

République Algérienne Démocratique et Populaire
Université Abderrahmane MIRA de Béjaïa
Faculté des Sciences Exactes

Département de Recherche Opérationnelle



Mémoire Présenté en vue de l'obtention du Diplôme de Master
en Mathématiques Appliquées

Spécialité : Sciences de Données et Aide à la Décision

**Prédiction de la demande d'électricité et équilibrage de
charge par classification intelligente**

Présenté par :
Saidene Samy
Bakour Yacine

Sous la direction de : Dr L. Asli

Défendu le 29/06/2025, devant le jury composé de :

M ^r B. Brahmi	M.C. classe/ A	Président de jury	UAMB - Bejaia.
M ^r A. Laouar	M.A. classe/ A	Examineur	UAMB - Bejaia
M ^{elle} I. Tetah	Doctorante	Examinatrice	UAMB - Bejaia.

Année Universitaire 2024 – 2025

Remerciements

Louange à Dieu, le miséricordieux, sans Lui rien de tout cela n'aurait pu être .

*On souhaite exprimer notre profonde gratitude à notre encadrant, **Monsieur ASLI Larbi**, pour sa bienveillance, sa rigueur scientifique et son accompagnement tout au long de ce travail. Sa disponibilité, ses conseils avisés et son exigence méthodologique ont constitué un appui fondamental dans la réalisation de ce mémoire.*

Nos remerciements vont également à l'ensemble des membres du jury pour l'intérêt porté à notre travail et pour les remarques constructives qui enrichissent cette recherche.

On remercie très sincèrement l'ensemble du corps enseignant du Master Science des données et aide à la décision pour la qualité des enseignements, leur dévouement et leur engagement à transmettre des compétences solides, aussi bien théoriques que pratiques, qui nous seront précieuses dans la suite de notre parcours.

On adresse également nos remerciements à tous nos camarades, en particulier ceux de notre spécialité, pour leur entraide, leurs échanges enrichissants et leur esprit de collaboration, qui ont rendu cette expérience académique encore plus humaine et stimulante.

Dédicaces

*À mes chers parents,
pour leur amour inconditionnel, leurs prières et leur soutien constant.*

*À mon petit frère et à ma sœur,
pour leur tendresse et leur encouragement.*

*À mes chers amis,
pour leur amitié sincère, leur aide et leur présence tout au long de ce parcours.
Ce travail vous est dédié avec amour, respect et gratitude.*

Samy.S

Je dédie ce modeste travail à toute ma famille, et en particulier :

*À ma mère, pour son amour inconditionnel, ses prières silencieuses et sa force discrète qui
m'ont porté jusqu'ici*

*À mon frère Ilyes, pour sa présence, son soutien loyal et ses encouragements constants dans
les moments décisifs*

Yacine.B

Table des matières

Remerciements	I
Dédicaces	II
Liste des figures	VI
Liste des tables	VII
Liste d'abréviations et notations	VIII
Introduction générale	1
1 Classification dans le domaine de l'énergie	3
Introduction	3
1.1 Concepts généraux de l'énergie électrique	3
1.2 Méthodes de production de l'énergie électrique	4
1.3 Évolution de la consommation d'énergie électrique	5
1.3.1 Évolution de la consommation d'énergie électrique par secteurs	6
1.4 Consommation d'énergie domestique	7
1.5 Approches mathématiques et algorithmiques pour l'équilibrage de charge	9
1.6 Principe de classification dans le domaine de l'énergie	10
1.7 Limitations et défis rencontrés dans la classification des foyers énergétiques	11
Conclusion	12
2 État de l'art pour la classification énergétique	13
Introduction	13
2.1 L'intelligence artificielle	14
2.2 Apprentissage automatique	15
2.2.1 Phases de l'apprentissage automatique	15
2.2.2 Limites de l'apprentissage automatique	16
2.2.3 Types d'apprentissage automatique	17
2.3 Travaux connexes	20
2.4 Algorithmes de l'apprentissage automatique	21
2.4.1 k-moyennes	22
2.4.2 Hiérarchique Agglomératif	23
2.4.3 DBSCAN	24
2.4.4 K-nearest neighbors (KNN)	25
2.4.5 Arbre de décision	28
2.4.6 Forêts aléatoires	30
2.5 Apprentissage profond	30
2.6 Métrique d'évaluation des modèles :	31

2.6.1	Accuracy (Précision globale) :	31
2.6.2	Precision (Précision positive) :	31
2.6.3	Recall (Rappel ou Sensibilité) :	31
2.6.4	F1-score :	31
2.6.5	AUC-ROC :	31
2.6.6	La métrique silhouette :	32
2.7	Optimisation et validation des modèles :	32
2.7.1	Validation croisée :	32
2.7.2	Optimisation avec GridSearchCV :	33
	Conclusion :	33
3	Étude de Cas et Résultats de Classification Énergétique	34
	Introduction :	34
3.1	Environnement de développement :	35
3.1.1	Langage de programmation :	35
3.1.2	Bibliothèques et frameworks pour la data science avec Python :	35
3.1.3	Environnement de développement intégré IDE :	37
3.2	Approche 1 : classification des foyers et des appareils pour l'optimisation de la consommation électrique résidentielle :	37
3.2.1	Collecte et chargement des données :	38
3.3	Modèles d'apprentissage automatique :	39
3.3.1	Classification des consommateurs selon leur consommation :	39
3.3.2	Équilibrage de la charge :	39
3.3.3	Décalage des machines flexibles vers les heures creuses :	40
3.4	Présentation des résultats :	40
3.4.1	Résultats de la Classification des consommateurs selon leur consommation :	41
3.4.2	Résultats de l'équilibrage de la charge :	41
3.5	Evaluation des modèles :	44
3.6	Approche 2 : Anticipation des surcharges électriques par classification supervisée et ajustement des usages flexibles :	45
3.7	Prétraitement des données :	45
3.7.1	Description des données initiales :	45
3.7.2	Nettoyage et transformation :	46
3.7.3	Statistiques descriptives :	46
3.7.4	Structuration finale du dataset :	46
3.8	Clustering horaire saisonnier :	47
3.8.1	Construction des profils horaires :	47
3.8.2	Méthode de clustering :	47
3.8.3	Résultats du clustering :	47
3.8.4	Visualisation et interprétation :	47
3.9	Segmentation des appareils électroménagers :	48
3.9.1	Construction des descripteurs :	48
3.9.2	Méthode de clustering :	49
3.9.3	Résultats du clustering :	49
3.9.4	Visualisation des clusters :	49

3.10	Classification de la surcharge des foyers	50
3.10.1	Objectif de la classification	50
3.10.2	Construction du dataset	50
3.10.3	Sélection des variables explicatives	51
3.10.4	Modèles de classification testés	51
3.10.5	Évaluation des performances	51
3.11	Recommandation de décalage d'activation d'appareils	52
3.11.1	Objectif	52
3.11.2	Règle heuristique de décalage	52
3.11.3	Visualisation de l'effet du décalage	53
3.12	Discussion	54
	Conclusion	54
4	dd	56
	Conclusion générale	57
	Bibliographie	60
	Résumé	61

Table des figures

1.1	Evolution de la consommation mondiale d'énergie électrique [14]	6
1.2	La consommation d'électricité finale par secteurs dans le monde en 2018 [14]	7
2.1	Relation entre IA, apprentissage automatique et apprentissage profond	14
2.2	Phases de l'apprentissage automatique	16
2.3	Types d'apprentissage automatique	17
2.4	Exemple des tâches de l'apprentissage supervisé	18
2.5	Exemple d'apprentissage automatique non supervisé	19
2.6	Exemple d'apprentissage automatique avec k-moyennes	23
2.7	Les étapes d'un algorithme de clustering hiérarchique agglomératif [8]	24
2.8	Exemple d'apprentissage automatique avec DBSCAN	25
2.9	Exemple d'un KNN	27
2.10	Exemple d'un arbre de décision	29
3.1	Logo de Python	35
3.2	Logo de Jupyterlab	37
3.3	Chargement des données	38
3.4	Illustration des profils de consommation des appareils par foyer.	39
3.5	Classification des consommateurs selon leur consommation avec k-means	41
3.6	Résultats de la classification des plages horaires	42
3.7	Résultats de la classification des machines	43
3.8	Résultats finale de la Stratégie d'équilibrage	43
3.9	Résultats d'Ajustement Énergétique	44
3.10	En haut : méthode du coude montrant que $k = 2$ est optimal pour le clustering horaire. En bas : visualisation des clusters horaires pour la saison d'automne distinguant les heures creuses et pleines.	48
3.11	Évolution du silhouette score en fonction du nombre de clusters.	49
3.12	Visualisation des clusters d'appareils électroménagers	50
3.13	Comparaison des performances des modèles de classification selon les principales métriques.	52
3.14	Impact de la stratégie de décalage sur la charge électrique	53

Liste des tableaux

2.1	Types d'apprentissage et leurs algorithmes associés	21
2.2	Configuration clé de GridSearchCV [19]	33
3.1	Évaluation comparative des modèles selon la résolution de problème	44

Liste d'abriviations et notations

IA	Intelligence Artificielle
ML	Machine Learning (Apprentissage automatique)
DL	Deep Learning (Apprentissage profond)
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
KNN	<i>k</i> -Nearest Neighbors
RF	Random Forest (Forêt aléatoire)
CSV	Comma-Separated Values (Format de données)
API	Application Programming Interface
kWh	Kilowattheure
RTE	Réseau de Transport d'Électricité
EDF	Électricité de France
SGBD	Système de Gestion de Base de Données
CPU	Central Processing Unit

Introduction générale

Dans un contexte mondial marqué par la transition énergétique, la maîtrise de la consommation d'électricité devient un enjeu stratégique majeur. L'augmentation constante de la demande, la variabilité des sources d'énergie renouvelables, ainsi que la pression environnementale imposent aujourd'hui une gestion plus intelligente, plus durable et plus efficace de l'énergie. Face à ces défis, les nouvelles technologies notamment l'intelligence artificielle et l'apprentissage automatique se positionnent comme des leviers puissants pour anticiper, optimiser et équilibrer les besoins énergétiques.

Jusqu'à présent, la majorité des efforts dans le domaine énergétique ont été concentrés sur la production. Or, cette approche montre ses limites : elle ne permet ni de réduire les gaspillages ni d'adapter la consommation aux capacités réelles du réseau. Il devient alors nécessaire d'intervenir du côté de la demande, en comprenant, modélisant et anticipant les comportements des consommateurs. Pourtant, malgré les avancées technologiques, plusieurs limites persistent dans la littérature : difficulté à identifier les profils de consommation, absence de modèles flexibles et adaptés aux foyers, et rareté de stratégies dynamiques permettant un déplacement intelligent de la charge électrique.

De nombreux travaux de recherche ont émergé pour répondre à ces enjeux, en explorant l'équilibrage de la charge et l'optimisation énergétique dans des contextes variés. Certains chercheurs ont recours à des modèles de programmation mathématique, tels que la programmation linéaire en variables mixtes ou la programmation quadratique, pour planifier les usages de manière efficiente. D'autres exploitent des techniques d'intelligence artificielle comme l'apprentissage par renforcement (DQN), permettant d'ajuster la consommation en fonction du comportement des utilisateurs et des conditions du réseau[13]. Par ailleurs, des approches basées sur la classification, comme le clustering (K-means, DBSCAN) ou les forêts aléatoires, sont utilisées pour segmenter les foyers selon leurs profils énergétiques, afin d'appliquer des stratégies d'optimisation ciblées[22]. Ces contributions offrent des solutions concrètes pour réduire les pointes de charge, lisser la consommation, et promouvoir une gestion énergétique plus intelligente, personnalisée et durable.

Ce mémoire s'inscrit dans cette problématique, en proposant une approche innovante basée sur l'utilisation d'algorithmes de classification intelligente pour segmenter les foyers selon leurs profils de consommation et optimiser l'équilibrage de la charge électrique. Il poursuit plusieurs objectifs :

- Analyser l'état de l'art des techniques de classification appliquées à la gestion énergétique ;
- Identifier les approches théoriques et pratiques les plus pertinentes pour la segmentation des foyers et des usages ;

- Évaluer l'impact des méthodes de classification sur l'élaboration de stratégies d'équilibrage de charge ;
- Proposer des recommandations concrètes pour une gestion plus intelligente et équilibrée de la consommation électrique domestique.

Ce mémoire est structuré en trois chapitres :

- **Le premier chapitre** présente une revue de la littérature sur l'énergie électrique, les méthodes de classification et les approches d'équilibrage de charge.
- **Le deuxième chapitre** expose les fondements théoriques des algorithmes utilisés ainsi que les principales métriques d'évaluation.
- **Le troisième chapitre** est consacré à l'étude de cas : collecte et traitement des données, expérimentation, analyse des résultats, et propositions d'améliorations.

1

Classification dans le domaine de l'énergie

L'électricité joue un rôle central dans notre quotidien, aussi bien à la maison qu'en milieu industriel. Facile à utiliser et à transporter, elle permet d'alimenter de nombreux appareils et d'automatiser des processus.

Aujourd'hui, son usage dépasse la simple consommation : elle est au cœur des enjeux liés à la transition énergétique, à la réduction des émissions de gaz à effet de serre et au développement des technologies intelligentes, comme les véhicules électriques, les bâtiments connectés ou les smart grids.

Comprendre sa production, son utilisation et les défis physiques associés est essentiel pour en faire un usage plus durable et contribuer à un avenir plus responsable.

Sommaire

Introduction	3
1.1 Concepts généraux de l'énergie électrique	3
1.2 Méthodes de production de l'énergie électrique	4
1.3 Évolution de la consommation d'énergie électrique	5
1.4 Consommation d'énergie domestique	7
1.5 Approches mathématiques et algorithmiques pour l'équilibrage de charge	9
1.6 Principe de classification dans le domaine de l'énergie	10
1.7 Limitations et défis rencontrés dans la classification des foyers énergétiques	11
Conclusion	12

1.1 Concepts généraux de l'énergie électrique

L'énergie électrique est une forme fondamentale d'énergie dans notre société moderne, omniprésente dans notre vie quotidienne. Au cœur de cette notion se trouvent plusieurs concepts clés :

la charge électrique : La charge électrique est une propriété inhérente aux particules sub-atomiques : les électrons, porteurs d'une charge négative, et les protons, dotés d'une charge positive.

Le courant électrique : représente le déplacement ordonné de ces charges à travers un conducteur, pouvant prendre la forme d'un flux continu (courant continu) ou alternatif (courant alternatif).

La tension électrique : ou différence de potentiel, correspond à la force motrice qui déplace les charges dans un circuit ; son unité de mesure est le volt (V).

La résistance électrique : traduit la capacité d'un matériau à s'opposer au passage du courant, une grandeur exprimée en ohms (Ω).

La puissance électrique La puissance électrique caractérise l'énergie transférée par unité de temps au sein d'un circuit, un paramètre clé pour analyser son efficacité et ses performances.

Ces concepts fondamentaux structurent notre compréhension de l'énergie électrique et de son utilisation dans un grand domaine d'applications.

1.2 Méthodes de production de l'énergie électrique

Selon [34], on peut identifier plusieurs modes de production d'énergie électrique, chacun présentant certains avantages et inconvénients. Voici quelques-uns des procédés les plus couramment utilisés :

Mise en œuvre d'énergie à partir de combustibles fossiles : dans ce cas, la méthode repose sur la combustion de combustibles fossiles tels que le charbon, le pétrole ou le gaz naturel afin de produire de la chaleur. Celle-ci est ensuite utilisée pour alimenter des turbines à vapeur ou des moteurs à combustion interne, générant ainsi de l'électricité. Bien que largement utilisée en raison de son coût relativement faible, cette méthode entraîne également l'émission de quantités importantes de méthane et d'autres gaz à effet de serre, contribuant de manière significative au changement climatique.

Énergie nucléaire : l'énergie nucléaire repose sur la fission des atomes, un processus qui libère une grande quantité de chaleur. Cette chaleur sert ensuite à actionner des turbines et des générateurs pour transformer l'énergie thermique en électricité. Bien que cette méthode émette très peu de gaz à effet de serre, elle soulève des inquiétudes, notamment en ce qui concerne la sécurité des installations et la gestion des déchets radioactifs

Énergie hydroélectrique : l'énergie hydroélectrique est faite en utilisant la force de l'eau pour faire tourner des turbines, souvent liées à des barrages avec eau. Cela peut rester et peut polluer ; mais il y a aussi cette manière d'endommager l'environnement, comme le

changement des écosystèmes d'eau et la submersion des terres.

Énergie du vent : les éoliennes servent à utiliser la force du vent pour faire tourner de grandes roues et transformer l'énergie du mouvement du vent en électricité. L'énergie du vent est un moyen de produire de l'électricité propre et renouvelable, mais elle varie selon le vent et le temps qu'il fait.

Énergie du solaire : les panneaux solaires transforment la lumière du soleil en électricité grâce aux cellules photovoltaïques. Le soleil est une ressource énergétique renouvelable, disponible, abondante et peu coûteuse à exploiter, avec un fort potentiel pour réduire les émissions de gaz à effet de serre. Cependant, cette forme d'énergie reste intermittente, car elle dépend du niveau de luminosité et des conditions d'ensoleillement.

Énergie géothermique : l'énergie géothermique utilise la chaleur naturelle de la Terre pour générer de l'électricité, souvent en faisant appel à la vapeur ou à l'eau chaude provenant des profondeurs géologiques pour faire fonctionner des turbines. Cette technique est durable et assez peu polluante ; cependant, elle est confinée à des zones géographiques spécifiques.

1.3 Évolution de la consommation d'énergie électrique

Les progrès technologiques et l'accès de plus en plus large à l'électricité ont joué un rôle clé dans l'augmentation de la consommation d'énergie. L'invention de nouveaux appareils et équipements électriques, ainsi que l'utilisation croissante de l'électricité dans divers secteurs, ont contribué à une forte hausse de la consommation mondiale d'électricité.

Par exemple, l'usage des climatiseurs et des réfrigérateurs dans les foyers, ainsi que l'utilisation des machines électriques dans l'industrie, ont joué un rôle important dans cette augmentation. De plus, la croissance de la population et l'extension des villes ont également contribué à la hausse de la demande en électricité.

À cause de tous ces facteurs, la consommation d'énergie électrique dans le monde augmente de manière continue depuis la fin du 19ème siècle et devrait continuer à croître dans les années à venir. Il est donc crucial de trouver des solutions pour répondre à cette demande croissante tout en favorisant des sources d'énergie durables, afin de limiter les impacts négatifs d'une trop grande dépendance aux énergies non renouvelables.

