

République Algérienne Démocratique et Populaire Ministère de
l'Enseignement Supérieur et de la Recherche Scientifique



Mémoire de Master

Présenté à l'Université 8 Mai 1945 de Guelma

Faculté des Mathématiques, Informatiques et Sciences de la Matière

Département d' : **Informatique**

Spécialité : **SIQ & STIC**

Présenté par : **CISSE Dramane & BEN ZIADI NABIL SKANDAR**

Indexation des Données dans les Réseaux Intelligents en utilisant l'Apprentissage Automatique

Encadrant : Pr. ZINEDDINE KOUAHLA

Examineur : TADJER HOUDA

Président : BENAMIRA ADEL

Représentant Pole Pro : ZAHRA ABBAS

Résumé

La transition énergétique et la digitalisation du réseau électrique ont donné naissance aux Smart Grids, réseaux intelligents capables de gérer efficacement la production, la distribution et la consommation d'énergie. Ce mémoire explore les problématiques liées à l'indexation et à la prédiction dans ce contexte. Une analyse approfondie des techniques d'indexation (séries temporelles, RMI, autoencodeurs, etc.) et des méthodes de prédiction (régression, forêts aléatoires, LSTM, etc.) est présentée. Un prototype est développé, combinant indexation intelligente et prédiction automatique à partir de données énergétiques simulées. Les résultats montrent une amélioration notable des temps de requête et de la précision des prédictions, démontrant la pertinence de l'approche pour la gestion intelligente de données massives dans les Smart Grids.

Abstract

The energy transition and the digitalization of electrical infrastructure have led to the emergence of Smart Grids—intelligent networks that optimize energy production, distribution, and consumption. This thesis addresses the dual challenges of data indexing and prediction within Smart Grids. It presents a comprehensive review of indexing methods (time series indexing, RMI, autoencoders) and predictive models (regression, random forests, LSTM). A prototype system combining intelligent indexing and automatic forecasting using energy time series data was implemented. The experimental results demonstrate significant improvements in query response time and prediction accuracy, validating the proposed approach as an effective solution for managing massive and heterogeneous energy data in Smart Grid systems.

Remerciement

Avant tout, nous rendons grâce à Dieu, Le Tout-Puissant, pour nous avoir accordé la santé, la patience et la persévérance tout au long de cette aventure intellectuelle. Sans Sa volonté et Son aide, rien n'aurait été possible.

Nous souhaitons ensuite exprimer nos profondes gratitude à Pr. Zineddine KOUAHLA, notre encadrant, pour son accompagnement rigoureux, sa disponibilité et ses précieux conseils. Sa vision et son encadrement bienveillant ont été une source constante de motivation et de structuration de ce travail.

Nous n'oublions pas d'adresser nos reconnaissances les plus sincères à nos familles, en particulier nos parents, pour leur amour inconditionnel, leur patience et leur soutien moral indéfectible, sans lesquels ce parcours n'aurait pas eu la même force.

À toutes celles et ceux qui ont, d'une manière ou d'une autre, contribué à l'élaboration de ce mémoire, qu'Allah vous récompense et vous comble de Ses bienfaits.

Chapitre 1 Table des matières

Introduction Générale	1
Contexte.....	1
Problématique	1
Objectifs du mémoire	1
Organisation du mémoire.....	2
Chapitre 1 – Les Réseaux Intelligents	4
1.1 Introduction.	4
1.2 Définition des Smart Grids	4
1.3 Objectifs (écologiques)	5
1.4 Principales composantes d'un smart grids	5
1.5 Différences avec les réseaux traditionnels	6
1.6 Évolution historique vers les Smart Grids	7
1.7 Conclusion	8
Chapitre 2 – Introduction à l'apprentissage automatique	9
2.1 Introduction	9
2.2 Principales Familles d'Apprentissage Automatique	10
2.2.1 Apprentissage supervisé	10
2.2.2 Apprentissage non supervisé.....	11
2.2.3 Apprentissage semi-supervisé.....	12
2.2.4 Apprentissage par renforcement	13
2.3 Évaluation des modèles	14
2.3.1 <i>Classification</i>	14
2.3.2 Régression.....	16
2.4 Conclusion	17
Chapitre 3 – Introduction à l'indexation des données dans les Smart Grids	18
3.1 Introduction	18
3.2 Techniques d'indexation traditionnelles	18
3.3 Techniques d'indexation intelligente	19
3.4 Techniques d'indexation dans les Smart Grids.....	20
3.5 Conclusion	22
Chapitre 4 – Système d'Indexation : Conception et Mise en Œuvre	23
4.1 Introduction	23

4.2 Énoncé du Problème	23
4.3 Solution Proposée	23
4.3.1 Vue d'Ensemble du Système	23
4.4 Paquets et Langage de Programmation.....	24
4.5 Carte de Développement du Modèle	24
4.5.1 Entraînement du Modèle	24
4.5.2 Métriques d'évaluation.....	25
4.5.3 Résultats obtenus	25
4.5.4 Analyse des résultats	25
4.6 Conclusion	26
Chapitre 5 – Système de Prédiction : Conception et Mise en Œuvre.....	27
5.1 Prévion de la Demande Énergétique : Conception et Mise en Œuvre ...	27
5.1.1 Introduction.....	27
5.1.2 Énoncé du Problème.....	27
5.1.3 Solution Proposée	27
5.1.3.1 Vue d'Ensemble du Système	28
5.1.4 Paquets et Langage de Programmation	30
5.1.5 Carte de Développement du Modèle	30
5.1.5.1 Collecte des Données.....	31
5.1.5.3 Ingénierie des Caractéristiques	34
5.1.5.4 Entraînement du Modèle	34
5.1.5.5 Résultats d'Évaluation.....	35
5.1.5.6 Déploiement du Modèle.....	36
5.1.5.7 Amélioration du Modèle	36
5.1.6 Conclusion	36
5.2 Conception et Mise en Œuvre : Prévion des Pannes des Lampadaires	36
5.2.1 Introduction.....	36
5.2.2 Énoncé du Problème.....	37
5.2.3 Solution Proposée	37
5.2.3.1 Vue d'Ensemble du Système	37
5.2.4 Paquets et Langage de Programmation	41
5.2.5 Carte de Développement du Modèle	41
5.2.5.1 Collecte de Données.....	41

5.2.5.2 Analyse Exploratoire des Données (AED)	42
5.2.5.3 Ingénierie des Caractéristiques	45
5.2.5.4 Entraînement des Modèles.....	45
5.2.5.5 Résultats d'Évaluation.....	46
5.2.5.6 Déploiement du Modèle.....	47
5.2.5.7 Amélioration du Modèle	48
5.2.6 Conclusion	48
Conclusion Générale.....	49
Annexe.....	50
Annexe I : Projet Smart Grid IA __ Création du StartUp	50
Annexe II : Interface Utilisateur.....	65
Références	71

Liste des figures

Figure 1: Exemple de Réseau Intelligent [W1].	5
Figure 2 : Réseau Traditionnel vs Intelligent [W2].	6
Figure 3 : Évolution du réseau électrique du traditionnel à l'intelligent 2.0 [W3].	8
Figure 4 : Taxonomie des techniques d'apprentissage automatique [W4] .	10
Figure 5: Vue d'Ensemble du Système d'Indexation .	24
Figure 6. Vue d'Ensemble du Système de Prédiction de consommation d'énergie [W5].	28
Figure 7: Flux de Travail du Développement du Modèle de Prédiction de Consommation d'Energie .	29
Figure 8: Module d'Ingénierie des Caractéristique .	29
Figure 9: Module d'Entraînement des Modèles .	30
Figure 10: Diagramme de Séquence des étapes de prétraitement des données. _	31
Figure 11. Consommation d'Énergie Horaire Moyenne pour le Bâtiment 41 - Montre les pics d'utilisation diurne, probablement dus à l'occupation et la climatisation. _	32
Figure 12. Consommation Moyenne Quotidienne et Mensuelle (2016–2018) - Illustre les tendances sur trois ans, avec des pics en été. _	32
Figure 13. Carte de Chaleur des Corrélations des Caractéristiques - Montre les corrélations, notamment entre énergie et température. _	33
Figure 14. Graphique de Dispersion Température, Humidité et Consommation d'Énergie - Montre la relation tridimensionnelle avec des pics à haute température. _	33
Figure 15. Historique d'Entraînement du LSTM - Présente la diminution de la perte au fil des époques d'entraînement. _	34
Figure 16: Diagramme de Classe. _	35
Figure 17. Comparaison de RMSE et R^2 entre les Modèles - Présente les métriques de performance pour différents modèles. _	35
Figure 18: Vue d'Ensemble su Système de Prédiction des Pannes [W5]. _	38
Figure 19: Flux de Travail de Développement du Système de Prédiction des Pannes .	39
Figure 20: Module de Prétraitement des Données .	39
Figure 21: Module d'Ingénierie des Caractéristiques .	40
Figure 22: Module d'Entraînement des Modèles .	40
Figure 23: Module d'Evaluation .	41
Figure 24. Distribution des Types de Pannes. _	42
Figure 25. Taux de Pannes par Heure de la Journée. _	43
Figure 26. Taux de Pannes Quotidiens sur le Temps .	43
Figure 27. Distributions des Caractéristiques Numériques .	44
Figure 28: Distribution des types de faute par Condition Environnementale. _	44
Figure 29: Corrélation des Caractéristiques Numériques .	45
Figure 30: Comparaison des Performances (Modèles de prédiction des pannes). _	47

Liste des tableaux

<i>Tableau 1: Principales Composantes d'un Smart Grid [6].</i>	5
<i>Tableau 2: Bref Comparaison entre Réseau Traditionnel et Intelligent [6].</i>	6
<i>Tableau 3: Tableau Comparatif des algorithmes d'apprentissage supervisés appliqués aux smart grid.</i>	10
<i>Tableau 4: Tableau Comparatif des algorithmes d'apprentissage non-supervisés appliqués aux smart grid.</i>	11
<i>Tableau 5: Tableau Comparatif des algorithmes d'apprentissage semi-supervisés appliquées aux smart grid.</i>	13
<i>Tableau 6: Tableau Comparatif des algorithmes d'apprentissage par renforcement applliquées aux smart grid.</i>	13
<i>Tableau 7: Tableau Comparatif des techniques d'indexation traditionnelles.</i>	19
<i>Tableau 8: Tableau Comparatif des techniques d'indexation intelligente.</i>	20
<i>Tableau 9: Techniques d'indexation classiques appliquées aux Smart Grids.</i>	21
<i>Tableau 10: Indexation spécifique aux séries temporelles.</i>	21
<i>Tableau 11: Techniques récentes et IA pour l'indexation.</i>	21
<i>Tableau 12: Performances des Modèles d'Indexation.</i>	25
<i>Tableau 13: Performances des Modèles de Prédiction pour consommation d'énergie.</i>	36
<i>Tableau 14: Performances des Modèles de prédiction des pannes (classification binaire).</i>	46
<i>Tableau 15: Performances des Modèles de prédiction des pannes (classification multiclasse).</i>	46

Liste des acronymes

- IA: Intelligence Artificielle
- ML: Machine Learning
- DL: Deep Learning
- RNN: Recurrent Neural Network
- LSTM: Long Short-Term Memory
- GRU: Gated Recurrent Unit
- CNN: Convolutional Neural Network
- XAI: eXplainable Artificial Intelligence
- XGBoost: Extreme Gradient Boosting
- KNN: K-Nearest Neighbors
- SVM: Support Vector Machine
- MAE: Mean Absolute Error
- RMSE: Root Mean Squared Error
- MAPE: Mean Absolute Percentage Error
- MSE: Mean Squared Error
- R^2 : Coefficient de Détermination
- AUC: Area Under Curve
- ROC: Receiver Operating Characteristic
- SCADA: Supervisory Control And Data Acquisition
- AML: Advanced Metering Infrastructure
- IoT: Internet of Things
- SIG: Système d'Information Géographique
- RMI: Recursive Model Index
- TSI: Time Series Index
- TS-Index : Time Series Index (technique spécifique)
- iSAX: Indexable Symbolic Aggregate approXimation
- XAI: Explainable Artificial Intelligence
- HVAC: Heating, Ventilation and Air Conditioning
- TD Learning: Temporal Difference Learning
- MADRL: Multi-Agent Deep Reinforcement Learning
- PPO: Proximal Policy Optimization
- NEAT: NeuroEvolution of Augmenting Topologies
- B+-Tree: Balanced Plus Tree
- SGBD : Système de Gestion de Base de Données
- OLAP: Online Analytical Processing
- Trie: Arbre de préfixes
- R-Tree: Rectangle Tree (index spatial)
- FAISS: Facebook AI Similarity Search

Introduction Générale

Contexte

La transition énergétique, conjuguée à la digitalisation des infrastructures électriques, a conduit à l'émergence des **Smart Grids**, ou réseaux électriques intelligents. Ces réseaux s'appuient sur une intégration poussée des technologies de l'information et de la communication (TIC) pour superviser, automatiser et optimiser les flux d'énergie. Dans ce nouveau paradigme, les compteurs intelligents, les capteurs IoT et les systèmes de gestion distribuée génèrent un **volume massif de données** en temps réel, hétérogènes et dynamiques.

L'exploitation efficace de ces données repose sur deux fonctions essentielles :

- L'indexation intelligente, qui permet de structurer et rechercher rapidement des informations dans des flux continus.
- La prédiction fiable, qui anticipe les besoins énergétiques, la production renouvelable ou les anomalies, afin de renforcer la stabilité et l'efficacité du réseau.

Cependant, la croissance exponentielle des données, leur complexité (non-linéarité, bruit, variabilité temporelle) et les exigences de performance en temps quasi réel imposent de dépasser les méthodes classiques. L'intelligence artificielle, en particulier l'**apprentissage automatique**, apparaît comme une réponse prometteuse pour relever ces défis.

Problématique

Face à cette complexité croissante, deux questions fondamentales guident ce mémoire :

- Comment indexer efficacement des flux de données massifs et hétérogènes provenant des Smart Grids pour permettre un accès rapide et pertinent ?
- Comment prédire avec précision, à partir de ces données, les comportements futurs (consommation, production, anomalies) pour améliorer la prise de décision ?

Objectifs du mémoire

Ce travail de recherche vise à :

- Réaliser un état de l'art approfondi sur les techniques d'indexation et de prédiction dans les Smart Grids.

- Proposer une architecture combinant indexation intelligente et prédiction automatique basée sur des algorithmes d'apprentissage automatique.
- Implémenter un prototype fonctionnel, appliqué à des séries temporelles issues de données énergétiques.
- Évaluer la performance de l'approche proposée selon des critères précis (temps de réponse, précision, robustesse).

Organisation du mémoire

Ce mémoire est structuré en **cinq chapitres principaux**, chacun abordant un aspect clé de la gestion intelligente des données dans les Smart Grids, depuis les fondements conceptuels jusqu'à l'implémentation pratique des solutions proposées.

- **Chapitre 1 : Les Réseaux Intelligents (Smart Grids)**

Ce chapitre introduit les Smart Grids en présentant leur architecture, leurs composantes essentielles, ainsi que les enjeux technologiques, écologiques et économiques liés à leur déploiement. Il établit le contexte global dans lequel s'inscrivent les problématiques d'indexation et de prédiction.

- **Chapitre 2 : Introduction à l'Apprentissage Automatique**

Une revue des principales approches d'apprentissage automatique est proposée, incluant les méthodes supervisées, non supervisées, semi-supervisées et par renforcement. Le chapitre met en lumière leur pertinence pour analyser les données complexes et dynamiques issues des Smart Grids.

- **Chapitre 3 : Techniques d'Indexation des Données dans les Smart Grids**

Ce chapitre explore les méthodes d'indexation, à la fois classiques et avancées, applicables aux séries temporelles issues des infrastructures intelligentes. Il souligne l'importance de l'indexation pour améliorer la rapidité d'accès aux données et optimiser leur traitement en temps réel.

- **Chapitre 4 : Conception et Mise en Œuvre d'un Système d'Indexation**

À travers ce chapitre, une architecture d'indexation intelligente est conçue, implémentée et testée sur des jeux de données énergétiques simulés. Les performances des différentes approches sont analysées à l'aide de métriques pertinentes (temps de réponse, précision, consommation mémoire).

- **Chapitre 5 : Conception et Mise en Œuvre de Systèmes de Prédiction**

Ce chapitre présente deux cas d'usage concrets de prédiction dans les Smart Grids : la prévision de la consommation énergétique d'un bâtiment, et la détection proactive des pannes de lampadaires. Divers modèles d'apprentissage automatique et profond sont comparés et évalués.

Introduction Générale

Enfin, une **conclusion générale** synthétise les contributions du mémoire, discute des limites rencontrées et propose des perspectives pour des travaux futurs.

Chapitre 1 – Les Réseaux Intelligents

1.1 Introduction.

Les réseaux électriques intelligents, ou *smart grids*, marquent une étape clé dans l'évolution du système électrique traditionnel. En intégrant les technologies numériques, la surveillance à distance et le pilotage en temps réel, ils permettent un échange bidirectionnel d'électricité et d'informations entre les producteurs et les consommateurs [1].

