

Charte Projet :

DIAGNOSTICS DE PERFORMANCE ÉNERGÉTIQUE



Réalisé par :

Gilles Adrien KOUEBOU DJONKO

Franck Russell FONGANG

Alaa MAZOUZ

Abrar BOUSLAHI

Sous la coordination de :

M. Abdelhadi MIFDAL,

M. Arnaud BOUJUT,

M. Michel ESTEVE

Table des matières

I.	Contexte du projet (CL)	2
1.	Contexte général	2
2.	Etude de l'existant	2
II.	Besoins exprimés	4
1.	Objectifs métiers et objectifs data science.....	4
2.	Livrables	5
3.	Métriques et règles de validation des livrables	5
III.	Contraintes & degrés de liberté	7
IV.	Risques, points durs.....	8
V.	Gestion de la charge de travail	9
VI.	Gestion du temps.....	10
VII.	Rôles	11
VIII.	Données d'entrée	12
IX.	Moyens mis à disposition	13
1.	Matériels du client	13
2.	Logiciels spécifiques	13
X.	Propriété intellectuelle	14
XI.	Pilotage & communication	15
XII.	Outil de gestion de projet	16

I. Contexte du projet (CL)

1. Contexte général

Dans le but d'atteindre l'un des objectifs de la stratégie Bas Carbone, consistant à réduire d'ici 2030 les émissions de gaz à effet de serre des bâtiments, la France, à travers l'entreprise ENEDIS (filiale d'EDF, chargée de la gestion et de l'aménagement de 95 % du réseau de distribution d'électricité), s'attaque au problème de la sobriété énergétique. Ce défi est amplifié par l'accélération des changements climatiques et la hausse des prix de l'énergie, rendant cruciale l'amélioration de la performance énergétique des bâtiments.

En effet, le diagnostic de performance énergétique (DPE) est un outil clé pour évaluer la consommation énergétique des bâtiments et leur impact environnemental. Comprendre la corrélation entre les classifications DPE et les consommations énergétiques réelles est donc primordial. Ce projet académique vise à approfondir cette relation et à développer des outils pour améliorer la précision des prédictions, facilitant ainsi la prise de décision pour les acteurs concernés.

2. Etude de l'existant

L'estimation des consommation électriques affichées dans les DPE repose sur des calculs conventionnels à l'instar de la méthode 3CL, et non sur les consommations réelles mesurées. Cette approche standardisée présente de nombreuses limites dues à sa non prise en compte du style de vie et du taux d'occupation des maisons ou appartements. Ainsi, dans l'optique d'améliorer l'évaluation du DPE plusieurs solutions et initiatives ont été développées par des entreprises et des organismes spécialisés parmi lesquelles, on peut citer :

- GreenFlex : qui fournit des solutions digitales pour analyser les données énergétiques et identifier les économies potentielles à travers l'utilisation de l'intelligence artificielle pour l'automatisation des recommandations en matière de rénovation énergétiques.
- Startup innovantes telles que :
 - DeepKi : qui propose des solutions SaaS pour optimiser la performance énergétique et suivre les données de consommation des bâtiments.
 - Effy : qui se spécialise dans l'accompagnement des travaux de rénovation énergétique, avec des outils numériques d'évaluation et de simulation.
- Université académique et laboratoire de recherche : Plusieurs projets universitaires ont exploré l'application de modèles de machine learning pour prédire les consommations énergétiques réelles à partir des données DPE. Les travaux incluent notamment l'intégration de données météorologiques et socio-économiques.

Malgré la diversification des approches, allant de la simple analyse statistique aux solutions complexes intégrant de l'intelligence artificielle, des défis subsistent :

- La fiabilité des données DPE : car elles sont généralement approximatives, pouvant biaiser les analyses et les prévisions.
- Un manque d'harmonisation des données : les différentes bases de données ne sont pas toujours compatibles.
- Un manque de prise en compte des spécificités des utilisateurs ou des bâtiments.
- Coût élevé des solutions avancées : Les plateformes SaaS ou les outils IA sont souvent coûteux.