D'après [?], la production mondiale d'électricité a considérablement augmenté au cours des trois dernières décennies, passant de moins de 12 000 térawattheures en 1990 à plus de 29 000 térawatt heures en 2022. Durant cette période, la production mondiale d'électricité n'a connu que deux baisses annuelles : en 2009, à la suite de la crise financière mondiale, et en 2020, en raison de la pandémie de coronavirus.

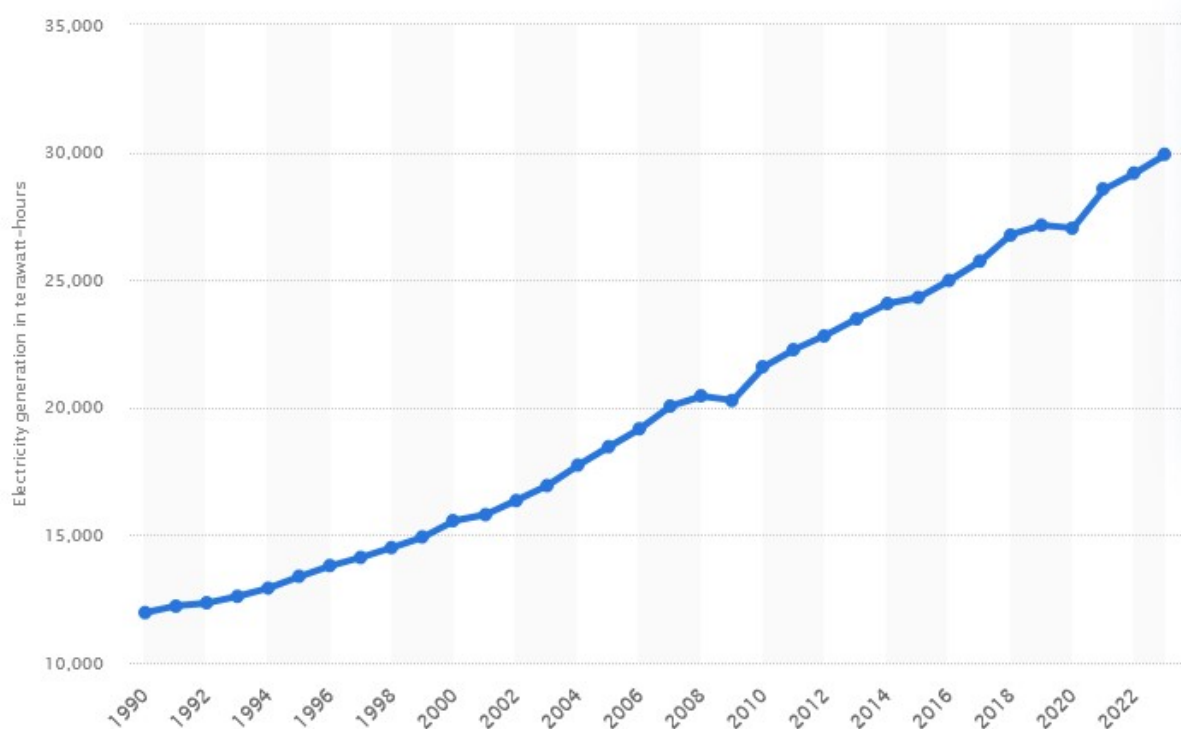


FIGURE 1.1 – Evolution de la consommation mondiale d'énergie électrique [14]

1.3.1 Évolution de la consommation d'énergie électrique par secteurs

Plusieurs secteurs entrent dans le champ de la consommation d'énergie électrique, et ce en raison de ses utilisations diverses et variées.

Les secteurs les plus importants dans lesquels l'énergie électrique apparaît sont les suivants :

Secteur résidentiel : l'énergie électrique est fournie à la population afin de satisfaire leurs exigences spécifiques liées à l'éclairage et au chauffage, tout en facilitant la réalisation des activités domestiques courantes.

Secteur des services publics : sa part d'énergie électrique est limitée au périmètre des services publics tels que les bâtiments commerciaux, les hôpitaux, les établissements d'enseignement, etc.

Secteur agricole : l'électricité sert à faire fonctionner des outils agricoles comme les faucheuses ou les pompes à eau, ainsi que d'autres machines utilisées au quotidien.

Secteur industriel : dans l'industrie, l'électricité est utilisée par les usines, les entreprises et les sites de production pour faire tourner les machines, les chaînes de montage, les systèmes automatisés et tous les équipements nécessaires à la fabrication de produits, au quotidien.

D'après [14], en 2018 la consommation d'électricité dans le monde était de 22 315 TWh. L'industrie était le plus grand consommateur avec 42%, suivie des habitations 27%. et des services 22%. Les autres usages représentaient 8%, et le transport seulement 2%. Cette répartition montre que l'électricité est surtout utilisée dans les usines et les maisons, et qu'il y a encore peu de transports électriques.



FIGURE 1.2 – La consommation d'électricité finale par secteurs dans le monde en 2018 [14]

1.4 Consommation d'énergie domestique

L'énergie que nous consommons chez nous n'est pas qu'un détail : elle pèse lourd dans la facture globale, autant pour la planète que pour notre portefeuille. Chauffage, éclairage, électroménager...

Ces besoins du quotidien ont un impact réel sur l'environnement, le coût de l'électricité et notre capacité à préserver les ressources sur le long terme. C'est là que la manière dont les foyers gèrent leur consommation entre en jeu. Plutôt que de subir, l'idée est d'anticiper et d'agir intelligemment : ajuster nos habitudes, optimiser l'usage des appareils ou encore adopter des technologies plus sobres. Une approche qui n'est pas seulement une contrainte, mais une vraie solution pour concilier confort et responsabilité. Au fond, c'est une façon de repenser notre rapport à l'énergie – pas seulement consommer moins, mais consommer mieux [29].

On utilise de plus en plus d'électricité chaque année, mais les causes de cette hausse ne sont pas les mêmes partout. Selon les pays ou les foyers, ça dépend de plein de choses : le coût de l'énergie, les innovations technologiques, les habitudes culturelles, l'organisation de la société, ou même la taille des familles. Alors, pourquoi cette tendance ? En gros, voici les principaux leviers qui influencent cette tendance :

Changement climatique : Vous l'avez sûrement remarqué : en été, quand le thermomètre s'affole, les climatisations tournent à fond ! Résultat, la consommation d'électricité grimpe en flèche pendant les canicules. À l'inverse, en hiver (du moins dans notre pays), on utilise moins d'énergie électrique pour se chauffer – mais gare aux pics si tout le monde allume ses radiateurs en même temps !

Croissance démographique : Plus de monde = plus de besoins. Quand de nouveaux quartiers sortent de terre, c'est logique : chaque foyer ajouté au réseau tire la demande vers le haut. Imaginez un immeuble neuf rempli de familles – les compteurs électriques s'activent en cascade !

Développement technologique : La technologie est partout maintenant ! Des usines automatisées aux objets connectés dans nos maisons, en passant par les data centers qui stockent nos vies numériques... Plus ces outils sont performants, plus ils paraissent économes, mais leur multiplication finit par alourdir la facture globale.

L'élévation et l'amélioration du niveau de vie : Qui dit meilleur salaire dit souvent plus d'appareils chez soi : clim' fixe, écran géant, robot cuiseur... Le "confort moderne" a un prix énergétique. Même le simple fait de s'équiper d'un deuxième frigo ou d'un home cinéma pèse sur la consommation.

Croissance économique : Usines, transports, chantiers... Quand l'économie carbure, les machines suivent le rythme ! Un secteur industriel en expansion, des camions électriques sur les routes, ou même des panneaux publicitaires lumineux – tout ça pompe des kilowatts à grande échelle.

Au final, la consommation d'énergie domestique repose sur une série de paramètres complexes, en tenant compte des comportements des occupants, de la qualité des appareils, des caractéristiques de l'isolation thermique et des conditions météorologiques.

1.5 Approches mathématiques et algorithmiques pour l'équilibrage de charge

Plusieurs travaux récents ont exploré le *load shifting* (déplacement de charge dans le temps) à l'aide de méthodes intelligentes issues des mathématiques appliquées, en particulier dans les domaines de l'énergie, du cloud computing et des systèmes distribués.

Dans le contexte des réseaux intelligents, Tushar *et al.* proposent une approche basée sur la théorie des jeux pour inciter les utilisateurs à modifier leur consommation [35]. Wang *et al.* formulent le problème comme un programme linéaire en variables mixtes [36].

Chen *et al.* appliquent l'apprentissage par renforcement profond (DQN) pour apprendre dynamiquement les décisions de déplacement de charge [13].

Dans les data centers, Sharma *et al.* intègrent la dimension thermique pour ordonnancer les tâches en fonction de la température des serveurs [31], et Zhou *et al.* utilisent des modèles prédictifs pour optimiser le *load shifting* en *edge cloud* [38].

Dans le secteur des transports, Liu *et al.* appliquent la programmation quadratique à la planification de la recharge des véhicules électriques [22].

Enfin, plusieurs travaux combinent classification intelligente et *load shifting*, notamment via des approches de *clustering* (K-means, SVM) pour adapter les stratégies selon les profils utilisateurs [3], ou encore des techniques de classification appliquées aux charges thermiques [?], permettant un délestage ciblé et optimisé.

En résumé, les approches de *load shifting* intelligentes s'appuient sur une combinaison de techniques issues des mathématiques appliquées, telles que la programmation mathématique, l'apprentissage automatique et la théorie des jeux.

L'intégration de méthodes de classification permet d'adapter dynamiquement les stratégies de déplacement de charge en fonction des profils utilisateurs ou des caractéristiques des équipements. Ces travaux offrent une base solide pour développer des solutions plus efficaces, flexibles et contextuelles, répondant aux enjeux croissants d'optimisation énergétique dans les systèmes complexes et distribués.

1.6 Principe de classification dans le domaine de l'énergie

Dans le secteur de l'énergie, classer les données, c'est un peu comme ranger une bibliothèque : on regroupe les informations par familles (consommation, production, coûts...) selon leurs caractéristiques communes. Ce tri intelligent permet de donner du sens à des chiffres qui, au premier abord, semblent abstraits. En organisant ces données, on peut repérer des motifs récurrents, comparer des scénarios, ou même anticiper des tendances – par exemple, comprendre pourquoi certains bâtiments gaspillent de l'énergie ou quelles sources renouvelables sont les plus adaptées à une région [32].

L'objectif? Transformer une montagne de chiffres en outils concrets pour les décideurs : ajuster les tarifs, améliorer l'efficacité des réseaux, ou guider les politiques publiques. Que ce soit pour réduire une facture d'électricité ou planifier une transition écologique, ce travail de catégorisation reste une base incontournable.

Voici les étapes principales à suivre pour mettre en place une classification efficace :

Sélection des caractéristiques : Identifier les attributs pertinents des données énergétiques (comme les habitudes de consommation, les profils d'usage, les facteurs environnementaux ou les données démographiques) est essentiel pour différencier efficacement les classes.

Préparation des données : Des étapes comme le nettoyage, la normalisation ou la réduction de dimensionnalité peuvent être nécessaires pour améliorer la qualité et l'utilité des données. Cela garantit un format adapté aux algorithmes de classification choisis.

Choix de l'algorithme de classification : La sélection dépend de la nature des données, de la complexité de la tâche et des résultats attendus. Les méthodes courantes incluent :

- Les arbres de décision.
- Les machines à vecteurs de support (SVM).
- Les k-plus proches voisins (k-NN).

Entraînement et évaluation du modèle : Le modèle apprend à partir de données étiquetées (où chaque entrée a une classe connue). Ses performances sont testées sur des données de test pour mesurer :

- F1-score
- AUC-ROC
- Exactitude (Accuracy)

Validation et optimisation : Le modèle est validé avec des données indépendantes pour vérifier sa fiabilité. Pour optimiser ses résultats, on ajuste ses paramètres via :

- La validation croisée.
- Grid Search

Interprétation des résultats : Comprendre comment le modèle prend ses décisions crucial. Des outils comme :

- L'analyse de l'importance des variables.
- La visualisation des résultats aident à expliquer ses conclusions et à identifier des tendances.

1.7 Limitations et défis rencontrés dans la classification des foyers énergétiques

La classification offre des perspectives prometteuses, elle reste confrontée à des obstacles techniques, éthiques et pratiques, parmi lesquels [33] :

Disponibilité et qualité des données : L'un des principaux obstacles réside dans l'accès à des données complètes et fiables sur la consommation énergétique des foyers. Ce défi est particulièrement accentué lorsque les infrastructures de comptage intelligent sont peu déployées ou que les méthodes de collecte manquent de standardisation.

Variabilité des données : La consommation énergétique varie fortement d'un foyer à l'autre, influencée par des facteurs tels que la taille du ménage, les habitudes d'occupation, les modes de vie et les caractéristiques du bâtiment. Cette grande hétérogénéité complique la création de modèles de classification capables de généraliser efficacement.

Sélection des caractéristiques : La sélection des variables pertinentes pour la classification est complexe. Si certaines informations comme la consommation quotidienne sont relativement faciles à obtenir, d'autres, liées au comportement des occupants ou à l'utilisation précise des appareils, sont difficiles à mesurer et à intégrer de manière fiable dans les modèles.

Complexité des modèles : Développer des modèles de classification performants requiert des algorithmes avancés et des ressources de calcul conséquentes. Plus les modèles deviennent sophistiqués, plus ils sont exposés aux risques de surapprentissage (overfitting) ou d'instabilité, notamment en présence de données bruitées ou limitées.

Problème de transparence : De nombreux modèles d'intelligence artificielle fonctionnent comme des « boîtes noires », rendant leurs décisions difficiles à interpréter. Pourtant, pour assurer la confiance dans les résultats, il est essentiel de comprendre et d'expliquer le raisonnement qui conduit à la classification.

Passage à grande échelle : L'analyse de données à grande échelle, couvrant de nombreux foyers, exige des algorithmes optimisés et des infrastructures informatiques distribuées, capables de traiter rapidement d'importants volumes d'informations.

Protection des données : L'exploitation des données de consommation énergétique soulève des enjeux cruciaux en matière de vie privée et de sécurité. Il est indispensable de trouver un équilibre entre la valorisation des données pour l'analyse et la protection des informations personnelles des habitants.

Conclusion

Ce chapitre a introduit les notions essentielles liées à l'énergie électrique et à sa consommation, en soulignant l'importance d'une gestion plus intelligente et efficace. La classification dans ce domaine se révèle être un levier stratégique d'analyse et d'optimisation. Le chapitre suivant s'inscrira dans cette continuité en présentant un état de l'art des approches existantes en intelligence artificielle appliquée à la classification énergétique.

2

État de l'art pour la classification énergétique

L'intelligence artificielle (IA), portée par les progrès de l'apprentissage automatique, joue un rôle central dans les transformations numériques actuelles. Elle permet de concevoir des systèmes capables d'analyser, d'apprendre et de prendre des décisions à partir de données.

Dans ce travail, j'ai appliqué des techniques telles que le clustering et la Random Forest, illustrant la diversité des approches d'analyse. Ce chapitre présente les fondements de l'apprentissage automatique et profond, leurs principales variantes, ainsi que leurs applications en prédiction. Il aborde également les enjeux liés à l'optimisation, à la validation des modèles et à l'évaluation de leurs performances, avant de conclure par un aperçu des avancées récentes dans le domaine.

Sommaire

Introduction	13
2.1 L'intelligence artificielle	14
2.2 Apprentissage automatique	15
2.3 Travaux connexes	20
2.4 Algorithmes de l'apprentissage automatique	21
2.5 Apprentissage profond	30
2.6 Métrique d'évaluation des modèles :	31
2.7 Optimisation et validation des modèles :	32
Conclusion	33

2.1 L'intelligence artificielle

D'après John McCarthy, l'un des pionniers de l'intelligence artificielle, cette discipline est « la science et l'ingénierie de créer des machines intelligentes, surtout des programmes informatiques intelligents » [24].

L'intelligence artificielle est une notion en informatique qui permet à un appareil ou un ordinateur de réfléchir et de résoudre des problèmes complexes, comme nous le faisons grâce à notre propre intelligence. Par conséquent, elle est employée pour reproduire les fonctions complexes liées à l'esprit humain, telles que la perception, le raisonnement et l'apprentissage. Quand nous réalisons une activité, des fautes peuvent se produire et nous en tirons ensuite des leçons. On s'attend également à ce que l'IA aborde un problème, commette des erreurs en tentant de le résoudre et apprenne de manière auto-correctrice dans le contexte de son auto-amélioration[2].

La figure 2.1 illustre l'écosystème de l'intelligence artificielle :

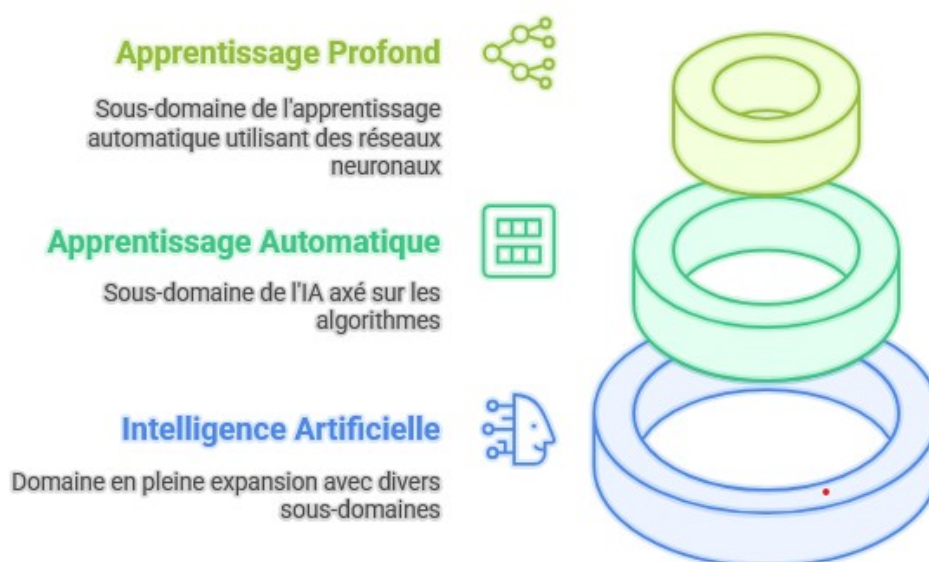


FIGURE 2.1 – Relation entre IA, apprentissage automatique et apprentissage profond

L'IA est un terme qui comprend un large éventail de technologies, notamment : le Machine Learning (ML), Deep Learning (DL).