Grâce à des capteurs performants et à des systèmes de communication avancés, les smart grids rendent le réseau plus souple, plus réactif, et capable de s'ajuster automatiquement en cas de problème. Ils facilitent aussi l'intégration des énergies renouvelables, comme le solaire et l'éolien, souvent variables et difficiles à prédire [1].

Concrètement, les smart grids offrent plusieurs avantages majeurs :

- a) Une meilleure fiabilité du réseau électrique, même en cas d'imprévu,
- b) Une gestion plus efficace de la production et de la distribution d'énergie,
- c) Une intégration plus fluide des sources renouvelables,
- d) Une contribution importante à la réduction des émissions de gaz à effet de serre [1].

Ces innovations s'inscrivent dans une démarche plus large de transition énergétique, en faveur d'un modèle plus durable. Les compteurs communicants, par exemple, jouent un rôle central : ils permettent de suivre la consommation en temps réel et ouvrent la voie à une gestion intelligente de la demande, adaptée aux habitudes des usagers [1].

En résumé, les smart grids apportent une nouvelle façon de penser le réseau électrique : plus souple, plus intelligent, et plus respectueux de l'environnement. Ils posent les bases d'un système énergétique moderne, capable de répondre aux défis actuels et futurs [1].

1.2 Définition des Smart Grids

Un Smart Grid désigne un réseau électrique nouvelle génération, enrichi par des technologies numériques capables de gérer intelligemment l'électricité à toutes les étapes : de la production à la consommation. Contrairement aux réseaux classiques, il repose sur une communication bidirectionnelle en temps réel entre les différents acteurs du système (producteurs, distributeurs et consommateurs) [2].

Concrètement, le Smart Grid s'appuie sur un ensemble de capteurs et de compteurs intelligents qui collectent et analysent en continu des données. Ces informations sont ensuite utilisées pour automatiser les décisions et optimiser l'ensemble de la chaîne énergétique : production, transport, distribution et usage final de l'électricité [3].

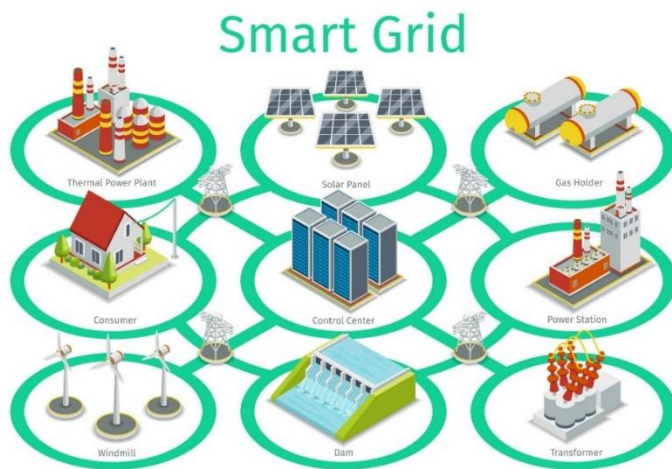


Figure 1: Exemple de Réseau Intelligent [W1].

1.3 Objectifs (écologiques)

L'un des objectifs majeurs des smart grids est de faciliter l'intégration des énergies renouvelables, comme le solaire ou l'éolien. Ces sources, bien qu'essentielles pour la transition énergétique, sont par nature variables et difficiles à prédire. Grâce à une meilleure gestion de cette variabilité, les smart grids contribuent à stabiliser le réseau tout en réduisant les émissions de gaz à effet de serre [4]. Par ailleurs, en limitant les pertes énergétiques tout au long de la chaîne, de la production à la consommation, et en optimisant l'usage de l'énergie via des systèmes intelligents de gestion de la demande, ces réseaux permettent une baisse significative des émissions de CO₂ [5].

1.4 Principales composantes d'un smart grids

Tableau 1: Principales Composantes d'un Smart Grid [6].

Composante	Description
Systèmes de Mesure Avancés (AMI)	Infrastructure pour la collecte, la gestion et la communication bidirectionnelle des données de consommation.
Capteurs et Automatismes de Distribution	Permettent la détection et la réponse rapide aux défauts et optimisent la performance du réseau.
Réseaux de Communication	Supportent la transmission fiable et sécurisée de données entre toutes les composantes du système.

Intégration des Sources d'Énergie Renouvelable	Permet l'intégration dynamique de l'éolien, du solaire et d'autres sources renouvelables.
Systèmes de Gestion de l'Énergie (EMS)	Optimisent la génération, la distribution et la consommation d'énergie.
Réseaux Intelligents Domestiques (HAN)	Facilitent l'interaction entre les appareils domestiques et le réseau électrique.
Centres de Contrôle Intelligents	Analytiques avancées pour surveiller, contrôler et prédire les états du réseau en temps réel.
Stockage d'Énergie	Permet de lisser la production variable et de stabiliser le réseau.
Véhicules Électriques et Infrastructure de Recharge	Interaction dynamique avec le réseau pour la recharge intelligente et le stockage temporaire.

1.5 Différences avec les réseaux traditionnels

Observation de la différence entre réseau électrique traditionnel et intelligent (smart grid) à partir des éléments suivants :

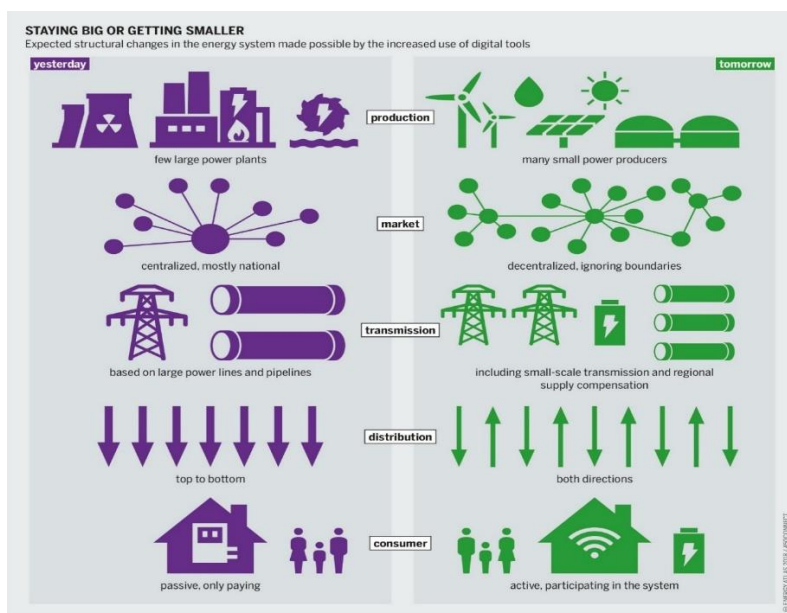


Figure 2 : Réseau Traditionnel vs Intelligent [W2].

Tableau 2: Bref Comparaison entre Réseau Traditionnel et Intelligent [6].

Critère	Réseau traditionnel	Smart Grid (réseau intelligent)

Technologie de base	Électromécanique	Numérique
Communication	Communication unidirectionnelle	Communication bidirectionnelle
Structure de production	Production centralisée	Production décentralisée
Capteurs	Peu de capteurs	Capteurs répartis dans tout le réseau
Surveillance	Surveillance manuelle	Auto-surveillance
Réparation	Rétablissement manuel	Auto-réparation (self-healing)
Gestion des incidents	Pannes et coupures	Adaptatif et pilotable
Contrôle	Contrôle limité	Contrôle omniprésent
Choix pour le client	Peu de choix pour les clients	Large éventail de choix pour les clients

1.6 Évolution historique vers les Smart Grids

Réseau électrique traditionnel (jusqu'au début du XXI^e siècle)

Le réseau électrique classique, souvent qualifié de "centralisé", a été conçu pour transporter l'électricité depuis de grandes unités de production (comme les centrales thermiques ou hydrauliques) vers les consommateurs finaux via unidirectionnelle : production → transport → distribution → consommation. Ce système est robuste mais manque de flexibilité, de visibilité sur la demande en temps réel, et de capacité à intégrer les énergies renouvelables [6].

Limites du système traditionnel

Le réseau souffre de pannes fréquentes, d'une faible efficacité énergétique, d'une réponse lente aux événements, et d'un contrôle très limité sur la consommation à distance. Il n'offre pas de retour d'information du consommateur, ni de gestion active de la demande [6].

Transition vers le Smart Grid (début des années 2000 – aujourd'hui)

Le Smart Grid (réseau intelligent) est né pour répondre à ces limites. Il intègre des technologies de communication bidirectionnelle, d'automatisation, de capteurs intelligents, de compteurs avancés (smart meters), et permet une intégration efficace des énergies renouvelables et des véhicules électriques. Le smart grid offre ainsi plus de résilience, d'efficacité, d'interactivité avec les utilisateurs, et un contrôle distribué [6].

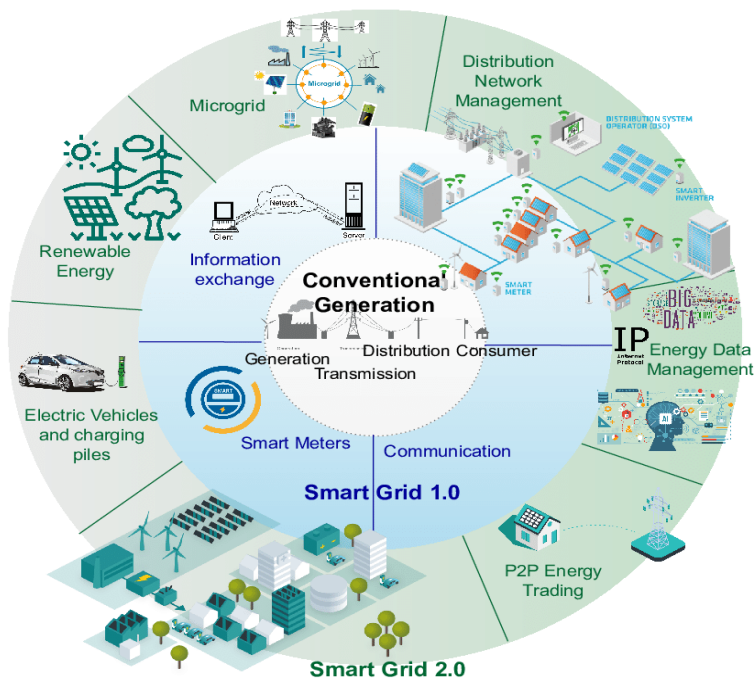


Figure 3 : Évolution du réseau électrique du traditionnel à l'intelligent 2.0 [W3].

1.7 Conclusion

Ce premier chapitre a posé les fondations conceptuelles du mémoire en introduisant la problématique de la transition énergétique et l'émergence des Smart Grids comme réponse technologique et durable aux enjeux contemporains. Les Smart Grids se distinguent des réseaux traditionnels par leur capacité à intégrer des sources d'énergie renouvelables, à assurer un pilotage dynamique de la production et de la consommation, et à exploiter les technologies de l'information pour une meilleure supervision du réseau électrique.

Le chapitre a également mis en lumière les principales composantes de ces réseaux intelligents – capteurs, compteurs communicants, systèmes de gestion, etc. – et les fonctions qu'ils remplissent en matière d'optimisation, d'équilibrage, de détection de pannes et de gestion prédictive.

Enfin, il a souligné le rôle central des données dans cette nouvelle architecture, montrant que leur collecte, leur traitement et leur analyse sont au cœur de l'intelligence des Smart Grids. Ces constats justifient pleinement l'étude et l'implémentation de solutions avancées d'indexation et de prédiction, qui seront approfondies dans les chapitres suivants.

Chapitre 2 – Introduction à l'apprentissage automatique

2.1 Introduction

L'émergence des Smart Grids (réseaux électriques intelligents) marque une révolution dans la gestion de l'énergie. Ces réseaux intègrent les technologies de l'information et de la communication (TIC) pour superviser, analyser et optimiser la production, la distribution et la consommation d'électricité de manière dynamique [6].

Face à la complexité croissante des systèmes énergétiques, caractérisée par l'intégration massive des énergies renouvelables, la variabilité de la demande, l'apparition des véhicules électriques, et la participation active des consommateurs, il devient indispensable d'utiliser des outils avancés capables de traiter des volumes massifs de données en temps réel [7], [8].

C'est dans ce contexte que l'apprentissage automatique (Machine Learning) s'impose comme une solution clé. En permettant aux systèmes de « s'auto-ajuster » à partir des données historiques et en cours, l'apprentissage automatique offre des capacités de prédiction, de classification, de détection d'anomalies et d'optimisation [9], [10].

Il permet ainsi d'automatiser de nombreuses tâches critiques dans le Smart Grid, telles que :

- a) La prévision de la consommation ou de la production d'énergie (notamment solaire ou éolienne) [11],
- b) L'optimisation de la charge des véhicules électriques [12],
- c) La détection de fraudes ou de cyberattaques [13],
- d) La gestion dynamique de la demande (Demand Response) [14],
- e) Le contrôle distribué de la fréquence et de la tension dans les réseaux [15].

Parmi les différentes approches de l'apprentissage automatique, on distingue principalement :

- 1) L'apprentissage supervisé, basé sur des données étiquetées pour la prédiction ou la classification [9],
- 2) L'apprentissage non supervisé, utilisé pour la segmentation de clients ou la détection d'événements inhabituels [10],
- 3) L'apprentissage par renforcement, qui permet à un agent d'interagir avec l'environnement du réseau pour apprendre des stratégies de gestion optimale à long terme [16].

Ainsi, l'intégration de l'apprentissage automatique au sein des Smart Grids vise non seulement à renforcer l'intelligence du réseau, mais aussi à garantir une résilience

accrue, une meilleure efficacité énergétique et une adaptation proactive aux évolutions technologiques et sociétales [6], [7].

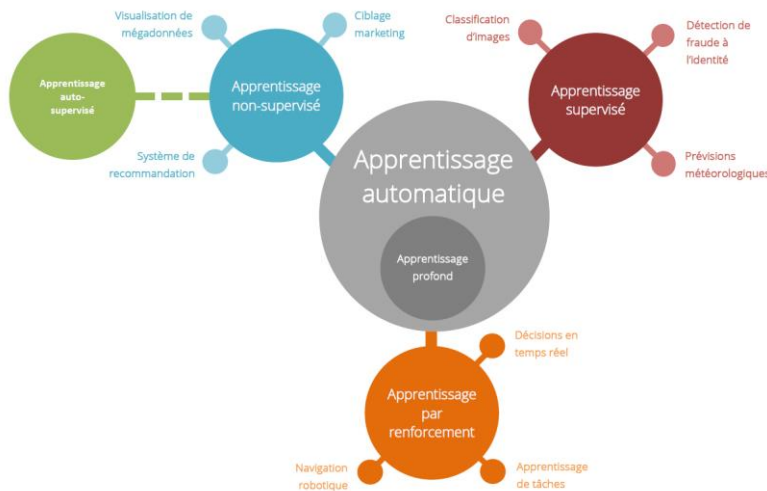


Figure 4 : Taxonomie des techniques d'apprentissage automatique [W4] .

2.2 Principales Familles d'Apprentissage Automatique

2.2.1 Apprentissage supervisé

L'apprentissage supervisé est une forme d'apprentissage automatique dans laquelle un modèle est entraîné à partir d'un ensemble de données étiquetées, c'est-à-dire où chaque exemple d'entrée est associé à une sortie (ou « étiquette ») connue. Le but est d'apprendre une fonction de prédiction qui, à partir de nouvelles données d'entrée, produira la sortie correspondante avec une précision maximale [17].

Tableau 3: Tableau Comparatif des algorithmes d'apprentissage supervisés appliqués aux smart grid.

Algorithme	Type de tâche	Avantages détaillés	Inconvénients détaillés	Applications dans les Smart Grids
Régression linéaire [18]	Régression	Simple, rapide, facilement interprétable, utile pour établir des tendances linéaires claires	Suppose une relation linéaire, mauvaise performance sur données non linéaires ou bruitées	Prévision de la charge, estimation de la production
Régression logistique [19]	Classification	Bonne performance pour classes binaires, résultats interprétables, entraînement rapide	Moins adaptée aux relations complexes, sensible aux corrélations entre variables indépendantes	Détection de fraude, classification de profils clients
Arbres de décision [20]	Classification, régression	Facile à visualiser, peu de préparation des données, gestion intuitive des décisions conditionnelles	Fortement sujet au surapprentissage, instable si les données changent légèrement	Segmentation client, prédiction de pannes

Chapitre 2 : Introduction à l'Apprentissage Automatique dans les Smart Grids

Random Forest [21]	Classification, régression	Très robuste, peu sensible au surapprentissage, gère bien les variables catégorielles et les données manquantes	Moins interprétable, plus lent que les arbres simples lors de l'inférence	Maintenance prédictive, prévision énergétique
SVM [22]	Classification	Très performant avec peu de données et données à haute dimension, flexible via les noyaux	Coût de calcul élevé, paramétrage complexe (choix du noyau, régularisation)	Détection d'événements ou d'anomalies
K-NN [23]	Classification, régression	Facile à implémenter, aucun entraînement préalable requis, fonctionne bien pour des données bien séparées	Lenteur à la prédiction, sensible au bruit et à la dimension (effet "curse of dimensionality")	Recommandation, classification des profils
ANN (Neural Nets) [24]	Classification, régression	Capable de modéliser des relations très complexes, adaptable à divers types de données	Nécessite beaucoup de données, peu interprétable, long à entraîner	Prévision fine de la demande, détection de défauts
Gradient Boosting [25]	Classification, régression	Haute précision, bien adapté aux données complexes et bruitées, très performant avec peu de réglages	Risque de surajustement si mal calibré, coûteux en temps de calcul, nécessite un bon réglage des hyperparamètres	Détection de fraude, prévision de production renouvelable

2.2.2 Apprentissage non supervisé

L'apprentissage non supervisé est une méthode d'apprentissage automatique où l'on entraîne un modèle sur un ensemble de données non étiquetées. L'objectif est de découvrir des structures cachées dans les données, comme des regroupements (clustering), des règles d'association ou des réductions de dimensions. Aucun signal de sortie explicite n'est fourni. [17]

Exemples :

- i) Regrouper des clients selon leurs comportements d'achat (clustering).
- ii) Détecter des anomalies dans des séries temporelles.
- iii) Réduire la dimensionnalité des données (ex. : PCA) [17].

