATION	GROUPE ELEC CATERPILAR/1C	1
GF1	2020	ALIMENTAIRE	GFCC	Gasoil	STATION	GROUPE ELEC VOLVO/150KV	1
GF1	2020	ALIMENTAIRE	GFCC	Essence	STATION	VEHICULE	1
GF1	2018	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	17510
GF1	2018	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	GROUPE ELEC PERKINS/105KI	113150
GF1	2018	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	VEHICULE	15330
GF1	2019	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	22794
GF1	2019	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	GROUPE ELEC PERKINS/105KI	113150
GF1	2019	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	VEHICULE	15330
GF1	2020	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	25027
GF1	2020	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	GROUPE ELEC PERKINS/105KI	113150
GF1	2020	ALIMENTAIRE	MBB	Gasoil	STATION TOTAL	VEHICULE	15330
GF1	2018	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	40609
GF1	2018	BOISSON	MEGA CONGO	Gasoil	STATION TOTAL	VEHICULE	219000
GF1	2019	BOISSON	MEGA CONGO	Gasoil	SEP	GROUPE ELEC POWER TECH	1
GF1	2019	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	104036
GF1	2019	BOISSON	MEGA CONGO	Gasoil	STATION TOTAL	VEHICULE	219000
GF1	2019	BOISSON	MEGA CONGO	Gasoil	SEP	GROUPE ELEC POWER TECH	1
GF1	2020	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	196363

Figure 3.6: Formulaire d'exportation de la table d'analyse

b) Formulaire d'ouverture d'Excel avec la macro complémentaire Tanagra

The screenshot shows an Excel spreadsheet with a data table and a dialog box. The spreadsheet has columns labeled A through K and rows numbered 1 through 31. The data table contains information about various entities, including their names, types, energy sources, and emissions. A dialog box is open in the center, titled 'X', with the text 'Dataset range (including the name of the attributes -- first row):' and a text input field containing '\$A\$1:\$K\$40'. The dialog box has 'OK' and 'Cancel' buttons.