L'intelligence artificielle est mise en œuvre de manière concrète dans divers secteurs et domaines, y compris, mais sans s'y limiter :

La santé : L'IA est mise à profit dans ce secteur pour effectuer des diagnostics médicaux, développer de nouveaux traitements et réaliser des analyses prévisionnelles concernant les maladies.

Le domaine financier : l'emploi de l'intelligence artificielle se reflète dans l'évaluation des crédits, la détection des fraudes et les prévisions financières

La vente au détail : L'IA est utilisée pour proposer des suggestions de produits, ajuster les tarifs et superviser la chaîne d'approvisionnement.

La production : L'IA garantit la qualité des produits, anticipe les pannes et améliore l'efficacité de la production.

Le transport : L'IA est exploitée dans les véhicules autonomes, et dans la prédiction et l'optimisation du trafic routier.

L'éducation : L'IA est appliquée pour individualiser l'apprentissage, mettre en place des évaluations adaptatives et développer des systèmes de tutorat intelligents.

2.2 Apprentissage automatique

Fabien Benureau [6] propose une définition de l'apprentissage qui s'applique aussi bien à un programme informatique, un robot, un animal domestique qu'à un individu humain : « L'apprentissage est une modification d'un comportement basé sur une expérience. »

On parle de machine learning (ML) dans le contexte d'un logiciel lorsqu'il est capable d'apprendre sans avoir été explicitement programmé pour chaque tâche. Cette définition a été formulée par Arthur Samuel 1959 [30].

L'apprentissage automatique (Machine Learning), sous-domaine de l'intelligence artificielle, permet aux ordinateurs d'apprendre à partir de données sans être explicitement programmés, il repose sur des algorithmes et des outils statistiques pour créer des modèles capables de prédire et d'analyser de nouvelles données. Cette technologie joue un rôle clé dans des domaines variés tels que la vision par ordinateur, la santé, l'économie ou encore la logistique [11].

2.2.1 Phases de l'apprentissage automatique

On distingue généralement deux grandes phases dans l'apprentissage automatique [11] :

La phase d'apprentissage (ou d'entraînement) : c'est une étape clé durant laquelle le modèle est exposé à de nombreux exemples issus des données d'entraînement. L'objectif est qu'il apprenne les relations ou règles sous-jacentes à partir de ces données, afin d'acquérir les connaissances nécessaires pour traiter des cas similaires à l'avenir.

La phase de test : une fois le modèle entraîné, cette phase consiste à l'évaluer sur de nouvelles données qu'il n'a jamais vues. Cela permet de vérifier s'il est capable de faire des prédictions fiables ou de prendre des décisions pertinentes dans des situations réelles, illustrant ainsi sa capacité à généraliser.

La figure 2.2 décrit les différentes phases du ML :

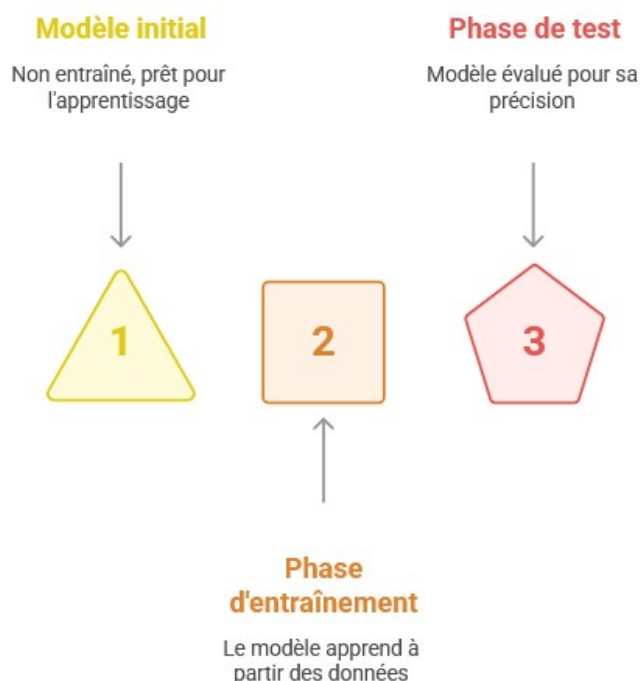


FIGURE 2.2 – Phases de l'apprentissage automatique

2.2.2 Limites de l'apprentissage automatique

L'apprentissage automatique présente plusieurs limites qu'il est important de prendre en compte [11] :

- Les prévisions ne sont jamais totalement fiables, ce qui peut poser problème dans des situations critiques.
- Certains algorithmes sont si complexes qu'il devient difficile d'expliquer clairement leurs décisions (problème de « boîte noire »).
- Détecter les erreurs et les anomalies n'est pas toujours évident, ce qui complique l'entraînement et l'amélioration des modèles.
- La complexité croissante des algorithmes oblige souvent à arbitrer entre précision et efficacité.
- La performance des modèles dépend largement des données d'entraînement, d'où la nécessité de sélectionner un jeu de données représentatif.
- L'augmentation du nombre de variables rend l'analyse plus difficile, nécessitant souvent des méthodes de réduction de dimensionnalité pour simplifier les travaux.

2.2.3 Types d'apprentissage automatique

Il existe plusieurs types d'apprentissage automatique, parmi lesquels se distinguent les quatre types les plus courants, comme illustré dans la figure 2.3 :

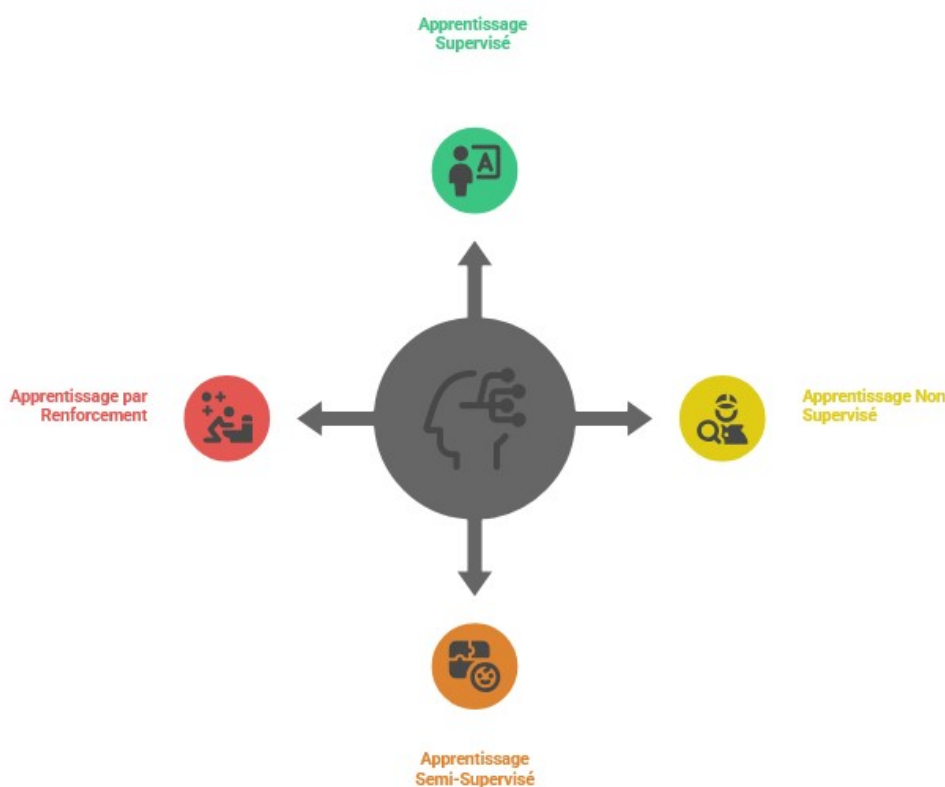


FIGURE 2.3 – Types d'apprentissage automatique

1. Apprentissage supervisé : Dans l'apprentissage supervisé, les algorithmes sont entraînés sur des jeux de données étiquetés, où chaque élément est associé à une étiquette (ou variable cible) explicite. Concrètement, ces algorithmes reçoivent des données accompagnées d'une « réponse correcte » qui indique comment elles doivent être interprétées. L'apprentissage supervisé a pour objectif de construire une fonction de prédiction $f : X \rightarrow Y$, reliant les variables explicatives x (entrées) à la variable cible y (sortie) [2]. Cette approche s'applique principalement à trois types de tâches :

- a) La classification :** Quand on veut ranger les données dans des cases bien précises.
- Classification binaire : C'est comme distinguer les chats des chiens sur des photos. Deux options, rien de plus.
 - Classification multiclasse : Là, c'est trier des fleurs en 10 espèces différentes. Plus de choix, mais le principe reste le même.

b) La régression : Quand on prédit un nombre continu, comme un prix ou une température.

- Régression linéaire : Imaginez tracer une droite pour prédire le prix d'une maison selon sa surface.
- Régression polynomiale : Si la droite ne suffit pas, on utilise une courbe pour suivre des tendances complexes.
- Régression logistique : Malgré son nom, elle sert surtout à classer (ex : "Ce mail est-il un spam ? Oui/Non").

c) Assemblage : on joue la carte de la collaboration : on combine plusieurs modèles (comme un groupe d'experts) pour obtenir des prédictions plus fiables. Par exemple, les forêts aléatoires, où plein d'arbres de décision votent pour un résultat final.

La figure 2.4 représente un exemple des tâches de l'apprentissage supervisé :

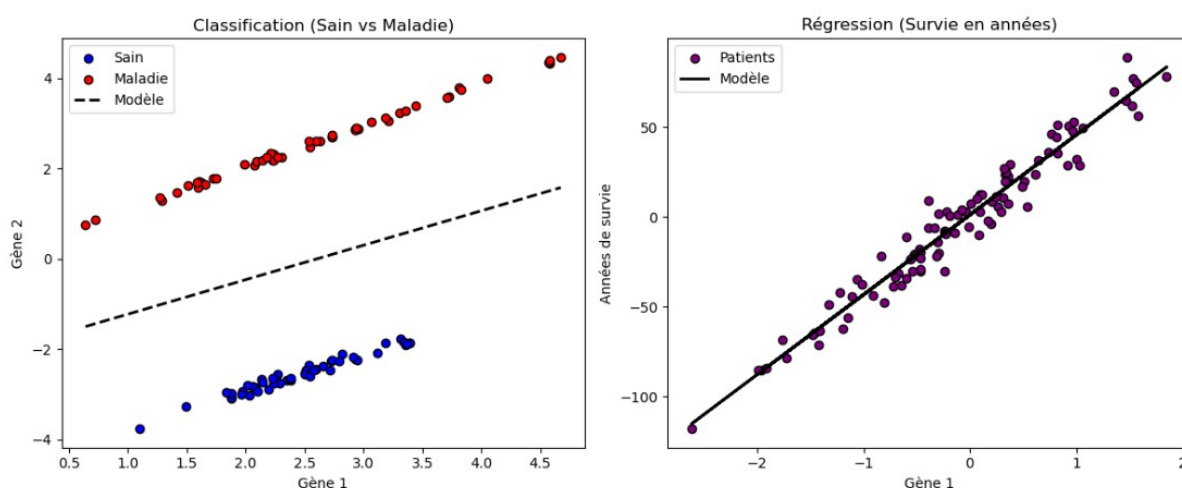


FIGURE 2.4 – Exemple des tâches de l'apprentissage supervisé

2. Apprentissage non-supervisé : Ce type d'apprentissage utilise des ensembles de données non étiquetées pour entraîner les algorithmes. Dans ce processus, l'algorithme est alimenté par des données qui ne comportent pas d'étiquettes, ce qui l'oblige à découvrir des modèles par lui-même, sans aucune aide extérieure [2].

la figure 2.5 représente l'apprentissage non supervisé :

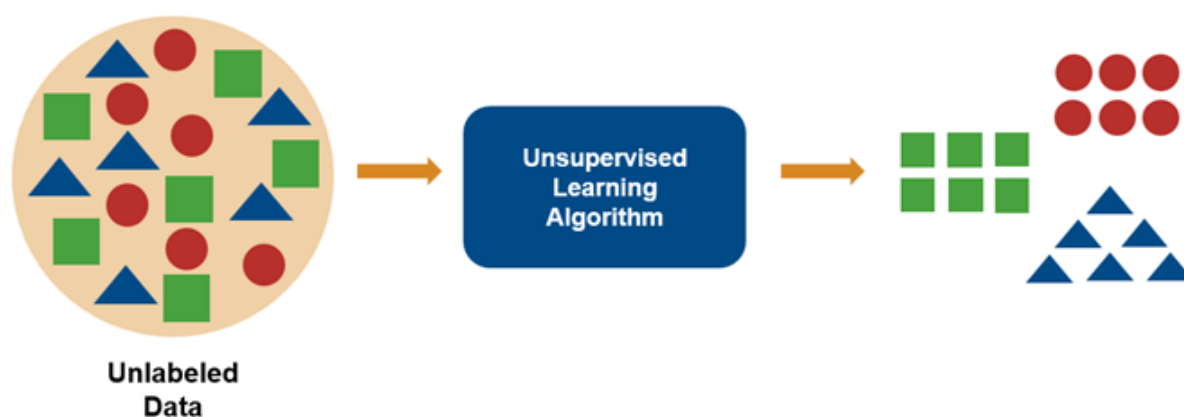


FIGURE 2.5 – Exemple d'apprentissage automatique non supervisé

Les algorithmes de cet apprentissage sont adaptés aux :

- a) **Clustering** : Division des données en clusters (groupes) selon leur similitude.
- b) **Détection d'anomalies** : Identification de points de données inhabituels dans un dataset.
- c) **Réduction de la dimensionnalité** : Diminution du nombre de variables dans un dataset.

3. Apprentissage semi-supervisé : L'apprentissage semi-supervisé utilise à la fois des données étiquetées et non étiquetées. Cela permet de créer des modèles plus précis et plus solides, tout en réduisant le coût et le temps nécessaires pour étiqueter les données.

Il s'appuie surtout sur des données non étiquetées, ce qui limite les erreurs liées aux jugements humains. Cette méthode est utilisée dans plusieurs domaines, comme la médecine, l'analyse des sentiments sur les réseaux sociaux ou encore la classification automatique de documents[2].

4. Apprentissage par renforcement : L'apprentissage par renforcement se fait quand un agent apprend en interagissant plusieurs fois avec son environnement, sans avoir besoin d'un superviseur. Il apprend grâce à l'expérience, en essayant différentes actions et en observant si celles-ci donnent une récompense ou une punition. Petit à petit, l'agent améliore sa façon d'agir pour obtenir le plus de récompenses possibles.

Ce type d'apprentissage permet donc à l'agent de choisir les bonnes actions pour atteindre ses objectifs, même dans des situations complexes et changeantes. On utilise cette méthode dans des domaines comme la robotique, les jeux vidéo ou encore les systèmes de recommandation[2].

2.3 Travaux connexes

Dans le cadre de cette étude, plusieurs recherches antérieures en lien avec la classification intelligente ont été examinées, mettant en œuvre diverses méthodes d'apprentissage automatique pour la gestion énergétique et la prédiction de la consommation.

Hippert et al. (2002) ont comparé les méthodes traditionnelles de prévision de charge, telles que les modèles ARIMA, à des approches de machine learning. Ils ont démontré que les techniques modernes permettent d'améliorer significativement la précision des prédictions, fournissant ainsi une base comparative pertinente pour notre projet [18].

Nguyen et al. (2010) se sont penchés sur la gestion intelligente de l'énergie dans les bâtiments résidentiels via un système de contrôle en temps réel. Leur solution, intégrant des capteurs sans fil, a permis une optimisation dynamique de la charge électrique, en cohérence avec les objectifs d'équilibrage de notre projet [28].

Amjady et al. (2011) ont développé des modèles hybrides combinant réseaux neuronaux et algorithmes génétiques pour la prévision à court terme de la demande électrique. L'intégration de données météorologiques et temporelles a permis d'augmenter la précision des prédictions, inspirant le choix d'algorithmes et l'optimisation des hyperparamètres dans notre approche [1].

Zoha et al. (2013) ont exploité les données de compteurs intelligents à l'aide de techniques telles que les SVM et Random Forests, pour modéliser les comportements de consommation énergétique. Leur approche soutient la phase d'entraînement et d'évaluation de nos modèles de classification [39].

Barbier (2017) a modélisé la consommation électrique en exploitant de grands volumes de données contextuelles (saisonnalité, météo, événements économiques). Cette étude souligne l'importance du big data dans l'élaboration de scénarios énergétiques, ce qui alimente notre réflexion sur l'intégration de variables externes [5].

Bessa et al. (2018) ont utilisé des modèles de deep learning combinant données météorologiques et socio-économiques pour prédire la demande énergétique dans les réseaux intelligents. Leur méthodologie met en lumière l'impact des facteurs contextuels complexes sur la consommation, en lien direct avec notre approche [7].

Enfin, Liu et al. (2019) ont appliqué des techniques de clustering (K-means, DBSCAN) pour segmenter les profils de consommation dans les réseaux intelligents. Cette segmentation a facilité l'équilibrage de charge et constitue une stratégie pertinente pour la classification intelligente des foyers dans notre projet [23].

2.4 Algorithmes de l'apprentissage automatique

Pour saisir la diversité des méthodes de l'apprentissage automatique, il est primordial d'étudier ses différents algorithmes, que nous résumons dans le tableau 2.1 [12] :

Type d'apprentissage	Tâches	Algorithmes
Supervisé	Classification	Régression logistique, arbres de décision, SVM, KNN, Réseaux de neurones et la classification naïve bayésienne.
	Régression	Régression linéaire, régression ridge, régression lasso, réseaux de neurones.
	Assemblage	Bagging, boosting, stacking.
Non supervisé	Clustering	K-means, DBSCAN.
	Réduction de dimension	Analyse en composantes principales (ACP).
	Détection d'anomalies	Isolation forest, One-Class SVM.
	Exploration d'associations	Apriori, FP-Growth.
Renforcement	Apprentissage par renforcement	Q-Learning.
Semi-supervisé	Classification et régression	Méthodes de propagation d'étiquettes.