Tableau 4: Tableau Comparatif des algorithmes d'apprentissage non-supervisés appliqués aux smart grid.

Algorithme	Type de tâche	Avantages détaillés	Inconvénients détaillés	Applications typiques dans les Smart Grids
K-Means [26], [27]	Clustering, segmentation	-Très rapide et facile à implémenter -Fonctionne bien sur de grands ensembles de données	-Le nombre de clusters K doit être défini à l'avance -Sensible aux valeurs aberrantes et au bruit	Segmentation de clients, profils de consommation

		-Résultats interprétables	-Mauvaise gestion des clusters non sphériques	
DBSCAN [28], [29]	Clustering, détection d'anomalies	-Identifie des clusters de forme arbitraire - Robuste au bruit et aux outliers -Pas besoin de spécifier le nombre de clusters	-Sensible aux paramètres eps et minPts -Moins performant en haute dimension	Détection de comportements inhabituels, anomalies de réseau
Clustering hiérarchique [30], [31]	Classification hiérarchique	-Pas besoin de spécifier le nombre de clusters -Produit une arborescence (dendrogramme) interprétable	-Peu scalable sur les grands volumes de données -Sensible au bruit et à la redondance	Groupement d'équipements, classification hiérarchique de capteurs
PCA [32]	Réduction de dimension	-Simple à mettre en œuvre -Améliore la performance des modèles en réduisant la redondance• Bon pour la visualisation 2D/3D	-Suppose une relation linéaire entre les variables -Perte d'interprétabilité des nouvelles composantes	Compression des séries temporelles, prétraitement pour l'analyse
Auto-Encodeur [33], [34]	Encodage, détection d'anomalies	-Capte des relations non linéaires complexes -Utilisable pour compression, reconstruction, détection -Puissant avec des séries	-Entraînement long et exigeant -Difficulté d'interprétation des résultats -Risque d'overfitting	Détection d'événements rares, prévision de consommation
SOM (Self-Organizing Map) [86], [35]	Clustering, visualisation	-Fournit une cartographie visuelle des données -Topologie préservée (voisins proches restent voisins) -Interprétation intuitive	-Temps d'entraînement long, surtout pour les grandes dimensions -Difficile à paramétrer (taille de grille, voisinage, etc.)	Visualisation des profils, analyse comportementale des compteurs

2.2.3 Apprentissage semi-supervisé

L'apprentissage semi-supervisé est une méthode d'apprentissage automatique qui exploite à la fois des données étiquetées et non étiquetées pour entraîner un modèle. Cette approche est particulièrement utile lorsque l'étiquetage des données est coûteux ou difficile, mais que l'on dispose d'un grand volume de données non étiquetées. L'objectif est d'améliorer la performance du modèle en utilisant intelligemment les données non étiquetées pour renforcer l'apprentissage supervisé. [36]

Exemple :

Utiliser quelques relevés de consommation électrique annotés (par exemple, avec des classes de comportement) et un grand nombre de relevés bruts non annotés pour améliorer la détection de profils d'usage dans un Smart Grid [36].

Chapitre 2 : Introduction à l'Apprentissage Automatique dans les Smart Grids

Tableau 5: Tableau Comparatif des algorithmes d'apprentissage semi-supervisés appliqués aux smart grid.

Méthode	Principe	Applications dans les Smart Grids	Avantages	Inconvénients
Self-Training [37], [38]	Réutilise les prédictions à haute confiance pour réentraîner le modèle.	NILM, détection d'événements locaux.	Simple à implémenter, extensible, peu de paramètres.	Propagation d'erreurs possibles, sur-apprentissage si mauvaises prédictions.
Co-Training [39], [40]	Deux modèles s'enseignent mutuellement à partir de vues différentes.	Classification multi-source (conso + météo, etc.).	Exploite redondance, robuste avec bonnes vues.	Requiert deux vues indépendantes ; instable si l'une est faible.
Propagation d'étiquettes [41], [42]	Diffuse les étiquettes connues via un graphe de similarité.	Détection FDI, anomalies locales.	Efficace avec très peu de labels, bien adapté aux données structurées.	Sensible à la qualité du graphe (erreurs de similarité).
Graph-Based SSL [43], [44]	Apprentissage direct sur un graphe global connectant tous les exemples.	Détection de cyberattaques, événements rares.	Capture bien les relations complexes, structure globale exploitée.	Coût de calcul élevé, nécessite un bon graphe.
Autoencodeurs semi-supervisés [45], [46]	Combine reconstruction (autoencodeur) et classification.	Détection de pertes, fraudes, maintenance prédictive.	Apprentissage profond efficace, encode bien les données.	Besoin de réglage fin, risque de surajustement.
SSL profond (Deep SSL) [47], [48]	Réseaux profonds (CNN/LSTM) avec régularisation semi-supervisée.	Prédiction de charge, détection d'anomalies de consommation.	Très bonne performance avec peu de labels, adapte bien les séries temporelles.	Exigeant en données, calcul intensif, tuning complexe.

2.2.4 Apprentissage par renforcement

L'apprentissage par renforcement est une branche de l'apprentissage automatique dans laquelle un agent apprend à prendre des décisions en interagissant avec un environnement. L'agent reçoit des récompenses (positives ou négatives) en fonction de ses actions, et son objectif est d'apprendre une politique qui maximise la récompense cumulée à long terme [49].

Élément central : Le processus est souvent modélisé comme un processus de décision de Markov (MDP).

Exemples :

- i) Optimiser la charge d'un réseau électrique.
- ii) Contrôler un robot mobile.
- iii) Jouer à des jeux (échecs, Go, etc.) [49].

Tableau 6: Tableau Comparatif des algorithmes d'apprentissage par renforcement appliqués aux smart grid.

Algorithme	Type	Principe de fonctionnement	Avantages	Inconvénients	Applications Smart Grid
------------	------	----------------------------	-----------	---------------	-------------------------

Chapitre 2 : Introduction à l'Apprentissage Automatique dans les Smart Grids

Q-Learning [50]	Sans modèle	Apprend une fonction de valeur $Q(s,a)$ pour estimer la récompense attendue d'une action dans un état donné.	Simple, convergence garantie	Inefficace pour grands espaces d'états	Charge des VE, gestion domestique
DQN [51]	Sans modèle, profond	Approche Q-Learning avec réseau de neurones profond, mémoire d'expérience et réseau cible.	Gère grands espaces d'états	Instabilité lors de l'apprentissage	Micro-réseaux, pics de demande
DDPG [52]	Acteur-Critique, continu	Algorithme acteur-critique pour actions continues, basé sur des réseaux neuronaux.	Actions continues, convergence rapide	Sensible aux hyperparamètres	Contrôle tension, répartition de puissance
PPO [53]	Acteur-Critique, politique	Mise à jour prudente de la politique pour stabiliser l'apprentissage, basé sur les gradients.	Stable, échantillonnage efficace	Moins adapté aux actions discrètes	Optimisation consommation des bâtiments
MADRL [54]	Multi-agent	Plusieurs agents interagissent dans un environnement partagé pour résoudre des problèmes distribués.	Scalabilité, décentralisation	Coordination complexe	Planification énergétique, gestion distribuée
TD Learning [55]	Sans modèle	Mise à jour des valeurs d'état par différences temporelles, sans attendre la fin de l'épisode.	Apprentissage rapide en ligne	Dépendance aux estimations initiales	Gestion énergétique des foyers
NEAT [56]	Évolutionnaire	Évolution conjointe des topologies et poids de réseaux de neurones via sélection naturelle.	Découverte automatique de topologies	Coût computationnel élevé	Contrôle adaptatif, stratégie énergétique

2.3 Évaluation des modèles

2.3.1 Classification

Pour évaluer les performances des modèles de classification (binaire ou multiclasse), plusieurs métriques sont couramment utilisées [57] :

- Exactitude (Accuracy) : proportion de prédictions correctes sur l'ensemble des prédictions effectuées.

$$[\text{Exactitude} = \frac{TP + TN}{TP + TN + FP + FN}]$$

où :

TP = vrais positifs

TN = vrais négatifs

FP = faux positifs

FN = faux négatifs

- Précision (Precision) : proportion de prédictions positives correctes parmi toutes les prédictions positives [57].

$$[\text{Précision} = \frac{TP}{TP + FP}]$$

où :

TP = vrais positifs

FP = faux positifs

- Rappel (Recall ou Sensibilité) : proportion de vrais positifs détectés parmi l'ensemble des cas positifs réels.

$$[\text{Rappel} = \frac{TP}{TP + FN}]$$

où :

TP = vrais positifs

FN = faux négatifs

- F-mesure (F1-score) : moyenne harmonique de la précision et du rappel, offrant un équilibre entre ces deux métriques, particulièrement utile en cas de déséquilibre des classes [57].

$$[F1 - score = 2 \times [Précision \times Rappel / Précision + Rappel]]$$

où :

$$Précision = TP / (TP + FP)$$

$$Rappel = TP / (TP + FN)$$

- Courbe ROC (Receiver Operating Characteristic) et AUC (Area Under the Curve) : la courbe ROC illustre le compromis entre le taux de vrais positifs et le taux de faux positifs, tandis que l'AUC quantifie la capacité du modèle à distinguer entre les classes [57].

$$[TPR = \frac{TP}{TP + FN} \quad (\text{Taux de vrais positifs, ou sensibilité})]$$

$$[FPR = \frac{FP}{FP + TN} \quad (\text{Taux de faux positifs})]$$

$$[AUC = \int_0^1 TPR(FPR) d(FPR)]$$

TPR (True Positive Rate) est l'ordonnée de la courbe.

FPR (False Positive Rate) est l'abscisse.

AUC (Area Under the Curve) correspond à l'aire sous la courbe ROC, variant entre 0 et 1.

2.3.2 Régression

Pour les tâches de régression, les métriques suivantes sont fréquemment utilisées [58] :

- Erreur quadratique moyenne (MSE - Mean Squared Error) : moyenne des carrés des écarts entre les valeurs prédites et les valeurs réelles, pénalisant davantage les grandes erreurs.

$$MSE = (1/n) \times \sum (y_i - \hat{y}_i)^2$$

- Erreur absolue moyenne (MAE - Mean Absolute Error) : moyenne des valeurs absolues des écarts entre les prédictions et les valeurs réelles, offrant une mesure plus robuste aux valeurs aberrantes.

$$MAE = (1/n) \times \sum |y_i - \hat{y}_i|$$

- Coefficient de détermination (R^2) : indique la proportion de la variance des données expliquée par le modèle. Un R^2 proche de 1 signifie que le modèle explique bien la variance observée.

$$R^2 = 1 - (\sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2)$$

Le choix de la métrique dépend de la nature du problème et des objectifs spécifiques. Par exemple, en cas de classes déséquilibrées, la F-mesure ou l'AUC sont préférables à l'exactitude [57].

2.4 Conclusion

Ce deuxième chapitre a permis de présenter les fondements de l'apprentissage automatique, pierre angulaire des systèmes intelligents intégrés dans les Smart Grids. Il a décrit les principales familles d'algorithmes – supervisé, non supervisé, semi-supervisé et par renforcement – en soulignant leurs mécanismes d'apprentissage, leurs domaines d'application et leurs avantages respectifs dans le contexte de la modélisation énergétique.

L'accent a été mis sur les techniques de classification et de régression, qui sont essentielles pour la prédiction d'événements (comme les pannes) et la quantification de variables continues (telles que la consommation ou la production). Le chapitre a également introduit des indicateurs clés de performance tels que la précision, le rappel, le score F1, l'AUC-ROC pour la classification, et le MSE, MAE, R^2 pour la régression, éléments indispensables pour l'évaluation rigoureuse des modèles.

En somme, ce chapitre a apporté une base théorique solide sur les méthodes d'apprentissage automatique, qui servira de socle pour les choix méthodologiques et algorithmiques détaillés dans les chapitres suivants, notamment en matière d'indexation intelligente et de prédiction dans les Smart Grids.

Chapitre 3 – Introduction à l'indexation des données dans les Smart Grids

3.1 Introduction

Avec l'émergence des Smart Grids (réseaux électriques intelligents), le secteur de l'énergie connaît une transformation profonde, intégrant des technologies numériques pour rendre le système électrique plus flexible, efficace et interactif [59]. Ces réseaux s'appuient sur des dispositifs tels que les compteurs intelligents, les capteurs IoT, ou encore les systèmes SCADA pour surveiller et gérer en temps réel la production, la distribution et la consommation d'énergie [60].

Ces infrastructures produisent un flux massif et continu de données hétérogènes : mesures de tension, de courant, historiques de consommation, prévisions météo, etc. [61]. L'exploitation efficace de ces données est essentielle pour assurer des fonctions critiques telles que la gestion de la demande, la prévention des pannes, ou l'intégration des énergies renouvelables [62].

Dans ce cadre, l'indexation des données joue un rôle central. Elle permet de structurer et référencer les données afin d'optimiser leur accès rapide, leur analyse en temps réel, et leur exploitation à grande échelle [80]. Une bonne indexation est cruciale pour améliorer la réactivité des systèmes, réduire les délais de traitement et accélérer les processus décisionnels automatisés, notamment ceux basés sur l'intelligence artificielle [63].

Cependant, les méthodes classiques d'indexation sont souvent inadaptées face aux caractéristiques spécifiques des données Smart Grids : volume élevé, nature temporelle, flux continu, et sources multiples [64]. Cela a conduit à l'adoption de techniques avancées issues du Big Data et du machine learning, permettant de construire des index dynamiques qui s'ajustent aux types de requêtes les plus fréquentes et aux variations des flux de données [65].

Ainsi, dans le contexte des Smart Grids, l'indexation n'est pas une simple opération d'optimisation des bases de données, mais constitue un levier stratégique pour la performance, la résilience et l'intelligence du réseau électrique [66].

3.2 Techniques d'indexation traditionnelles

L'indexation est une technique fondamentale en informatique qui permet d'accélérer l'accès aux données dans les systèmes de gestion de bases de données (SGBD). Elle repose sur la création de structures auxiliaires facilitant la recherche, le tri et l'accès rapide à l'information. Depuis les premières bases de données relationnelles, plusieurs techniques d'indexation dites *traditionnelles* ont été développées, chacune ayant ses avantages selon le type de données et les cas d'usage.

Tableau 7: Tableau Comparatif des techniques d'indexation traditionnelles.

Technique	Avantages	Inconvénients	Cas d'application	Temps de recherche
B+-Tree [67]	- Recherche rapide- Ordonné- Bon pour les plages de valeurs	- Moins efficace pour faible cardinalité- Mise à jour coûteuse	Bases de données relationnelles, index primaire/secondaire	$O(\log n)$
Bitmap [68]	- Très compact pour faible cardinalité- Opérations logiques rapides	- Inefficace si haute cardinalité- Peu adapté aux mises à jour	Entrepôts de données, champs booléens ou catégoriels	$O(1)$ pour les opérations logiques
Hash [69]	- Recherche exacte très rapide- Simple à implémenter	- Pas de recherche par intervalle- Sensible aux collisions	Tables mémoire, bases NoSQL, recherche clé-valeur	$O(1)$ (amorti)
Trie [70]	- Efficace pour recherche par préfixe- Bon pour les chaînes	- Gourmand en mémoire- Moins efficace pour de petites données	Moteurs de recherche, autocomplétion, dictionnaires	$O(m)$ (m = longueur du mot)
Index inversé [71]	- Rapide pour recherche textuelle- Optimisé pour documents non structurés	- Nécessite un traitement préalable (tokenisation)- Moins adapté au numérique	Recherche plein texte, moteurs de recherche web	$O(1)$ à $O(\log n)$ selon l'implémentation
R-Tree [72]	- Adapté aux données spatiales- Gère les requêtes multidimensionnelles	- Moins efficace si les objets sont très denses ou imbriqués	SIG, bases de données spatiales, applications géographiques	$O(\log n)$

Les techniques d'indexation traditionnelles ont été conçues pour répondre à des besoins variés en matière de structuration et d'accès aux données. Si elles restent efficaces dans de nombreux cas, les nouveaux contextes de traitement massif, comme le Big Data, les séries temporelles ou les données non structurées, posent de nouveaux défis. Cela justifie le développement de méthodes d'indexation avancées et adaptatives, que l'on retrouve notamment dans les systèmes distribués ou basés sur l'intelligence artificielle.