A	B	C	D	E	F	G	H	I	J	K
Année de Référence	Secreur	Nom	Type/Energies	Fournisseur	Equipement Utilise	Quantite-Consomme	Unité	Facteur Emission	Emission (kgCO2e/An)	Classe
2018	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	40609	Kwh/An	0,003	121,827	Renouvelable
2018	BOISSON	MEGA CONGO	Gasol	STATION TOTAL	VEHICULE	219000	Litres/An	3,17	694230	Non Renouvelable
2018	BOISSON	MEGA CONGO	Gasol	SEP	GRUPE ELEC POWER TECH	1	Litres/An	3,17	3,17	Non Renouvelable
2019	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	104006	Kwh/An	0,003	312,108	Renouvelable
2019	BOISSON	MEGA CONGO	Gasol	STATION TOTAL	VEHICULE	219000	Litres/An	3,17	694230	Non Renouvelable
2019	BOISSON	MEGA CONGO	Gasol	SEP	GRUPE ELEC POWER TECH	1	Litres/An	3,17	3,17	Non Renouvelable
2020	BOISSON	MEGA CONGO	Electricité	SNEL	RESIDENCE	589089	Kwh/An	0,003	589,089	Renouvelable
2020	BOISSON	MEGA CONGO	Gasol	STATION TOTAL	VEHICULE	694230	Litres/An	3,17	694230	Non Renouvelable
2020	BOISSON	MEGA CONGO	Gasol	SEP	RESIDENCE	1	Litres/An	3,17	3,17	Non Renouvelable
2018	ALIMENTAIRE	PAIN DE VILLE	Electricité	SNEL	STATION TOTAL	VEHICULE	1	Litres/An	0,003	21,6
2018	ALIMENTAIRE	PAIN DE VILLE	Gasol	STATION TOTAL	VEHICULE	1	Litres/An	3,17	69423	Non Renouvelable
2019	ALIMENTAIRE	PAIN DE VILLE	Electricité	SNEL	STATION TOTAL	VEHICULE	1	Litres/An	0,003	21,6
2019	ALIMENTAIRE	PAIN DE VILLE	Gasol	STATION TOTAL	VEHICULE	1	Litres/An	3,17	69423	Non Renouvelable
2020	ALIMENTAIRE	PAIN DE VILLE	Electricité	SNEL	STATION TOTAL	VEHICULE	1	Litres/An	0,003	21,6
2020	ALIMENTAIRE	PAIN DE VILLE	Gasol	STATION TOTAL	VEHICULE	1	Litres/An	3,17	69423	Non Renouvelable
2018	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	17510	Kwh/An	0,003	52,53	Renouvelable
2018	ALIMENTAIRE	MBB	Gasol	STATION TOTAL	GRUPE ELEC PERKINS /DSKV	113150	Litres/An	3,17	358665,5	Non Renouvelable
2018	ALIMENTAIRE	MBB	Gasol	STATION TOTAL	VEHICULE	15330	Litres/An	3,17	48596,1	Non Renouvelable
2019	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	22794	Kwh/An	0,003	66,382	Renouvelable
2019	ALIMENTAIRE	MBB	Gasol	STATION TOTAL	GRUPE ELEC PERKINS /DSKV	113150	Litres/An	3,17	358665,5	Non Renouvelable
2019	ALIMENTAIRE	MBB	Gasol	STATION TOTAL	VEHICULE	15330	Litres/An	3,17	48596,1	Non Renouvelable
2020	ALIMENTAIRE	MBB	Electricité	SNEL	RESIDENCE	25027	Kwh/An	0,003	75,081	Renouvelable
2020	ALIMENTAIRE	MBB	Gasol	STATION TOTAL	GRUPE ELEC PERKINS /DSKV	113150	Litres/An	3,17	358665,5	Non Renouvelable
2020	ALIMENTAIRE	MBB	Gasol	STATION TOTAL	VEHICULE	15330	Litres/An	3,17	48596,1	Non Renouvelable
2018	ALIMENTAIRE	GFCO	Electricité	CABINE PRIVEE	RESIDENCE	1	Kwh/An	0,003	0,003	Renouvelable
2018	ALIMENTAIRE	GFCO	Gasol	STATION	GRUPE ELEC CATERPILAR /1000KV	1	Litres/An	3,17	3,17	Non Renouvelable
2018	ALIMENTAIRE	GFCO	Gasol	STATION	GRUPE ELEC VOLVO /150KV	1	Litres/An	3,17	3,17	Non Renouvelable
2018	ALIMENTAIRE	GFCO	Gasol	STATION	GRUPE ELEC PERKINS /DSKV	1	Litres/An	3,17	3,17	Non Renouvelable
2018	ALIMENTAIRE	GFCO	Essence	CABINE PRIVEE	VEHICULE	1	Litres/An	2,79	2,79	Non Renouvelable
2019	ALIMENTAIRE	GFCO	Electricité	CABINE PRIVEE	RESIDENCE	1	Kwh/An	0,003	0,003	Renouvelable