TABLE 2.1 – Types d'apprentissage et leurs algorithmes associés

Dans les sections qui suivent, nous présentons et expliquons certains de ces algorithmes d'apprentissage automatique.

2.4.1 k-moyennes

La méthode des k-moyennes, également connue sous son nom anglais k-means, est une technique de classification non supervisée (clustering) visant à répartir un ensemble de données en k groupes distincts, appelés clusters. Chaque donnée est associée au groupe dont le centre, appelé centroïde, est le plus proche, généralement en se basant sur la distance euclidienne. L'objectif est de regrouper ensemble les données présentant des caractéristiques similaires afin de faire apparaître des structures homogènes au sein du jeu de données [10].

L'algorithme K-means, c'est un peu comme un organisateur de données ! Son but est de regrouper des points similaires en K groupes (clusters), sans savoir à l'avance comment les séparer. Imaginez que vous avez une collection de points dispersés sur une feuille : le K-means va les organiser en groupes compacts, en suivant ces étapes simples [10] :

Fonctionnement de l'algorithme k-moyennes

1. Initialisation des centroïdes :

- On choisit aléatoirement K points (appelés centroïdes) comme centres initiaux des clusters.

2. Affectation des points aux clusters :

- Chaque point est attribué au cluster dont le centroïde est le plus proche.

Formule :

$$\text{Cluster du point } \mathbf{x}_i = \arg \min_k \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2 \quad (2.1)$$

3. Recalcul des centroïdes :

- On met à jour la position de chaque centroïde en calculant la moyenne des points de son cluster.

Formule :

$$\boldsymbol{\mu}_k = \frac{1}{|\text{Cluster}_k|} \sum_{x_i \in \text{Cluster}_k} x_i \quad (2.2)$$

4. Répétition :

- On répète les étapes 2 et 3 jusqu'à ce que les centroïdes stabilisent (plus de changement significatif).

La figure 2.6 représente un exemple d'un modèle k-moyennes :

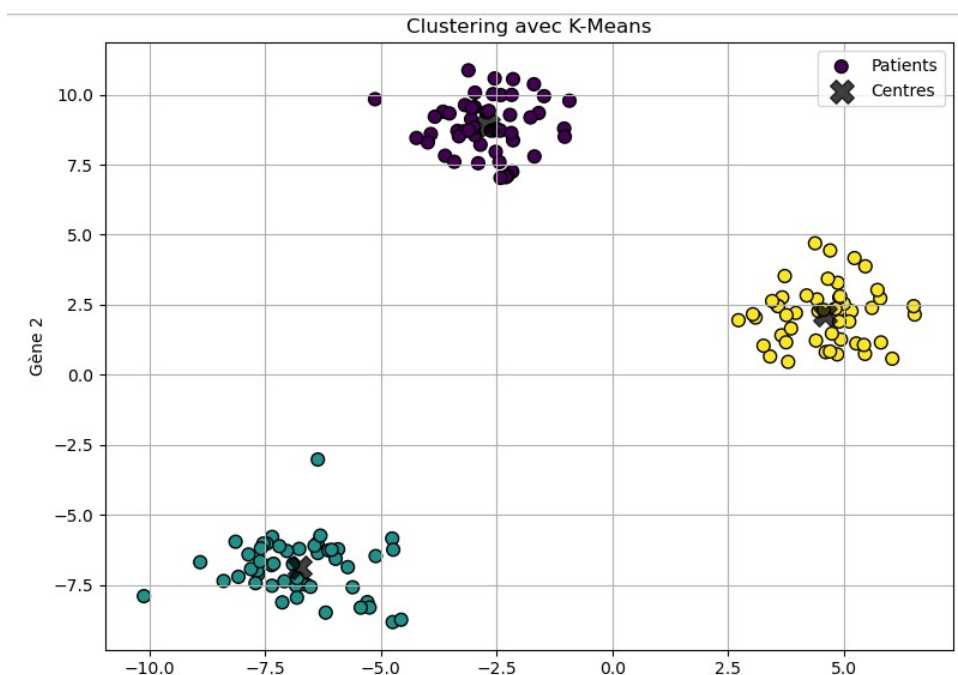


FIGURE 2.6 – Exemple d'apprentissage automatique avec k-moyennes

2.4.2 Hiérarchique Agglomératif

L'algorithme agglomératif est une méthode de clustering hiérarchique, utilisée pour regrouper des objets similaires en formant une hiérarchie d'ensembles. Contrairement à d'autres méthodes de regroupement comme K-means, cet algorithme ne nécessite pas de spécifier à l'avance le nombre de clusters. Il est très utilisé en analyse de données exploratoire, en biologie (phylogénie), en marketing, et en traitement d'image [26].

Fonctionnement de l'algorithme agglomératif

L'algorithme agglomératif fonctionne en suivant un processus itératif de fusion, où les éléments les plus similaires sont regroupés étape par étape pour construire une hiérarchie de clusters [26].

Les étapes principales sont les suivantes [26] :

1. **Initialisation** : chaque observation est considérée comme un cluster individuel.
2. **Calcul des distances** : une matrice de distances est établie entre tous les clusters existants.
3. **Fusion** : les deux clusters les plus proches (selon une métrique choisie) sont fusionnés.
4. **Mise à jour** : la matrice des distances est mise à jour pour refléter la nouvelle structure.
5. **Répétition** : le processus est répété jusqu'à ce qu'un seul cluster regroupe toutes les données ou jusqu'à atteindre un nombre de clusters souhaité.

La figure 2.7 représente Les étapes d'un algorithme de clustering hiérarchique agglomératif :

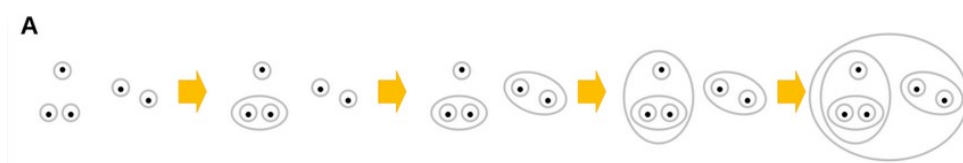


FIGURE 2.7 – Les étapes d'un algorithme de clustering hiérarchique agglomératif [8]

2.4.3 DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) est un algorithme de clustering non supervisé. Plus précisément, il regroupe les points de données denses et proches les uns des autres, tout en identifiant les points isolés comme du bruit. Contrairement à des méthodes comme k-means, il n'a pas besoin de connaître à l'avance le nombre de groupes à former et s'adapte naturellement à des formes complexes [20].

Principes clés :

Voici les concepts clés de l'algorithme DBSCAN [20] :

- **Points essentiels (core points) :** points ayant au moins un nombre minimal de voisins (MinPts) dans un rayon (epsilon).
- **Points frontières (border points) :** Points frontières (border points) : points à une distance d'un point essentiel mais qui n'ont pas eux-mêmes assez de voisins pour être essentiels.
- **Points de bruit (noise points) :** points qui ne sont ni essentiels ni frontières, considérés comme des outliers

Fonctionnement de l'algorithme DBSCAN

L'algorithme DBSCAN fonctionne selon les étapes suivantes [20] :

1. Choix des paramètres

- (epsilon) : distance maximale pour considérer deux points comme voisins.
- MinPts : nombre minimal de points pour former une région dense (cœur d'un cluster).

2. Processus étape par étape

- L'algorithme parcourt chaque point non visité du jeu de données.
- Pour chaque point, il récupère tous les voisins situés à une distance ϵ .
- Si le nombre de voisins est inférieur à $MinPts$, le point est marqué comme bruit.
- Sinon, un nouveau cluster est créé et le point devient un point essentiel.
- Les voisins de ce point sont ajoutés au cluster, et si l'un d'eux est également essentiel, ses propres voisins sont ajoutés récursivement (expansion du cluster).

Ce processus continue jusqu'à ce que tous les points soient visités.

La figure 2.8 représente un exemple d'un modèle DBSCAN

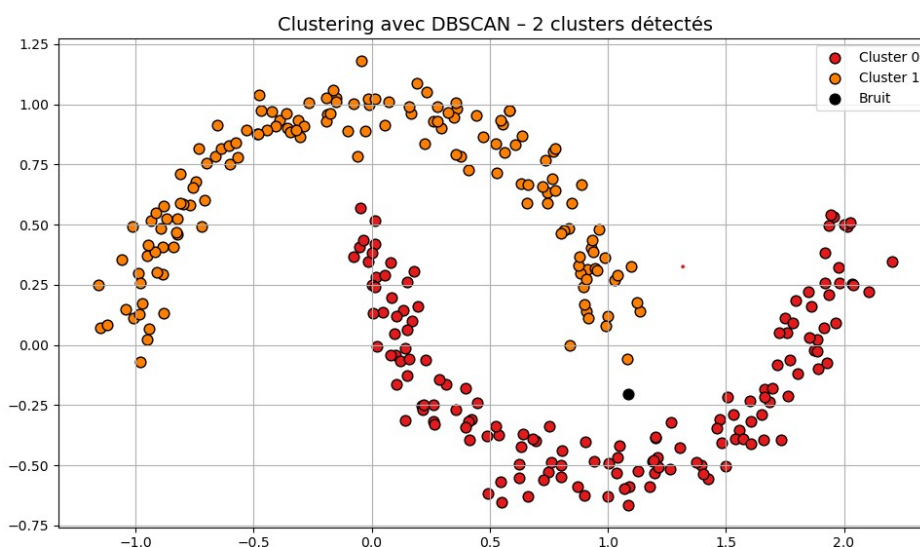


FIGURE 2.8 – Exemple d'apprentissage automatique avec DBSCAN

2.4.4 K-nearest neighbors (KNN)

L'algorithme des K plus proches voisins ou Knearest neighbors (KNN) est un modèle d'apprentissage supervisé non paramétrique et non linéaire, qui permet de classer des éléments ou d'estimer des valeurs en fonction de leur ressemblance avec des exemples connus. Son principe est simple : lorsqu'un nouvel élément doit être analysé, on examine les k éléments les plus proches (et non "proches") parmi les données déjà enregistrées [17].

Pour une classification : l'élément est attribué à la catégorie la plus fréquente parmi ses voisins.

Pour une estimation (régression) : la valeur attribuée est calculée en faisant la moyenne des valeurs de ces voisins.

Fonctionnement des KNN

L'algorithme vise à classer une nouvelle observation en fonction des étiquettes des k éléments les plus proches situés autour du point de données cible. Dans un problème de

classification, on utilise un vote majoritaire pour lui attribuer une étiquette, correspondant à la classe dominante parmi celles de ses k voisins [17].

L'algorithme KNN fonctionne selon les étapes suivantes :

- 1. Déterminer le nombre de voisins (k) :** Choisir une valeur appropriée pour k , c'est-à-dire le nombre de voisins à prendre en compte.
- 2. Choisir le nombre de voisins (k) :** Déterminer combien de voisins seront pris en compte pour faire la prédiction.
- 3. Mesurer la distance :** Sélectionner une méthode pour calculer la distance entre les points (par exemple, la distance euclidienne ou de Manhattan)

Distance euclidienne :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (2.3)$$

Distance de Manhattan :

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n| \quad (2.4)$$

Comparer la nouvelle donnée avec chaque point du jeu de données pour mesurer à quel point ils sont proches.

Formulation mathématique

Le KNN évalue la similarité entre points via une mesure de distance (euclidienne, Manhattan...). Ce choix critique dépend de la structure des données : chaque métrique capture différemment les relations spatiales [2].

Cette version garde l'essence technique tout en étant plus accessible pour un lecteur non expert.

Voici comment KNN effectue ses prédictions :

- Pour un problème de classification
- Calculer la probabilité d'appartenance de la nouvelle observation à chaque classe c , basée sur la fréquence des classes parmi les k plus proches voisins.
- Attribuer à la nouvelle observation la classe avec la probabilité d'appartenance la plus élevée. l'étiquette prédite $f(x)$ est :

$$f(x) = \arg \max_c \sum_{i \in \mathcal{N}_k(x)} \mathbf{1}(y_i = c) \quad (2.5)$$

Dans cette équation :

- $f(x)$ représente la classe prédite pour un point de donnée x .
- $\arg \max$ indique que nous cherchons la classe c qui maximise la somme.
- $\mathcal{N}_k(x)$ représente l'ensemble des k plus proches voisins de x .
- $\mathbf{1}(y_i = c)$ est la fonction indicatrice, qui vaut 1 si $y_i = c$ et 0 sinon.
- c parcourt l'ensemble des classes.

La figure 2.9 représente un exemple d'un modèle KNN :

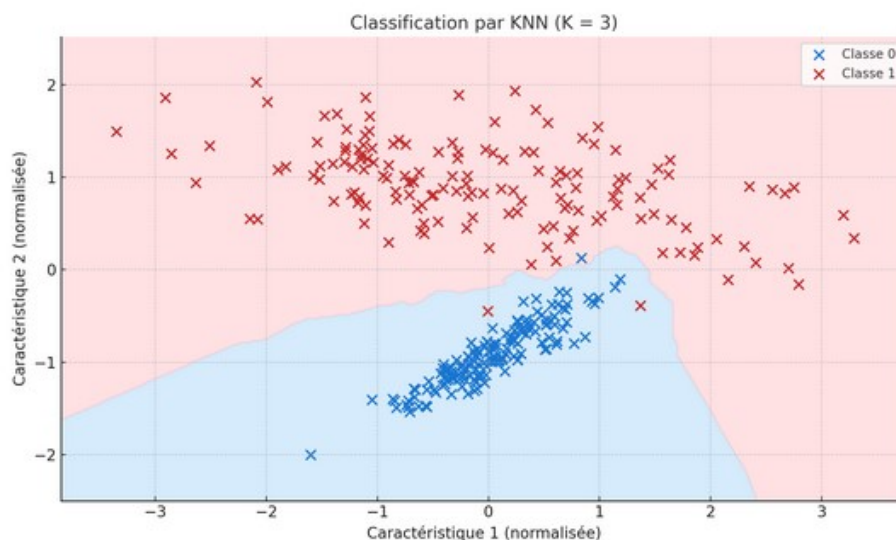


FIGURE 2.9 – Exemple d'un KNN

2.4.5 Arbre de décision

Un arbre de décision est un outil visuel qui ressemble à un schéma avec des embranchements, où chaque décision possible mène à des résultats concrets. Imaginez un guide pas-à-pas qui pose des questions ou évalue des critères pour arriver à une conclusion. C'est une méthode d'apprentissage automatique simple mais puissante, souvent utilisée pour prédire des catégories (classification) ou des valeurs numériques (régression). Son grand avantage ? Il est facile à comprendre, même pour des non-experts, car ses règles ressemblent à une logique « si... alors... » du quotidien. De plus, il s'adapte bien aux données complexes, fonctionne avec des bases de données variées, et son fonctionnement transparent en fait un choix populaire pour des décisions critiques et explicables [11].

Fonctionnement des arbres de décision

Un arbre de décision fonctionne comme un schéma en trois parties :

- **Le nœud racine :** C'est le point de départ, là où l'arbre commence à poser des questions sur les données (exemple : « Quel est l'âge du client ? »).
- **Les nœuds internes :** Ce sont des étapes intermédiaires, comme des carrefours où l'arbre prend une décision en fonction d'une caractéristique (exemple : « Si le revenu est supérieur à X, alors... »). Chaque choix génère une branche, qui guide vers la prochaine question ou conclusion.
- **Les feuilles :** Ce sont les résultats finaux, sans autre embranchement. Elles donnent la réponse prédite : une catégorie (comme « Oui/Non ») ou un nombre (comme un prix estimé).

En pratique, l'algorithme découpe les données en groupes homogènes en suivant des règles simples (« si... alors... »), ce qui le rend intuitif à interpréter, comme un guide pas-à-pas pour prendre des décisions basées sur des critères clés.

Formulation mathématique

L'indice de Gini et l'entropie croisée sont deux outils pour évaluer à quel point les boîtes sont "mal triées" (impureté). Voici leur rôle en termes simples [11] :

- **L'indice de Gini :** L'indice de Gini mesure la probabilité qu'un élément choisi au hasard soit mal classé si l'on attribue à cet élément une étiquette selon la distribution des classes dans le nœud.

La formule est :

$$\text{Gini}(t) = 1 - \sum_{i=1}^C p_i^2 \quad (2.6)$$

•**L'entropie (ou entropie croisée) :** L'entropie mesure l'incertitude associée à la distribution des classes. Elle est issue de la théorie de l'information.

La formule est :

$$\text{Entropie}(t) = - \sum_{i=1}^c p_i \log_2(p_i) \tag{2.7}$$

La figure 2.10 illustre un arbre de décision :



FIGURE 2.10 – Exemple d'un arbre de décision [15].

2.4.6 Forêts aléatoires

Random forest (forêt aléatoire) est un modèle d'apprentissage supervisé utilisé en classification et en régression. Il s'appuie sur la méthode du bagging [9].

Le bagging est une méthode d'ensemble qui vise à améliorer la précision et la stabilité des algorithmes d'apprentissage automatique. Elle consiste à générer plusieurs modèles à partir de sous-échantillons aléatoires du jeu de données d'entraînement, puis à combiner leurs prédictions (par exemple par vote majoritaire ou moyenne) pour obtenir un résultat plus fiable [11]

Fonctionnement de forêt aléatoire

la forêt aléatoire fournit une prédiction finale robuste, réduisant le risque de sur-apprentissage. Ce modèle est efficace pour gérer des données complexes et bruitées [9]

1. Construction des arbres :

- Un nombre fixe d'arbres n est défini.
- Pour chaque arbre T_i :
 - Un échantillon bootstrap D_i est tiré aléatoirement avec remplacement à partir de l'ensemble de données D .
 - Un arbre de décision T_i est construit en utilisant cet échantillon D_i .
 - À chaque nœud de l'arbre, une sous-sélection de m caractéristiques est choisie aléatoirement parmi les p caractéristiques disponibles.
 - La meilleure division est choisie parmi ces m caractéristiques.

2. Prédiction :

- Pour une nouvelle observation x :
 - Chaque arbre T_i fournit une prédiction y_i .
 - Pour la classification, la prédiction finale est obtenue par **vote majoritaire** parmi les y_i .
 - Pour la régression, la prédiction finale est la **moyenne** des y_i .

2.5 Apprentissage profond

Apprentissage profond également connue sous son nom anglais Deep Learning, et une branche de l'apprentissage automatique, un domaine de pointe en informatique, a suscité un grand intérêt dans les universités et les entreprises grâce à sa capacité à analyser d'énormes quantités de données avec des résultats très précis [37].