II. Besoins exprimés

1. Objectifs métiers et objectifs data science

Les objectifs métiers (BO : Business Objectives) et objectifs data sciences (DSO : Data Science Objectives) permettent de lier les finalités stratégiques d'un projet à ses réalisations techniques. Tandis que les BO définissent les objectifs métiers répondant à des besoins concrets, les DSO traduisent ces objectifs en tâches spécifiques réalisables à l'aide de l'analyse et de la modélisation des données.

➤ BO1 :

Eclairer les décisions de rénovation et de transition énergétique : en quantifiant l'impact des rénovations énergétiques (amélioration du DPE sur les consommations électriques) afin de réduire les consommations énergétiques. Ceci se fera, à travers les objectifs data sciences suivantes :

- **DSO1-** Analyser l'impact économique des classes DPE sur les consommations réelles d'électricité : Explorer les distributions et identifier les corrélations clés entre les classes DPE et les coûts énergétiques réels.
- **DSO2-** Créer un modèle de prédiction de la consommation électrique des foyers français en fonctions du DPE : Implémenter des modèles supervisés pour estimer les consommations énergétiques selon les différents profils de logements ; prendre en compte les spécificités telles que la taille des logements, le type de chauffage, et les équipements électroménagers.
- **DSO3-** Créer un modèle de prédiction des économies faites suite à une rénovation : Intégrer les données météorologiques et socio-économiques pour affiner les prédictions et quantifier les bénéfices ; tester différents scénarios de rénovation pour évaluer leur impact potentiel.

➤ BO2 :

Valider la conformité au réel des estimations de consommation électrique présentées dans le DPE : en comparant les données de consommations réelles et celles estimées par le DPE, et en quantifiant la variabilité due aux comportements individuels. Les objectifs data science qui y découlent sont les suivantes :

- **DSO4-** Effectuer une segmentation des logements selon les critères spécifiques : Inclure des variables telles que l'année de construction, le type de chauffage, la qualité de l'isolation et la région climatique ; Identifier les groupes les plus susceptibles de présenter des écarts importants entre consommation réelle et estimée.
- **DSO5-** Illustrer le lien les données de consommations réelles et celle estimée par le DPE : Créer des visualisations explicites (graphiques comparatifs, heatmaps) pour appuyer les analyses et fournir des insights clairs ; Identifier les biais potentiels dans les estimations du DPE et proposer des solutions correctives.

Objectifs data science	Priorités
------------------------	-----------

DSO1	1
DSO2	3
DSO3	5
DSO4	4
DSO5	2

2. Livrables

Ils sont regroupés en deux groupes :

- Les livrables fonctionnels :
 - Model de prédiction.
 - Dashboard interactifs permettant la visualisation des tendances et corrélations, ainsi que des statistiques descriptives de consommations
- Les livrables documentaires :
 - Rapports techniques : présentant la documentation des modèles prédictifs et des hypothèses utilisées, ainsi que des outils techniques utilisé
 - Rapport final : présentant les résultats clés du projet, les méthodologies utilisées, et finalement la conclusion générale du projet.

3. Métriques et règles de validation des livrables

Il s'agit des indicateurs permettant de mesurer un objectif. Ainsi, ils évaluent la qualité, la pertinences et l'impact des travaux réalisé en fonctions des objectifs définis. Ainsi on a :

Objectifs	Métriques	Descriptions	Justificatifs
DSO1/ DSO5	Variance et écart-type.	Différence moyenne estimée avant/après rénovation supérieure à 15 % d'économie d'énergie.	Afin d'être cohérents avec les objectifs de réduction d'émissions inscrit dans les stratégies bas carbone européennes et françaises.
DSO2 / DSO3	RMSE, MAPE, R^2	Permet d'évaluer la précision des prédictions. <ul style="list-style-type: none"> - Le RSME doit être le plus proche de zéro possible. - Le MAPE < 10 % - $R^2 > 85$ % 	
DSO4	Score de silhouette	Mesure la performance des modèles de clustering (segmentation)	

Aussi, les livrables doivent respecter certains critères :

- Analyses exploratoires (Dashboard) :
 - Les distributions des données et les corrélations clés sont clairement mises en évidence.
 - Les visualisations incluent des comparatifs pertinents (par classe DPE, par type de logement, etc.).
 - Les résultats doivent être compréhensibles pour un public non technique grâce à des graphiques et des explications détaillées.
- Les rapports :
 - Usage d'un langage compréhensible pour un public non technique.