Figure 3.7: Interface d'ouverture d'Excel avec la macro complémentaire Tanagra

### c) Formulaire d'analyse avec Tanagra et création de notre premier jeu de données (Dataset)

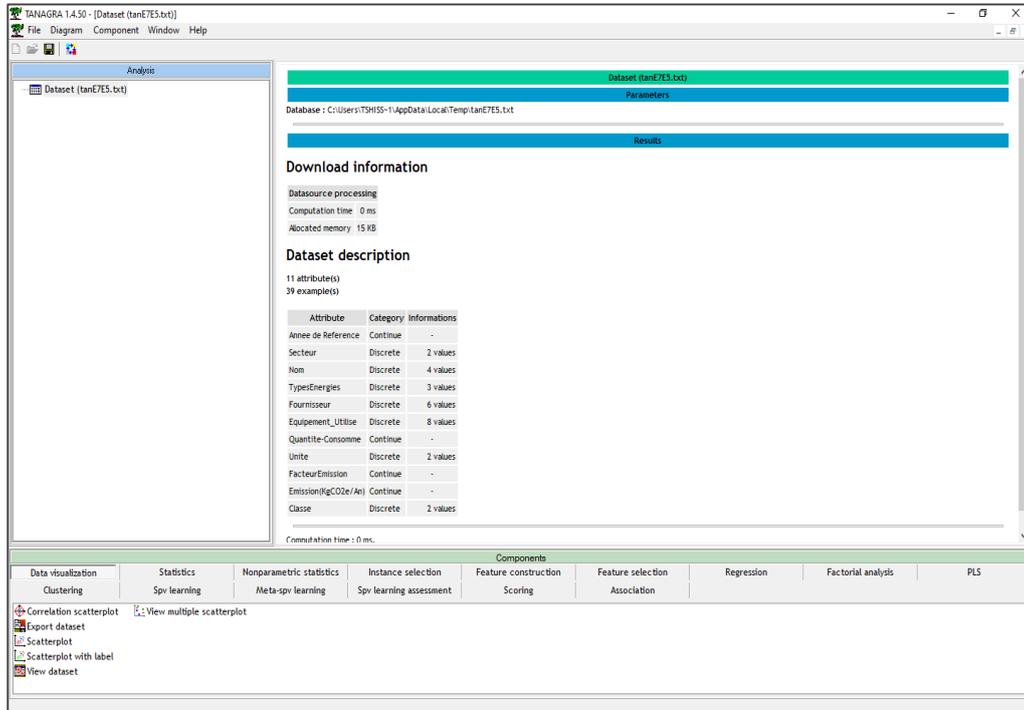


Figure 3.8: Formulaire d'ouverture d'Excel avec la macro complémentaire Tanagra

**Tableau 4 :** Répartition des effectifs en pourcentages de variables qualitatives  
 Secteur, Types d'énergie et Classe et leurs histogrammes après  
 avoir eu appliqué la statistique univariée qualitative.

**RESULTS**

Attribute	Gini	Distribution			
Secteur	0,3550	Values	Count	Percent	Histogram
		BOISSON	9	23,08 %	
		ALIMENTAIRE	30	76,92 %	
TypesEnergies	0,5207	Values	Count	Percent	Histogram
		Electricité	12	30,77 %	
		Gasoil	24	61,54 %	
		Essence	3	7,69 %	
Classe	0,4260	Values	Count	Percent	Histogram
		Renouvelable	12	30,77 %	
		Non Renouvelable	27	69,23 %	

**d) Exploration multi variable**

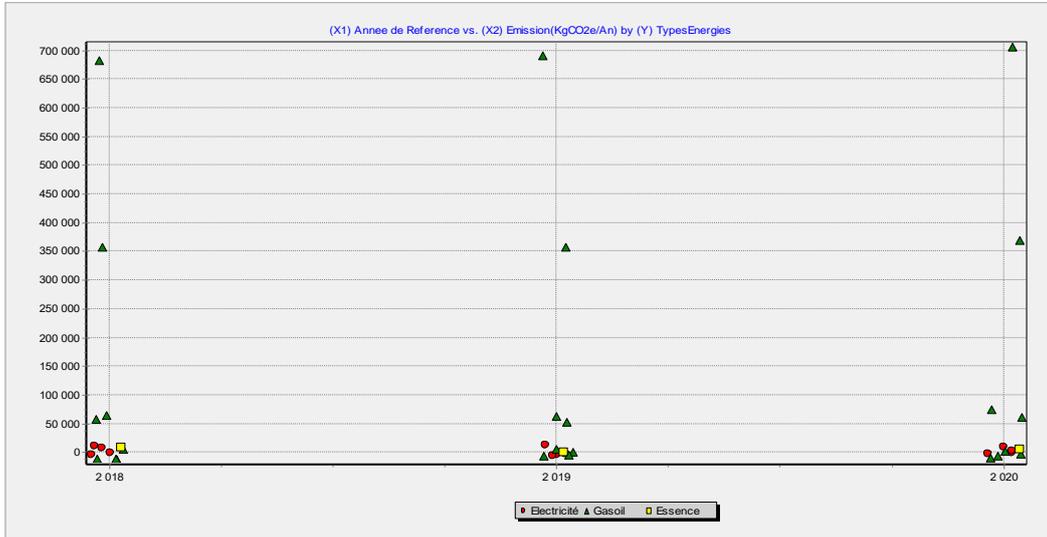


Figure 3.9: Exploration multi variable appliquée sur les variables émission en KgCO<sub>2</sub>équiv, Année de référence et Types d'énergie.

