



Statistique pour l'Evaluation et la
Prévision

PROJET DIGITAL



DPE

Equipe 3 :

BAROIN Guillaume

SEJDI Benjamin

GRILO Mathias

THIERUS Armand

PIMPERNELLE Léa

ZEHOUNKPE Graziana

2023 - 2024

SOMMAIRE

INTRODUCTION	3
I- OBJECTIFS DU PROJET	6
II- METHODOLOGIE.....	7
1.1 Collecte de Données sur la Consommation d'Énergie	7
1.2 Nettoyage de la base de données.....	8
1.3 Algorithme et modèle de prédiction.....	9
1.4 Description des variables exploités	11
1.5 Conception de l'interface utilisateur et de sortie	13
1.6 Intégration de l'algorithme dans l'interface d'assistance à la décision.....	14
III- RESUTATS.....	15
3.1 Modèle de prédiction	15
3.2 Interface de saisie	16
IV- DIFFICULTES	17
V- RECOMMANDATION.....	19
BIBLIOGRAPHIE	20
ANNEXES	21

INTRODUCTION

En France, le DPE est particulièrement essentiel à la lumière des préoccupations croissantes liées au changement climatique et à l'engagement du pays en faveur de la réduction des émissions de gaz à effet de serre. La France a adopté des objectifs ambitieux en matière de lutte contre le réchauffement climatique, et le secteur du bâtiment, en tant que grand consommateur d'énergie, joue un rôle majeur dans la réalisation de ces objectifs.

Mis en place en 2006, le DPE, ou Diagnostic de Performance Énergétique, est un document obligatoire en France lors de la vente ou de la location d'un bien immobilier. Son objectif est d'informer les futurs acquéreurs ou locataires sur la performance énergétique d'un logement, en évaluant sa consommation d'énergie dont son impact environnemental par l'émission de gaz à effet de serre. Le DPE se présente sous la forme d'une étiquette énergétique, similaire à celles que l'on trouve sur les appareils électroménagers, avec une échelle de lettres (de A à G) et de couleurs (du vert au rouge), ainsi qu'une estimation de la consommation d'énergie du bien en kilowattheures par mètre carré et par an (kWh/m².an). Le DPE est valable pendant 10 ans, sauf en cas de modifications substantielles du bâtiment.

Le DPE vise à sensibiliser les propriétaires, les locataires et les acheteurs aux performances énergétiques des bâtiments, en mettant en évidence leur consommation d'énergie et les émissions de gaz à effet de serre associées. Dans un contexte où les défis climatiques exigent des mesures concrètes, le DPE est un outil qui incite les propriétaires à investir dans des améliorations de l'efficacité énergétique de leurs bâtiments.

L'entrée en vigueur du nouveau DPE et son lot de changements en juillet 2021 (rectifié par un arrêté en octobre 2021) s'accompagne de nouvelles mesures pour les propriétaires bailleurs.

- Un DPE pleinement opposable ;
- Un DPE plus fiable ;
- Un DPE prenant mieux en compte les enjeux climatiques : le calcul des étiquettes de performance énergétique intègre non seulement la consommation d'énergie primaire du logement mais tient également compte des émissions de gaz à effet de serre.

Le calendrier mis en place par l'État vise à interdire progressivement la mise en location de bâtiment et logements trop énergivores. Dans un futur proche, il est prévu que les logements classés G ne seront plus autorisés sur le marché locatif dès 2025, suivis des logements classés F à partir de 2028, et enfin des logements classés E à compter de 2034. Progressivement, ces logements seront considérés comme non conformes à la loi, donnant au locataire le droit d'exiger des travaux. Cet élément peut dissuader des candidats à la location, soucieux de l'environnement et de la maîtrise de leurs dépenses énergétiques. La location de logements à forte consommation d'énergie, qualifiés de "passoires thermiques", est désormais prohibée. À partir du 1er janvier 2023, en France métropolitaine, la limite maximale de consommation d'énergie finale autorisée pour un logement est établie à 450 kWh/m². Les propriétaires bailleurs ont donc tout intérêt à améliorer la performance énergétique de leurs biens. Ils pourront ainsi non seulement proposer ces biens à des loyers plus élevés, mais aussi les vendre à des prix plus avantageux.

La validité du Diagnostic de Performance Énergétique (DPE) est de 10 ans. Si un bien est mis en location et que son DPE date de plus de 10 ans, il doit être renouvelé. Depuis 2006, la fourniture de ce diagnostic est une obligation lors de la location d'un bien, qu'il soit meublé ou non, y compris pour les locations saisonnières occupées plus de 4 mois par an. En général, le coût du DPE varie de 100 à 250 € pour un logement selon l'ADEME.

Deux approches distinctes sont utilisées pour calculer le DPE : la méthode conventionnelle et la méthode basée sur les factures d'énergie. La méthode basée sur les factures d'énergie suit un processus en trois étapes, comprenant la description du bâtiment, des systèmes de chauffage et de production d'eau chaude sanitaire, ainsi qu'une analyse de trois années consécutives de factures d'énergie, couvrant la consommation de gaz et d'électricité.

Cependant, cette méthode est amenée à être éliminée car elle dépend du taux d'occupation et du mode de vie des résidents. Elle sera progressivement supprimée à partir du 1er juillet 2021 avec l'introduction du nouveau DPE.

La méthode conventionnelle du DPE est un processus standardisé qui repose sur une série d'étapes rigoureuses pour évaluer la consommation d'énergie et les émissions de gaz à effet de serre associées à un bâtiment. Elle est couramment utilisée pour évaluer la performance énergétique des bâtiments. Cette méthode repose sur des critères standardisés et est indépendante du mode de vie des occupants ou du taux d'occupation de la propriété. L'estimation du DPE en

utilisant cette méthode est basée sur une température intérieure de 19°C, composée de 18°C pour le chauffage et 1°C pour les apports par les personnes et les appareillages.

Cette méthode se déroule en trois étapes. Tout d'abord, une description complète du bâtiment est réalisée. Ensuite, une campagne de mesure des superficies déperditives est entreprise, englobant des éléments tels que la superficie des fenêtres, leur orientation, l'épaisseur et le type d'isolant, ainsi que des informations sur le système de ventilation. Enfin, des données précises sur le système de chauffage et de production d'eau chaude sanitaire sont collectées, incluant des détails sur le type de chaudière, l'année d'installation, etc. En outre, toutes les informations financières et énergétiques du DPE sont calculées en supposant une température intérieure de 19°C. Ce processus est réalisé par des diagnostiqueurs immobiliers certifiés et vise à fournir aux propriétaires, aux locataires et aux acheteurs des informations essentielles sur la performance énergétique du bâtiment. Au terme de ce processus, deux diagrammes sont produits : le diagramme de l'énergie, qui représente la consommation d'énergie du bâtiment, et le diagramme des Gaz à Effet de Serre, reflétant les émissions de gaz à effet de serre liées à la consommation d'énergie.

Le diagramme de l'énergie quantifie la consommation annuelle d'énergie d'une propriété en kilowattheures par mètre carré par an. La première note de ce diagramme évalue la consommation annuelle d'énergie du bâtiment liée au chauffage et à la production d'eau chaude sanitaire. La valeur indiquée dans ce diagramme est obtenue en divisant la consommation totale d'énergie annuelle par la superficie de la propriété. Par exemple, si une maison de 100 mètres carrés consomme 17 000 kWh par an, il suffit de diviser 17 000 par 100 pour obtenir une note de consommation de 170 kWh, ce qui correspondrait à une classification en catégorie D.