Cette méthode fonctionne bien dans plusieurs domaines, comme la reconnaissance d'images ou le traitement du langage. Toutefois, le deep learning nécessite souvent de grandes quantités de données et une puissance de calcul importante, ce qui peut en limiter l'utilisation dans certaines situations [27].

2.6 Métrique d'évaluation des modèles :

Parmi les métriques d'évaluation couramment utilisées en classification [4] :

2.6.1 Accuracy (Précision globale) :

L'accuracy mesure la proportion de bonnes prédictions parmi toutes les observations. C'est la métrique la plus simple [4].

$$\text{Accuracy} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{VN} + \text{FP} + \text{FN}} \quad (2.8)$$

2.6.2 Precision (Précision positive) :

La précision évalue la qualité des prédictions positives : parmi tout ce que le modèle a prédit comme appartenant à la classe positive, combien sont corrects ? [4].

$$\text{Précision} = \frac{\text{VP}}{\text{VP} + \text{FP}} \quad (2.9)$$

2.6.3 Recall (Rappel ou Sensibilité)

Le rappel mesure la capacité du modèle à détecter toutes les occurrences de la classe positive : parmi tous les exemples positifs réels, combien sont correctement prédits ? [4].

$$\text{Rappel} = \frac{\text{VP}}{\text{VP} + \text{FN}} \quad (2.10)$$

2.6.4 F1-score

Le F1-score est la moyenne harmonique entre la précision (precision) et le rappel (recall), deux métriques issues de la matrice de confusion. Il fournit un compromis utile entre les faux positifs et les faux négatifs [4].

$$\text{F1-score} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (2.11)$$

2.6.5 AUC-ROC

est une représentation graphique qui permet d'évaluer les performances d'un modèle de classification binaire en fonction de différents seuils de décision [4]

L'aire sous la courbe (AUC) résume tout en un chiffre entre 0.5 et 1 :

$$\text{AUC} = \begin{cases} 0.5 & \text{(aléatoire)} \\ 1.0 & \text{(parfait)} \end{cases} \quad (2.12)$$

2.6.6 La métrique silhouette

La métrique silhouette est une mesure utilisée pour évaluer la qualité d'un partitionnement en clustering. Elle quantifie à quel point un point de données est bien assigné à son cluster par rapport aux autres clusters [26].

Pour un point i , on définit :

- $a(i)$: la distance moyenne entre i et tous les autres points de son propre cluster (cohésion).
- $b(i)$: la distance moyenne entre i et tous les points du cluster le plus proche autre que le sien (séparation).

Le score de silhouette $s(i)$ est alors calculé par la formule :

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.13)$$

Ce score varie entre -1 et $+1$:

- $s(i) \approx +1$: le point est bien assigné à son cluster, il est plus proche des points de son cluster que des autres.
- $s(i) \approx 0$: le point est proche de la frontière entre deux clusters.
- $s(i) \approx -1$: le point est mal assigné, il serait plus proche d'un autre cluster.

La silhouette moyenne sur tous les points donne une indication globale de la qualité du clustering.

2.7 Optimisation et validation des modèles :

Cette section se concentre sur les techniques d'optimisation et de validation des modèles d'apprentissage automatique. Nous examinerons les méthodes utilisées pour améliorer les performances des modèles, ainsi que pour garantir leur robustesse et leur généralisation à des données inconnues.[19]

2.7.1 Validation croisée

Pour valider nos modèles et sélectionner les paramètres optimaux, nous avons eu recours à la **validation croisée**. Cette méthode d'évaluation permet de mesurer la capacité d'un modèle à généraliser ses prédictions sur de nouvelles données. Son principe consiste à [19] :

- Diviser le jeu de données en k sous-ensembles (*folds*)
- Entraîner le modèle successivement sur $k - 1$ sous-ensembles
- Tester ses performances sur le sous-ensemble restant

Dans le cas spécifique des **séries temporelles**, la validation croisée classique présente un écueil majeur : elle ignore la chronologie des données, risquant ainsi de fausser les résultats. Nous avons donc utilisé *TimeSeriesSplit* [19], une variante qui préserve l'ordre temporel en créant des plages de validation toujours postérieures aux données d'entraînement.

2.7.2 Optimisation avec GridSearchCV

La méthode *GridSearchCV* automatise la recherche des meilleures combinaisons d’hyperparamètres grâce à [19] :

- Une exploration systématique de toutes les possibilités d’une grille prédéfinie
- L’évaluation objective des performances via la métrique *neg_mean_squared_error* (erreur quadratique moyenne négative)
- L’intégration native de notre stratégie *TimeSeriesSplit* pour le traitement des séries chronologiques

TABLE 2.2 – Configuration clé de GridSearchCV [19]

Paramètre	Rôle
<code>estimator</code>	Algorithme ML à optimiser (ex : régression linéaire)
<code>param_grid</code>	Espace de recherche des hyperparamètres (valeurs discrètes)
<code>cv</code>	Stratégie de validation croisée (<i>TimeSeriesSplit</i> dans notre cas)
<code>scoring</code>	Métrique d’évaluation des performances
<code>n_jobs</code>	Accélération des calculs via parallélisation (utilisation de plusieurs cœurs CPU)

Cette approche rigoureuse permet d’éviter le surapprentissage tout en identifiant les paramètres qui maximisent la pertinence des prédictions sur des données futures.

Conclusion

En conclusion, ce chapitre nous a permis d’explorer en profondeur les fondements de l’apprentissage automatique, en présentant plusieurs types de modèles ainsi que leur fonctionnement dans différents contextes. Chaque approche possède ses avantages et ses limites, et le choix du modèle dépend fortement de la nature du problème à résoudre et des données disponibles.

Dans le chapitre suivant, nous allons nous concentrer sur méthodologie de l’étude et résultats, en abordant les différentes étapes pour analyser les données et appliquer les techniques les plus appropriées pour résoudre les défis identifiés.

3

Étude de Cas et Résultats de Classification Énergétique

Dans la continuité des fondements théoriques abordés précédemment, ce chapitre présente la démarche méthodologique adoptée pour la mise en œuvre de notre étude sur la classification énergétique des foyers. L'objectif est de structurer le processus permettant de passer de la donnée brute à une analyse intelligente et exploitable, en s'appuyant sur des outils et techniques issus de l'apprentissage automatique.

Nous exposerons ainsi la logique générale suivie, depuis la préparation des données jusqu'à l'application des modèles de classification. Sans entrer dans les détails techniques, ce chapitre vise à montrer comment ces méthodes permettent de mieux comprendre les profils de consommation et d'orienter les décisions en matière de gestion énergétique.

Il sert également de base pour interpréter les résultats qui seront présentés par la suite, et pour évaluer la pertinence des choix effectués au regard des objectifs fixés.

Sommaire

Introduction	34
3.1 Environnement de développement	35
3.2 Approche 1 : classification des foyers et des appareils pour l'optimisation de la consommation électrique résidentielle	37
3.3 Modèles d'apprentissage automatique	39
3.4 Présentation des résultats	40
3.5 Evaluation des modèles	44
3.6 Approche 2 : Anticipation des surcharges électriques par classification supervisée et ajustement des usages flexibles	45
3.7 Prétraitement des données	45
3.8 Clustering horaire saisonnier	47
3.9 Segmentation des appareils électroménagers	48

3.10 Classification de la surcharge des foyers	50
3.11 Recommandation de décalage d'activation d'appareils	52
3.12 Discussion	54
Conclusion	54

3.1 Environnement de développement

Cette section décrit l'environnement de développement utilisé pour la mise en œuvre des modèles de prédiction de la demande d'électricité et d'équilibrage de charge. Nous aborderons les outils, les bibliothèques et les technologies employés pour développer, tester et valider les modèles de classification intelligente.

3.1.1 Langage de programmation

Pour mettre en œuvre notre projet, nous avons choisi d'utiliser Python. Ce langage de programmation open source, interprété et de haut niveau, créé par Guido van Rossum et lancé en 1991, se distingue par sa lisibilité et sa grande polyvalence.

Il permet d'exprimer des concepts complexes avec moins de lignes de code qu'en C++ ou Java, tout en supportant divers paradigmes de programmation, tels que l'orienté objet, l'impératif, le fonctionnel et le procédural.

En science des données et en apprentissage automatique, Python est particulièrement prisé pour sa syntaxe claire et sa communauté dynamique, qui enrichit constamment une vaste gamme de bibliothèques et de frameworks adaptés à la manipulation de données, à la modélisation statistique, à l'apprentissage automatique et à l'apprentissage profond[25].

La figure 3.1 représente le logo Python



FIGURE 3.1 – Logo de Python

3.1.2 Bibliothèques et frameworks pour la data science avec Python

Les outils suivants constituent les piliers fondamentaux de la science des données avec Python, fournissant des compétences clés pour manipuler efficacement les données, créer des visualisations pertinentes et concevoir des modèles d'apprentissage automatique et profond[25].

Les piliers fondamentaux

- **NumPy** : C'est une bibliothèque qui offre des tableaux multidimensionnels puissants et des fonctions mathématiques pour la manipulation efficace de données numériques. Il est indispensable pour le traitement de grands ensembles de données.
- **Pandas** : Cette bibliothèque offre des structures de données optimisées pour la science des données, telles que les DataFrames, ainsi que des outils performants pour l'analyse et la manipulation des données. Elle simplifie les tâches de nettoyage, de tri, d'agrégation et de visualisation des données.



Visualisation informative de données

- **Matplotlib** : Bibliothèque de base pour la création de graphiques et de visualisations en Python, proposant une vaste sélection de types de graphiques et de possibilités de personnalisation.
- **Seaborn** : C'est une bibliothèque, extension de Matplotlib, extension de Matplotlib qui construit sur ses fonctionnalités pour créer des visualisations statistiques attrayantes et informatives. Il est particulièrement utile pour la visualisation de données complexes.
- **Plotly** : Bibliothèque de visualisation interactive permettant de créer des graphiques riches, personnalisables et dynamiques. Elle propose une large variété de types de graphiques, incluant des visualisations en 3D et des graphiques interactifs, avec de nombreuses options de personnalisation.



Apprentissage automatique et apprentissage profond

- **Scikit-learn** : Bibliothèque complète d'algorithmes d'apprentissage automatique couvrant des tâches telles que la classification, la régression et le clustering. Elle offre une interface simple et intuitive pour la mise en œuvre de modèles de machine learning.
- **TensorFlow/PyTorch** : Frameworks de calcul numérique avancés conçus pour la création et le déploiement de modèles d'apprentissage profond. Ils fournissent des fonctionnalités robustes pour le développement de réseaux neuronaux et les applications de deep learning.



Bibliothèques statistiques

- **SciPy** : Il s'agit d'une bibliothèque de visualisation offrant des fonctions statistiques avancées, notamment pour l'analyse de distributions de probabilité, les tests d'hypothèse et l'analyse statistique.



3.1.3 Environnement de développement intégré IDE

JupyterLab est un environnement de développement interactif qui permet de créer, modifier et exécuter des documents Jupyter, comme les notebooks ou les fichiers texte.

Il est particulièrement apprécié des data scientists, chercheurs et développeurs, grâce à ses nombreuses extensions personnalisables et son intégration avec la ligne de commande.

JupyterLab propose des fonctionnalités avancées telles que la visualisation de données, la coloration syntaxique et la complétion automatique pour plusieurs langages, notamment Python, R et Julia [25].

La figure 3.2 représente le logo de Jupyterlab.



FIGURE 3.2 – Logo de Jupyterlab

3.2 Approche 1 : classification des foyers et des appareils pour l'optimisation de la consommation électrique résidentielle

Cette section est consacrée à l'étude de cas portant sur la prévision de la demande en électricité et l'optimisation de la répartition de charge à l'aide de méthodes de classification intelligente.

Nous décrivons d'abord les étapes de collecte et de préparation des données, en précisant la manière dont les informations ont été obtenues, nettoyées et structurées pour l'analyse.

Nous présentons ensuite les différentes stratégies de modélisation adoptées pour anticiper la consommation énergétique et équilibrer la charge, en exposant les algorithmes d'apprentissage automatique retenus ainsi que les justifications de leur choix.

Dans le cadre de cette étude, la résolution du problème a reposé sur une démarche structurée fondée sur la classification et l'analyse intelligente des données :

- **Classification des consommateurs :** Les utilisateurs ont été segmentés selon leur profil de consommation énergétique. Cette classification permet de regrouper les comportements similaires afin d'adapter les stratégies d'optimisation.

- **Classification des plages horaires** : La journée a été divisée en trois catégories temporelles heures creuses, heures pleines et heures de pointe afin d'identifier les périodes critiques et les opportunités de déplacement de charge.
- **Typologie des machines** : Les équipements ont été classifiés en deux groupes :
 - *Machines flexibles* : leur fonctionnement peut être déplacé sans impact majeur ;
 - *Machines non flexibles* : leur fonctionnement est contraint dans le temps.
- **Stratégie d'optimisation** : Une réallocation des machines flexibles vers les heures creuses a été effectuée afin de lisser la courbe de demande et de réduire les pics de consommation, contribuant ainsi à un meilleur équilibrage de la charge sur le réseau.

3.2.1 Collecte et chargement des données

Les données de consommation électrique ont été extraites à partir du fichier Excel fourni par la plateforme Zenodo, dans le cadre d'une étude portant sur une communauté énergétique résidentielle. Ce jeu de données couvre une période d'un an et contient des informations détaillées sur la consommation horaire de 51 bâtiments, répartis entre 50 foyers et une bibliothèque municipale. Les données sont désagrégées par type d'appareil électroménager et incluent également la production photovoltaïque individuelle de chaque bâtiment.

Ces informations sont organisées selon différents profils d'utilisation, permettant une analyse fine des habitudes de consommation énergétique. Chaque bâtiment est identifié, et les données sont structurées pour permettre une exploitation directe en apprentissage automatique, notamment pour des tâches de classification ou de prédiction.

Nous avons utilisé la fonction `pd.read_excel()` de la bibliothèque `pandas` pour charger les données à partir du fichier Excel correspondant.

La figure 3.3 illustre l'étape de chargement des données.

```
Colonnes du dataset : ['Consumer1', 'Consumer2', 'Consumer3', 'Consumer4', 'Consumer5', 'Consumer6', 'Consumer7', 'Consumer8', 'Consumer9', 'Consumer10', 'Consumer11', 'Consumer12', 'Consumer13', 'Consumer14', 'Consumer15', 'Consumer16', 'Consumer17', 'Consumer18', 'Consumer19', 'Consumer20', 'Consumer21', 'Consumer22', 'Consumer23', 'Consumer24', 'Consumer25', 'Consumer26', 'Consumer27', 'Consumer28', 'Consumer29', 'Consumer30', 'Consumer31', 'Consumer32', 'Consumer33', 'Consumer34', 'Consumer35', 'Consumer36', 'Consumer37', 'Consumer38', 'Consumer39', 'Consumer40', 'Consumer41', 'Consumer42', 'Consumer43', 'Consumer44', 'Consumer45', 'Consumer46', 'Consumer47', 'Consumer48', 'Consumer49', 'Consumer50']
Nombre de lignes : 96
```

FIGURE 3.3 – Chargement des données

Pour chaque consommateur, nous disposons des mesures de consommation associées à différents appareils électroménagers, tels que les climatiseurs (AC1 à AC4), le chauffe-eau, la télévision, le micro-ondes, la bouilloire, l'éclairage et le réfrigérateur. Ces données ont été extraites et exploitées à l'aide de la bibliothèque `pandas`.

La figure 3.4 illustre le Profil de consommation des appareils par foyer.

Unnamed: 0	Periods	Total Consumption	AC1	AC2	AC3	AC4	Water heater	\
0	0	1	2.964	0.0	0.0	0.0	0	0.0
1	1	2	2.584	0.0	0.0	0.0	0	0.0
2	2	3	3.071	0.0	0.0	0.0	0	0.0
3	3	4	2.694	0.0	0.0	0.0	0	0.0
4	4	5	2.569	0.0	0.0	0.0	0	0.0

	TV	Microwave	Kettle	Lighting	Refrigerator
0	0.0	0.0	0.0	0.0	0.00
1	0.0	0.0	0.0	0.0	0.00
2	0.0	0.0	0.0	0.0	0.00
3	0.0	0.0	0.0	0.0	0.14
4	0.0	0.0	0.0	0.0	0.00

FIGURE 3.4 – Illustration des profils de consommation des appareils par foyer.

3.3 Modèles d'apprentissage automatique

Après avoir bien analysé notre dataset et comparé entre divers algorithmes de classification, nous avons désigné les plus adaptés à notre problématique. On cite :

- kmeans clustering
- Forêt aléatoire
- agglomerative clustering

3.3.1 Classification des consommateurs selon leur consommation

Le modèle K-means a été retenu car il s'agit d'un algorithme de clustering à la fois simple, rapide et efficace pour regrouper mes 50 consommateurs selon leur profil de consommation énergétique. Sa simplicité d'application permet de segmenter les données en groupes distincts par exemple : Faible, Moyen et Élevé tout en minimisant la variance intra-cluster. Cette approche convient parfaitement à mes données de consommation, préalablement normalisées, et le choix de 3 clusters a été fixé pour refléter ces niveaux de consommation.

Création et entraînement du modèle

Le modèle *K-means* a été initialisé via `KMeans()`, puis entraîné sur les données grâce à la fonction `fit()` de l'objet `kmeans_model`.

3.3.2 Équilibrage de la charge

Dans le cadre de l'équilibrage de la charge énergétique, une double classification a été réalisée : les machines ont été distinguées selon leur flexibilité d'usage, et les plages horaires selon la demande (creuses, pleines, pointe).

Cette approche a permis de déplacer l'utilisation des équipements flexibles vers les heures creuses, réduisant ainsi les pics de consommation et améliorant la répartition de la charge tout

en favorisant des économies d'énergie.

- **Classification des plages horaires**

Le modèle Random Forest a été retenu pour classer les plages horaires (creuses, pleines, pointe) en raison de sa robustesse, de sa capacité à modéliser des relations non linéaires et de sa bonne performance sur des jeux de données de taille moyenne. En combinant plusieurs arbres de décision, il offre une prédiction précise et limite le surapprentissage. Il exploite des variables telles que l'heure, le jour ou la consommation moyenne pour détecter des schémas complexes et classer efficacement les périodes selon leur niveau de demande énergétique.