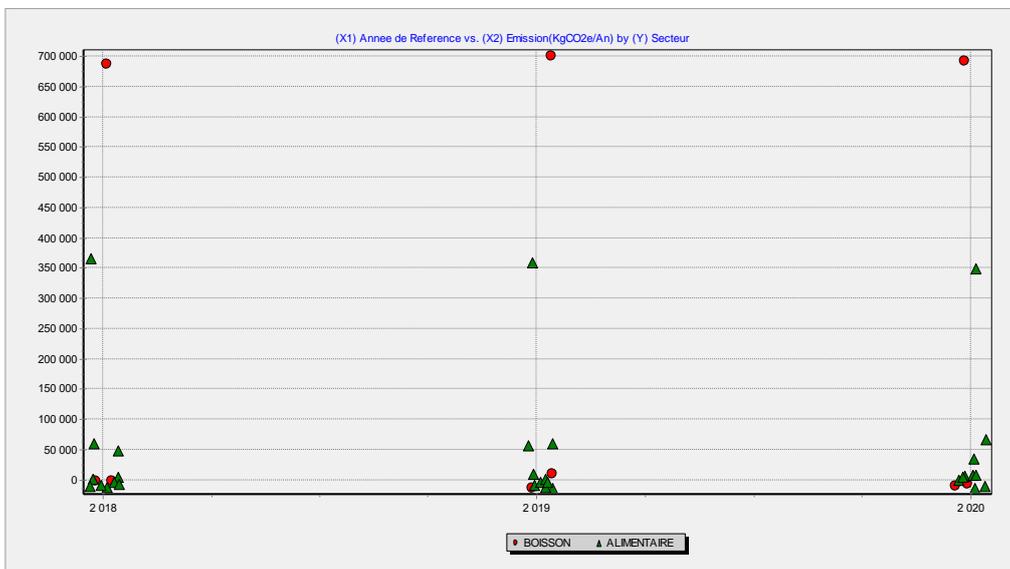


Figure 3.10: Exploration multi variable appliquées sur les variables émission en KgCO<sub>2</sub>équiv, Année de référence et Secteur.

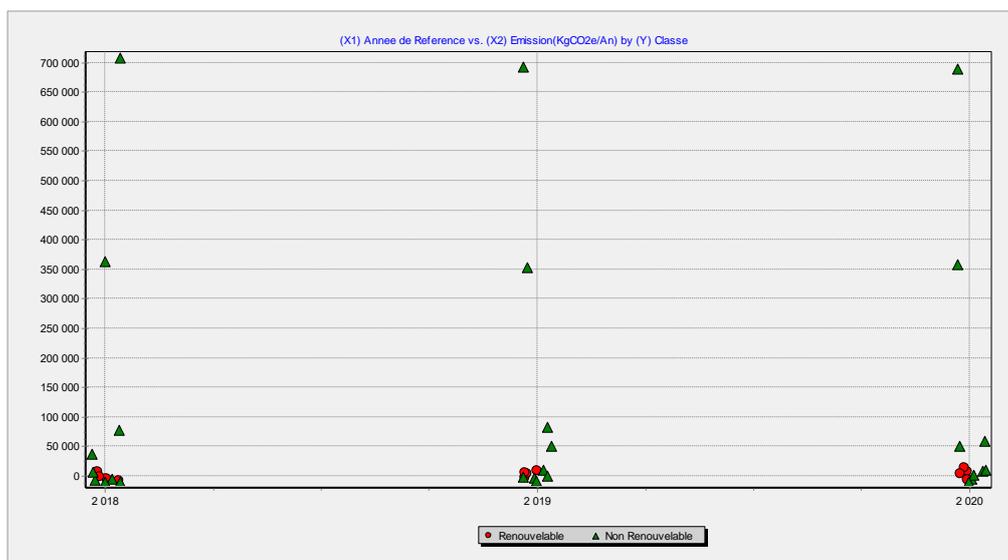


Figure 3.11: Exploration multi variable appliquées sur les variables émission en KgCO<sub>2</sub>équiv, Année de référence et Classe.