3.3 Techniques d'indexation intelligente

L'explosion des volumes de données, en particulier dans les systèmes temps réel et les environnements Big Data, a mis en évidence les limites des techniques d'indexation traditionnelles. Face à la complexité croissante des données (volumineuses, hétérogènes, temporelles, non structurées), des approches dites "intelligentes" ont émergé, s'appuyant sur des méthodes issues du machine learning, de l'apprentissage automatique ou des systèmes adaptatifs. Ces techniques permettent de créer des index dynamiques, contextuels, voire auto-optimisants, capables de s'adapter aux types de requêtes, aux modèles de données ou aux évolutions dans le temps.

Chapitre 3 : Introduction à l'Indexation des Données dans les Smart Grids

Tableau 8: Tableau Comparatif des techniques d'indexation intelligente.

Technique	Avantages	Inconvénients	Cas d'application	Temps de recherche (moyen)
Learned Indexes (RMI) [73]	- Compact- Très rapide- Prédicatif	- Sensible à la distribution- Difficulté de mise à jour	Bases clé-valeur, systèmes NoSQL	$O(1)$ à $O(\log n)$
Adaptive Indexing (Cracking) [74]	- Construction progressive- Auto-ajustement aux requêtes	- Moins performant pour requêtes aléatoires- Instabilité initiale	Systèmes OLAP, requêtes analytiques	Variable (s'améliore avec le temps)
iSAX / TS-Index [75], [76], [77]	- Compression symbolique- Rapide pour requêtes temporelles	- Requier normalisation préalable- Moins flexible	Bases temporelles, séries IoT, Smart Grids	$O(\log n)$
LSH (Locality-Sensitive Hashing) [78], [79]	- Très rapide pour données de haute dimension- Approximation scalable	- Résultats non exacts- Sensible aux paramètres	Recherche de similarité, reconnaissance de formes	$O(1)$ amorti
Autoencodeurs (Deep Indexing) [80]	- Réduction de dimension automatique- Bonne généralisation	- Modèle coûteux à entraîner- Moins interprétable	Indexation multimédia, séries complexes, Big Data	$O(1)$ pour recherche latente
Hyperspace / Spark Indexing [81]	- Scalable- Intégration Big Data- Suivi des requêtes	- Overhead mémoire- Complexité de gestion distribuée	Data Lakes, bases distribuées, DataFrame Spark	$O(\log n)$ ou dépendant du moteur

Les **techniques d'indexation intelligente** représentent une évolution essentielle pour répondre aux défis posés par la **croissance exponentielle des données**, en particulier dans les **systèmes temps réel** (comme les Smart Grids) et les **environnements Big Data**. Contrairement aux index traditionnels, ces approches :

- ✓ **S'adaptent** dynamiquement aux données et aux requêtes,
- ✓ Exploitent l'**apprentissage machine** pour optimiser les structures internes,
- ✓ Améliorent la **recherche par similarité**, la **recherche floue** ou **contextuelle**,
- ✓ Permettent une **scalabilité** sur des architectures distribuées.

Elles ne sont pas destinées à **remplacer systématiquement les index traditionnels**, mais à les **compléter ou les surpasser** dans des cas spécifiques où la **complexité**, la **dynamique** et la **volumétrie** des données rendent les index classiques inefficaces.

3.4 Techniques d'indexation dans les Smart Grids

Les Smart Grids génèrent d'immenses volumes de données issues de sources hétérogènes (capteurs, compteurs intelligents, équipements de contrôle, etc.). Ces données, souvent sous forme de séries temporelles, doivent être efficacement indexées pour permettre une recherche rapide, une analyse en temps réel et une prédiction fiable. L'indexation devient ainsi une composante critique pour la scalabilité des systèmes Smart Grids.

Chapitre 3 : Introduction à l'Indexation des Données dans les Smart Grids

Tableau 9: Techniques d'indexation classiques appliquées aux Smart Grids.

Technique	Avantages	Inconvénients	Cas d'usage typiques	Référence
Indexation par arbre (Tree-based)	- Efficace pour les données multidimensionnelles. - Prise en charge des requêtes spatio-temporelles.	- Performances dégradées en haute dimension. - Coût de mise à jour élevé avec des flux continus.	- Indexation des capteurs répartis géographiquement. - Requêtes sur l'évolution spatiale de la consommation/production.	[82], [83]
Indexation symbolique (iSAX)	- Compression efficace des séries temporelles. - Scalabilité à très grande échelle.	- Perte de précision due à la discrétisation.	- Requêtes de similarité dans les séries de consommation énergétique. - Détection de motifs similaires.	[84], [85]
Hashing (LSH)	- Très rapide pour des requêtes de similarité approximative.	- Non adapté aux recherches exactes. - Nécessite un prétraitement lourd.	- Recherche de motifs récurrents dans les séries de charge. - Indexation pour détection d'événements anormaux.	[79]

Tableau 10: Indexation spécifique aux séries temporelles.

Technique	Avantages	Inconvénients	Cas d'usage typiques	Référence
DFT Indexing	- Capture les caractéristiques globales. - Réduction de la dimensionnalité.	- Mauvaise précision locale.	- Recherche de tendances saisonnières ou périodiques.	[86]
DWT Indexing	- Bonne localisation dans le temps et la fréquence.	- Complexité de calcul.	- Compression et indexation des données de tension/fréquence.	[87]
PAA et iSAX	- Représentation simple et rapide. - Compatible avec les arbres.	- Perte d'information sur les détails fins.	- Classification rapide des profils de consommation.	[84]

Tableau 11: Techniques récentes et IA pour l'indexation.

Technique	Avantages	Inconvénients	Cas d'usage typiques	Référence
Autoencoders	- Représentations compactes et non linéaires. - Résistance au bruit.	- Complexité élevée. - Coût computationnel important.	- Indexation de séries multi-variables (météo, consommation, etc.).	[88], [89]
Recursive Model Index (RMI)	- Index prédictif rapide et hiérarchique.	- Moins robuste pour des requêtes aléatoires.	- Indexation de séries chronologiques avec timestamp.	[90]

Les Smart Grids exigent des systèmes d'indexation capables de traiter des données massives, hétérogènes, et évolutives. Les approches classiques restent utiles pour des contextes simples ou bien structurés, tandis que les techniques modernes basées sur l'apprentissage profond (autoencoders, RMI) apportent une flexibilité et une efficacité accrues dans les environnements dynamiques. Le choix de la

technique dépendra du volume de données, du type de requêtes attendues et des contraintes en temps réel.

3.5 Conclusion

Le troisième chapitre s'est attaché à explorer les techniques d'indexation des données, une étape cruciale pour assurer la performance des systèmes de traitement dans les Smart Grids. Face à la volumétrie, la variété et la vélocité des données générées par les infrastructures intelligentes, il est impératif de recourir à des méthodes d'indexation efficaces, capables d'assurer une récupération rapide et pertinente de l'information.

Dans ce cadre, plusieurs approches classiques ont été rappelées, telles que les arbres B+, les tables de hachage ou les structures Trie, avant de mettre l'accent sur des méthodes plus adaptées aux données hétérogènes et temporelles : indexation vectorielle (k-d trees, ball trees), techniques basées sur la distance (k-NN), indexation probabiliste (LSH), ainsi que des méthodes hybrides exploitant l'apprentissage automatique, comme les autoencodeurs ou les modèles RMI.

Ce panorama des méthodes d'indexation a permis de souligner leurs avantages, leurs limites et leurs conditions d'applicabilité selon la nature des données et les contraintes des Smart Grids. Il fournit ainsi un socle décisionnel essentiel pour le développement d'une solution d'indexation intelligente intégrée à un pipeline de prédiction, tel que proposé dans les chapitres suivants.

Chapitre 4 – Système d'Indexation : Conception et Mise en Œuvre

4.1 Introduction

Ce chapitre décrit la conception, l'implémentation et l'évaluation d'un système d'indexation efficace pour les données énergétiques des Smart Grids.

4.2 Enoncé du Problème

Le module a été conçu pour :

- i) Indexer efficacement les séries temporelles électriques (puissance, tension, etc.)
- ii) Permettre une recherche rapide par timestamp ou par similarité (voisins proches)
- iii) Comparer la performance de différents algorithmes d'indexation

4.3 Solution Proposée

4.3.1 Vue d'Ensemble du Système

Le système est organisé en trois modules principaux :

- i) Module de prétraitement : nettoyage des données, normalisation, extraction des features temporelles.
- ii) Module d'indexation : construction des index avec plusieurs structures (KD-Tree, Ball Tree, KNN, LSH).
- iii) Module d'évaluation : mesure du temps de construction, de la latence des requêtes, de la précision, et de la mémoire.

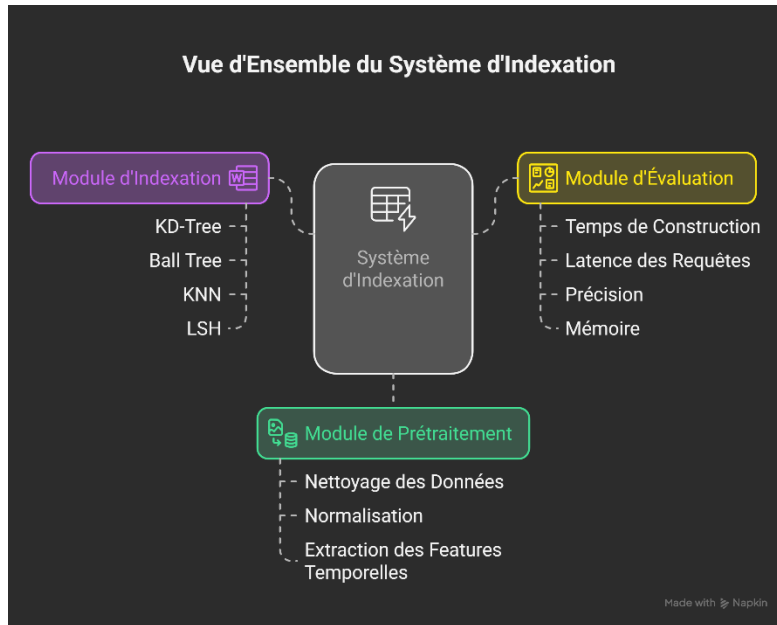


Figure 5: Vue d'Ensemble du Système d'Indexation .

4.4 Paquets et Langage de Programmation

Le système a été développé en Python 3, choisi pour ses vastes bibliothèques d'apprentissage automatique et ses capacités de traitement de données. Les principaux paquets incluent :

pandas et numpy : Manipulation des données et opérations numériques.

scikit-learn : Algorithmes d'apprentissage automatique et métriques d'évaluation.

xgboost : Cadre de boosting par gradient pour des performances optimisées.

matplotlib et seaborn : Visualisation des données.

joblib : Sérialisation des modèles pour le chargement des modèles pré-entraînés.

L'analyse a été réalisée dans des notebooks Jupyter, garantissant la reproductibilité et la modularité.

datasketch : Bibliothèque utilisée pour l'indexation intelligente des séries temporelles, basée sur des techniques de hachage de similarité (MinHash), afin d'optimiser les requêtes énergétiques dans les grandes bases de données.

4.5 Carte de Développement du Modèle

Les méthodes d'indexation ont été implémentées et testées sur le même jeu de données utilisé pour le *système de prévision des pannes de lampadaires* dans le chapitre 5, pour plus d'information sur le jeu de données (voir chapitre 5).

4.5.1 Entraînement du Modèle

Les méthodes suivantes ont été implémentées et testées :

KD-Tree : arbre binaire optimisé pour les requêtes de type k-NN dans les espaces euclidiens.

Ball Tree : structure alternative à KD-Tree pour les données à plus haute dimension.

KNN (sans index) : méthode de référence, effectue une recherche exhaustive.

LSH (Locality Sensitive Hashing) : technique d'approximation pour les recherches par similarité à grande échelle.

4.5.2 Métriques d'évaluation

Les critères suivants ont été utilisés pour comparer les performances :

- a) Temps de construction (s)
- b) Temps moyen de requête (ms)
- c) Précision (vs. référence KNN)
- d) Mémoire utilisée (MB)

4.5.3 Résultats obtenus

Les performances des méthodes sont synthétisées ci-dessous :

Tableau 12: Performances des Modèles d'Indexation.

Méthode	Construction (s)	Requête (ms)	Précision (%)	Mémoire (MB)
KD-Tree	2.1	1.8	100	45
Ball Tree	2.3	2.1	100	48
KNN (k=5)	-	12.5	100	60
LSH (MinHash)	1.5	0.9	92	30

4.5.4 Analyse des résultats

KD-Tree et Ball Tree : excellentes performances sur les requêtes exactes avec des temps de réponse très faibles et une précision parfaite.

KNN : méthode de base, très précise mais lente, peu adaptée aux exigences temps réel. **LSH** : compromis entre rapidité et précision, efficace pour les grands ensembles de données où une approximation est acceptable.

4.6 Conclusion

Ce chapitre a présenté la conception, l'implémentation et l'évaluation d'un système d'indexation dédié aux séries temporelles issues des Smart Grids. L'objectif était de permettre un accès rapide et pertinent aux données énergétiques massives, tout en comparant différentes structures d'index.

Les tests menés ont permis de valider l'efficacité des approches KD-Tree et Ball Tree, qui ont montré une précision de 100 % avec des temps de requête très faibles, idéaux pour des applications à faible latence. La méthode LSH, quant à elle, offre un compromis pertinent entre vitesse et approximation, particulièrement adaptée aux contextes à très grand volume où la précision absolue est moins critique.

Enfin, la comparaison avec la méthode de base (k-NN sans index) a démontré les gains substantiels en performances grâce à l'utilisation de structures d'indexation intelligentes.

Ce module constitue une brique essentielle pour toute architecture de gestion intelligente des données Smart Grid, en apportant scalabilité, rapidité et modularité.

Chapitre 5 – Système de Prédiction : Conception et Mise en Œuvre

5.1 Prévision de la Demande Énergétique : Conception et Mise en Œuvre

5.1.1 Introduction

Cette section décrit la conception et la mise en œuvre d'un système de prévision de la demande énergétique pour le bâtiment 41, utilisant un ensemble de données intégrant la consommation d'énergie et les données météorologiques. L'objectif est de développer un modèle prédictif robuste pour anticiper la consommation énergétique, permettant une gestion efficace de l'énergie et une optimisation des coûts. Le système exploite des techniques d'apprentissage automatique et d'apprentissage profond, avec un accent sur l'analyse de séries temporelles pour capturer les motifs temporels et environnementaux.

5.1.2 Énoncé du Problème

La prévision précise de la demande énergétique est essentielle pour optimiser l'allocation des ressources, réduire les coûts et soutenir des pratiques énergétiques durables. Le défi réside dans la modélisation de l'interaction complexe entre la consommation d'énergie et les facteurs externes, tels que les conditions météorologiques et les motifs temporels. L'ensemble de données, comprenant 26 303 observations horaires de consommation d'énergie et de variables météorologiques, présente une variabilité significative due aux fluctuations saisonnières et quotidiennes, nécessitant des techniques de modélisation avancées pour obtenir des prédictions fiables.

5.1.3 Solution Proposée

La solution proposée consiste en un pipeline d'apprentissage automatique comprenant le prétraitement des données, l'ingénierie des caractéristiques, l'entraînement du modèle et l'évaluation. Le système utilise à la fois des modèles d'apprentissage automatique traditionnels (par exemple, Random Forest, XGBoost) et des modèles d'apprentissage profond (par exemple, LSTM) pour capturer les dépendances non linéaires et temporelles dans les données. Le pipeline garantit une gestion robuste des séries temporelles, avec un focus sur la scalabilité et l'adaptabilité pour une utilisation réelle.

5.1.3.1 Vue d'Ensemble du Système

Le système comprend plusieurs composants :

Ingestion des Données : Chargement et nettoyage de l'ensemble de données (merged_consumption_weather.csv).

Prétraitement : Gestion des valeurs manquantes, ingénierie des caractéristiques et mise à l'échelle des données.

Entraînement du Modèle : Entraînement de plusieurs modèles, incluant des méthodes d'apprentissage automatique traditionnelles et LSTM, avec un réglage des hyperparamètres.

Évaluation : Évaluation des performances du modèle à l'aide de métriques comme MAE, RMSE et R^2 .

Visualisation : Génération de graphiques pour visualiser les tendances, les performances du modèle et les prédictions.

Déploiement : Préparation du meilleur modèle pour son intégration dans des systèmes de gestion énergétique.

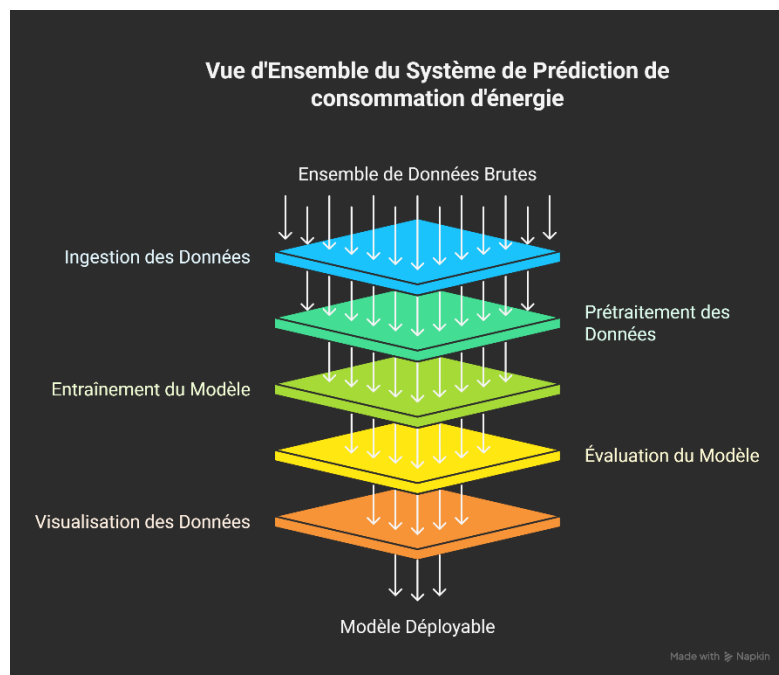


Figure 6. Vue d'Ensemble du Système de Prédiction de consommation d'énergie.