Le besoin des clients en matière de DPE est de disposer d'informations claires, fiables et objectives sur la performance énergétique des bâtiments. Cela leur permet de prendre des décisions éclairées, de maîtriser leurs coûts énergétiques, de se conformer à la réglementation et de contribuer à la réduction de l'impact environnemental. Le DPE joue un rôle essentiel en fournissant ces informations essentielles à une variété d'acteurs du secteur immobilier.

I- OBJECTIFS DU PROJET



Dans le cadre de notre étude, l'objectif principal est de développer une interface destinée à évaluer et à attribuer des classes énergétiques aux appartements afin d'obtenir un DPE approximatif à partir des informations disponibles dans notre base de données.

Le Diagnostic de Performance Énergétique (DPE) est un élément clé dans la promotion de l'efficacité énergétique des bâtiments en France, et notre recherche s'inscrit dans le contexte de la mise en place de solutions modernes pour faciliter cette évaluation.

Alors que le DPE est actuellement réalisé par des diagnostiqueurs immobiliers certifiés, notre démarche vise à automatiser et à simplifier ce processus en utilisant des données spécifiques aux appartements, dans le but de fournir aux propriétaires, aux locataires et aux acheteurs une évaluation rapide et précise de la performance énergétique de leur logement. De manière spécifique, il s'agira de :

- ✚ Récupérer une base de données récente contenant des données historiques sur la consommation d'énergie ;
- ✚ Créer un formulaire VBA interactif et intuitif pour collecter les données d'entrée des utilisateurs, en garantissant une expérience utilisateur optimale ;
- ✚ Livrer une interface de sortie des résultats, comprenant un diagramme des catégories, un diagramme en bâtons des facteurs significatifs et les probabilités d'appartenance à chaque catégorie.

II- METHODOLOGIE

I.1 Collecte de Données sur la Consommation d'Énergie

L'Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME) occupe une position centrale dans la collecte de données sur la consommation d'énergie en France. Sur son site web, l'ADEME propose un processus de collecte de données pour élaborer le DPE (Diagnostic de Performance Énergétique), un document essentiel pour gérer la performance énergétique des bâtiments. Les informations utilisées dans ce projet sont celles rassemblées par l'ADEME après l'arrêté de Juillet 2021. Ces données sont classées par département.

Au démarrage du projet, les données collectées étaient spécifiquement axées sur la région du Grand Est. Cependant, lors de l'analyse initiale des données pour le Diagnostic de Performance Énergétique (DPE), il est apparu que le nombre d'appartements classés dans la catégorie A était très limité, se comptant seulement à 05 dans cette région. Cette rareté des données dans cette catégorie particulière a soulevé des préoccupations quant à la représentativité de l'échantillon pour cette classification énergétique spécifique.

Dans le but d'obtenir une meilleure représentativité des différentes catégories du DPE, notamment la catégorie A, une décision stratégique a été prise pour étendre la collecte des données à l'échelle de la France métropolitaine. Cette expansion a permis d'inclure un échantillon plus vaste et diversifié d'appartements, offrant ainsi une vue plus complète et plus représentative des évaluations énergétiques, en particulier pour la catégorie A.

Cette extension de la collecte des données à l'échelle nationale a permis d'enrichir significativement notre jeu de données, augmentant le nombre d'appartements classés dans la catégorie A et renforçant ainsi la fiabilité et la pertinence de nos analyses.

Les données sont spécifiques aux appartements, car ils sont souvent situés dans des zones urbaines où la densité de population est plus élevée. Cela signifie que les actions visant à améliorer l'efficacité énergétique des appartements peuvent avoir un impact considérable sur la consommation d'énergie globale du pays. Aussi, les appartements tendent à avoir des caractéristiques qui diffèrent par rapport à d'autres types de biens immobiliers comme les maisons, ce qui facilite la création d'une interface pouvant être plus largement appliquée.

La base de données brute téléchargée contient un total de 350 colonnes avec des colonnes en double et 823 662 lignes, parmi lesquelles plusieurs présentent des valeurs manquantes (NA). Cette ampleur de données offre une riche source d'informations, bien que la présence de valeurs manquantes nécessite une attention particulière.

I.2 Nettoyage de la base de données



Le nettoyage des données est une étape essentielle dans l'analyse de données, en particulier dans le contexte du Diagnostic de Performance Énergétique (DPE). La gestion des données manquantes (NA), est l'une des étapes cruciales dans le processus d'analyse de données.

Etant donné que Python offre des bibliothèques telles que Pandas avec des outils pour gérer les valeurs manquantes, le nettoyage de la base de données a été effectué avec ce dernier. Pour cela, on a repéré les valeurs manquantes dans la base de données du DPE en explorant chaque colonne. Ne pouvant pas juste supprimer toutes les lignes contenant les NA, seules celles représentant réellement un manque d'information ont été supprimées et les autres ont été imputées.

Un algorithme de nettoyage offrant une solution flexible pour la maintenance de bases de données a été programmé. Peu importe la base de données utilisée, cet algorithme est conçu pour s'adapter sans nécessiter de modifications majeures. Ce nettoyeur permet de spécifier les colonnes à conserver, simplifiant ainsi le processus de nettoyage. Il faudrait indiquer les noms des colonnes à préserver, et l'algorithme se chargera de supprimer les autres, garantissant ainsi que seules les données pertinentes sont conservées.

Dans la base, certaines colonnes renvoient une liste de modalités (Comme la variable Orientation mur exposés : nord, sud, est, ouest). Ces modalités ont été scindées en 4 colonnes et renvoi 1 si la ligne qui contient la modalité correspondante à la colonne, et 0 sinon. La colonne initiale est alors supprimée. Les variables dont les modalités sont trop proches, ont été fusionnées.











Pour les NA qui ne peuvent pas être juste supprimées, elles ont été remplacées par les valeurs souhaitées.

Dans le processus de prétraitement des données, un code a été ajouté pour effectuer une transformation des données, éliminant ainsi les chaînes de caractères et conservant uniquement les valeurs numériques.

A terme, l'algorithme de nettoyage génère une base de données propre et exploitable, prête à être utilisée pour les analyses ultérieures. Cette base de données est constituée de 59 colonnes et de 360 321 lignes. Pour que la base de données nettoyée soit constituée de chaque catégorie de manière relativement équilibrée, on est alors passé à une base de 59 colonnes et de 4130 lignes. Ce processus de nettoyage a permis d'obtenir une répartition homogène des différentes catégories au sein de la base de données, offrant ainsi une représentation équitable des éléments étudiés.