### 1.9. Résultats obtenus avec Sipina

La figure 3.12 ci-dessous décrit l'arbre de décision construit à partir de la variable cible ou variable à prédire et les variables explicatives ou variables prédictives. La variable à prédire : Classe ; et les variables prédictives : Emission en KgCO<sub>2</sub>équiv et Types d'énergies. Nous cherchons à déterminer l'appartenance des objets d'une classe à expliquer ou d'une variable à expliquer pour notre cas la classe d'énergie, à partir des variables explicatives notamment, types d'énergies et quantité d'émissions en KgCO<sub>2</sub>équiv/ An.

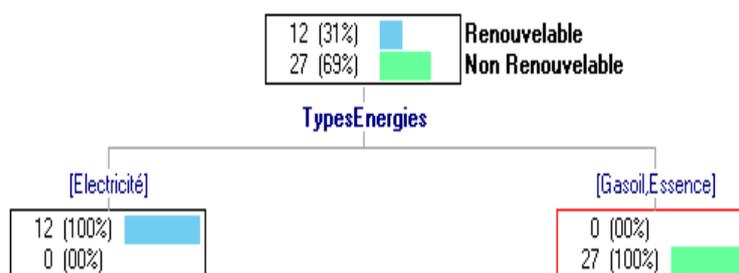


Figure 3.12. Arbre de décision

### **1.10. Interprétation de résultats**

Le tableau 4 et le graphique de la figure 3.5 présentent les constats suivants :

Le secteur des boissons est faiblement représenté, car les données n'ont été collectées que dans une seule industrie de fabrication de boissons, ce qui représente une contribution de 23 %. En revanche, le secteur alimentaire est davantage représenté, avec des données recueillies auprès de quatre industries, représentant une contribution de 77 %. Il ressort de ces secteurs une faible exploitation de l'électricité fournie par la SNEL, avec une contribution de 12 %, et une forte utilisation du gasoil, représentant 61 %.

Les industries de fabrication de boissons et de produits alimentaires utilisent majoritairement des énergies non renouvelables (gasoil ou essence) dans leurs chaînes d'activités, avec une contribution de 69 %, contre une faible utilisation d'énergies renouvelables (électricité fournie par la SNEL), qui représente 31 %. Les graphiques des figures 3.9, 3.10 et 3.11 illustrent les observations suivantes :

Pour les années de référence 2018, 2019 et 2020, on observe de fortes émissions de CO<sub>2</sub> dues à l'utilisation excessive de gasoil, et des émissions plus faibles liées à l'utilisation de l'électricité fournie par la Société Nationale d'Électricité (SNEL). Dans les secteurs des boissons et des produits alimentaires, les émissions élevées sont toujours associées à l'utilisation de gasoil, tandis que les émissions plus faibles sont liées à l'utilisation de l'électricité dans la catégorie des énergies non renouvelables. Le graphique de la figure 3.12 montre que, quel que soit le type d'énergie utilisée par les industries de fabrication de boissons et de produits alimentaires, et que leurs émissions soient très élevées ou faibles, ces énergies seront toujours classées comme renouvelables ou non renouvelables, comme l'indique l'arbre de décision construit.

## **2. DISCUSSION DES RÉSULTATS**

Dans cette section, nous centrons notre discussion sur les principaux résultats obtenus au cours de cette étude. La création d'une nouvelle base de données pour l'identification et la quantification des activités émettrices de CO<sub>2</sub> par les industries, grâce à la modélisation réalisée avec la méthode MERISE et d'autres méthodes associées, nous a permis de surmonter le défi majeur auquel nous faisons face : l'absence d'une base de données exploitable. La modélisation du système d'identification et de quantification des activités émettrices de CO<sub>2</sub> des industries de fabrication de boissons et de produits alimentaires a ainsi permis de pallier ce manque de données. La modélisation

utilisée dans cette étude, bien que n'étant pas un concept nouveau <sup>21</sup>, s'est avérée particulièrement utile pour surmonter certaines difficultés rencontrées lors de la collecte de données existantes. Le choix de la méthode MERISE se justifie par sa simplicité à représenter les données selon plusieurs formalismes conceptuels, logiques et physiques, ce qui a facilité l'implémentation de notre base de données pour l'identification et la quantification des activités émettrices de CO<sub>2</sub> par les industries de fabrication de boissons et de produits alimentaires.<sup>22</sup>