III. Contraintes & degrés de liberté

Les différentes contraintes techniques de notre projet sont :

- Normalisation des adresses via l'API BAN : En effet, Les données DPE et celles d'Enedis sont collectées et structurées selon des méthodologies différentes, rendant leur fusion complexe sans un travail préalable d'harmonisation. Ainsi, il faut au préalable normaliser les données d'adresse sous une forme standard.
- Utilisation d'une base de données de DPE postérieur à juillet 2021. En effet, ces données ne regroupent que 6 Millions de logement soit 20%, du parc résidentiel. Ce qui ne le rend pas très représentatifs. Cette limite est due à l'absence de DPE pour certains logements, notamment ceux non vendus ou non loués.
- Étude restreinte aux consommations électriques. Car en l'absence de données sur d'autres sources d'énergie (comme le gaz) à la maille des adresses, seules les consommations électriques seront analysées.
- Données de consommation approximatives. Car les données de consommations individuelles étant personnelles, dans un souci de respect de la vie privée, les données de consommation par adresse fournies ne comportent que les adresses regroupant au moins 10 logements, excluant ainsi les maisons individuelles et petits immeubles. Cela réduit le périmètre d'étude et peut biaiser les résultats pour certains types de logements.
- Ressources matérielles : L'analyse des données et l'entraînement des modèles demandent d'importantes ressources matérielles. Ainsi, il est nécessaire de travailler dans un environnement (cloud ou local) fournissant au minimum 8Go de RAM et une bonne carte graphique.

IV. Risques, points durs

Les différents défis pouvant être observés lors de la réalisation de ce projet sont :

- **Croiser les données de consommations avec la base des DPE.** En effet, cette tâche est d'une grande complexité car elle nécessite :
 - Une très bonne compréhension des différents jeux de données,
 - Une capacité à fusionner des données de différentes sources et de méthodologies différentes,
 - La mise en œuvre d'un redressement des données pour corriger les éventuelles incohérences ou données manquantes.
- **Vérifier la fiabilité des données historiques.** Car aucune garantie de fiabilité n'est donnée pour les données de consommation électrique annuelle ; aussi les données DPE collectées avant juillet 2021 peuvent ne pas être représentatives ou cohérentes avec les données récentes.
- **Changements imprévus dans les jeux de données.** Risque de modification partielle ou totale des jeux de données, nécessitant des ajustements dans le traitement et l'analyse.

V. Gestion de la charge de travail

Point	Valeur
Nombre de séances projet	55 séances
Effectif de l'équipe	4 personnes
Capital-temps	$55 \text{ séances} * 1\text{h}30 * 4 \text{ personnes} =$ 330 heures.homme
Part de ce capital-temps à consacrer :	
200 heures.homme	200 heures.homme
- À la gestion du projet	60 heures.homme
- Aux livrables documentaires	70 heures.homme
Commentaire	Estimation basée sur l'importance du développement de la solution.

VI. Gestion du temps

- Nombre de sprints : 4

Sprints	Objectifs	Livrables
Sprint 1	Finaliser le cahier de charges du projet	Cahier des charges complet
Sprint 2	Commencer l'analyse des impacts des classes DPE et développer un modèle prédictif	Analyses des corrélations entre DPE et coûts énergétiques Première version du modèle de prédiction de consommation électrique
Sprint 3	Approfondir les modèles prédictifs et segmenter les logements selon des critères spécifiques	Modèle de prédiction des économies réalisées après rénovation Rapport de segmentation des logements
Sprint 4	Créer des visualisations pour comparer les consommations réelles et estimées, et identifier les biais	Graphiques comparatifs, heatmaps et recommandations basées sur les visualisations pour améliorer le DPE

VII. Rôles

Les différents rôles sont :

- L'agile master : Gilles Adrien KOUEBOU DJONKO
- L'assistant PO : Franck Russell FONGANG
- Les développeurs :
 - Alaa MAZOUZ
 - Abrar BOUSLAHI
 - Franck Russell FONGANG
 - Gilles Adrien KOUEBOU DJONKO