1.3 Algorithme et modèle de prédiction

En attendant de disposer de la base de données exploitable, des données fictives ont été générées pour débiter l'entraînement du modèle. Ainsi, pour simuler les attributs des biens immobiliers, nous allons générer un ensemble de données aléatoires. Ce jeu de données simulées comprend les éléments suivants :

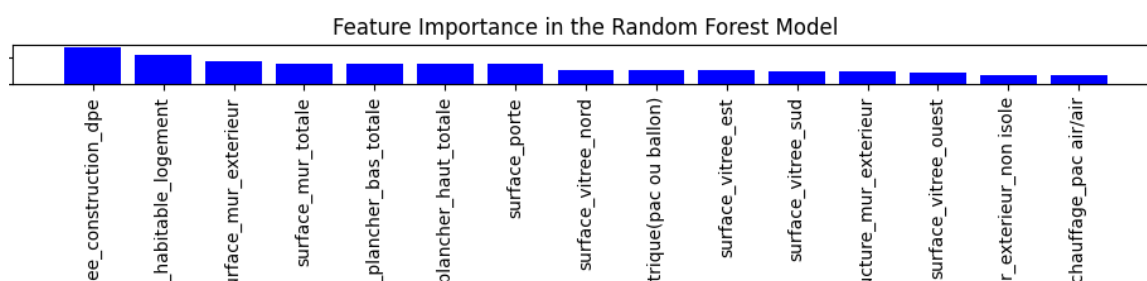
-  Le Diagnostic de Performance Énergétique (DPE), noté de A à G.
-  Le type de propriété, qui peut être une maison ou un appartement.
-  La surface en mètres carrés du bien.
-  Le nombre de pièces.
-  La consommation énergétique en kWh/m²/an.
-  Les émissions de gaz à effet de serre en kgCO₂/m²/an.
-  Le type de toiture, qui peut être plat, à deux pans, mansardé, monopente, à quatre pans ou en forme de pavillon.
-  Le matériau de revêtement de toiture.
-  Le type de revêtement, qui peut être en brique, en enduit, en bois, en béton ou en pierre.
-  Le type d'isolation, qui peut être en laine de verre, en polystyrène expansé, en laine de roche, en laine de bois ou en polyuréthane.

Chacune de ces caractéristiques est générée de manière aléatoire pour chaque entrée du jeu de données, garantissant ainsi une représentation variée des biens immobiliers simulés.

Une fois que la base de données a été nettoyée, elle a servi de fondement pour la création d'un modèle de prédiction. À cet effet, un algorithme de K_neighbors, également connu sous le nom de k plus proches voisins (k-NN), a été mis en œuvre. Le modèle k-NN repose sur un concept qui prédit une valeur en se basant sur les données de ses "k" voisins les plus proches. Pour ce faire, il mesure la proximité des voisins selon une métrique définie, puis effectue une agrégation pondérée pour déterminer la prédiction finale. Le k-NN a été ajusté et validé sur la base de données nettoyée pour s'assurer qu'il puisse fournir des prédictions pertinentes dans le contexte du projet.

Pour optimiser l'ergonomie du produit, une analyse approfondie a été réalisée à l'aide d'un modèle Random Forest. L'objectif est d'identifier les variables qui jouent un rôle déterminant dans la prédiction de la classe de DPE (Diagnostic de Performance Énergétique) et de réduire le risque de surapprentissage. Pour garantir une interface intuitive et efficace, nous avons choisi de nous concentrer sur les 15 meilleures variables prédictives.

Graphique I : Random Forest graph



Source : Les auteurs à partir des données de l'ADEME, 2023

Ce graphe présente les 15 premières variables sélectionnées par le modèle de Random Forest pour prédire les caractéristiques du logement. Ces variables ont été identifiées comme ayant le plus grand impact sur la prédiction des données du Diagnostic de Performance Énergétique (DPE). Elles représentent des aspects tels que la surface habitable du logement, l'année de construction du DPE, la ventilation, les surfaces murales et vitrées, ainsi que les différents types de générateurs d'énergie.

Cette visualisation met en évidence l'importance de ces variables dans la prédiction des caractéristiques énergétiques du logement évalué, offrant ainsi un aperçu des facteurs clés à

prendre en considération pour améliorer l'efficacité énergétique et les performances du bien immobilier.

Dans le cadre de l'amélioration du processus d'analyse, plusieurs étapes ont été entreprises :

- ✚ **Intégration de la Fonction de Génération de Graphiques** : Une nouvelle fonctionnalité a été intégrée pour générer des graphiques représentant les facteurs les plus significatifs.
- ✚ **Implémentation de la sélection de variables dans l'algorithme K-NN** : Le processus d'analyse par l'algorithme K-NN a été amélioré en intégrant la sélection dynamique de variables.
- ✚ **Tests rigoureux des fonctions et transformations** : Toutes les fonctions et transformations ont été soumises à des tests approfondis pour s'assurer de leur bon fonctionnement dans divers scénarios.

1.4 Description des variables exploités

Pour mener à bien ce projet, il est important de comprendre les variables qui seront exploitées dans notre processus d'analyse. Ces variables sélectionnées, jouent un rôle central dans la mise en oeuvre de stratégies visant à réduire la consommation d'énergie tout en maintenant un niveau de confort et d'efficacité. La variable expliquée dans notre modèle est la "**Classe du Diagnostic de Performance Énergétique (DPE)**". C'est une variable qualitative (A,B,C,D,E,F,G) qui reflète la catégorie attribuée aux logements en fonction de la synthèse des consommations d'énergie primaire et des émissions de CO2 dans les cinq principaux domaines d'utilisation : ECS, chauffage, climatisation, éclairage et auxiliaires.

Outre cette variable, les principaux facteurs que nous considérerons pour mieux prédire la Classe du DPE sont ceux retenues grâce au Random Forest. Il s'agit de :

- **Surface habitable en m²** : Contrairement à la surface totale d'un bâtiment, la surface habitable ne comprend que les zones où les occupants résident et effectuent leurs activités quotidiennes. Les espaces non habitables tels que les murs extérieurs, les sous-sols non aménagés, les greniers inutilisables, les garages et les dépendances ne sont pas inclus dans le calcul de la surface habitable.

-
- **Surface totale des murs en m²** : La variable fait référence à la superficie totale de l'ensemble des murs d'un logement ou d'un bâtiment, quelle que soit leur orientation (murs extérieurs, murs intérieurs, etc.). Elle représente la somme des superficies de tous les murs présents à l'intérieur du logement.
 - **Surface des murs extérieur en m²** : La variable fait référence à la superficie totale des murs extérieurs d'un logement ou d'un bâtiment. Ces murs extérieurs séparent l'intérieur de l'extérieur du bâtiment et contribuent à l'enveloppe du bâtiment.
 - **Surface des portes en m²** : La variable fait référence à la superficie totale des portes d'un logement ou d'un bâtiment. Cette mesure inclut la somme des surfaces de toutes les portes présentes dans l'habitation.
 - **Surface totale des hauts planchers en m²** : La variable fait référence à la superficie totale de l'ensemble des planchers hauts d'un logement ou d'un bâtiment. Les planchers hauts sont les surfaces qui séparent l'intérieur du logement du niveau supérieur ou du toit.
 - **Surface totale des bas planchers en m²** : La variable fait référence à la superficie totale de l'ensemble des planchers bas d'un logement ou d'un bâtiment. Les planchers bas sont les surfaces qui séparent l'intérieur du logement du sol, du sous-sol ou d'autres espaces situés en dessous du logement.
 - **Surface des baies vitrées sud/ nord/est/ouest en m²** : La variable fait référence à la superficie totale des baies vitrées orientées vers le sud/nord/est/ouest d'un logement ou d'un bâtiment. Les baies vitrées sud/nord/est/ouest sont les fenêtres, portes-fenêtres ou autres ouvertures vitrées qui font face au sud/nord/est/ouest
 - **Epaisseur moyenne de la structure d'un mur** : L'épaisseur moyenne de la partie structure d'un mur se réfère à la mesure moyenne de la partie du mur qui constitue sa structure de base, excluant tout matériau d'isolation ajouté ou doublage supplémentaire
 - **Année de construction de 2023 à 1595** : représente l'année à laquelle le logement ou le bâtiment a été construit. Il s'agit d'une donnée quantitative qui indique depuis combien d'années le logement existe.
 - **Générateur ballon accumulation électrique** : Un ballon à accumulation est un réservoir d'eau chaude confiné par un isolant thermique permettant de stocker de l'eau chaude pour la délivrer à une installation sanitaire ou de chauffage.