Les figures suivantes illustrent le flux de travail après la solution proposée :

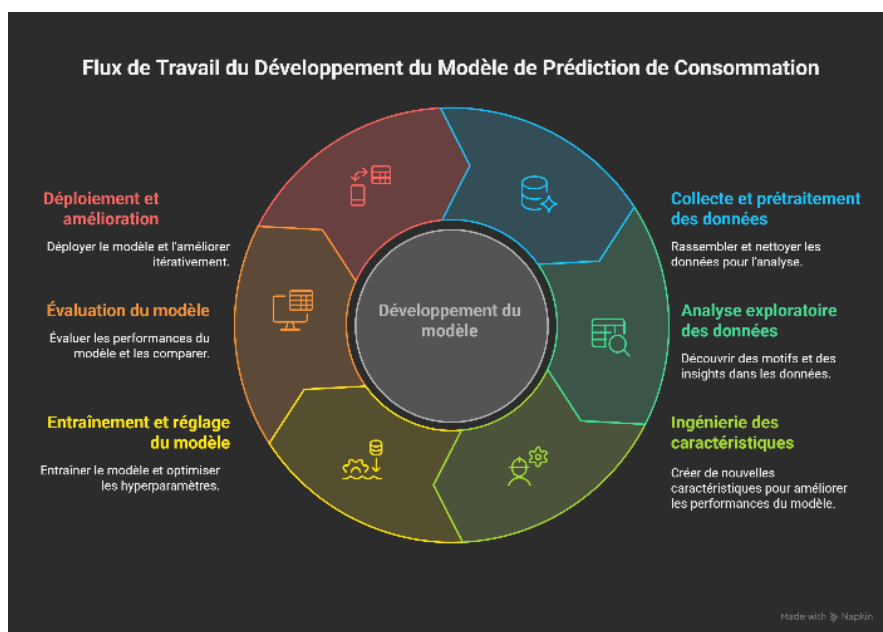


Figure 7: Flux de Travail du Développement du Modèle de Prédiction de Consommation d'Energie .

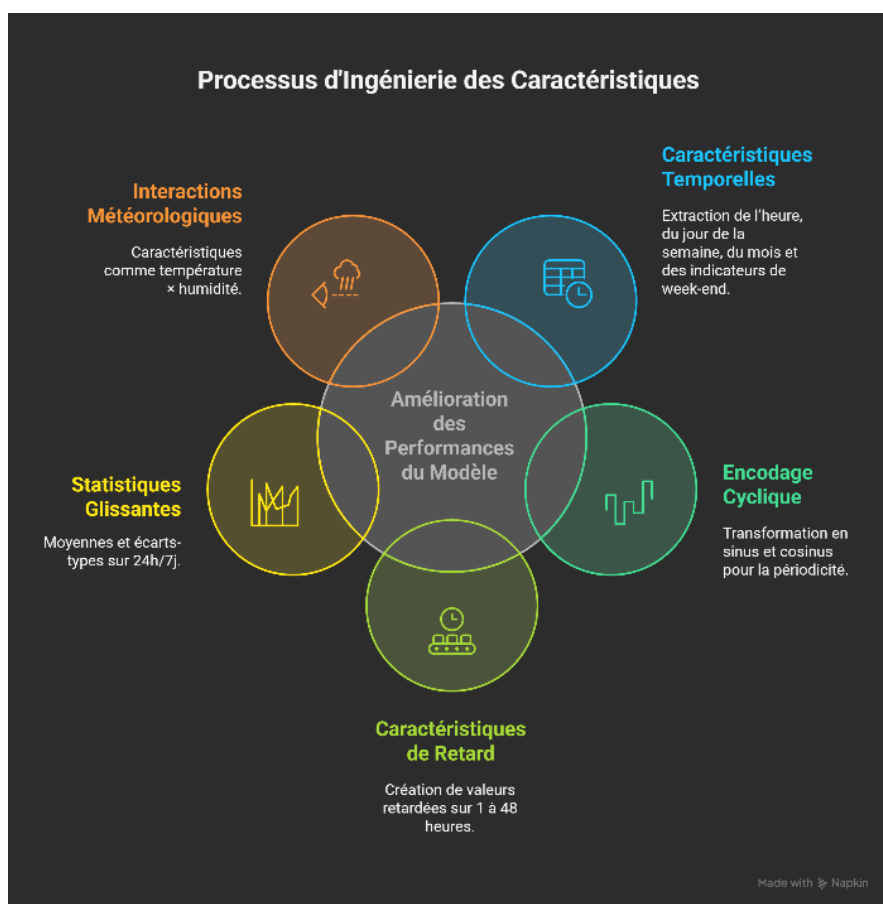


Figure 8: Module d'Ingénierie des Caractéristique .

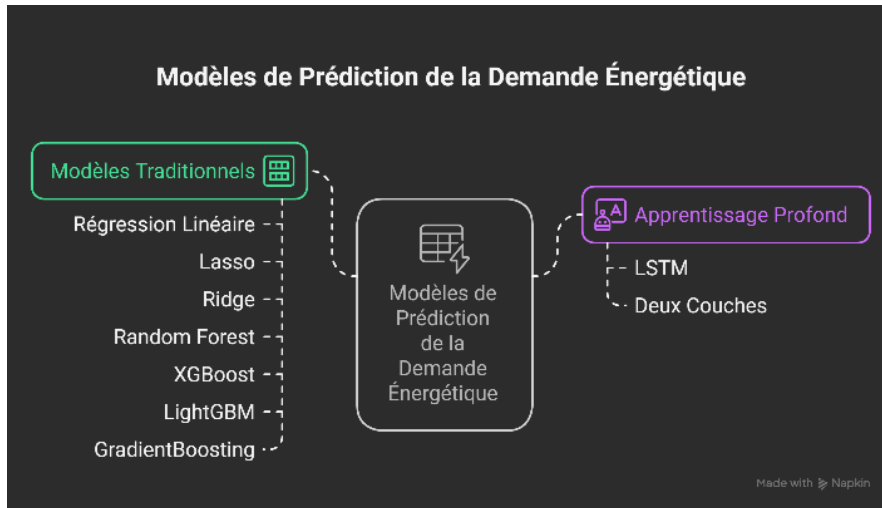


Figure 9: Module d'Entraînement des Modèles .

5.1.4 Paquets et Langage de Programmation

Le système est implémenté en Python, en utilisant les paquets suivants :

Pandas : Pour la manipulation et l'analyse des données.

NumPy : Pour les calculs numériques.

Scikit-learn : Pour les modèles d'apprentissage automatique traditionnels et les métriques d'évaluation.

TensorFlow/Keras : Pour la construction et l'entraînement du modèle LSTM.

Matplotlib/Seaborn : Pour la visualisation des données.

XGBoost/LightGBM : Pour des modélisations d'ensemble avancées.

Ces paquets assurent un traitement efficace des données, le développement du modèle et la visualisation, Python offrant un environnement flexible et largement supporté.

5.1.5 Carte de Développement du Modèle

Le processus de développement du modèle suit un pipeline structuré, illustré dans la Figure 7, qui décrit le flux de travail du traitement des données brutes au modèle final. Cette figure représente le pipeline complet du projet, de la collecte des données à l'amélioration itérative du modèle. Elle montre les étapes séquentielles et les boucles de rétroaction pour l'optimisation. Le processus comprend :

1. Collecte et prétraitement des données.
2. Analyse exploratoire des données pour comprendre les motifs.
3. Ingénierie des caractéristiques pour améliorer les performances du modèle.
4. Entraînement du modèle et réglage des hyperparamètres.
5. Évaluation et comparaison des performances du modèle.

6. Déploiement et amélioration itérative.

5.1.5.1 Collecte des Données

L'ensemble de données, stocké dans merged_consumption_weather.csv, contient 26 303 observations horaires de la consommation d'énergie pour le bâtiment 41 et des variables météorologiques associées. Les données incluent :

Consommation d'Énergie : Enregistrée dans la colonne building 41 (float64, probablement en kWh).

Caractéristiques Temporelles : Time (datetime), month, HH (heure), TD (différence de température ou point de rosée).

Caractéristiques Météorologiques : Temp (température, °C), RH (humidité relative, %), Q (rayonnement solaire), DR (durée de précipitation), FF (vitesse du vent), FX (vitesse maximale du vent), P (pression atmosphérique).

Aucune valeur manquante n'a été détectée. Le diagramme de séquence dans la Figure 10 illustre les étapes détaillées du prétraitement des données, y compris le nettoyage et la mise à l'échelle. Il met en évidence l'ordre des opérations pour préparer les données à l'entraînement. Les données couvrent environ trois ans, capturant des motifs saisonniers et quotidiens.

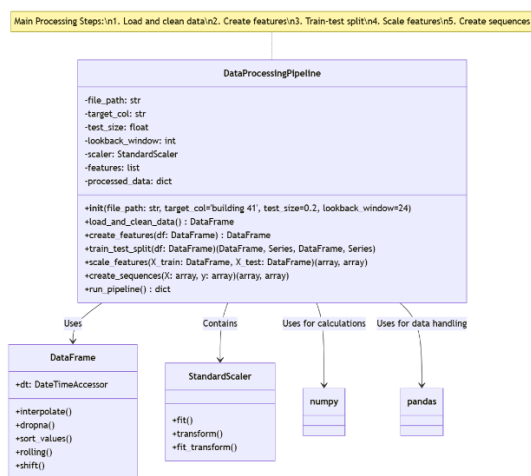


Figure 10: Diagramme de Séquence des étapes de prétraitement des données.

5.1.5.2 Analyse Exploratoire des Données (AED)

Une analyse exploratoire des données a été menée pour comprendre les caractéristiques de l'ensemble de données et identifier les motifs. Les principaux résultats incluent :

Résumé de l'Ensemble de Données : 26 303 entrées avec 13 colonnes, consommation moyenne de 25,69 kWh.

Valeurs Manquantes : Aucune valeur manquante détectée.

Motifs Saisonniers : Consommation plus élevée en été (31,16–32,74 kWh).

Variabilité : Plus élevée en été (6,24–6,60 kWh) qu'en hiver (2,30–2,71 kWh).

Motifs Horaires : Pic entre 10h et 14h (25–28 kWh), illustré dans la Figure 11, reflétant l'occupation diurne. Cette figure montre la variation moyenne de la consommation d'énergie horaire pour le bâtiment 41.

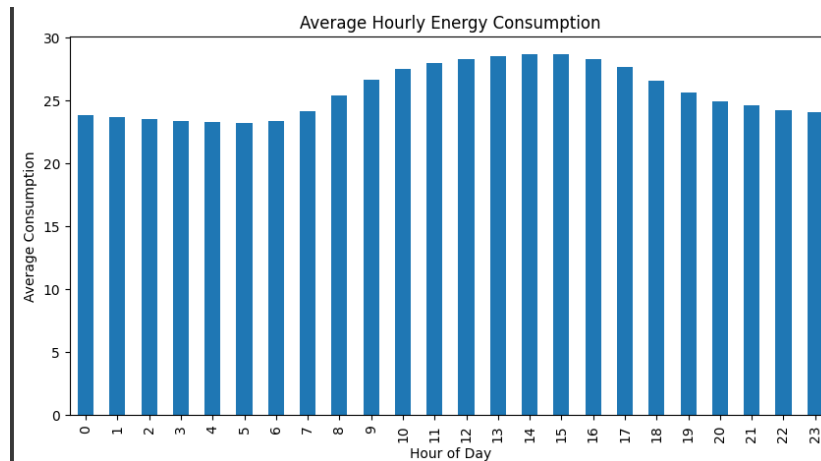


Figure 11. Consommation d'Énergie Horaire Moyenne pour le Bâtiment 41 - Montre les pics d'utilisation diurne, probablement dus à l'occupation et la climatisation.

Tendances à Long Terme : Pics quotidiens et saisonniers en été, comme montré dans la Figure 12. Cette figure présente les tendances de consommation moyenne sur trois ans.

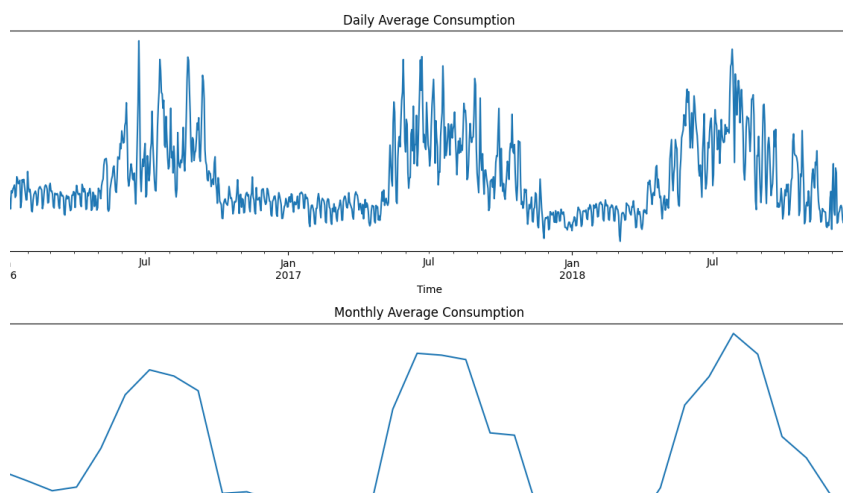


Figure 12. Consommation Moyenne Quotidienne et Mensuelle (2016–2018) - Illustre les tendances sur trois ans, avec des pics en été.

Corrélations : Fortes corrélations avec la température (0,79) et l'humidité (0,41),

visualisées dans la Figure 13. Cette carte de chaleur met en évidence les relations entre les variables.

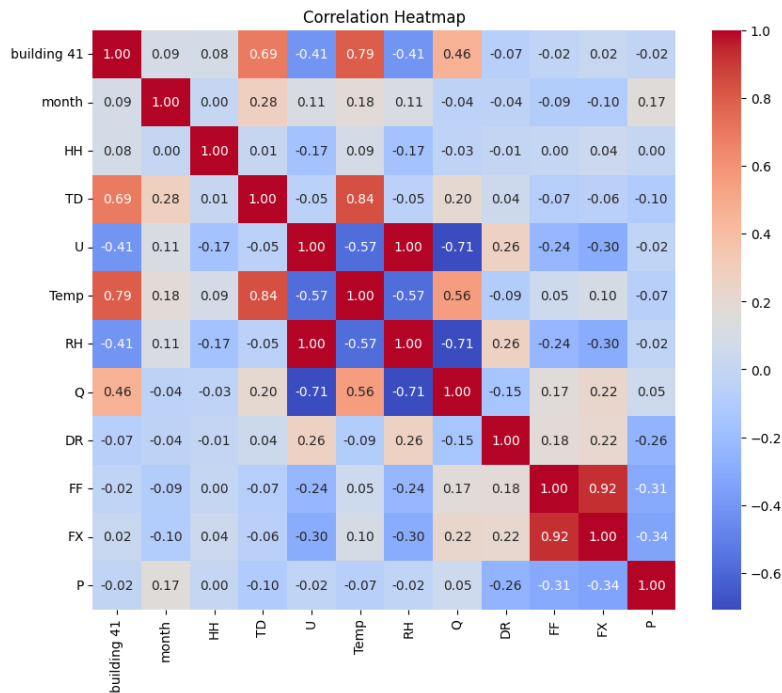


Figure 13. Carte de Chaleur des Corrélations des Caractéristiques - Montre les corrélations, notamment entre énergie et température.

Relation 3D : Consommation plus élevée à haute température/humidité, comme dans la Figure 14. Ce graphique en 3D illustre la relation entre température, humidité et énergie.

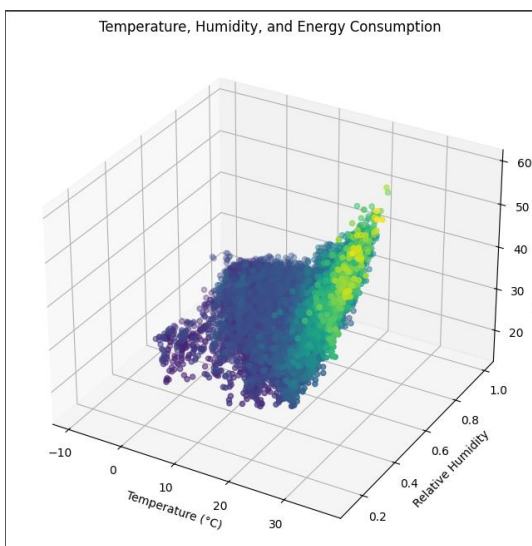


Figure 14. Graphique de Dispersion Température, Humidité et Consommation d'Énergie - Montre la relation tridimensionnelle avec des pics à haute température.