En plus de la méthode MERISE, la quantification des activités émettrices de CO<sub>2</sub> des entreprises a été réalisée en utilisant l'approche de Niveau 1 proposée par les lignes directrices du Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC). Cette méthode permet de calculer les émissions liées aux combustibles stationnaires, c'est-à-dire les émissions produites au sein des entreprises. Nous avons appliqué cette méthode dans la base de données afin d'obtenir automatiquement les estimations des émissions de CO<sub>2</sub>.<sup>23</sup>

À partir de la base de données créée, nous avons appliqué des outils de data mining (Tanagra, Sipina) pour explorer les données <sup>24</sup>. Les résultats obtenus après cette exploration ont mis en évidence les différentes tendances existantes entre les variables, notamment les émissions de CO<sub>2</sub>, les types d'énergie, les secteurs d'activités, les années de référence et les catégories d'énergie. Il en ressort que, dans le secteur des boissons et des produits alimentaires, les émissions élevées sont principalement dues à l'utilisation de gasoil, tandis que les émissions plus faibles sont associées à l'utilisation de l'électricité fournie par la SNEL.

Enfin, à partir du même jeu de données extrait de la base de données pour l'identification et la quantification des activités émettrices de CO<sub>2</sub> des industries de fabrication de boissons et de produits alimentaires, nous avons construit un modèle de classification automatique basé sur les arbres de décision. Nous avons opté pour cette méthode arborescente car elle simplifie et facilite la prise de décision. Les résultats obtenus à partir de l'arbre de

---

<sup>21</sup> Frendi Mohammed, Modélisation de Système d'Information, Thèse, Mars 2013. <https://theses.univ-oran1.dz/document/TH4020.pdf>

<sup>22</sup> Laila Alami, Système d'information et Base de données, 2015-2016. [https://horizon.documentation.ird.fr/exl-doc/pleins\\_textes/pleins\\_textes\\_4/sci\\_hum/24939.pdf](https://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_4/sci_hum/24939.pdf)

<sup>23</sup> Lignes directrices 2006 du GIEC pour les inventaires nationaux des gaz à effet de serre, Chapitre 2 Combustion stationnaire. [https://www.ipccnggip.iges.or.jp/public/2006gl/french/pdf/2\\_Volume2/V2\\_2\\_Ch2\\_Stationary\\_Combustion.pdf](https://www.ipccnggip.iges.or.jp/public/2006gl/french/pdf/2_Volume2/V2_2_Ch2_Stationary_Combustion.pdf)

<sup>24</sup> Stéphane Tuffery, Exploration de données et statistiques pour la prise de décision, Université de Rennes, France, 2011.

décision montrent que les industries de fabrication de boissons et de produits alimentaires utilisent majoritairement des énergies non renouvelables (gasoil ou essence) dans leurs chaînes d'activités, avec une contribution de 69 %, contre 31 % pour les énergies renouvelables (électricité fournie par la SNEL).

Les applications du data mining sont nombreuses dans divers domaines, produisant des résultats intéressants selon les techniques employées. Voici quelques exemples notables<sup>25</sup>:

- Le data mining dans le secteur bancaire : utilisé pour prédire la réaction des clients aux changements des taux d'intérêt, identifier les clients les plus réceptifs aux nouvelles offres de produits, repérer les clients fidèles et déterminer ceux qui présentent le risque le plus élevé de ne pas respecter leurs engagements de prêt.
- Le data mining dans la bio-informatique et la biotechnologie : utilisé pour prédire les structures de différentes protéines et déterminer la complexité des structures de plusieurs médicaments.
- Le data mining dans la détection de fraude : utilisé pour détecter les fraudes liées aux cartes de crédit, identifier les fraudes dans les listes électorales en combinant les réseaux de neurones avec le data mining, détecter les fraudes dans les demandes de passeport grâce à un système de diagnostic par apprentissage en ligne, et repérer les fausses demandes de remboursement médical.
- Le data mining dans la médecine et la pharmacie : utilisé pour prédire la présence de maladies et/ou de complications, choisir un traitement pour le cancer, sélectionner les antibiotiques pour certaines infections, choisir une technique particulière (comme les sutures ou le matériel de suture) lors de procédures chirurgicales, et gérer l'approvisionnement en médicaments les plus fréquemment prescrits.

---

<sup>25</sup> Chami Djazia, une plateforme orientée agent pour le data mining, Thèse, 2009-2010.  
<http://eprints.univ-batna2.dz/193/1/Djazia%20CHAMI.pdf>

## CONCLUSION

Cette recherche a permis de démontrer l'application du data mining (exploration ou fouille de données) aux informations relatives aux activités émettrices de CO<sub>2</sub> recueillies dans les industries de fabrication de boissons et de produits alimentaires de la commune de Limete. À travers cette étude, nous avons réussi à modéliser un système d'identification et de quantification des activités émettrices de CO<sub>2</sub> des entreprises, en mettant l'accent sur celles du secteur des boissons et des produits alimentaires, en utilisant la méthode Merise. Cette approche a abouti à la création d'une base de données relationnelle pour l'identification et la quantification de ces activités, implémentée à l'aide d'un système de gestion de base de données. Cette base de données a été essentielle pour collecter et structurer les informations provenant des industries concernées.

Les activités émettrices de CO<sub>2</sub> de ces entreprises ont été quantifiées en appliquant l'approche de niveau 1 des lignes directrices du Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC), une méthode servant à calculer les émissions liées aux combustibles stationnaires. En explorant ces données à l'aide des outils de data mining (Tanagra, Sipina), nous avons constaté que, pour les années de référence 2018, 2019 et 2020, les fortes émissions de CO<sub>2</sub> étaient principalement dues à l'usage intensif du gasoil, tandis que les émissions plus faibles étaient associées à l'utilisation de l'électricité produite par la Société Nationale d'Électricité (SNEL). Il ressort également que, dans les secteurs des boissons et de l'alimentaire, les émissions les plus élevées sont systématiquement liées à l'utilisation du gasoil, une énergie non renouvelable, alors que les émissions les plus faibles sont liées à l'électricité, une énergie renouvelable.

Enfin, avec le même ensemble de données (dataset), nous avons élaboré un modèle prédictif de classification automatique basé sur des arbres de décision. Il est donc important de noter que, dans l'industrie de fabrication de boissons et de produits alimentaires, l'utilisation excessive des énergies non renouvelables (gasoil ou essence) dans les chaînes d'activités représente 69 % des émissions, contre 31 % pour l'usage des énergies renouvelables (électricité fournie par la SNEL). L'arbre de décision montre que, quel que soit le type d'énergie utilisé par ces industries et le niveau de leurs émissions, ces énergies seront systématiquement classées comme renouvelables ou non renouvelables.