- **Générateur thermodynamique électrique** : Ces centrales convertissent généralement l'énergie thermique en électricité en utilisant la vapeur d'eau pour faire tourner une turbine connectée à un générateur électrique afin de produire de l'eau.
- **Ventilation mécanique à extraction hygroréglable** : La ventilation mécanique à extraction hygroréglable est un système de ventilation mécanique qui ajuste son débit en fonction de l'humidité ambiante à l'intérieur du bâtiment. Ce type de ventilation est particulièrement utile dans des espaces tels que les salles de bains et les cuisines, où l'humidité peut s'accumuler rapidement.

1.5 Conception de l'interface utilisateur et de sortie

La conception de l'interface utilisateur et de sortie avec VBA (Visual Basic for Applications) implique de créer une interface conviviale pour les utilisateurs et de prévoir la manière dont les résultats seront affichés. Pour ce faire, on a :

- ✚ Identifier dans un premier temps les informations que les utilisateurs devront saisir. Ces informations sont relatives aux variables qui seront dans notre base de données ;
- ✚ Déterminer la catégorie de données qui seront générées (qualitatives/quantitatives) et quel format elles auront en sortie (les graphiques adaptés à chaque type de données) ;
- ✚ Créer une esquisse de l'interface utilisateur pour visualiser l'emplacement des composants tels que les boutons, les zones de texte, les listes déroulantes, etc...

L'interface de sortie ¹des résultats du Diagnostic de Performance Énergétique (DPE) est conçue pour fournir une présentation claire et informative des évaluations énergétiques. Cette interface dynamique a été créée à l'aide d'une macro VBA et des fonction Excel, garantissant une analyse précise et détaillée.

- ✚ L'interface commence par la catégorisation du Diagnostic de Performance Énergétique (DPE) selon les données fournies par l'utilisateur. D'abord, une fonction Excel détermine la classe énergétique ayant la proportion la plus élevée parmi les trois classes générées par l'algorithme. Ensuite, une formule RechercheH est utilisée pour identifier la lettre

¹ Cf. annexe 3 : Interfaces – Figure 5 : Interface de résultat

correspondant à cette proportion dans le tableau des résultats. Et enfin, le résultat est affiché.

- ✚ Un diagramme en anneau complète la catégorisation en illustrant visuellement la répartition des différentes catégories du DPE. Des pourcentages indiquent la probabilité d'appartenance à chaque catégorie. Cette information permet d'apporter une nuance supplémentaire et d'approfondir la compréhension des évaluations.
- ✚ L'interface présente également un diagramme en bâtons mettant en lumière les facteurs (variables) qui ont le plus contribué à l'obtention de la catégorie. Cette représentation visuelle permet de cibler les éléments clés ayant influencé la classification du DPE.
- ✚ Un petit texte qui donne à l'utilisateur une estimation basée sur les recommandations de l'ADEME, offrant une projection financière pour leurs éventuelles améliorations en matière d'efficacité énergétique. Cette estimation se base sur un logement de 75m².

1.6 Intégration de l'algorithme dans l'interface d'assistance à la décision

Pour automatiser le processus de visualisation des résultats, on relie le formulaire de saisie utilisateur en VBA à un script Python. Cette intégration se fera en plusieurs étapes clés.

- ✚ Dans une première étape, après avoir conçu le formulaire de saisie en VBA, l'utilisateur soumet les informations nécessaires qui seront automatiquement enregistrées dans une feuille Excel dédiée.
- ✚ Une fois les données enregistrées, Excel exécute le script python qui après avoir effectué le traitement nécessaire, crée un fichier texte avec la catégorie du DPE (lettre). Le script renvoie dans la feuille source trois (03) pourcentages associés à un chiffre (0=A, ...). Puis avec une fonction Excel on renvoie la classe sous forme de lettre (A, B, C, ...).
- ✚ Une fois les données enregistrées, Excel exécute le script python qui après avoir effectué le traitement nécessaire, crée un fichier texte avec la catégorie du DPE (lettre).
- ✚ Pour une visualisation efficace des résultats, nous utiliserons les données mises à jour dans Excel pour créer des graphiques représentatifs de l'information.

III- RESUTATS

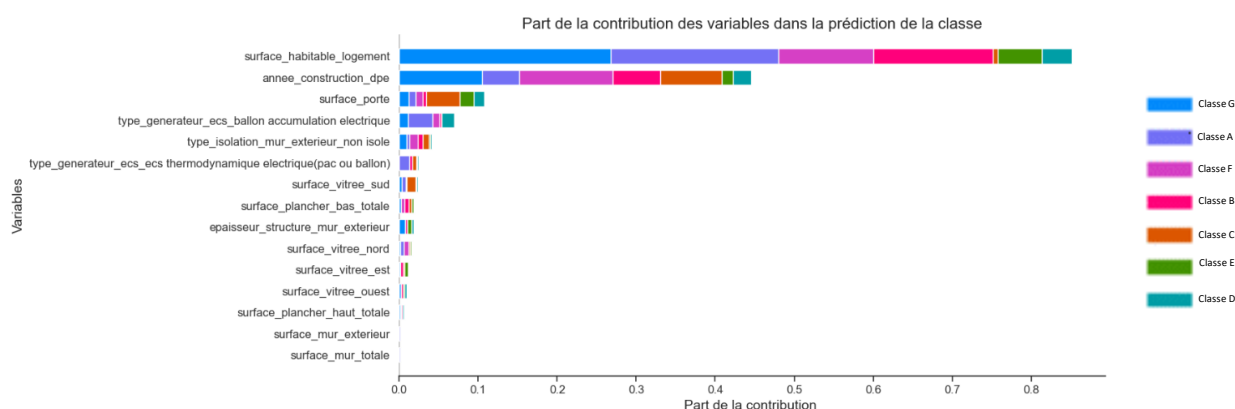
3.1 Modèle de prédiction

Le modèle de prédiction adopté est le k-NN. Cette méthode repose sur la similitude entre les échantillons dans l'espace des caractéristiques. En repérant les voisins les plus proches de notre échantillon, le modèle attribue la catégorie la plus fréquemment rencontrée parmi ces voisins. Les prédictions du modèle k-NN ont révélé des indications pertinentes concernant les catégories du DPE, offrant ainsi une vue sur les possibilités d'amélioration de l'efficacité énergétique. Ce modèle a révélé une précision de prédiction de 45% pour une classe spécifique, tandis que pour une prédiction sur trois classes, cette précision varie entre 70% et 100%. Il convient de souligner que ce modèle présente une précision plus précise pour les classes A, B et C.

Cette disparité dans les performances de prédiction entre les différentes classes met en évidence une particularité significative du modèle. Bien que la classe spécifique ait démontré une précision plus modeste de 45%, les classes A, B et C ont montré une précision remarquable. Cette tendance suggère une capacité accrue du modèle à prédire avec une grande exactitude les éléments appartenant à ces catégories spécifiques. Après avoir exploré les prédictions fournies par le modèle KNN, une analyse approfondie des impacts spécifiques des variables sur les prédictions individuelles a été réalisée en utilisant la méthode SHAP (SHapley Additive exPlanations).