5.1.5.3 Ingénierie des Caractéristiques

Pour améliorer les performances du modèle, les caractéristiques suivantes ont été enrichi avec l'ingénierie des caractéristiques:

Caractéristiques Temporelles : Extraction de l'heure, du jour de la semaine, du mois et des indicateurs de week-end.

Encodage Cyclique : Transformation en sinus et cosinus pour la périodicité.

Caractéristiques de Retard : Création de valeurs retardées sur 1 à 48 heures.

Statistiques Glissantes : Moyennes et écarts-types sur 24h/7j.

Interactions Météorologiques : Caractéristiques comme température × humidité.

Ces caractéristiques ont été mises à l'échelle avec StandardScaler et séquencées (24h de regard en arrière) pour les modèles LSTM.

5.1.5.4 Entraînement du Modèle

Plusieurs modèles ont été entraînés pour prédire la demande énergétique :

Modèles Traditionnels : Régression Linéaire, Lasso, Ridge, Random Forest, etc.

Apprentissage Profond : LSTM avec deux couches, illustré dans la Figure 15. Cette figure montre la courbe de perte pendant l'entraînement du LSTM, indiquant une convergence efficace.

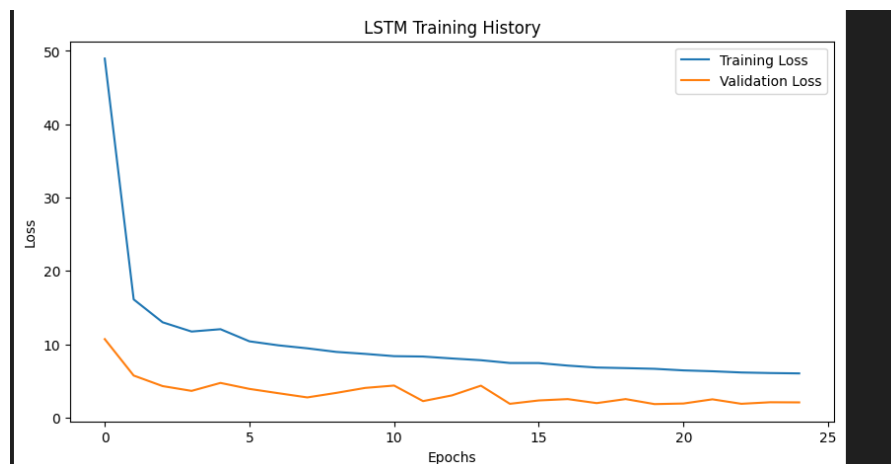


Figure 15. Historique d'Entraînement du LSTM - Présente la diminution de la perte au fil des époques d'entraînement.

Le diagramme de classes dans la Figure 16 illustre la structure d'implémentation. Cette figure montre les classes principales, leurs attributs et relations dans le système.

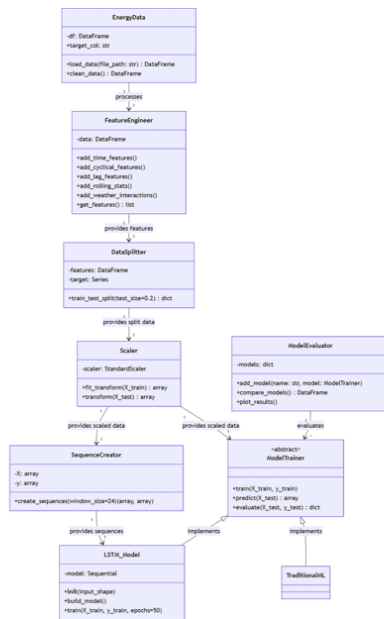


Figure 16: Diagramme de Classe.

Un réglage des hyperparamètres a été effectué avec RandomizedSearchCV.

5.1.5.5 Résultats d'Évaluation

Les performances ont été évaluées avec MAE, RMSE et R^2 , comme comparé dans la Figure 17. Cette figure compare les performances des modèles en termes de RMSE et R^2 , mettant en évidence les meilleurs modèles. Les visualisations incluent des graphiques en barres et actualisé vs prédit.

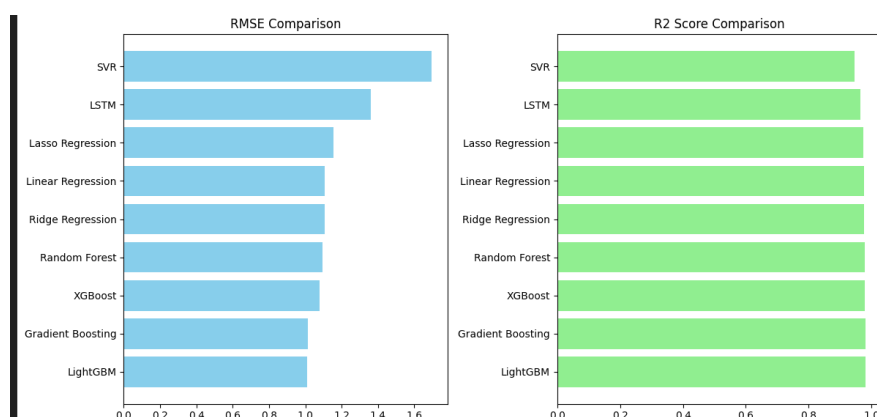


Figure 17. Comparaison de RMSE et R^2 entre les Modèles - Présente les métriques de performance pour différents modèles.

Les modèles les plus performants sont :

Tableau 13: Performances des Modèles de Prédiction pour consommation d'énergie.

Modèle	MAE	RMSE	R ²
LightGBM	0.676698	1.009965	0.980818
LSTM	0.987602	1.360910	0.964954

5.1.5.6 Déploiement du Modèle

Le meilleur modèle est prêt pour un déploiement en temps réel dans un système de gestion énergétique.

5.1.5.7 Amélioration du Modèle

Améliorations proposées : analyse d'importance des caractéristiques, réglage du LSTM, validation croisée, étude des motifs hivernaux.

5.1.6 Conclusion

La modélisation de la demande énergétique du bâtiment 41 a démontré l'efficacité d'un pipeline combinant prétraitement avancé, ingénierie des caractéristiques et apprentissage automatique. Les modèles testés – notamment LightGBM et LSTM – ont permis d'obtenir des prédictions précises, avec des métriques de performance élevées (RMSE < 1.4, R² > 0.96).

L'analyse exploratoire a mis en évidence des variations saisonnières et horaires significatives, fortement corrélées à des facteurs environnementaux tels que la température et l'humidité. L'intégration de ces variables dans le modèle a considérablement amélioré la précision des prévisions.

Ce cas d'usage valide l'hypothèse selon laquelle une modélisation fondée sur les données peut optimiser la gestion énergétique d'un bâtiment, anticiper les pics de consommation, et faciliter les stratégies de réponse à la demande.

Des pistes d'amélioration restent ouvertes, notamment :

- La prise en compte de comportements utilisateurs en temps réel,
- L'enrichissement du modèle LSTM par des techniques d'attention ou multi-head,
- Et l'intégration dans un outil décisionnel pour les gestionnaires de bâtiments.

5.2 Conception et Mise en Œuvre : Prévision des Pannes des Lampadaires

5.2.1 Introduction

Le développement d'un système efficace de prévision des pannes des lampadaires répond au besoin d'une maintenance proactive dans les infrastructures urbaines.

Cette section décrit la conception et la mise en œuvre d'une solution basée sur l'apprentissage automatique pour prédire les pannes des lampadaires, en s'appuyant sur un ensemble de données de métriques opérationnelles et environnementales. Le système vise à améliorer la fiabilité, réduire les coûts de maintenance et garantir la sécurité publique en identifiant les pannes potentielles avant qu'elles ne surviennent.

5.2.2 Énoncé du Problème

Les lampadaires sont essentiels à la sécurité et à la mobilité urbaine, mais les inspections manuelles pour détecter les pannes sont laborieuses et réactives. Les pannes, allant des défaillances électriques à la dégradation des ampoules, peuvent perturber le fonctionnement et présenter des risques pour la sécurité. Le défi consiste à développer un modèle prédictif qui identifie avec précision les occurrences de pannes (classification binaire) et les types de pannes (classification multiclasse) en utilisant des données opérationnelles, telles que la consommation d'énergie et les niveaux de tension, ainsi que des facteurs environnementales comme la température et les conditions météorologiques. L'ensemble de données présente un déséquilibre des classes, avec environ 70,3 % des instances sans panne, ce qui complique la détection des types de pannes rares.

5.2.3 Solution Proposée

La solution proposée est un pipeline d'apprentissage automatique qui traite les données des lampadaires, effectue une ingénierie des caractéristiques, entraîne des modèles prédictifs et évalue leurs performances. Le système utilise des algorithmes d'apprentissage supervisé, notamment Random Forest, XGBoost, Gradient Boosting et Régression Logistique, pour répondre aux tâches de classification binaire et multiclasse.

5.2.3.1 Vue d'Ensemble du Système

Le système comprend les composants suivants :

Collecte de Données : Un ensemble de 34 310 enregistrements, incluant des identifiants de lampadaires, des horodatages, la consommation d'énergie, les niveaux de tension, les fluctuations de courant, la température, les conditions environnementales et les types de pannes.

Analyse Exploratoire des Données (AED) : Résumés statistiques et visualisations pour comprendre les distributions et les relations des données.

Ingénierie des Caractéristiques : Création de caractéristiques temporelles, glissantes et dérivées pour améliorer les performances des modèles.

Entraînement des Modèles : Entraînement de plusieurs classificateurs avec des données prétraitées.

Évaluation : Évaluation des modèles à l'aide de l'exactitude, de la précision, du

rappel, du score F1 et de l'AUC-ROC.

Déploiement : Recommandations pour intégrer le modèle dans un système de surveillance en temps réel.

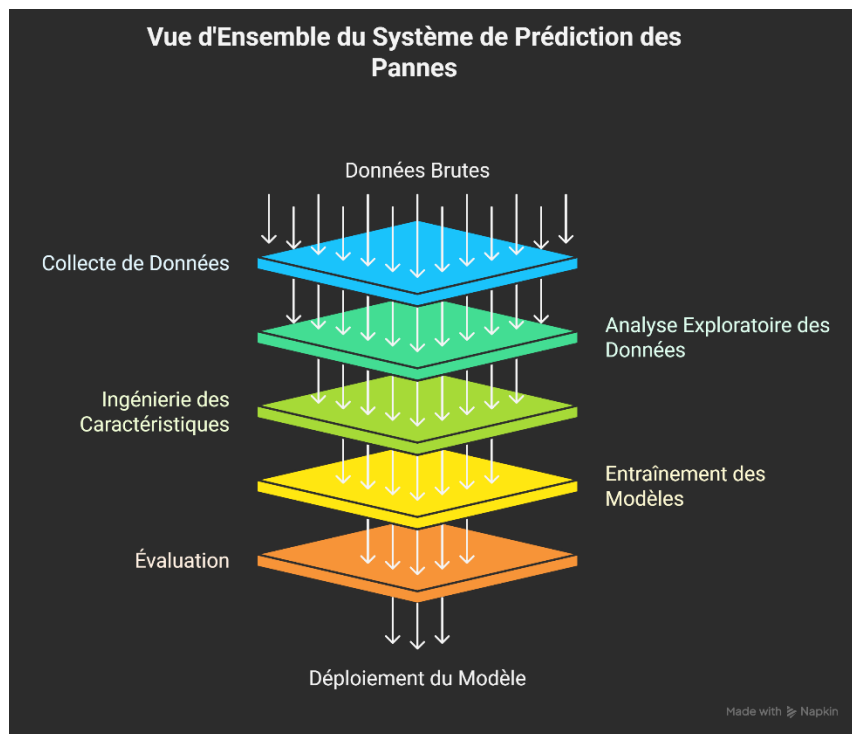


Figure 18: Vue d'Ensemble du Système de Prédiction des Pannes [W5].

Les figures suivantes illustrent le flux de travail après la solution proposée :

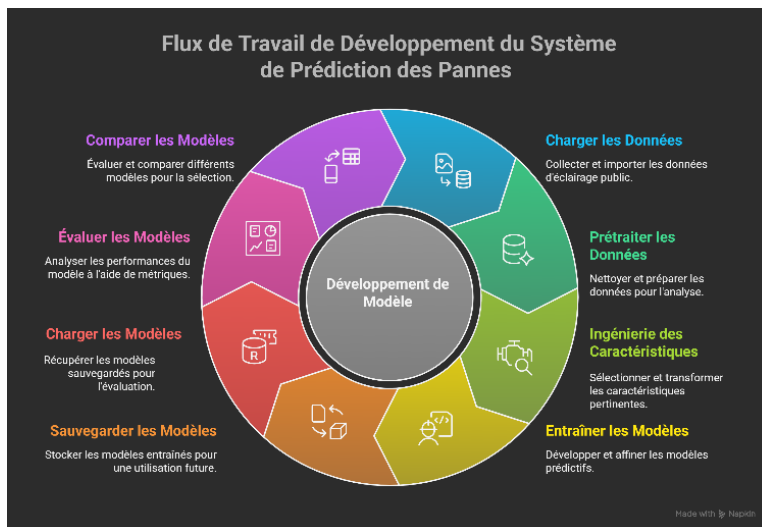


Figure 19: Flux de Travail de Développement du Système de Prédiction des Pannes .

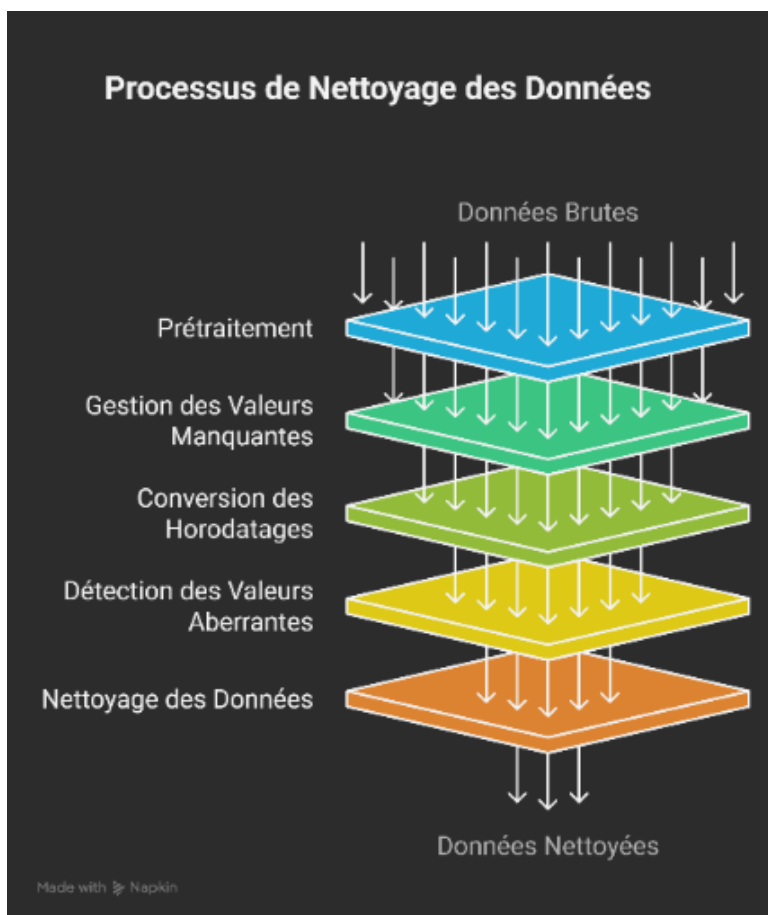


Figure 20: Module de Prétraitement des Données .

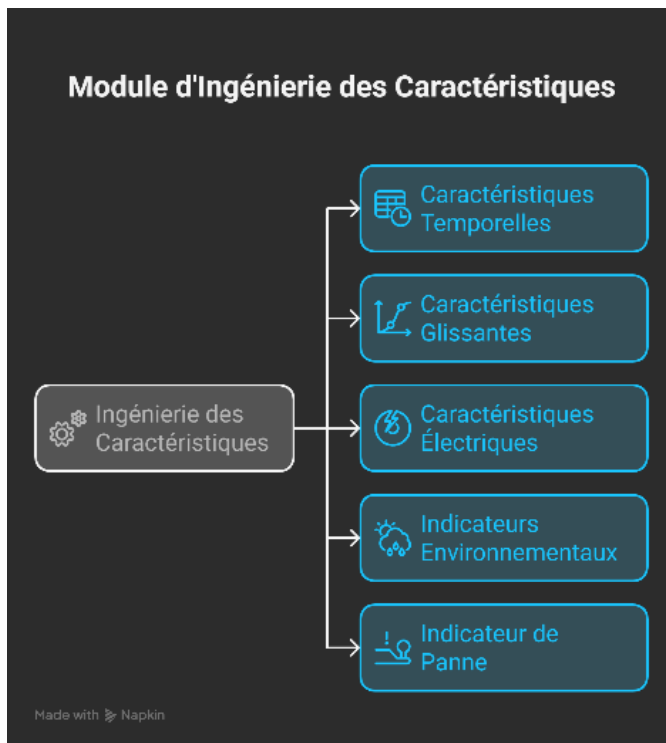


Figure 21: Module d'Ingénierie des Caractéristiques .