VIII. Données d'entrée

Les données fournies par le client sont :

- Un cahier de charge très détaillé, présentant le projet dans sa globalité, les objectifs à atteindre, le but recherché.
- Un catalogue de jeu de données parmi lesquelles on a :
 - **DPE Logement existants depuis juillet 2021 collecté par l'ADEME.** Ce jeu donné de données fournis les informations sur les caractéristiques (la surface, l'orientation, les matériaux, murs, fenêtres, etc.) du logement ; ses équipements (chauffage, eau chaude, ventilation) ; consommation énergétique et émissions de gaz à effet de serre.
 - **DPE tertiaire existant depuis 2021 collecté par l'ADEME.** Il comporte les mêmes données que le jeu de données précédent, cependant, le secteur tertiaire est subdivisé en 4 catégories : les bâtiments à usage principal de bureau, d'administration ou d'enseignement ; bâtiment à occupation continue (par exemple : hôpitaux, hôtels, internats, maisons de retraite, etc.) ; autres cas (par exemple : théâtres, salles de sport, restauration, commerces individuels, etc.) et les centres commerciaux.
 - **DPE logements neufs existant depuis juillet 2021 collecté par l'ADEME.** Il est différent de celui pour les appartements anciens. Le bâtiment n'ayant jamais été habité, il s'agit d'une estimation.
 - **DPE tertiaire** avant juillet 2021 collecté par l'ADEME.
 - **BDNB** (Base de Données National des Bâtiments) collecté par la CSTB (Centre Scientifique et Technique du Bâtiment). Il s'agit d'une cartographie du parc de bâtiments existant. Elle est construite par croisement géospatial d'une vingtaine de base de données issues d'organismes publics.
 - **Consommation d'électricité annuelle résidentielle par adresse collecté depuis 2021.** Base de données conçus par ENEDIS. Il fournit la consommation électrique annuelle pour les adresses de 10 logements ou plus.
 - **BAN (Base Adresse Nationale).** Il s'agit d'une base de données de référence officielle pour l'adressage en France. Elle regroupe et normalise toutes les adresses du territoire national. En tant que référentiel officiel, elle sert de socle pour de nombreuses politiques publiques et infrastructures gouvernementales.

A chaque donnée de ce catalogue collecté par EDEME, s'ajoute des documents d'accompagnement :

- Un dictionnaire de données précisant à quoi correspond chaque champ
- Une note technique précisant les informations essentielles à connaître afin d'interpréter correctement les données fournies.
- Une description des contrôles de cohérence effectuées sur les données des DPE

IX. Moyens mis à disposition

1. Matériels du client

L'open data University a mis à disposition un environnement de travail entièrement configuré sur le Datalab SSP Cloud de l'INSEE. Les différentes données fournies y sont déjà chargées. Il propose un ensemble de services destinés aux traitements statistiques de données ainsi qu'à la gestion complète des projets de data science.

2. Logiciels spécifiques

Pour une exécution en local, les logiciels nécessaires pour ce projet sont :

- Base de données : Postgress ou MySQL
- Analyse de données : Jupiter, pycharm
- Visualisation : Power BI
- Collaboration : GitHub, Teams.

X. Propriété intellectuelle

Les données fournies pour ce projet sont strictement à usage académique et nullement professionnelle.

Quant aux résultats produit à la fin de ce projet,

- Les différents rapports et les graphiques de visualisations seront publics étant que ce projet se porte dans le cadre d'un défi universitaire publique.
- Les API des modèles quant à eux ne le seront pas, et seront juste partagé avec nos encadreurs académiques.

XI. Pilotage & communication

- **Engagements du client** : Le client doit garantir sa disponibilité pour les réunions périodiques et coopérer dans la validation des livrables à chaque étape clé.
- **Livrables de gestion de projet** : Comptes-rendus de réunions détaillant les discussions et décisions prises, ainsi qu'une roadmap actualisée.
- **Prise de décision** : En cas de difficultés techniques, de conflits internes ou de retards, des décisions seront prises collectivement ou avec le client, et documentées dans un log des risques ou des actions correctives.

XII. Outil de gestion de projet

Le logiciel qui sera utilisé pour la gestion de projet est Jira.