SHAP est une méthode d'interprétabilité des modèles qui permet d'expliquer les prédictions de manière individuelle. Dans le cadre de notre projet, l'utilisation de SHAP est particulièrement pertinente pour comprendre l'importance des variables dans les prédictions spécifiques faites par le modèle entraîné. En appliquant SHAP à un individu test, on obtient des informations sur la contribution de chaque variable à la prédiction faite pour cet individu. Cette méthode quantifie l'impact de chaque caractéristique sur la prédiction en prenant en compte toutes les combinaisons possibles de caractéristiques.

Graphique 2 : Contribution de chaque variable par classe de DPE



Source : Les auteurs à partir des données de l'ADEME, 2023

3.2 Interface de saisie

Le formulaire de saisie de données élaboré sous VBA constitue un élément important du projet d'aide à l'optimisation de la consommation d'énergie. Sa conception, les variables sélectionnées et les restrictions intégrées garantissent la collecte et cohérente des données nécessaires à l'analyse. Le formulaire assure une expérience utilisateur intuitive et efficace². Des éléments tels que des zones de saisie et des listes déroulantes ont été intégrés pour faciliter la collecte des données.

Chaque variable nécessaire à l'analyse énergétique a été identifiée et intégrée dans le formulaire. Ces variables ont été catégorisées en deux types :

- **Variables avec Liste** : Ces variables sont associées à des options prédéfinies pour garantir la cohérence des données. Par exemple, la zone « Année de construction » est sélectionnée à partir d'une liste déroulante proposant différentes années.
- **Variables par Saisie d'Information de l'Utilisateur** : Ces variables requièrent une entrée directe de l'utilisateur. Par exemple, la surface du bâtiment doit être saisie en mètres carrés.

Dans le formulaire, deux boutons ont été intégrés pour offrir une expérience utilisateur plus fluide et permettre une gestion simplifiée des informations dans le formulaire : le bouton "**Fermer**" et le bouton "**Effacer**".

² Cf. Annexe 3 : Interfaces - Figure 3 : Interface de saisie utilisateur

Le bouton "**Fermer**" a pour fonction de refermer le formulaire sans enregistrer les modifications apportées. Il permet à l'utilisateur de sortir du formulaire sans sauvegarder les données actuellement saisies.

Quant au bouton "**Effacer**", son rôle est de réinitialiser le formulaire en effaçant toutes les données préalablement saisies. Il offre à l'utilisateur la possibilité de nettoyer le formulaire pour une nouvelle saisie, annulant ainsi toute information précédemment renseignée.

Aussi, des restrictions ont été mis en place pour garantir l'uniformité des données saisies. Des vérifications sont effectuées pour s'assurer que les valeurs sont cohérentes et respectent les caractères acceptables. Par exemple, le champ dédié aux surfaces ne peut contenir que des chiffres et est limité à 5 caractères maximum. La virgule est le séparateur décimal exigé pour ce champ. Si un utilisateur saisit une lettre ou utilise un point comme séparateur décimal, le formulaire ne pourra pas être validé, et les cellules concernées seront mises en surbrillance rouge ³.

Il est important de noter que tous les champs du questionnaire sont obligatoires. Ainsi, si l'utilisateur ne renseigne pas tous les champs requis, le formulaire signalera une erreur lors de la validation ⁴.

Une fois que les informations sont validées, elles sont enregistrées dans une feuille Excel dédiée. Une fois le traitement achevé, ces données sont supprimées pour assurer la confidentialité et la sécurité des informations collectées.

IV- DIFFICULTES

La difficulté majeure rencontrée résidait dans la rareté des données correspondant à la catégorie A du Diagnostic de Performance Énergétique (DPE). La rareté des données relatives à la classe énergétique A a nécessité des ajustements dans la base de données pour assurer une représentation équilibrée de toutes les catégories du Diagnostic de Performance Énergétique. Afin de pallier cette lacune, la base de données a fait l'objet de trois révisions successives. Chaque

³ Cf. annexe 3 : Interfaces – Figure 4 : Messages d'erreur

⁴ Cf. annexe 3 : Interfaces – Figure 4 : Messages d'erreur

itération a visé à accroître la représentativité de la classe A en incorporant des données supplémentaires.

Aussi, la variabilité des formats, la qualité variable des données et les valeurs manquantes ont nécessité des efforts importants pour assurer la cohérence et la fiabilité des données.

Un autre des défis a été d'harmoniser la sélection des variables avec les résultats obtenus par le modèle de Random Forest. La sélection des variables influençant le modèle, notamment celles identifiées comme les plus prédictives, a été cruciale pour obtenir des prédictions précises. Cependant, il était tout aussi essentiel de déterminer si ces variables étaient facilement compréhensibles et exploitables pour les utilisateurs de l'outil. La question clé a été de savoir si les utilisateurs finaux de l'outil auront accès à ces informations sur les variables prédictives.

V- RECOMMANDATION

Pour optimiser l'expérience et maximiser les résultats, voici quelques recommandations qu'on vous propose :

- Il pourrait être envisagé d'enrichir les données en incluant des variables liées au "climat", à la "géographie" ou à la "température minimale", en effectuant une recherche détaillée pour chaque zone géographique. Par exemple, étant donné que l'adresse des individus était une variable, une approche intéressante aurait pu consister à utiliser des outils tels que QGIS pour géolocaliser les logements. Cela aurait permis d'utiliser différentes cartes pour obtenir des informations supplémentaires sur chaque logement, ou même sur la région environnante (ville, campagne, montagne, bord de mer).
- Les classes qui ont montré une précision modeste nécessitent une attention particulière pour comprendre les raisons sous-jacentes de cette performance inférieure. Il serait judicieux de mener des analyses approfondies et des ajustements ciblés pour améliorer la capacité prédictive de ce modèle pour ces classes spécifiques.
- D'autres modèles peuvent être envisagés pour ce cas précis. Par exemple, le Support Vector Machine présente une flexibilité intéressante, permettant d'ajuster les noyaux pour mieux s'adapter aux données. En revanche, bien que les Arbres de Décision et la Forêt Aléatoire soient puissants, leur mise en œuvre peut s'avérer plus complexe lorsqu'il est nécessaire de définir des frontières précises, telles que celles observées dans notre cas, car cela peut altérer aléatoirement la sélection des variables.