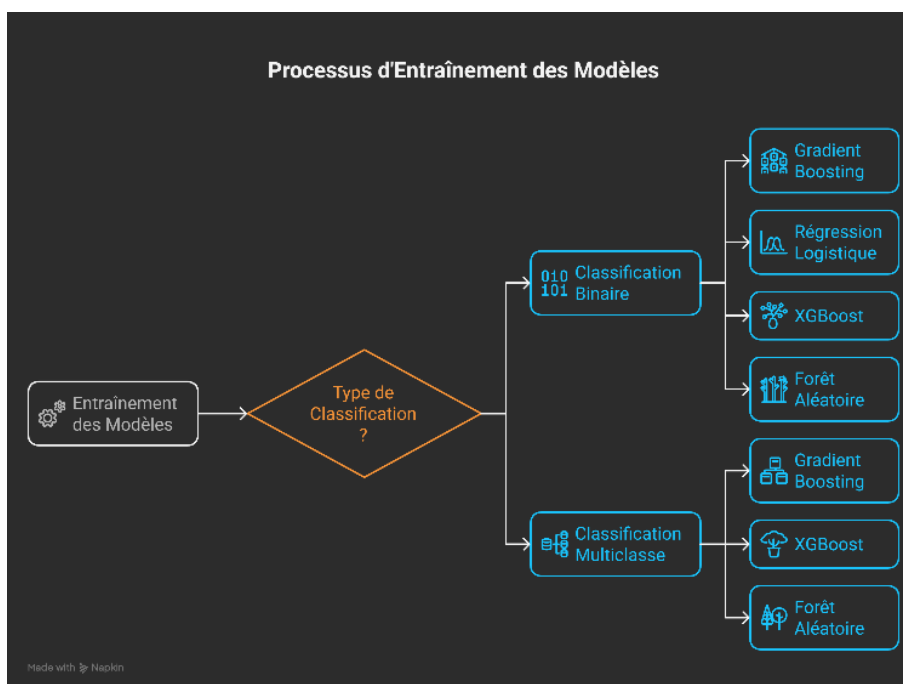


Figure 22: Module d'Entraînement des Modèles .

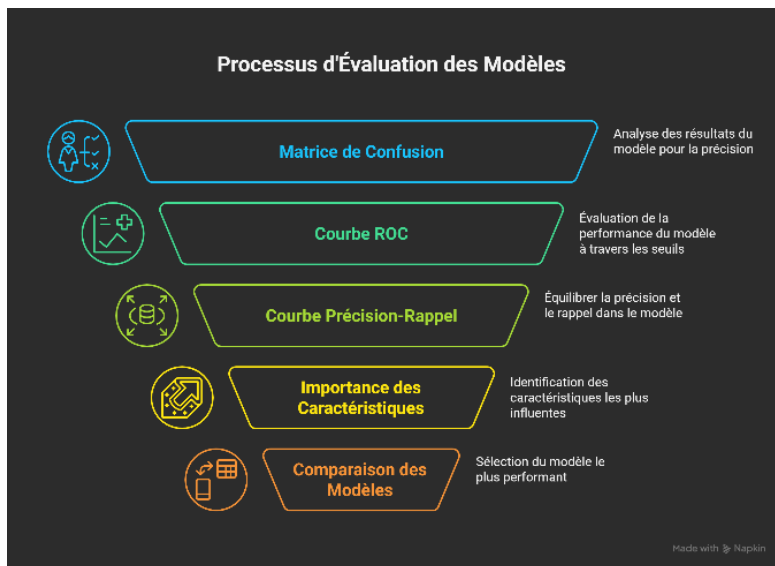


Figure 23: Module d'Evaluation .

5.2.4 Paquets et Langage de Programmation

Le système a été développé en Python 3, choisi pour ses vastes bibliothèques d'apprentissage automatique et ses capacités de traitement de données. Les principaux paquets incluent :

pandas et numpy : Manipulation des données et opérations numériques.

scikit-learn : Algorithmes d'apprentissage automatique et métriques d'évaluation.

xgboost : Cadre de boosting par gradient pour des performances optimisées.

matplotlib et seaborn : Visualisation des données.

joblib : Sérialisation des modèles pour le chargement des modèles pré-entraînés.

L'analyse a été réalisée dans des notebooks Jupyter, garantissant la reproductibilité et la modularité.

5.2.5 Carte de Développement du Modèle

Le processus de développement a suivi une approche structurée, détaillée ci-dessous.

5.2.5.1 Collecte de Données

L'ensemble de données, stocké sous `street_light_fault_prediction_dataset.csv`, contient 34 310 enregistrements avec 9 caractéristiques :

bulb_number : Entier (1–94).

timestamp : Chaîne de caractères (non analysée dans l'AED).

power_consumption (Watts) : Flottant (21,52–200).

voltage_levels (Volts) : Flottant (210–259,31).

current_fluctuations (Amperes) : Flottant (0–9,96).

temperature (Celsius) : Flottant (10–30).

environmental_conditions : Catégorique (Clair, Pluvieux, Nuageux).

current_fluctuations_env (Amperes) : Flottant (0–9,88).

fault_type : Entier (0 pour aucune panne, 1–4 pour les types de pannes).

Aucune valeur manquante n'a été trouvée, assurant l'intégrité des données.

5.2.5.2 Analyse Exploratoire des Données (AED)

L'AED a révélé des informations clés à travers des résumés statistiques et des visualisations :

Distribution des Types de Pannes : Environ 70,3 % des instances sont sans panne (24 111), avec des types de pannes 1 à 4 variant de 5,0 % (1 706) à 9,9 % (3 414), indiquant un déséquilibre des classes. Cela est illustré dans la Figure 24, montrant un décompte dominant pour le type 0 par rapport aux décomptes mineurs pour les types 1 à 4.

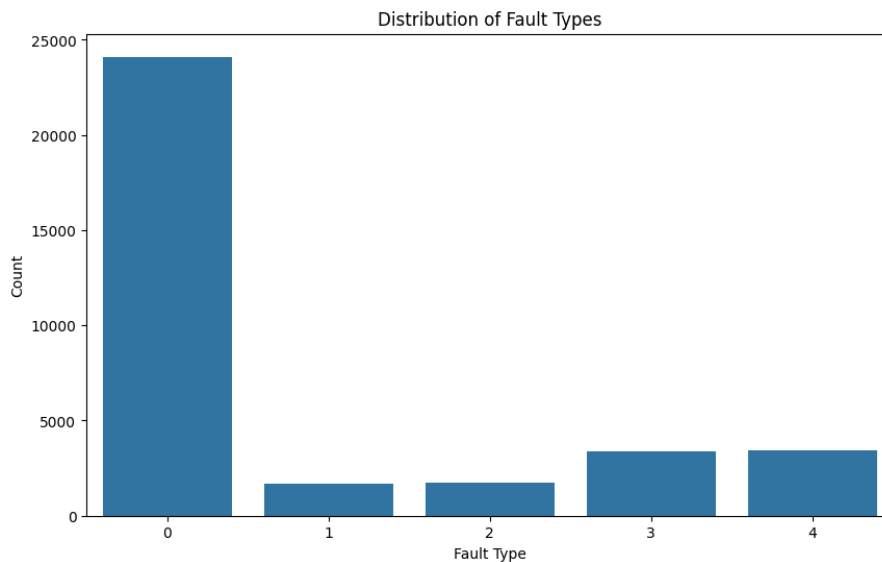


Figure 24. Distribution des Types de Pannes.

Taux de Pannes par Heure de la Journée : La Figure 25 illustre la probabilité de pannes sur 24 heures, avec un pic autour de 0,35 à l'heure 1 et des probabilités relativement stables (0,25–0,30) tout au long de la journée, suggérant une plus grande probabilité de pannes tôt le matin.

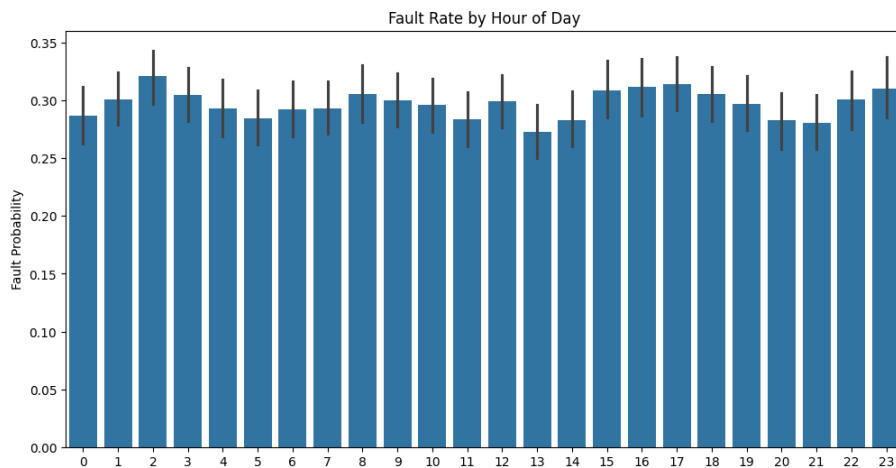


Figure 25. Taux de Pannes par Heure de la Journée.

Taux de Pannes Quotidiens sur le Temps : La Figure 26 montre les tendances du taux de pannes de janvier 2023 à janvier 2024, avec des fluctuations entre 0,25 et 0,45, et des pics notables à la mi-2023 et début 2024, indiquant des schémas saisonniers ou périodiques.

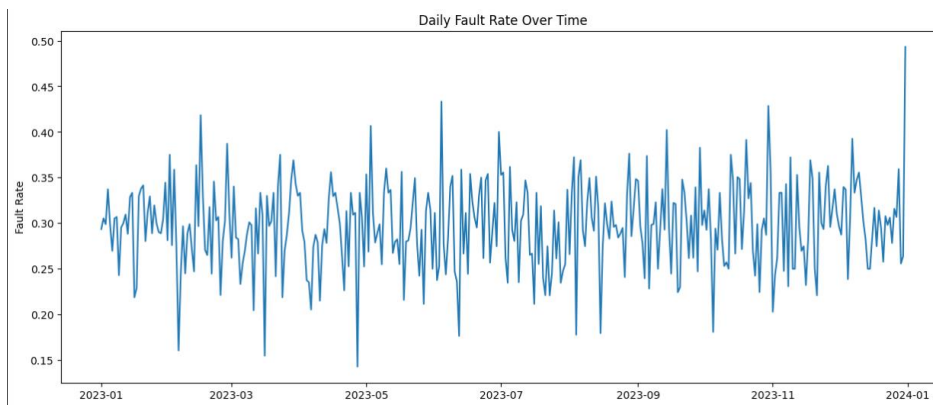


Figure 26. Taux de Pannes Quotidiens sur le Temps .

Distributions des Caractéristiques Numériques : La Figure 27 présente des histogrammes de power_consumption, voltage_levels, current_fluctuations et temperature. Power_consumption culmine autour de 125 Watts, voltage_levels se centre à 225–230 Volts, current_fluctuations est biaisé vers les valeurs basses (0–4 Ampères), et temperature suit une distribution normale autour de 20°C, reflétant des conditions environnementales stables.

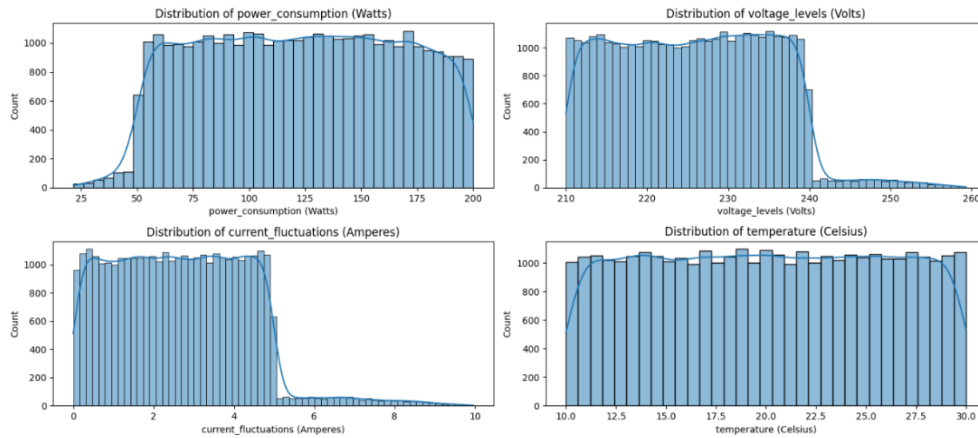


Figure 27. Distributions des Caractéristiques Numériques .

Conditions Environnementales : Presque équilibrées (Clair : 35,8 %, Pluvieux : 34,3 %, Nuageux : 29,8 %).

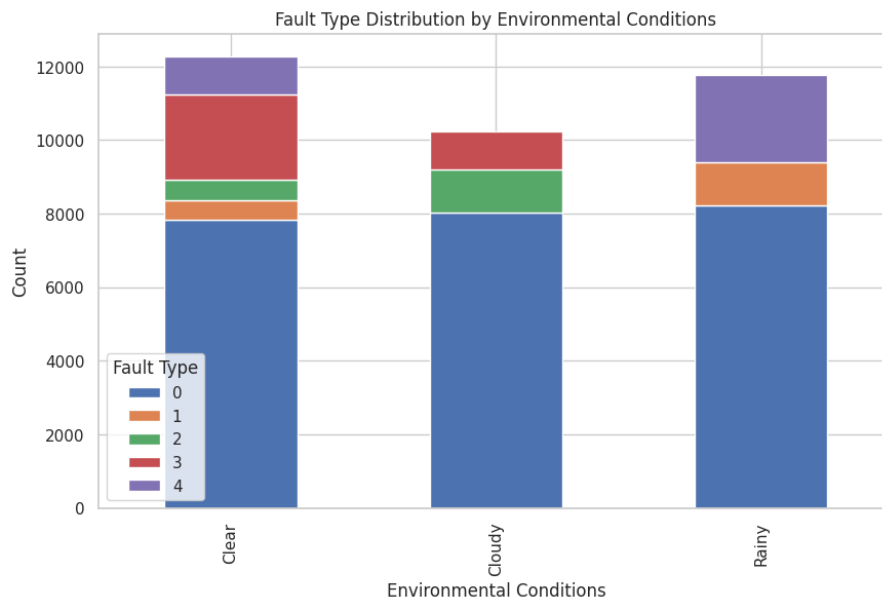


Figure 28: Distribution des types de faute par Condition Environnementale.

Analyse de Corrélation : Les caractéristiques numériques ont montré de faibles corrélations ($< 0,02$), suggérant une indépendance.

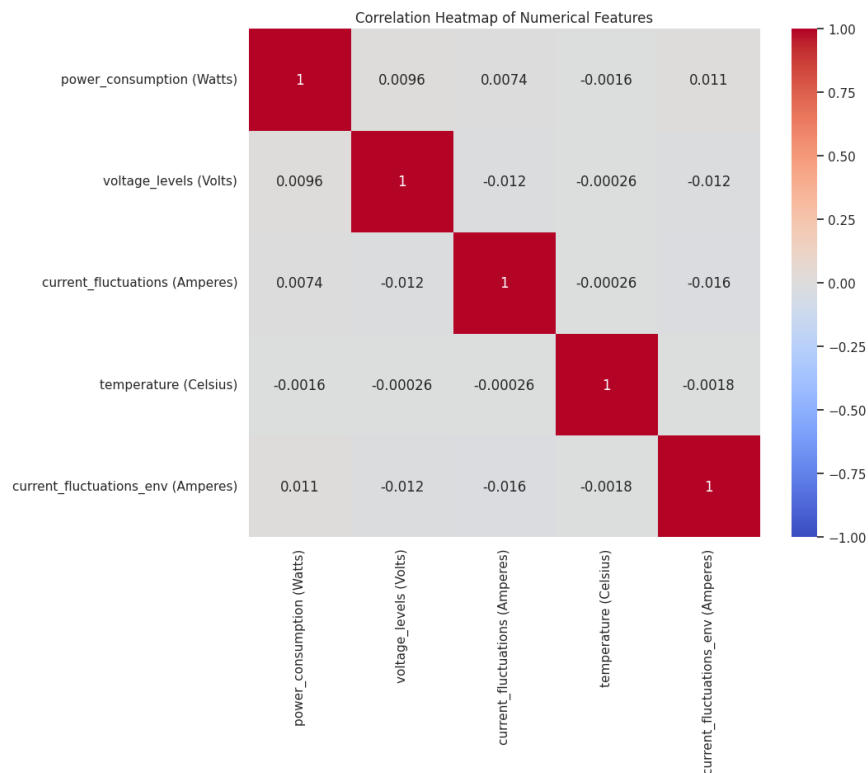


Figure 29: Corrélation des Caractéristiques Numériques .

5.2.5.3 Ingénierie des Caractéristiques

L'ingénierie des caractéristiques a enrichi l'ensemble de données avec 21 caractéristiques, incluant :

Caractéristiques Temporelles : Heure, jour de la semaine, mois et indicateur de week-end à partir de timestamp.