## RÉFÉRENCE BIBLIOGRAPHIE

1. Annuaire statistique 2015, Mars 2017. <https://www.ins.cd/wp-content/uploads/2021/04/Annuaire-statistique-2015-Web.pdf>
2. Benamar Houmadi, étude exploratoire d'outils pour le data mining, Thèse, Avril 2007, <http://depot-e.uqtr.ca/id/eprint/1423/1/030000495.pdf>
3. Cigref, Valorisation des données dans les grandes entreprises, Maturité, pratiques et modèle, Novembre 2016. <https://www.cigref.fr/wp/wp-content/uploads/2016/11/CIGREF-Valorisation-des-donnees-Pratiques-Modele-2016.pdf>
4. Chami Djazia, une plateforme orientée agent pour le data mining, Thèse, 2009-2010. <http://eprints.univ-batna2.dz/193/1/Djazia%20CHAMI.pdf>
5. Changement climatique, genre et santé, Organisation mondiale de la Santé, 2016. [https://apps.who.int/iris/bitstream/handle/10665/204177/9789242508185\\_fre.pdf](https://apps.who.int/iris/bitstream/handle/10665/204177/9789242508185_fre.pdf)
6. Dominique Crié, de l'extraction des connaissances au knowledge management, 2003. <https://www.cairn.info/revue-francaise-de-gestion-2003-5-page-59.htm>
7. Frendi Mohammed, Modélisation de Système d'Information, Thèse, Mars 2013. <https://theses.univ-oran1.dz/document/TH4020.pdf>
8. Georges Gardarin, bases de données, Eyrolles, 2003. [http://georges.gardarin.free.fr/Livre\\_BD\\_Contentu/XX-TotalBD.pdf](http://georges.gardarin.free.fr/Livre_BD_Contentu/XX-TotalBD.pdf)
9. Harat Ahmed, Effet de gaz à effet de serre sur la santé humaine, Avril 2017. [https://www.researchgate.net/publication/316739125\\_EFFECT\\_DE\\_GAZ\\_A\\_EFFECT\\_DE\\_SERRE\\_SUR\\_LA\\_SANTE\\_HUMAINE](https://www.researchgate.net/publication/316739125_EFFECT_DE_GAZ_A_EFFECT_DE_SERRE_SUR_LA_SANTE_HUMAINE)
10. Laila Alami, Système d'information et Base de données, 2015-2016. [https://horizon.documentation.ird.fr/exl-doc/pleins\\_textes/pleins\\_textes\\_4/sci\\_hum/24939.pdf](https://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_4/sci_hum/24939.pdf)
11. Le changement climatique, ADEME, Juin 2012. <https://www.ademe.fr/sites/default/files/assets/documents/guide-changement-climatique.pdf>
12. Lignes directrices 2006 du GIEC pour les inventaires nationaux des gaz à effet de serre, Chapitre 2 Combustion stationnaire. [https://www.ipcc-nggip.iges.or.jp/public/2006gl/french/pdf/2\\_Volume2/V2\\_2\\_Ch2\\_Stationary\\_Combustion.pdf](https://www.ipcc-nggip.iges.or.jp/public/2006gl/french/pdf/2_Volume2/V2_2_Ch2_Stationary_Combustion.pdf)
13. Merati Medjeded, bases de données avancées, polycopié de cours, 2019-2020. <http://fmi.univ-tiaret.dz/images/Polycopies/Bases-de-Donnes-Avanc.pdf>
14. Nicolas Larrousse, Création de bases de données, Pearson Education France, 2009. <http://tony3d3.free.fr/files/Creation-de-bases-de-donnees.pdf>

15. Programme – Pays de la RDC au Fonds Vert Climat (GCF), RDC, 20 Février 2019. <http://www.gip-recherche-justice.fr/wp-content/uploads/2020/01/17.05-RF-contentieux-climatiques.pdf>
16. Stéphane Tuffery, *Exploration de données et statistiques pour la prise de décision*, Université de Rennes, France, 2011.
17. <http://www.gip-recherche-justice.fr/wp-content/uploads/2020/01/17.05-RF-contentieux-climatiques.pdf>
18. <https://fr.wikipedia.org/wiki/Limete>
19. [https://www.congovirtuel.com/page\\_rapport\\_travaux/rapport\\_unikin/rapport\\_ndjodji\\_df](https://www.congovirtuel.com/page_rapport_travaux/rapport_unikin/rapport_ndjodji_df)
20. <https://www.espace-aubade.fr/blog/chauffage/se-chauffer-energies-renouvelables.html>
21. [https://eric.univ-lyon2.fr/ricco/tanagra/fichiers/le\\_logiciel\\_tanagra\\_egc\\_2005.pdf](https://eric.univ-lyon2.fr/ricco/tanagra/fichiers/le_logiciel_tanagra_egc_2005.pdf)
22. <https://www.rocq.inria.fr/axis/modulad/archives/numero-33/tutorial-rakotomalala-33/rakotomalala-33-tutorial.pdf>