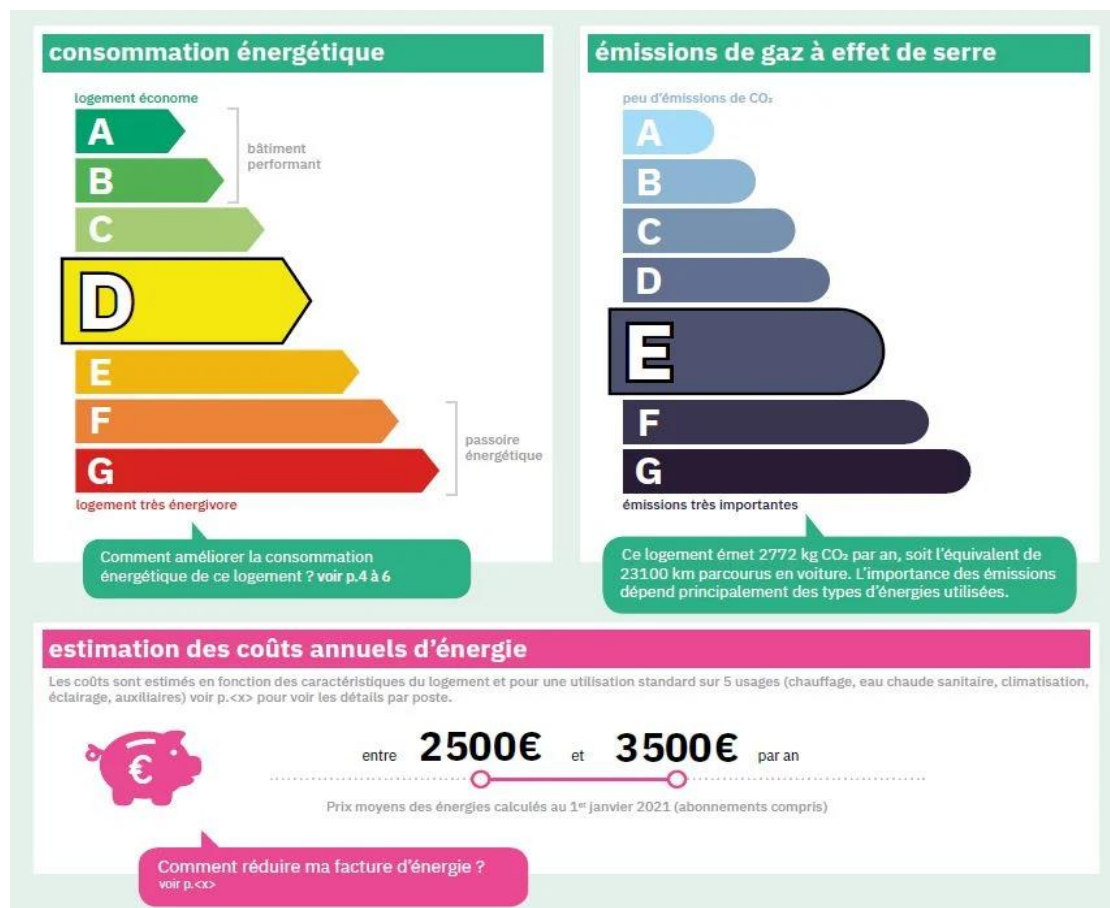
BIBLIOGRAPHIE

1. Agence de la transition écologique (2023). Disponible sur *expertises.ademe.fr*
2. Insee (2010). Les chiffres du logement de la Marne
3. Merly-Alpa, T. (2023). Comment mesurer ? L'exemple de la rénovation énergétique des logements. *Statistique et Société*, 11(1), 11-22.
4. Site de Service-Public : Diagnostic immobilier : diagnostic de performance énergétique (DPE). Disponible à <https://www.service-public.fr/particuliers/vosdroits/F16096>.
5. Site de Gouvernement : Interdiction à la location des logements avec une forte consommation d'énergie.
Disponible à <https://www.gouvernement.fr/actualite/interdiction-a-la-location-des-logements-avec-une-forte-consommation-d-energie-des-2023>

ANNEXES

Annexe I : Quelques données

Figure I : Classe de diagnostic de performance énergétique et émission de gaz à effet de serre



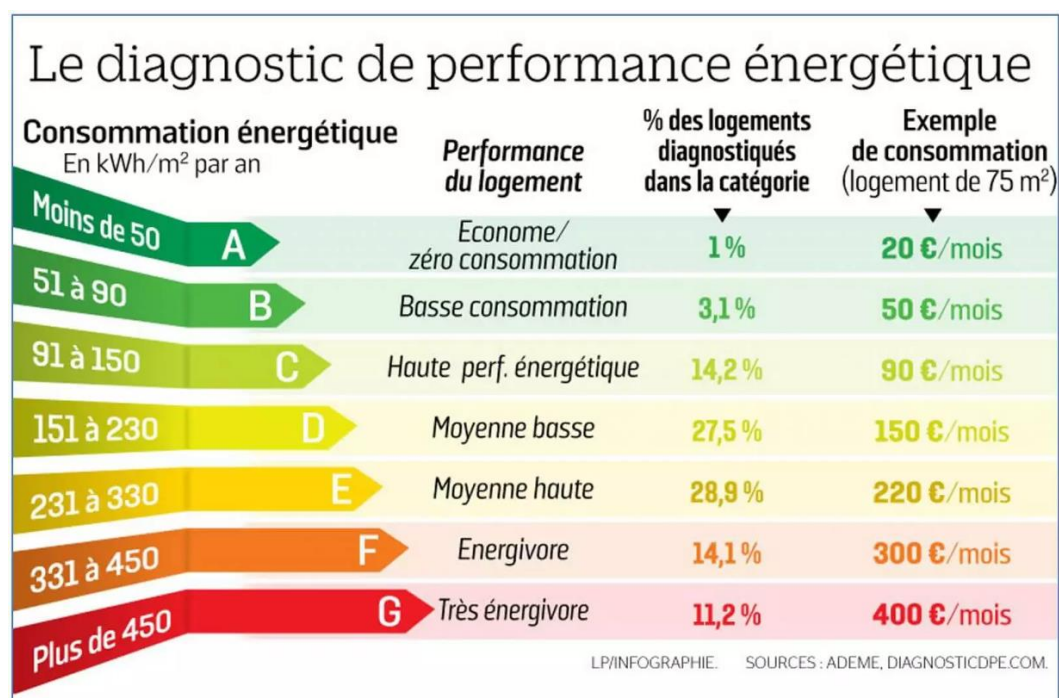
Source : Diagnostic de performance énergétique (DPE), 2021

Tableau I : Récapitulatif de la méthode à utiliser pour la réalisation du DPE

	Bâtiment à usage principal d'habitation						Bâtiment ou partie de bâtiment à usage principal autre que d'habitation
	DPE pour un immeuble ou une maison individuelle		Appartement avec système collectif de chauffage ou de production d'ECS sans comptage individuel quand un DPE a été réalisé à l'immeuble	DPE non réalisé à l'immeuble		Appartement avec système collectif de chauffage ou de production d'ECS sans comptage individuel	
				Appartement avec systèmes individuels de chauffage et de production d'ECS ou collectifs et équipés comptages individuels			
				Bâtiment construit avant 1948	Bâtiment construit après 1948		
Calcul conventionnel		X	A partir du DPE à l'immeuble		X		
Utilisation des factures	X			X		X	X

Source : Calcul DPE : Méthodes, infos, estimations, tout savoir, 2023. <https://www.dimo-diagnostic.net/actualite-diagnostic-immobilier/calcul-dpe#:~:text=Le%20chiffre%20figurant%20dans%20ce,logement%20sera%20de%20170%20k%20Wh.>

Figure 2 : Exemple de consommation de logement de 75m² en fonction de la classe de DPE



Sources : ADEME, expertises.ademe.fr

Annexe 2 : Récapitulatif des variables

Variable	Description	Type
Variables dépendantes		
Classe du Diagnostic de Performance Énergétique (DPE)	La catégorie attribuée aux logements en fonction de la synthèse des consommations d'énergie primaire et des émissions de CO ₂ dans les cinq principaux domaines d'utilisation : ECS, chauffage, climatisation, éclairage et auxiliaires.	Variable qualitative (A ; B ; C ; D ; E ; F ; G)
Variables indépendantes		
Année de Construction du Logement	l'année de construction du logement	Variable quantitative
Version du Diagnostic de Performance Énergétique (DPE)	la version spécifique du Diagnostic de Performance Énergétique, conformément à l'arrêté de 2021.	Variable quantitative
Surface Habitable du Logement	représente la superficie habitable du logement telle que définie dans le cadre du DPE, mais elle n'est pas renseignée pour les DPE réalisés au niveau de l'immeuble	Variable quantitative