Création et entraînement du modèle : Le modèle *Forêt Aléatoire* pour la classification a été initialisé à l'aide de `RandomForestRegressor()`, puis entraîné sur les données d'apprentissage à l'aide de la méthode `fit()` de l'objet modèle `_rf_model`.

- **Classification des machines**

L'algorithme de clustering agglomératif a été choisi pour sa capacité à détecter des structures complexes sans fixer le nombre de groupes à l'avance. Il permet de visualiser les similarités entre machines via un dendrogramme, facilitant ainsi la distinction entre équipements flexibles et non flexibles sur la base de leurs profils de consommation. Cette classification a été affinée par une analyse comparative avec des seuils spécifiques (écart-type $\leq 0,2$ kWh et corrélation avec l'indicateur `IsPeak` $\geq 0,5$), permettant d'identifier des machines telles que AC1, AC2, AC3, Water Heater et Microwave comme flexibles, et TV, Lighting, Refrigerator comme non flexibles.

Le modèle a ensuite été appliqué aux données normalisées des machines (`X_scaled_appliance`) à l'aide de la méthode `fit_predict()`, permettant d'attribuer automatiquement chaque machine à un cluster en fonction de son profil de consommation énergétique.

3.3.3 Décalage des machines flexibles vers les heures creuses

Une stratégie de décalage a été mise en œuvre pour réduire les pics de consommation en déplaçant la charge des machines flexibles vers les heures creuses. La journée a été divisée en trois plages (creuse, pleine, pointe), et seules les machines flexibles ont été ciblées.

La moitié de leur consommation en heures de pointe est transférée vers les heures creuses du matin, l'autre moitié vers celles du soir.

La répartition est faite de façon équilibrée pour chaque consommateur, sans modifier la consommation totale, ce qui est vérifié par une comparaison entre la charge initiale et ajustée.

3.4 Présentation des résultats

Cette section présente les résultats des modèles de prévision de la demande en électricité :

3.4.1 Résultats de la Classification des consommateurs selon leur consommation

La classification des consommateurs d'électricité consiste à regrouper les utilisateurs en fonction de leur niveau de consommation énergétique. Cette démarche est essentielle pour analyser les comportements de consommation, adapter les politiques énergétiques et optimiser la gestion du réseau électrique.

En pratique, cette classification permet d'identifier trois grandes catégories :

- **Faible consommation** : utilisateurs économes ou peu actifs, souvent dotés d'équipements limités ou d'une faible présence à domicile.
- **Consommation moyenne** : utilisateurs au profil standard, représentant la majorité des foyers avec un usage régulier.
- **Forte consommation** : foyers à forte demande énergétique, possiblement liée à des équipements énergivores, des surfaces plus grandes ou une occupation continue.

Classification des 50 Consommateurs par Consommation Totale". Il montre trois groupes : Faible (vert), Moyen (bleu) et Élevé (orange).

la figure 3.5 démontre la classification des consommateurs

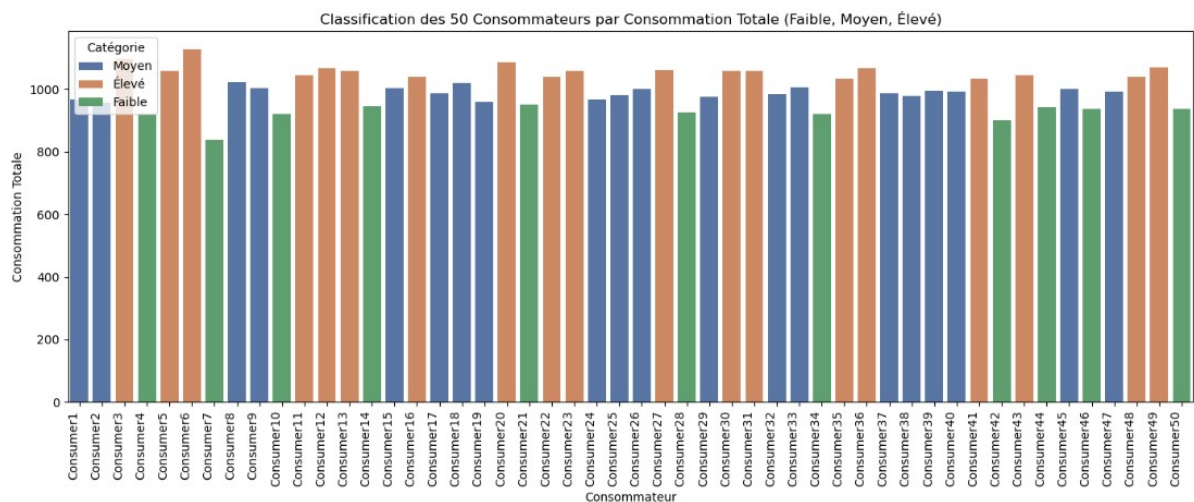


FIGURE 3.5 – Classification des consommateurs selon leur consommation avec k-means

3.4.2 Résultats de l'équilibrage de la charge

• résultats Classification des plages horaires

Le graphique présenté met en évidence la classification des **96 périodes journalières** (de 15 minutes chacune) en **trois grandes catégories horaires**, selon leur niveau moyen de consommation électrique :

- **Heures creuses (vert)** : Ce sont les périodes où la demande en électricité est la plus faible, principalement localisées entre **minuit et 6h du matin**, puis entre **22h et minuit**. Ces plages horaires sont idéales pour le *déplacement de la charge*, notamment pour l'utilisation des *machines flexibles*, afin d'alléger la pression sur le réseau.
- **Heures pleines (jaune)** : Elles correspondent à une consommation **modérée**, observée généralement en **début de matinée (6h–8h)**, en **milieu de journée (12h–17h)** et en **début de soirée (21h–22h)**. Elles représentent des périodes de transition énergétique entre les creux et les pics de consommation.
- **Heures de pointe (rouge)** : Ces périodes sont marquées par une **consommation maximale**, typiquement entre **8h et 12h** puis entre **17h et 21h**. Ce sont les moments les plus critiques pour le réseau électrique, en raison d'une forte sollicitation des équipements par les foyers.

la figure 3.6 démontre les résultats de la classification des plages horaires.

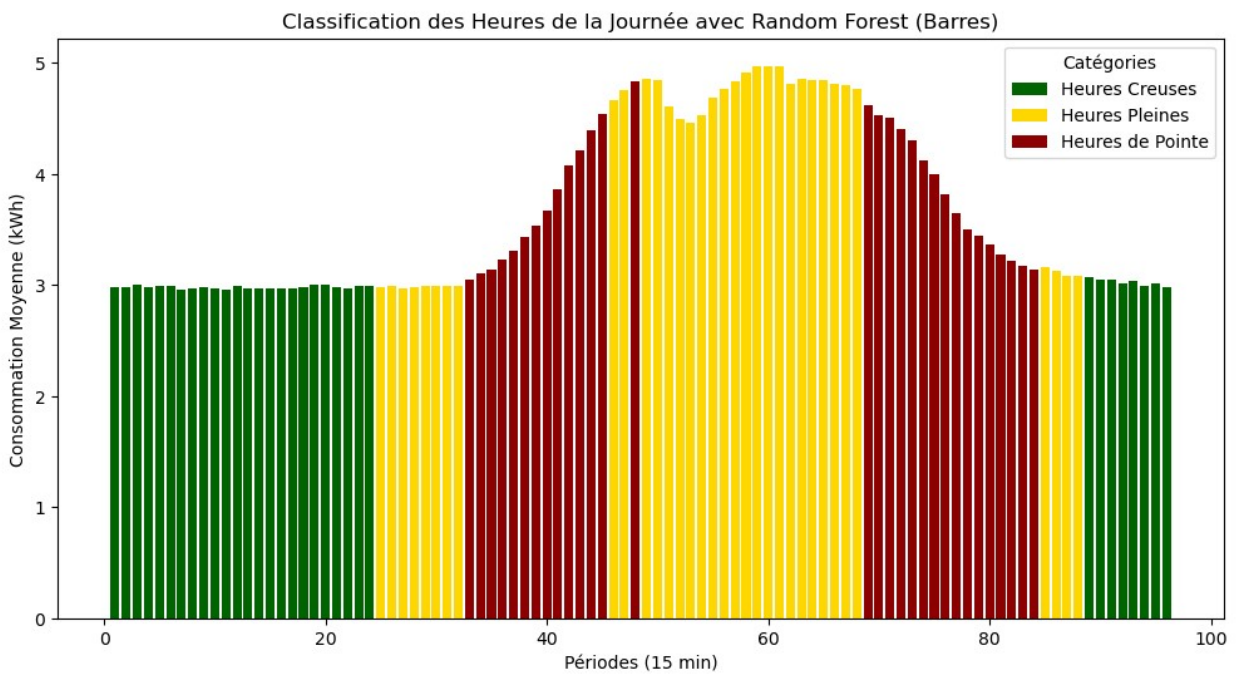


FIGURE 3.6 – Résultats de la classification des plages horaires

- Résultats de la classification des machines Les équipements ont été classés en deux catégories :

Les équipements électriques ont été classés en deux catégories principales selon leur capacité à être déplacés dans le temps :

- **Machines flexibles** : Ce sont des appareils pouvant être déplacés dans le temps sans impact majeur sur leur fonctionnement. Par exemple, les lave-linges ou les chauffe-eaux

programmables.

- **Machines non flexibles** : Ce sont des équipements dont le fonctionnement est fixe et difficilement déplaçable, comme les réfrigérateurs ou les systèmes de chauffage central.

La **classification hiérarchique ascendante** (*Agglomerative Clustering*) a été utilisée pour structurer cette séparation, en se basant sur les caractéristiques d'utilisation de chaque machine. la figure 3.7 démontre les résultats de la classification des machines.

```

Classification basée sur les Clusters :
AC1: Flexible
AC2: Flexible
AC3: Flexible
Water heater: Flexible
TV: Non Flexible
Microwave: Flexible
Kettle: Flexible
Lighting: Non Flexible
Refrigerator: Non Flexible

```

FIGURE 3.7 – Résultats de la classification des machines

• Résultats finale de la Stratégie d'équilibrage

La stratégie d'équilibrage appliquée, en décalant les machines flexibles vers les heures creuses, a permis de réduire la charge pendant les périodes de pointe et d'améliorer la répartition de la consommation sur l'ensemble de la journée.

la figure 3.8 démontre les résultats de la Stratégie d'équilibrage

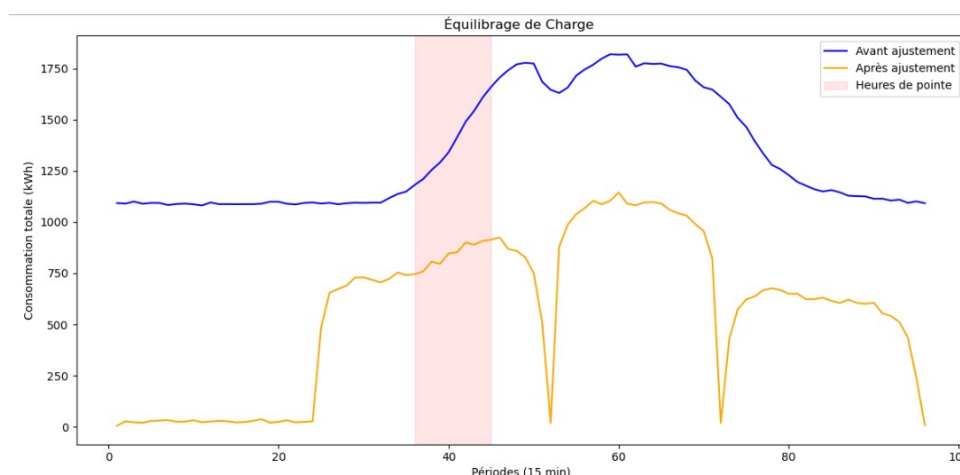


FIGURE 3.8 – Résultats finale de la Stratégie d'équilibrage

L'évaluation révèle une charge totale initiale de 128302,86 kWh réduite à 54577,16 kWh après ajustement, avec une différence de 73725,70 kWh suggérant une possible erreur ou une intention de réduire la consommation, malgré une charge déplacée limitée à 49,30 kWh. La charge moyenne de pointe passe de 1399,81 kWh à 841,80 kWh, soit une réduction de 558,01 kWh, tandis que la variance diminue de 7936,89 kWh² à 1473,25 kWh² (réduction de 67938,50 kWh²), indiquant un équilibrage efficace des heures de pointe et une distribution plus homogène.

la figure 3.9 démontre les résultats de la Stratégie d'équilibrage

```

Évaluation de l'Équilibrage de Charge :
Charge totale initiale : 128302.86 kWh
Charge totale après ajustement : 54577.16 kWh
Différence de charge totale : 73725.70 kWh (devrait être proche de 0)
Charge totale déplacée : 49.30 kWh
Charge moyenne de pointe initiale (périodes 36-45) : 1399.81 kWh
Charge moyenne de pointe ajustée (périodes 36-45) : 841.80 kWh
Réduction de charge de pointe : 558.01 kWh
Variance initiale de la consommation par période : 79386.89 kWh2
Variance ajustée de la consommation par période : 147325.39 kWh2
Réduction de la variance : -67938.50 kWh2 (plus élevée, meilleur est l'équilibrage)

```

FIGURE 3.9 – Résultats d'Ajustement Énergétique

3.5 Evaluation des modèles

Après l'entraînement, nous avons évalué les modèles pour résoudre des problèmes spécifiques à l'aide de plusieurs métriques classiques, telles que l'Accuracy et le F1-Score pour la classification des plages horaires, ainsi que le Score de Silhouette pour la classification des consommateurs et des machines.

Le tableau suivant présente les résultats obtenus :

TABLE 3.1 – Évaluation comparative des modèles selon la résolution de problème

Modèle	Résolution de problème	Métrique	Valeur
K-Means	Classification des consommateurs	Score de Silhouette	0.8649
Agglomératif	Classification des machines	Score de Silhouette	0.8300
Random Forest	Classification des plages horaires	Accuracy	0.9000
		F1-Score	0.8980

3.6 Approche 2 : Anticipation des surcharges électriques par classification supervisée et ajustement des usages flexibles

Dans le cadre de ce projet, nous avons développé une approche alternative intelligente de prédiction de la demande d'électricité et d'équilibrage de charge à l'échelle résidentielle. L'objectif principal est d'anticiper les situations de surcharge au niveau des foyers et de proposer des stratégies de décalage ciblé de la consommation électrique, notamment à travers les appareils électroménagers flexibles.

L'approche adoptée repose sur une succession d'étapes structurées, allant du prétraitement des données brutes jusqu'à la recommandation finale d'actions de décalage. Chaque étape a été pensée pour exploiter au mieux les caractéristiques temporelles, saisonnières et contextuelles des données issues de 500 foyers, tout en assurant une prise de décision individualisée.

L'originalité de cette alternative réside dans l'association de plusieurs techniques complémentaires :

- Le **clustering horaire saisonnier** permet d'identifier les périodes de forte et de faible demande à partir des profils moyens de consommation.
- Une **segmentation des appareils électroménagers** selon leur comportement permet de distinguer les équipements les plus propices au décalage.
- Une **classification supervisée locale** permet de prédire, pour chaque foyer, les situations de surcharge électrique en fonction du contexte horaire et de la consommation observée.
- Enfin, une **règle heuristique simple mais efficace** permet de recommander le décalage des appareils flexibles vers des périodes moins critiques.

L'ensemble de ces étapes constitue une chaîne de traitement intelligente visant à simuler un système de gestion active de la demande au sein de foyers résidentiels, sans recours à des approches de régression classique, souvent sujettes à sur-apprentissage dans ce type de données complexes.

3.7 Prétraitement des données

Cette étape vise à transformer les données brutes collectées en un format exploitable pour les analyses ultérieures, tout en assurant leur qualité et leur cohérence.

3.7.1 Description des données initiales

Le jeu de données comprend 100 000 enregistrements correspondant aux activations d'appareils électroménagers dans différents foyers. Chaque ligne représente une consommation énergétique ponctuelle d'un appareil à un instant donné, avec les informations suivantes : un identifiant unique du foyer (*Home ID*), le type d'appareil électroménager activé (*Appliance Type*), la consommation instantanée en kWh (*Energy Consumption*), les informations temporelles (*Time* et *Date*), la température extérieure en degrés Celsius (*Outdoor Temperature*), la saison (*Season*) ainsi que le nombre de personnes dans le foyer (*Household Size*).

Le jeu de données ne comporte aucune valeur manquante, ce qui a simplifié les étapes de prétraitement. De plus, les types de données sont cohérents avec leur signification : les variables numériques concernent principalement la consommation et la température, tandis que les variables catégorielles décrivent les types d'appareils, les saisons et les identifiants de foyers.

3.7.2 Nettoyage et transformation

Les étapes suivantes ont été appliquées :

- **Conversion du temps** : fusion des colonnes Date et Time pour obtenir un timestamp unique en format datetime.
- **Tri temporel** : les données ont été triées par timestamp afin d'assurer une cohérence temporelle des analyses.
- **Renommage des colonnes** : uniformisation des noms de colonnes pour éviter les caractères spéciaux (par exemple, `energy_consumption_kwh`).
- **Ajout de colonnes temporelles dérivées** : heure, timestamp, etc., ont été extraits pour enrichir les analyses contextuelles.
- **Filtrage éventuel des valeurs aberrantes** : les valeurs extrêmes de consommation et de température ont été analysées, bien qu'aucune valeur aberrante significative n'ait été détectée à ce stade.

3.7.3 Statistiques descriptives

Les statistiques globales montrent une moyenne de consommation énergétique par activation de **1,50 kWh**, avec une valeur maximale de **5 kWh**. La température extérieure varie entre **-10°C** et **40°C**, reflétant une couverture annuelle complète. La taille moyenne des foyers est d'environ **3 personnes**, ce qui suggère une population représentative de foyers standards.

3.7.4 Structuration finale du dataset

À l'issue du prétraitement, un jeu de données nettoyé et enrichi a été constitué. Il servira de base aux étapes suivantes de clustering, de classification et de recommandation. Les colonnes principales de ce DataFrame final incluent : `home_id`, `appliance_type`, `energy_consumption_kwh`, `timestamp`, `season`, `household_size`, `outdoor_temperature`, ainsi que les variables temporelles dérivées.

3.8 Clustering horaire saisonnier

Afin d'identifier les plages horaires typiques de consommation (heures pleines vs heures creuses), nous avons appliqué une méthode de clustering non supervisé basée sur les profils horaires moyens.

3.8.1 Construction des profils horaires

Pour chaque saison (*Winter, Spring, Summer, Fall*), la consommation d'énergie a été agrégée par heure sur l'ensemble des foyers. Cela a permis de construire, pour chaque saison, un vecteur de 24 moyennes correspondant à la consommation moyenne à chaque heure de la journée.

Ces profils horaires saisonniers permettent de mettre en évidence les variations de charge quotidiennes typiques selon les saisons.

3.8.2 Méthode de clustering

Le clustering a été réalisé à l'aide de l'algorithme **K-Means**, appliqué séparément pour chaque saison. L'objectif est de segmenter les 24 heures en classes homogènes de consommation (par exemple, heures creuses et heures pleines).

Pour déterminer le nombre optimal de clusters, nous avons utilisé la **méthode du coude** (*Elbow Method*), basée sur l'évolution de l'inertie intra-cluster. Cette méthode a montré de manière cohérente que **2 clusters** étaient suffisants pour chaque saison, ce qui correspond intuitivement à deux niveaux de consommation distincts : *heures creuses* et *heures pleines*.

3.8.3 Résultats du clustering

Chaque heure de la journée a ainsi été étiquetée, pour chaque saison, comme appartenant soit à une période de **forte consommation (peak)**, soit à une période de **faible consommation (offpeak)**. Ce découpage horaire typologique a été intégré dans le dataset principal sous la forme d'une nouvelle variable catégorielle : `period_type`.

3.8.4 Visualisation et interprétation

Des courbes horaires moyennes par saison ont été tracées avec les clusters superposés, ce qui a permis de visualiser les découpages temporels obtenus. Ces clusters reflètent des comportements typiques attendus, comme des pointes de consommation en soirée en hiver, ou des creux en fin de nuit.

Ces informations joueront un rôle essentiel dans les étapes ultérieures, notamment pour la prédiction des surcharges et la recommandation de décalage vers les heures les moins sollicitées.

Les figures 3.10 suivantes illustrent la sélection du nombre optimal de clusters et la segmentation horaire résultante :

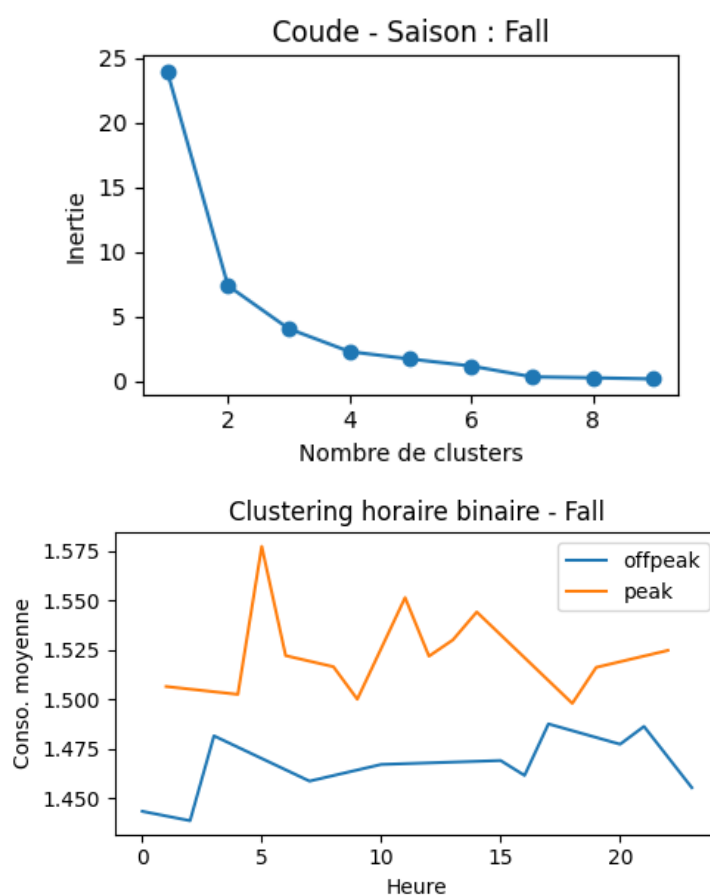


FIGURE 3.10 – En haut : méthode du coude montrant que $k = 2$ est optimal pour le clustering horaire. En bas : visualisation des clusters horaires pour la saison d’automne distinguant les heures creuses et pleines.

3.9 Segmentation des appareils électroménagers

Dans le but d’identifier les équipements électroménagers susceptibles d’être décalés dans le temps sans altérer le confort des utilisateurs, une segmentation non supervisée des appareils a été réalisée.

3.9.1 Construction des descripteurs

Chaque type d’appareil a été représenté par un ensemble de descripteurs statistiques construits à partir des historiques d’activations :

- `mean_energy` : consommation moyenne d’énergie (kWh),
- `std_energy` : écart-type de la consommation,
- `usage_count` : fréquence d’utilisation totale,
- `avg_hour` : heure moyenne d’utilisation,
- `std_hour` : dispersion horaire des activations.

Ces variables ont été extraites pour chaque type d'appareil, puis normalisées afin de garantir l'équité entre les différentes échelles de mesure.

3.9.2 Méthode de clustering

Un clustering a été réalisé à l'aide de l'algorithme **K-Means**. Pour évaluer le nombre optimal de clusters, nous avons utilisé le **silhouette score**, une métrique qui mesure la qualité de séparation entre les clusters.

La figure 3.11 suivante illustre l'évolution du silhouette score pour différentes valeurs de k :

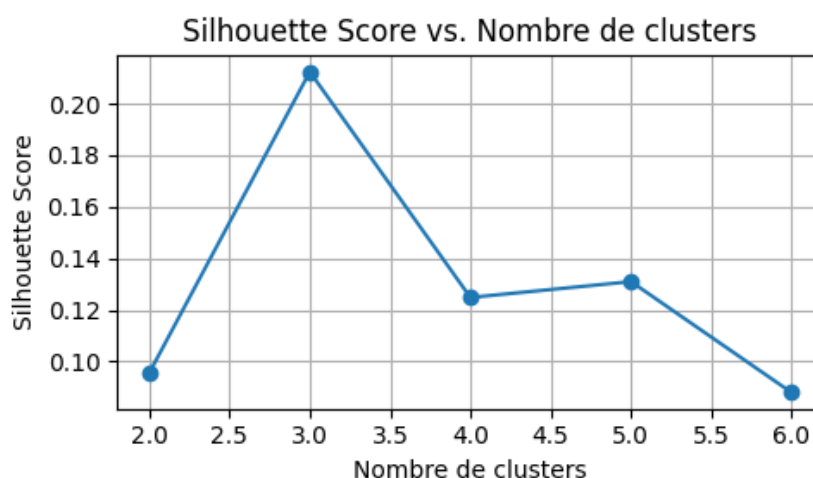


FIGURE 3.11 – Évolution du silhouette score en fonction du nombre de clusters.

La valeur optimale a été atteinte pour $k = 3$, ce qui a guidé le choix du paramètre dans l'algorithme K-Means.

3.9.3 Résultats du clustering

Les appareils ont ainsi été répartis en trois groupes :

- **Cluster 0 (flexibles)** : appareils à faible consommation, grande variabilité horaire, et utilisations fréquentes (ex. : micro-ondes, machine à laver),
- **Cluster 1 (non-flexibles)** : appareils à consommation élevée, souvent utilisés à des heures fixes (ex. : lumières),
- **Cluster 2 (semi-flexibles)** : appareils intermédiaires, tels que le chauffage et le climatiseur.

3.9.4 Visualisation des clusters

Afin de visualiser les résultats du clustering dans un espace à deux dimensions, une **Analyse en Composantes Principales (PCA)** a été appliquée sur les descripteurs des appareils. Cette méthode de réduction de dimension permet de projeter les appareils dans un plan 2D tout en conservant l'essentiel de la variance du jeu de données.

Cette projection facilite l'interprétation des groupes formés par l'algorithme de clustering. Les appareils appartenant à un même cluster se retrouvent proches dans cet espace réduit, ce qui permet d'observer leur séparation naturelle :

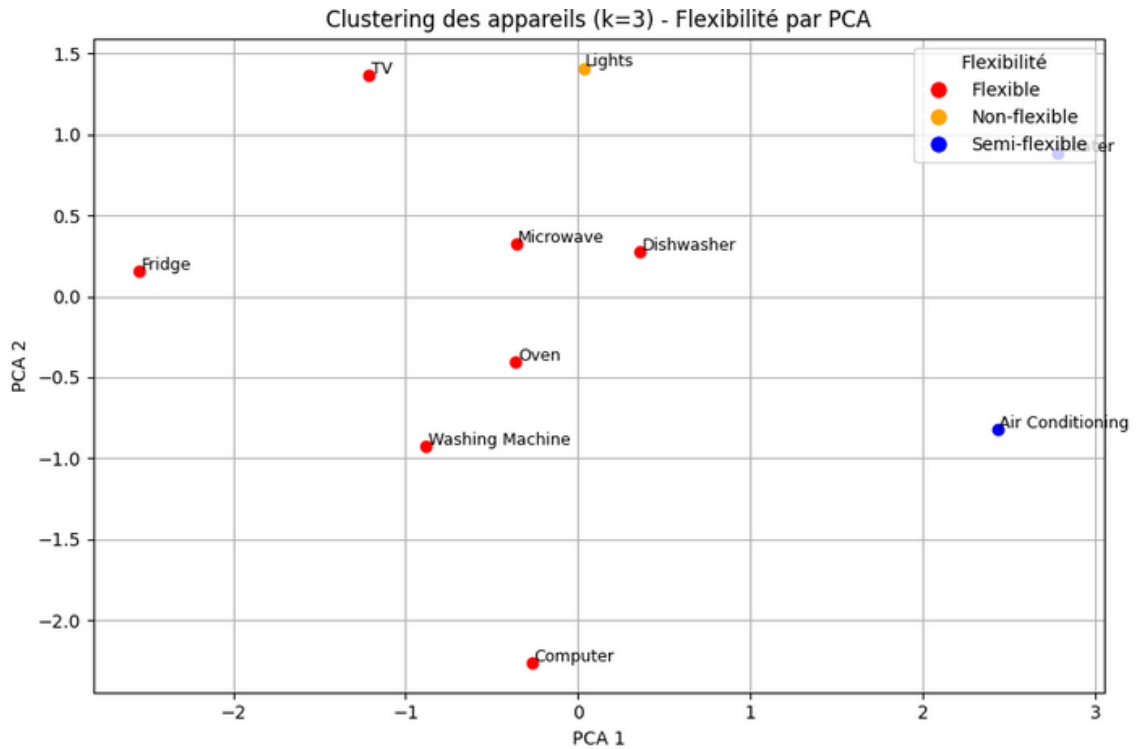


FIGURE 3.12 – Visualisation des clusters d'appareils électroménagers

3.10 Classification de la surcharge des foyers

3.10.1 Objectif de la classification

L'objectif de cette étape est de prédire, à un instant donné, si un foyer est en situation de **surcharge électrique**. Pour cela, nous avons défini une variable binaire `is_overloaded` : un foyer est considéré comme surchargé lorsque sa consommation totale dépasse le **60^e percentile** de sa propre distribution de consommation.

3.10.2 Construction du dataset

Le dataset utilisé pour la classification est construit à partir des activations agrégées par pas de temps fixe (15 minutes), en regroupant toutes les consommations d'un foyer à un instant donné. Chaque ligne représente donc un point temporel d'un foyer, enrichi par des variables contextuelles :

- `hour` : heure de la journée,
- `season_enc` : saison encodée,
- `outdoor_temperature_(°c)` : température extérieure,

- `household_size` : taille du foyer,
- `flexibility_level` : type d'appareils utilisés à cet instant (issu du clustering précédent),
- `period_type_enc` : type de plage horaire (`peak` / `offpeak`) encodé.

3.10.3 Sélection des variables explicatives

Trois méthodes complémentaires de sélection de variables ont été testées pour identifier les variables les plus pertinentes :

- **Corrélation absolue** : mesure la corrélation linéaire (positive ou négative) entre chaque variable et la variable cible.
- **Random Forest Importance** : importance des variables mesurée via l'impact sur les arbres de décision (réduction de l'impureté moyenne).
- **Mutual Information** : mesure la dépendance non linéaire entre les variables et la cible.

En nous basant sur la méthode la plus robuste (Random Forest), nous avons retenu les variables suivantes pour l'entraînement final :

```
outdoor_temperature_(°c), flexibility_level, hour, household_size,  
season_enc
```

3.10.4 Modèles de classification testés

Deux modèles supervisés ont été comparés :

- **Régression logistique** : modèle linéaire avec régularisation $L2$, optimisé par validation croisée via `GridSearchCV`.
- **K-Nearest Neighbors (KNN)** : modèle non paramétrique basé sur les plus proches voisins, avec recherche des meilleurs hyperparamètres (distance, nombre de voisins, pondération).

Résultats optimaux obtenus après optimisation :

- **Logistic Regression** : `C = 0.01`, `penalty = 'l2'`, `solver = 'lbfgs'`
- **KNN** : `n_neighbors = 9`, `metric = 'manhattan'`, `weights = 'distance'`

3.10.5 Évaluation des performances

Les performances ont été évaluées à l'aide de validation croisée (`GridSearchCV`, 5 folds), en utilisant plusieurs métriques : **accuracy**, **precision**, **recall**, **F1-score**.

Résultats pour la régression logistique (optimisée) :

- Accuracy : **76%**
- F1-score global : **0.75**
- Rappel (Surchargé) : **0.87**
- Précision (Surchargé) : **0.76**

Résultats pour KNN (optimisé) :

- Accuracy : **71%**
- F1-score global : **0.71**
- Rappel (Surchargé) : **0.77**

— Précision (Surchargé) : **0.75**

Ces résultats montrent que la régression logistique offre de meilleures performances globales sur notre dataset, notamment en termes de précision et de rappel sur la classe minoritaire.

La figure 3.13 suivante compare les performances des deux modèles à l'aide des principales métriques d'évaluation (accuracy, precision, recall, F1-score) :

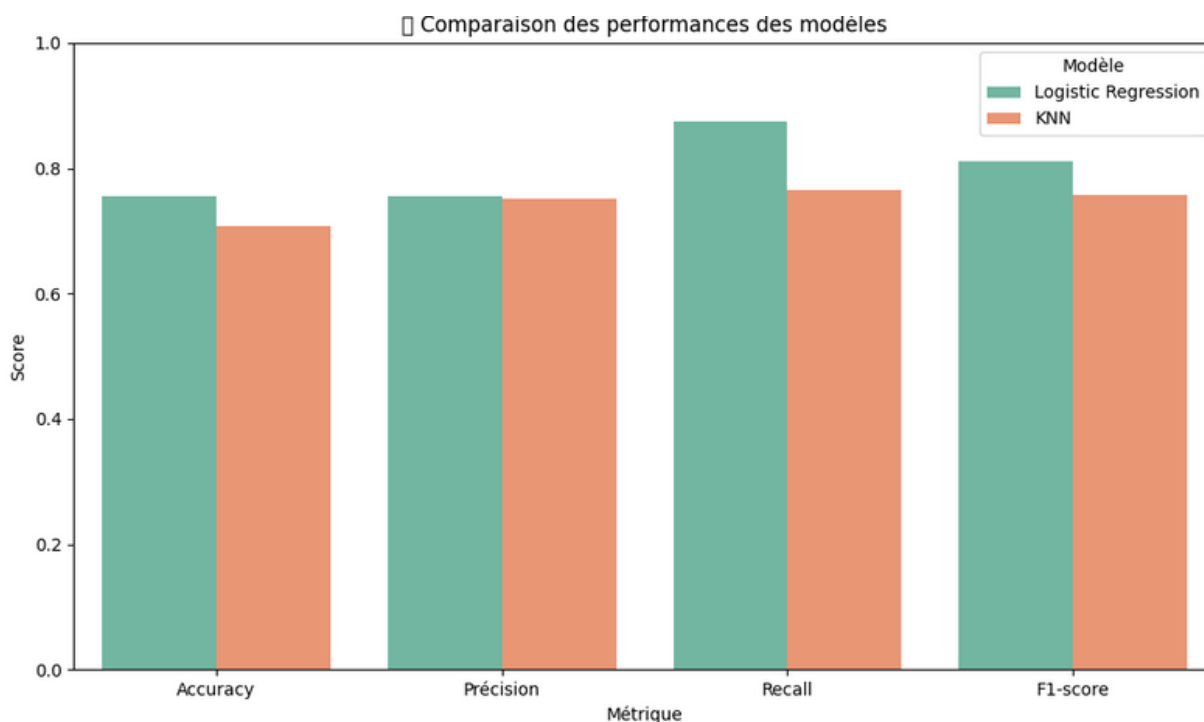


FIGURE 3.13 – Comparaison des performances des modèles de classification selon les principales métriques.

3.11 Recommandation de décalage d'activation d'appareils

3.11.1 Objectif

L'objectif de cette étape est de proposer une action corrective simple en cas de surcharge électrique d'un foyer, via le décalage intelligent de certains appareils électroménagers. L'approche choisie repose sur une **règle heuristique** efficace et facilement applicable dans un système automatisé.

3.11.2 Règle heuristique de décalage

La règle mise en place s'appuie sur les résultats précédents (classification de la surcharge, typologie horaire et flexibilité des appareils). Elle se formule comme suit :

Si un foyer est en surcharge ($is_overloaded = 1$) et que l'activation se produit en heure de pointe ($period_type = peak$), alors :

- *Identifier les appareils actifs à cet instant,*
- *Si un ou plusieurs de ces appareils sont de type flexible ou semi-flexible,*
- *Recommander le décalage de leur activation vers une heure creuse ou normale.*

Cette stratégie vise à réduire la consommation instantanée sans affecter les appareils critiques (non-flexibles). Elle privilégie les appareils dont le report dans le temps est acceptable.