Consommation Annuelle des 5 Usages (Énergie Primaire)	Consommation annuelle 5 usages (ecs/chauffage/climatisation/éclairage/auxiliaires) en énergie primaire (déduit de la production pv autoconsommée) (kWh _{ep} /m ² /an). valable uniquement pour les DPE appliquant la méthode de l'arrêté du 31 mars 2021 (en vigueur actuellement)	Variable quantitative
Consommation Annuelle des 5 Usages (Énergie Finale)	Consommation annuelle 5 usages (ecs/chauffage/climatisation/éclairage/auxiliaires) en énergie finale (déduit de la production pv autoconsommée) (kWh _{ef} /m ² /an). valable uniquement pour les DPE appliquant la méthode de l'arrêté du 31 mars 2021 (en vigueur actuellement)	Variable quantitative
Émission de Gaz à Effet de Serre (GES) des 5 Usages par mètre carré	Elle évalue les émissions totales de gaz à effet de serre en kilogrammes de CO ₂ par mètre carré par an, provenant des 5 principaux usages (ECS, chauffage, climatisation, éclairage, auxiliaires).	Variable quantitative
Type d'Installation de Chauffage	Détermine si le chauffage est de type collectif ou individuel.	Variable qualitative
Type d'Énergie du Générateur de Chauffage Principal	Type d'énergie pour le générateur de chauffage principal (enum version simplifiée BDNB)	Variable qualitative
Type de Générateur de Chauffage Principal	Précise le type de générateur de chauffage utilisé en tant que source principale de chaleur dans le logement.	Variable qualitative
L'ancienneté du Générateur de Chauffage Principal	Renseigne sur la durée d'utilisation du générateur de chauffage.	Variable qualitative
Nombre de Générateurs de Chauffage	Indique combien de générateurs de chauffage sont présents dans le logement.	Variable quantitative
Nombre d'Installations de Chauffage	Il révèle combien d'installations de chauffage existent dans le logement	Variable quantitative
Type de Générateur de Climatisation Principal	Elle précise le type de générateur de climatisation utilisé comme source principale de refroidissement	Variable qualitative
l'Ancienneté du Générateur de Climatisation Principal	Evaluer la durée d'utilisation du générateur de climatisation principal	Variable qualitative
Type d'Installation d'Eau Chaude Sanitaire (ECS)	Elle spécifie si l'ECS est de type collectif ou individuel	Variable qualitative
Type d'Énergie du Générateur d'Eau Chaude Sanitaire (ECS) Principal	Renseigne sur la source d'énergie utilisée par le générateur d'ECS principal	Variable qualitative
Type de Générateur d'Eau Chaude Sanitaire (ECS) Principal	Précise le type de générateur utilisé pour la production d'ECS	Variable qualitative
Ancienneté du Générateur d'Eau	Evaluer la durée d'utilisation de cet équipement	Variable qualitative

Chaud Sanitaire (ECS) Principal		
Présence d'ECS Solaire	Indique si le logement est équipé d'un système d'eau chaude sanitaire solaire	Variable qualitative
Nombre de Générateurs d'ECS	Spécifie combien de générateurs d'ECS sont présents dans le logement	Variable quantitative
Nombre d'Installations d'ECS	indique combien d'installations d'ECS existent dans le logement	Variable quantitative
Présence de Plusieurs Façades Exposées au Vent	Indique si le logement dispose de plusieurs façades exposées au vent.	Variable quantitative
Type de Ventilation	Spécifie le type de système de ventilation en place.	Variable quantitative
Type de Production d'Énergie Renouvelable	Précise le type de production d'énergie renouvelable utilisé dans le cadre du DPE.	Variable quantitative
Type de Vitrage Principal	Précise le type de vitrage des baies vitrées du logement.	Variable quantitative
Type de Matériaux de Menuiserie Principal	Indique le matériau principal utilisé pour les menuiseries des baies vitrées du DPE.	Variable quantitative
Type de Gaz Injecté dans la Lame Entre les Vitrages des Baies Vitrées	Spécifie le type de gaz utilisé dans la lame entre les vitrages des baies vitrées du DPE. (air sec ; argon ou krypton)	Variable quantitative
Type de Fermeture Principale des Baies Vitrées	Précise le type de fermeture installé sur les baies vitrées du DPE, qu'il s'agisse de volets, de persiennes, ou d'autres dispositifs.	Variable quantitative
Vitrage Traité avec un Traitement à Isolation Renforcée	indique si le vitrage a été soumis à un traitement d'isolation renforcée	Variable quantitative
Surface de Baies Vitrées Orientées au Nord	représente la somme des surfaces des baies vitrées orientées vers le nord.	Variable quantitative
Surface de Baies Vitrées Orientées au Sud	Représente la somme des surfaces des baies vitrées orientées vers le Sud.	Variable quantitative
Surface de Baies Vitrées Orientées à l'Ouest	Représente la somme des surfaces des baies vitrées orientées vers l'Ouest.	Variable quantitative
Surface de Baies Vitrées Orientées à l'Est	Représente la somme des surfaces des baies vitrées orientées vers l'Est.	Variable quantitative
L'Indicateur du Côté Traversant du Logement	Indique si le logement est traversant, c'est-à-dire s'il possède une exposition des deux côtés.	Variable quantitative
Coefficient de Transmission Thermique des Baies Vitrées (Ujn)	Quantifie la capacité des baies vitrées à transférer la chaleur. Cela inclut le calcul de la résistance additionnelle des fermetures.	Variable quantitative

Facteur de Transmission du Flux Solaire par la Baie Vitrée	Mesure le coefficient de transmission du flux solaire par la baie vitrée, un nombre compris entre 0 et 1.	Variable quantitative
Présence de Balcons Identifiés par Analyse des Coefficients de Masques Solaires	Indique si des balcons ont été identifiés par l'analyse des coefficients de masques solaires du DPE.	Variable qualitative
Liste des Orientations des Baies Vitrées	Spécifie les orientations des baies vitrées	Variable qualitative
Type d'Isolation Principale des Murs Donnant sur l'Extérieur	Précise le type d'isolation utilisé pour les murs extérieurs du DPE.	Variable qualitative
Matériau ou Principe Constructif Principal des Murs Extérieurs	Renseigne sur le matériau ou le principe constructif principal utilisé pour les murs extérieurs.	Variable qualitative
L'Épaisseur Moyenne de la Partie Structure des Murs Extérieurs	indique l'épaisseur moyenne de la partie structure des murs extérieurs, sans inclure l'isolation rapportée ni les doublages.	Variable quantitative
Surface de Murs Totale	représente la somme des surfaces totales des murs dans le logement.	Variable quantitative
Surface de Murs Donnant sur l'Extérieur	indique la somme des surfaces de murs en contact avec l'extérieur.	Variable quantitative
Surface de Murs Déperditifs	représente la somme des surfaces de murs donnant sur des locaux non chauffés et sur l'extérieur, c'est-à-dire les surfaces déperditives.	Variable quantitative
Liste des Locaux Non Chauffés en Contact avec les Murs	Elle quantifie la surface totale des planchers hauts du logement.	Variable qualitative
Liste des Orientations des Murs Donnant sur l'Extérieur	précise les orientations des murs donnant sur l'extérieur.	Variable qualitative
Type d'Isolation Principale des Planchers Bas Déperditifs	qui précise le type d'isolation utilisé pour les planchers bas déperditifs.	Variable qualitative
Matériau ou Principe Constructif Principal des Planchers Bas	indique le matériau ou le principe constructif principal utilisé pour les planchers bas.	Variable qualitative
Surface de Planchers Bas Totale	représente la somme des surfaces totales des planchers bas dans le logement.	Variable quantitative
Surface de Planchers Bas Déperditifs	Indique la somme des surfaces de planchers bas donnant sur des locaux non chauffés et sur l'extérieur, c'est-à-dire les surfaces déperditives.	Variable quantitative
Locaux Non Chauffés en Contact avec les Planchers Bas	Liste les locaux non chauffés en contact avec les planchers bas, une information essentielle pour	Variable qualitative