3.11.3 Visualisation de l'effet du décalage

La figure 3.14 suivante illustre la différence de charge d'un foyer avant et après l'application de la règle de décalage :

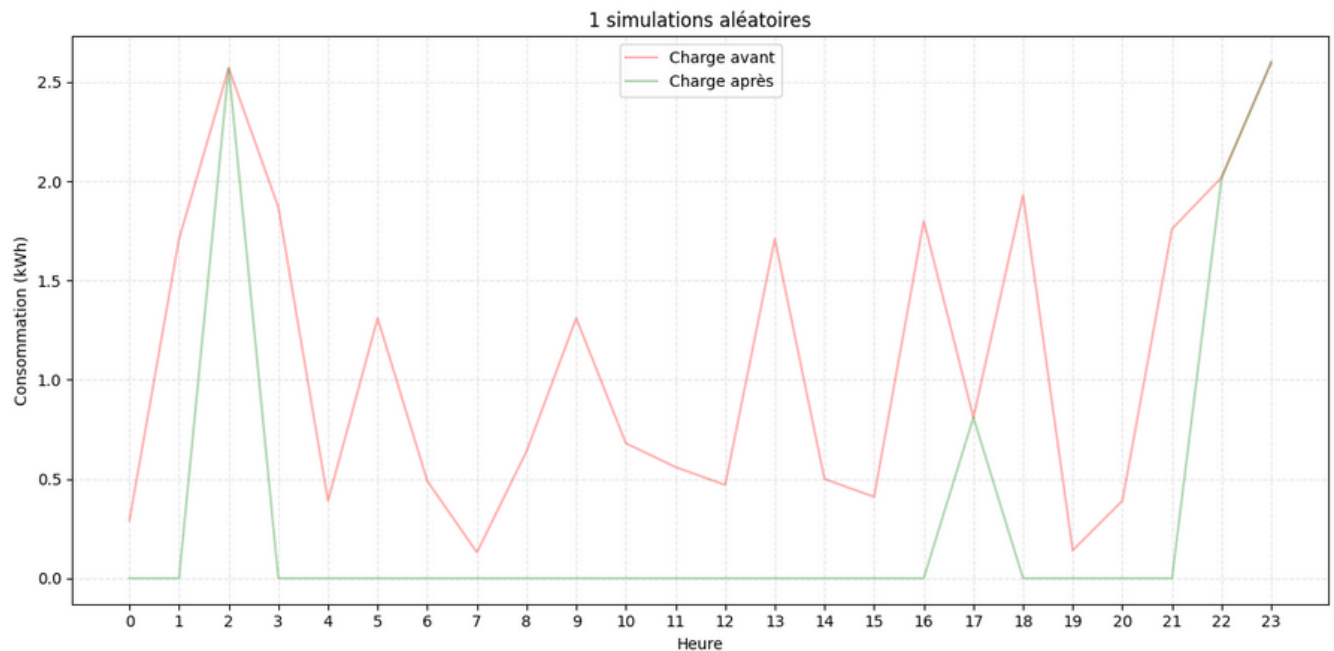


FIGURE 3.14 – Impact de la stratégie de décalage sur la charge électrique

On y observe une baisse significative de la charge (énergie gagnée : 18.49 kWh) et une réduction des pics de consommation.

3.12 Discussion

Dans le cadre de ce travail, deux approches ont été développées afin d'optimiser la gestion énergétique résidentielle et d'anticiper les situations de surcharge dans un contexte intelligent.

La première approche s'inscrit dans une logique d'optimisation de la consommation électrique, à travers une classification multi-niveaux appliquée aux foyers, aux plages horaires et aux appareils électroménagers. Elle repose sur un enchaînement de techniques adaptées à chaque sous-problème. Tout d'abord, une classification des foyers a été réalisée par clustering K-means, permettant de segmenter les utilisateurs selon leur niveau de consommation (basse, moyenne, élevée). Ensuite, pour identifier les plages horaires critiques (heures pleines et heures creuses), une classification supervisée a été effectuée à l'aide de l'algorithme Random Forest, qui a permis de prédire les périodes de forte charge en fonction de caractéristiques temporelles et saisonnières. Enfin, pour distinguer les appareils flexibles (dont l'usage peut être décalé) des non flexibles, un clustering hiérarchique agglomératif a été utilisé, en se basant sur des descripteurs comme la puissance, la durée d'utilisation et la fréquence d'activation. Cette approche combinée permet non seulement d'identifier les comportements énergétiques, mais aussi d'élaborer une stratégie de déplacement intelligent de la charge, en réaffectant les usages flexibles aux heures les moins critiques. Elle est particulièrement efficace pour répartir la demande de manière proactive sans perturber le confort des usagers.

La deuxième approche, complémentaire, adopte une démarche purement prédictive, visant à anticiper les surcharges électriques. Contrairement à la première, elle repose sur des techniques de classification supervisée, et utilise deux modèles : le K-Nearest Neighbors (KNN) et la régression logistique. Après un important travail de prétraitement, incluant la structuration des profils horaires et la sélection de variables explicatives (nombre d'appareils actifs, durée d'utilisation, saison, etc.), ces modèles ont été entraînés pour prédire la probabilité d'apparition d'une surcharge dans un foyer donné. KNN permet de classer un nouveau cas en fonction de ses voisins les plus proches, tandis que la régression logistique fournit une estimation probabiliste et interprétable. Une fois la surcharge détectée ou prédite, une règle heuristique de décalage a été définie, afin de proposer aux utilisateurs un réajustement de leurs usages vers des plages moins sollicitées. Cette approche permet une gestion dynamique des pics de consommation, en s'appuyant sur des décisions contextuelles basées sur les données en temps réel.

En définitive, ces deux approches, bien que différentes sur le plan algorithmique, sont complémentaires dans leurs finalités. La première, orientée vers la planification et l'optimisation en continu, repose sur une stratégie de segmentation intelligente et de réaffectation des usages ; la seconde, centrée sur la prévision et la réaction rapide, permet une anticipation des anomalies et une réponse ciblée. Leur intégration dans un même système énergétique intelligent permettrait de concilier efficacité, adaptabilité et résilience face aux contraintes croissantes des réseaux électriques.

Conclusion

Ce chapitre a présenté l'application concrète de notre approche fondée sur l'intelligence artificielle pour une gestion optimisée de la consommation électrique. Grâce à l'utilisation d'algorithmes de classification, nous avons pu segmenter les consommateurs, identifier les périodes critiques et simuler un déplacement de charge.

Les résultats obtenus montrent une amélioration significative de l'équilibrage énergétique, confirmant la pertinence de notre méthode. Cette démarche ouvre des perspectives prometteuses pour une gestion plus efficace et durable de la demande en électricité.

4
dd

Conclusion générale

Face aux défis croissants de gestion énergétique dans un monde de plus en plus interconnecté, ce mémoire a proposé une démarche innovante articulée autour de la classification intelligente des foyers pour un équilibrage optimal de la charge électrique. En nous inscrivant dans une logique d'intelligence artificielle appliquée à la transition énergétique, nous avons développé deux approches complémentaires qui, chacune à leur manière, ont apporté des réponses concrètes à la problématique étudiée.

La première approche a permis de segmenter finement les consommateurs et leurs appareils, révélant des profils types sur lesquels des stratégies ciblées d'optimisation ont pu être construites. Cette vision granulaire de la consommation a ouvert la voie à des ajustements précis et adaptatifs, contribuant ainsi à un allègement des charges aux heures critiques, sans compromettre le confort des utilisateurs.

La seconde approche, davantage tournée vers l'anticipation, a introduit un mécanisme prévisionnel capable de détecter les situations de surcharge potentielle. En s'appuyant sur des modèles de classification supervisée, elle a montré qu'il est possible non seulement de diagnostiquer les risques, mais aussi de recommander intelligemment des ajustements d'usage, renforçant ainsi la résilience du système électrique résidentiel.

Ces résultats, confirment l'efficacité de notre démarche et l'importance d'une modélisation centrée sur les comportements réels des consommateurs. Ils illustrent la capacité des outils d'analyse avancée à transformer des données brutes en leviers d'action intelligents, au service d'une gestion énergétique plus fine et plus durable.

Au-delà des résultats obtenus, ce travail ouvre des perspectives de recherche prometteuses. L'intégration de données en temps réel via les compteurs intelligents, l'extension de notre méthodologie à d'autres types d'environnements (industriels ou tertiaires), ou encore l'exploration de modèles hybrides combinant apprentissage automatique et optimisation multi-objectifs, constituent autant de voies à approfondir. De plus, la prise en compte de dimensions socio-comportementales ou économiques pourrait enrichir la personnalisation des stratégies proposées.

En définitive, ce mémoire se veut une contribution engagée à la convergence entre science des données et transition énergétique. Il affirme que l'intelligence artificielle, lorsqu'elle est orientée vers des objectifs durables et humains, peut devenir un puissant catalyseur de changement dans les usages quotidiens de l'énergie.

Bibliographie

- [1] AMJADY, N., KEYNIA, F., AND ZAREIPOUR, H. Short-term load forecasting of power systems by combination of neural networks and genetic algorithms. *IEEE Transactions on Power Systems* 26, 3 (2011), 1591–1598.
- [2] AZENCOTT, C. A. *Introduction au machine learning*, 2^e éd. ed. Dunod, 2022.
- [3] BALAKUMAR, P., VINOPRABA, T., AND CHANDRASEKARAN, K. Deep learning based real time demand side management controller for smart building integrated with renewable energy and energy storage system. *Journal of Energy Storage* 58 (2023), 106412.
- [4] BALDI, P., BRUNAK, S., CHAUVIN, Y., ANDERSEN, C. A. F., AND NIELSEN, H. Assessing the accuracy of prediction algorithms for classification : an overview. *Bioinformatics* 16, 5 (2000), 412–424.
- [5] BARBIER, T. *Modélisation de la consommation électrique à partir de grandes masses de données pour la simulation des alternatives énergétiques du futur*. Thèse de doctorat, PSL Research University, 2017.
- [6] BENUREAU, F. *Self Exploration of Sensorimotor Spaces in Robots*. Thèse de doctorat, Université de Bordeaux, 2015.
- [7] BESSA, R. J., TRINDADE, A., AND MIRANDA, V. Spatial-temporal solar power forecasting for smart grids. *Renewable and Sustainable Energy Reviews* 73 (2018), 655–666.
- [8] BIOINFO-FR.NET. La diversité des algorithmes de clustering. <https://bioinfo-fr.net/la-diversite-des-algorithmes-de-clustering>, 2025.
- [9] BREIMAN, L. Forêts aléatoires. *Apprentissage automatique* 45 (2001).
- [10] BURKARDT, J. K-means clustering. *Virginia Tech, Advanced Research Computing, Interdisciplinary Center for Applied Mathematics* (2009), 5.
- [11] CELIK, O. A research on machine learning methods and its applications. *Journal of Educational Technology and Online Learning* 1, 3 (2018), 25–40.
- [12] CHEN, T., AND GUESTRIN, C. Xgboost : Un système de boosting d’arbres évolutif. In *Actes de la 22^e conférence internationale ACM SIGKDD sur la découverte des connaissances et l’exploration de données* (août 2016), pp. 785–794.
- [13] CHEN, Y., ZHANG, Z., AND LI, F. Residential load shifting using deep reinforcement learning. *IEEE Access* 9 (2021), 106904–106914.
- [14] EDF. La consommation d’électricité en chiffres. <https://www.edf.fr/groupe-edf/comprendre/electricite-au-quotidien/usages/consommation-electricite-en-chiffres>, 2024.

- [15] FERECATU, M. Apprentissage, réseaux de neurones et modèles graphiques (rcp209) : Arbres de décision. <http://cedric.cnam.fr/vertigo/Cours/ml2/>, année du cours (par exemple 2025). Cours, Département Informatique, CNAM, Paris, France.
- [16] GUERMEUR, Y. *SVM multiclasses, théorie et applications*. Thèse de doctorat, Université Henri Poincaré-Nancy I, 2007.
- [17] GÉRON, A. *Apprentissage automatique pratique avec scikit-learn, Keras et TensorFlow*, 2^e éd. ed. Éditions O'Reilly, 2022.
- [18] HIPPERT, H. S., PEDREIRA, C. E., AND SOUZA, R. C. Neural networks for short-term load forecasting : A review and evaluation. *IEEE Transactions on Power Systems* 16, 1 (2002), 44–55.
- [19] HYNDMAN, R. J., AND ATHANASOPOULOS, G. *Forecasting : Principles and Practice*, 2^e éd. ed. OTexts, Melbourne, Australie, 2018.
- [20] KERROUMI, S. *Extraction des paramètres et classification dynamique dans le cadre de la détection et du suivi de défaut de roulements*. PhD thesis, Université de Reims Champagne-Ardenne, 2016.
- [21] LASGOUTTES, J.-M. Introduction au boosting. Cours de Mastère spécialisé ESD, Institut National des Sciences Appliquées (INSA) de Rouen, INRIA de Paris, 2023. Méthodes d'arbres en apprentissage statistique, année 2023–2024.
- [22] LIU, H., WANG, Y., LUO, X., AND SHAO, M. Optimal load shifting for electric vehicle charging via quadratic programming. *International Journal of Electrical Power & Energy Systems* 131 (2021), 107003.
- [23] LIU, Y., ZHANG, D., AND GOOI, H. B. Clustering-based load balancing in smart grids. In *2019 IEEE International Conference on Energy Internet (ICEI)* (2019), IEEE, pp. 123–128.
- [24] MCCARTHY, J. What is artificial intelligence? <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>, 2007.
- [25] MCKINNEY, W. *Python for Data Analysis*, 3 ed. O'Reilly Media, 2022.
- [26] MÜLLER, A. C., AND GUIDO, S. *Introduction to Machine Learning with Python : A Guide for Data Scientists*. O'Reilly Media, 2016.
- [27] NENE, S. Deep learning for natural language processing. *International Research Journal of Engineering and Technology (IRJET)* 4, 11 (2017), 933. Dept. of Computer Engineering, VESIT, Maharashtra, Inde.
- [28] NGUYEN, N. H., LEGER, J.-M., TRAN, Q.-T., AND VUONG, T.-P. Intelligent energy management in residential buildings with a real-time control & wireless meter-recording system. In *Fourteenth International Middle East Power Systems Conference (MEPCON'10)* (Cairo, Egypt, 2010), pp. 1–6.
- [29] REINDERS, A., VOSS, K., AND BERKEL, K. *Domestic Energy Consumption : Theory and Application of Parameter Estimation*. Earthscan, Londres, 2013.
- [30] SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development* 3, 3 (1959), 210–229.

- [31] SHARMA, A., SAHU, P. K., AND SAHU, P. K. Thermal-aware task scheduling for energy-efficient cloud computing. *ACM Transactions on Modeling and Performance Evaluation of Computing Systems (TOMPECS)* 4, 3 (2019), 1–23.
- [32] SMITH, J., AND JONES, E. Towards trustworthy energy disaggregation : A review of challenges, methods, and perspectives for non-intrusive load monitoring. <https://arxiv.org/abs/2207.02009>, 2022. arXiv preprint arXiv :2207.02009.
- [33] SRINIVASAN, D. S., ET AL. A comparison of energy consumption prediction models based on neural networks and support vector machines. *Journal of Intelligent Manufacturing* 29, 6 (2018), 1355–1368.
- [34] TEBANI, H. *Production de l'énergie électrique*, 2^e éd. ed. Librairie El-Djazair, Alger, Algérie, 2019.
- [35] TUSHAR, W., YUEN, C., POOR, H. V., SAHA, T., AND WOOD, K. L. Transforming energy networks via peer-to-peer energy trading : The potential of game-theoretic approaches. *IEEE Signal Processing Magazine* 35, 4 (2018), 90–111.
- [36] WANG, Y., ZHANG, N., ZHUO, Z., KANG, C., AND KIRSCHEN, D. Mixed-integer linear programming-based optimal configuration planning for energy hub. *Applied Energy* 261 (2020), 114305.
- [37] YU, J., DE ANTONIO, A., AND VILLALBA-MORA, E. Deep learning (cnn, rnn) applications for smart homes : a systematic review. *Computers* 11, 2 (2022), 26.
- [38] ZHOU, X., QIN, Y., LIU, Z., AND CHEN, M. Energy-efficient load shifting in edge clouds with predictive modeling. *IEEE Transactions on Cloud Computing* (2022). Early Access.
- [39] ZOHA, A., GLUHAK, A., IMRAN, M. A., AND RAJASEGARAR, S. Non-intrusive load monitoring approaches for disaggregated energy sensing : A survey. *Sensors* 12, 12 (2013), 16838–16866.

Résumé

Dans un contexte de transition énergétique, la gestion intelligente de la consommation d'électricité représente un enjeu crucial. Ce mémoire explore l'optimisation de la demande énergétique à travers une classification intelligente des foyers et un équilibrage dynamique de la charge. Contrairement aux approches classiques axées sur la prévision globale de la demande, notre démarche s'appuie sur une segmentation fine des profils de consommation dans le but de déplacer certaines charges vers les heures creuses, sans compromettre le confort des usagers.

Pour cela, nous avons mis en œuvre plusieurs algorithmes d'apprentissage automatique, notamment K-means, l'agglomératif hiérarchique et Random Forest. Ces techniques ont permis de classer les foyers selon leurs habitudes de consommation, d'identifier les périodes critiques, et de distinguer les appareils électriques flexibles de ceux qui ne le sont pas. Cette classification a servi de base à une simulation de déplacement de charge visant à équilibrer la consommation énergétique.

Les résultats obtenus montrent que notre approche contribue efficacement à la réduction des pics de consommation, à l'amélioration de la stabilité du réseau, ainsi qu'à une utilisation plus rationnelle de l'électricité au niveau domestique.

Mots-clés : Transition énergétique, Apprentissage automatique (Machine Learning), Classification des foyers, Équilibrage de charge, K-means, Random Forest, Clustering hiérarchique, Consommation d'électricité.

Abstract

In the context of the energy transition, the intelligent management of electricity consumption represents a critical challenge. This thesis explores the optimization of energy demand through intelligent household classification and dynamic load balancing. Unlike traditional approaches focused on forecasting overall demand, our approach relies on a detailed segmentation of consumption profiles with the goal of shifting certain loads to off-peak hours, without compromising user comfort.

To achieve this, we implemented several machine learning algorithms, including K-means, hierarchical agglomerative clustering, and Random Forest. These techniques made it possible to classify households based on their consumption habits, identify critical time periods, and distinguish flexible electrical appliances from non-flexible ones. This classification served as the basis for simulating load shifting aimed at balancing energy consumption.

The results show that our approach effectively contributes to reducing consumption peaks, improving grid stability, and promoting more efficient use of electricity at the household level.

Keywords : Energy transition, Machine learning, Household classification, Load balancing, K-means, Random Forest, Hierarchical clustering, Electricity consumption.