	évaluer l'efficacité énergétique des planchers bas, conformément à la classification DPE 2021.	
Type d'Adjacence Principale des Planchers Bas	Renseigne sur l'adjacence principale des planchers bas, indiquant s'ils sont en contact avec l'extérieur ou avec des locaux non chauffés	Variable qualitative
Type d'Isolation Principal des Planchers Hauts Déperditifs	Le type d'isolation principalement utilisé pour les planchers hauts	Variable qualitative
Matériau ou Principe Constructif Principal des Planchers Haut	Cette variable qualitative définit le matériau ou le principe constructif principal utilisé pour les planchers hauts.	Variable qualitative
Surface de Plancher Haut Totale	Quantifie la surface totale des planchers hauts du logement. Une grande surface de plancher haut peut indiquer une plus grande zone nécessitant une isolation adéquate pour optimiser l'efficacité énergétique.	Variable quantitative
Surface de Plancher Haut Donnant sur des Locaux Non Chauffés et sur l'Extérieur (Surfaces Déperditives)	Quantifie la superficie des planchers hauts en contact avec des locaux non chauffés et exposés à l'extérieur.	Variable quantitative
Locaux Non Chauffés en Contact avec les Planchers Hauts	<i>Liste les locaux non chauffés en contact avec les planchers hauts</i>	Variable qualitative
Type d'adjacence Principale des Planchers Hauts	Renseigne sur l'adjacence principale des planchers hauts, indiquant s'ils sont en contact avec l'extérieur ou avec des locaux non chauffés	Variable qualitative
Type de Porte du DPE	Le type de porte utilisé pour le DPE	Variable qualitative
Surface de Portes du DPE	Elle quantifie la surface totale des portes du DPE.	Variable quantitative
Classe d'Inertie du DPE	Définit la classe d'inertie attribuée au DPE, selon la classification BDNB (légère, moyenne, lourde, très lourde)	Variable qualitative

Source : Les auteurs 2023

Annexe 3 : Interfaces

Figure 3 : Interface de saisie utilisateur

DPE
Diagnostic Performance Energétique

Diagnostic de Performance Energétique

Prévision sur les appartements du Grand-Est


Logement économe

3-50 A	XXX Kwh/m.an
51 à 90 B	
91 à 150 C	
151 à 230 D	
231 à 330 E	
331 à 450 F	
> 450 G	
Logement énergivore	

Ce formulaire a été conçu pour prédire le DPE d'un appartement du Grand-Est. Vous retrouverez une notice utilisateur en documentation.

Vous aurez besoin de prendre des mesures dans votre appartement pour remplir ce formulaire. Une fois le formulaire rempli, vous aurez une prédiction du DPE ainsi que l'estimation plus précise dans 3 classes.

De plus, vous aurez des informations supplémentaire sur les potentiels travaux à effectuer



Formulaire de saisie

Votre appartement

Général

Surface habitable : m2

Année de construction :

Murs

Surface mur total : m2

Surface mur extérieur : m2

Epaisseur mur extérieur : cm

Type ventilation

Ventilation mécanique à extraction hydropneumatique :

Type générateur Eau Chaude Sanitaire

ECS thermodynamique électrique (pac ou ballon) :

ECS ballon accumulation électrique :

Porte

Surface porte : m2

Plancher

Surface plancher haut total : m2

Surface plancher bas total : m2

Vitre

Surface vitre sud : m2

Surface vitre nord : m2

Surface vitre est : m2

Surface vitre ouest : m2

Fermer

Effacer

Valider

Source : Les auteurs 2023

Figure 4 : Messages d'erreur

Type générateur Eau Chaude Sanitaire

0 = Non

ECS thermique

ECS solaire

Porte

Surface

Plancher

Surface plancher haut total : m2

Surface plancher bas total : m2

Vitre

Surface vitre sud : m2

Surface vitre nord : m2

Surface vitre est : m2

Surface vitre ouest : m2

Fermer Effacer Valider

Type générateur Eau Chaude Sanitaire

0 = Non

ECS thermique

ECS solaire

Porte

Surface

Plancher

Surface plancher haut total : m2

Surface plancher bas total : m2

Vitre

Surface vitre sud : m2

Surface vitre nord : m2

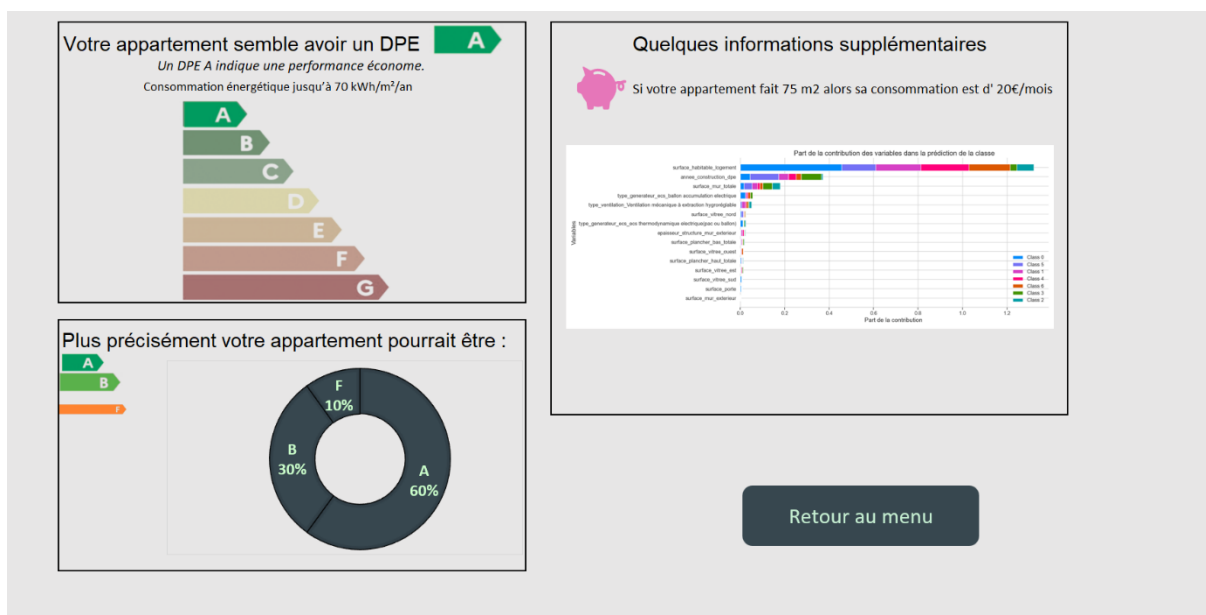
Surface vitre est : m2

Surface vitre ouest : m2

Fermer Effacer Valider

Source : Les auteurs 2023

Figure 5 : Interface de résultat



Source : Les auteurs 2023

Annexe 4 : Roadmap du projet & Notice d'utilisateur