Statistiques Glissantes : Moyennes et écarts-types glissants sur 5 enregistrements pour power_consumption, voltage_levels, et current_fluctuations, regroupés par bulb_number.

Caractéristiques Dérivées : Rapport puissance-tension, déséquilibre de courant, jours depuis le dernier enregistrement, et indicateurs binaires pour les conditions environnementales (par ex., is_rainy).

Gestion des Valeurs Aberrantes : Limitation des caractéristiques numériques au 99e percentile.

Ces caractéristiques ont capturé des schémas temporels, des anomalies opérationnelles et des impacts environnementaux.

5.2.5.4 Entraînement des Modèles

Quatre modèles ont été entraînés pour la classification binaire (has_fault : 0 ou 1) et trois pour la classification multiclasse (fault_type : 0–4) :

— Random Forest : Ensemble d'arbres de décision, robuste au déséquilibre.

- XGBoost : Boosting par gradient avec performances optimisées.
 - Gradient Boosting : Implémentation standard du boosting par gradient.
 - Régression Logistique : Modèle linéaire uniquement pour la classification binaire.
- Les modèles ont été entraînés sur des données prétraitées, avec un prétraitement (par ex., mise à l'échelle, encodage) intégré dans des pipelines chargés via joblib.

5.2.5.5 Résultats d'Évaluation

Les performances des modèles ont été évaluées à l'aide de l'exactitude, de la précision, du rappel, du score F1 (pondéré) et de l'AUC-ROC.

Résultats de la Classification Binaire :

Tableau 14: Performances des Modèles de prédiction des pannes (classification binaire).

Modèle	Exactitude	Précision	Rappel	F1	AUC-ROC
Random Forest	0,968	0,970	0,968	0,968	0,989
XGBoost	0,867	0,885	0,867	0,854	0,927
Gradient Boosting	0,846	0,874	0,847	0,827	0,807
Régression Logistique	0,757	0,745	0,757	0,726	0,722

Observations Clés :

Random Forest a surpassé tous les modèles, atteignant des métriques presque parfaites (exactitude : 0,968, F1 : 0,967–0,968, AUC-ROC : 0,988–0,989).

XGBoost et Gradient Boosting ont montré des performances modérées, avec un rappel plus faible pour les types de pannes rares (par ex., type 3).

Régression Logistique était le plus faible, avec une détection de pannes médiocre (rappel : 0,33).

Le déséquilibre des classes a affecté le rappel pour les classes minoritaires, notamment les types de pannes 1 et 2.

Résultats de la Classification Multiclasse :

Tableau 15: Performances des Modèles de prédiction des pannes (classification multiclasse).

Modèle	Exactitude	Précision	Rappel	F1	AUC-ROC
Random Forest	0,968	0,970	0,968	0,967	0,988
XGBoost	0,873	0,892	0,873	0,851	0,967
Gradient Boosting	0,849	0,876	0,849	0,817	0,895

Observations clés :

Le modèle Random Forest est le plus robuste et fiable pour ce jeu de données, offrant des performances très homogènes et précises.

Bien que XGBoost reste performant, il est moins équilibré. Il peut être un bon choix si l'objectif est la généralisation, mais au détriment d'une légère perte de précision globale.

Gradient Boosting affiche des performances correctes mais nettement inférieures. Il pourrait être optimisé, mais n'est pas le meilleur choix en l'état actuel.

Comparaison de la performance des modèles :

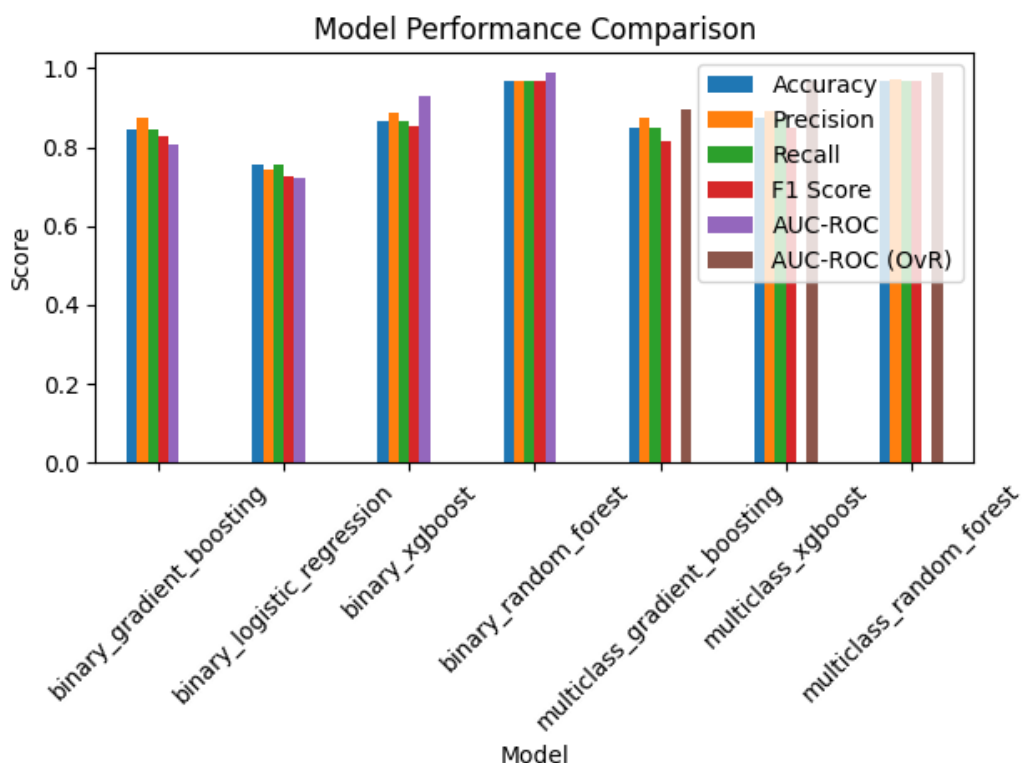


Figure 30: Comparaison des Performances (Modèles de prédiction des pannes).

5.2.5.6 Déploiement du Modèle

Pour le déploiement, le modèle Random Forest est recommandé en raison de ses performances supérieures. Les considérations clés incluent :

Prévision en Temps Réel : Intégrer le modèle dans un système de surveillance qui traite les données des capteurs en direct.

Scalabilité : Déployer sur des plateformes cloud (par ex., AWS, Google Cloud) avec des API pour l'ingestion de données.

Interprétabilité : Utiliser SHAP ou LIME pour fournir des prédictions interprétables aux équipes de maintenance.

Surveillance : Mettre en place une détection de dérive pour gérer les changements dans les distributions des données au fil du temps.

5.2.5.7 Amélioration du Modèle

Pour améliorer les performances, les améliorations suivantes sont proposées :

Gestion du Déséquilibre : Appliquer SMOTE ou des poids de classe pour améliorer le rappel des classes minoritaires.

Sélection des Caractéristiques : Utiliser les valeurs SHAP ou l'élimination récursive des caractéristiques pour prioriser les caractéristiques impactantes.

Ajustement des Hyperparamètres : Optimiser les paramètres du modèle à l'aide d'une recherche en grille.

Apprentissage Profond : Expérimenter avec des réseaux neuronaux pour la reconnaissance de schémas complexes.

Analyse des Erreurs : Étudier les échantillons mal classés pour affiner les caractéristiques ou les modèles.

5.2.6 Conclusion

Ce deuxième cas d'étude a porté sur la prédiction des pannes des lampadaires, problématique essentielle pour les villes intelligentes souhaitant optimiser leur maintenance préventive et réduire les coûts d'intervention.

En utilisant des données environnementales et opérationnelles, plusieurs modèles de classification ont été testés pour prédire les défaillances, tant en mode binaire (panne ou non) qu'en classification multiclassées (types de pannes). Les performances observées sont satisfaisantes, avec des scores encourageants, en particulier sur les classes dominantes.

Les résultats montrent que l'analyse conjointe de données historiques et de conditions externes (humidité, température, heure) permet d'anticiper des pannes récurrentes. Cela ouvre la voie à des interventions ciblées, réduisant les délais d'action et améliorant la sécurité urbaine.

Cependant, plusieurs améliorations sont envisageables :

- L'intégration de séries temporelles pour capter la dynamique des pannes,
- Le recours à des techniques de deep learning plus robustes pour les cas déséquilibrés,
- Et la collaboration avec les services techniques pour enrichir les données avec des annotations humaines.

En somme, cette étude prouve la faisabilité d'une approche prédictive simple mais efficace, transposable à d'autres équipements urbains dans un cadre Smart City.

Conclusion Générale

Ce mémoire a permis d'explorer de manière approfondie les enjeux liés à la gestion intelligente des données énergétiques dans les réseaux électriques intelligents (Smart Grids). Dans un contexte marqué par la transition énergétique, la montée en puissance des énergies renouvelables et la digitalisation des infrastructures, l'exploitation efficace de données massives, hétérogènes et en temps réel représente un levier stratégique pour renforcer la résilience, l'efficacité et la durabilité des systèmes électriques.

À travers une revue complète des fondements technologiques des Smart Grids et des techniques d'apprentissage automatique, nous avons identifié les défis principaux liés à l'indexation et à la prédiction de séries temporelles issues de capteurs. L'étude a permis de comparer plusieurs approches : des méthodes classiques (arbres B+, hachage, k-NN), jusqu'aux techniques avancées fondées sur l'apprentissage automatique (LSH, RMI, autoencodeurs), en évaluant leur pertinence selon la nature des données et les contraintes temps réel.

Un prototype fonctionnel a été conçu et testé, combinant :

- Une couche d'indexation intelligente (KD-Tree, Ball Tree, LSH) pour accélérer les requêtes sur des séries de données,
- Et une couche de prédiction automatique (Random Forest, XGBoost, LSTM) pour anticiper la consommation ou détecter les anomalies.

Les résultats expérimentaux montrent des gains significatifs en matière de précision, de latence et de robustesse, validant la faisabilité technique du système proposé.

Néanmoins, plusieurs limites subsistent :

- La dépendance à des données simulées, limitant la représentativité,
- L'absence de validation sur un environnement embarqué réel (edge computing),
- Et le besoin d'interprétabilité renforcée pour une adoption opérationnelle (ex. maintenance prédictive).

Perspectives

Plusieurs axes peuvent prolonger ce travail :

- Tests sur données réelles issues de bâtiments publics ou industriels,
- Intégration avec des plateformes d'IA embarquée (Edge AI),
- Déploiement dans une architecture distribuée avec gestion du scaling,
- Valorisation industrielle à travers un modèle SaaS ou embarqué dans un projet de startup.

Annexe

Annexe I : Projet Smart Grid IA __ Création du StartUp

Premièrement : le problème principal

Le projet **Smart Grid IA** serait motivé par un besoin critique : l'inefficacité actuelle dans la gestion de l'énergie des bâtiments et infrastructures. Les gestionnaires d'immeubles et responsables énergétiques peinent à collecter des données précises en temps réel et à prendre des décisions optimales pour réduire la consommation. Les systèmes existants sont souvent cloisonnés (chauffage, climatisation, éclairage, etc.) et ne communiquent pas entre eux, rendant difficile l'optimisation de la demande ou l'intégration de renouvelables. Cette fragmentation des solutions entrave la capacité à répondre aux objectifs de durabilité et de performance économique. En appliquant une démarche centrée utilisateur, on constaterait que le déficit d'une plateforme intelligente unifiée fait obstacle à la maîtrise des coûts énergétiques et à la réduction de l'empreinte carbone, malgré l'essor des technologies connectées. Par exemple, des études ont montré que des systèmes de gestion du bâtiment (BMS) bien conçus peuvent réduire la facture énergétique d'environ 30 %; l'absence de tels outils intelligents constitue donc un frein majeur.

Personas types

Pour cibler précisément les utilisateurs d' Smart Grid IA, plusieurs personas représentatifs pourraient être définis :

Client type Persona 1

	<u>Objectifs et ambitions</u> - Réduire fortement l'empreinte carbone du site industriel - Se conformer aux normes ISO 50001 - Démontrer l'engagement environnemental aux investisseurs ESG - Anticiper les obligations réglementaires à venir sur la performance énergétique	<u>Problèmes et difficultés</u> - Difficulté à surveiller les équipements énergivores - Reporting énergétique trop manuel et chronophage - Coûts élevés de l'énergie (gaz, électricité) sur site industriel
---	--	--

<u>Informations personnelles :</u>	<u>La vie quotidienne</u>	<u>Loisirs</u>
Nom : Samir Yasin Âge : 50 ans Profession : Directeur RSE d'un grand groupe industriel État civil : Marié – vit à Skikda	- Alterne entre le siège administratif et l'usine - Gère le reporting RSE trimestriel et supervise les indicateurs de performance - Analyse des rapports techniques et financiers pour planifier les investissements verts - Échange régulièrement avec d'autres directeurs lors de salons ou webinaires internationaux	- Golf le week-end - Lecture d'ouvrages sur la transition énergétique, les normes environnementales, ou la finance verte - Abonnement à des revues techniques internationales (environnement, industrie, IA)

Client type Persona 2

	<u>Objectifs et ambitions</u> - Réduire les factures d'énergie - Garantir le confort des usagers - Optimiser l'exploitation technique des bâtiments - Intégrer progressivement des solutions basées sur l'IA pour améliorer la performance énergétique	<u>Problèmes et difficultés</u> - Manque de visibilité sur la consommation en temps réel - Absence d'alertes automatiques en cas de dérive - Contraintes budgétaires limitant les investissements dans des outils modernes.
<u>Informations personnelles :</u> Nom : Ahmed Bourou Âge : 45 ans Profession : Gestionnaire technique d'immeuble État civil : Marié – vit à Alger	<u>La vie quotidienne</u> - Présent sur le terrain et au bureau - Suit les indicateurs techniques manuellement ou par tableurs - Communique souvent par téléphone avec les équipes techniques et administratives - Participe à des réunions internes pour justification des dépenses énergétiques.	<u>Loisirs</u> - Lecture de revues techniques spécialisées en bâtiment et efficacité énergétique - Participation à des conférences professionnelles (GTB, Smart Building, maintenance prédictive).

Annexe

Chaque persona aurait un ensemble de besoins et de frustrations propres, mais tous partageraient la nécessité d'une solution technologique avancée permettant d'agréger et d'analyser les données énergétiques pour orienter leurs décisions.

Carte d'empathie (pour Ahmed Bourou)

Une **carte d'empathie** permettra de visualiser ce qu'un utilisateur type (ici Ahmed) entend, voit, pense/ressent, dit/fait, ainsi que ses **peines** (peurs/frustrations) et **gains** (besoins/attentes) . Par exemple :

Catégorie	Contenu
Entend	- Des collègues parler de la hausse des factures d'énergie - Des échanges sur les nouvelles obligations réglementaires (labels, aides, décret Tertiaire)
Voit	- Des bâtiments voisins ou concurrents équipés de capteurs IoT et de solutions cloud - Des dashboards modernes dans des présentations techniques
Dit / Fait	- « Il faut réduire cette facture énergétique ! » - Réalise des audits ponctuels - Consulte les factures manuellement - Échange avec des fournisseurs d'énergie - Rencontre des intégrateurs techniques - Lit des cas d'étude de bâtiments performants
Pense / Ressent	- Ressent du stress face à l'augmentation continue des coûts - Se sent frustré de ne pas recevoir d'alertes automatiques - Souhaite mieux maîtriser les dépenses pour convaincre sa direction
Peines (Pains)	- Informations de consommation dispersées entre différents systèmes - Chauffage/climatisation non centralisés - Crainte de sanctions ou reproches en cas de dépassement de seuils
Gains (Besoins / Attentes)	- Espère rendre les bâtiments plus économes - Souhaiterait valoriser les économies réalisées auprès de ses supérieurs - A besoin d'un outil visuel, clair, avec alertes - Attente d'un bon ROI pour justifier l'investissement

Cette carte d'empathie illustre qu'Ahmed valoriserait une solution qui lui simplifie la collecte et l'analyse des données, en lui apportant confiance (gains) et en éliminant les frustrations actuelles (peines) .

Parcours client (Customer Journey)

Le **parcours client** d'Ahmed, dans le contexte de l'adoption d' Smart Grid IA, s'articulerait en cinq étapes clés.

Annexe

Étape	Déclencheur / Contexte	Actions	Critères / Attentes	Résultat attendu
1. Prise de conscience	Hausse de facture, nouvelles obligations réglementaires, benchmarking avec bâtiments voisins	Consulte ses factures, analyse des écarts, s'informe via ses pairs ou des rapports	Comprendre l'origine de la surconsommation Identifier les lacunes de son système actuel	Réalise qu'il a besoin d'une solution plus avancée qu'une GTB de base
2. Réflexion	Besoin d'amélioration, veille technologique (webinaires, articles, retours d'expérience d'autres gestionnaires)	Recherche sur Internet, compare capteurs IoT, logiciels IA, GTB avancées ; échange avec confrères et experts	Simplicité d'usage Compatibilité avec l'existant Fiabilité des prestataires ROI rapide	Établit une shortlist de solutions potentielles
3. Décision	Résultats de la comparaison, retours positifs sur la solution retenue	Analyse des offres finales, vérifie les garanties de performance, lit les CGV, consulte le service client	Clarté du contrat SaaS Suivi post-installation Capacité d'adaptation aux besoins spécifiques	Choisit un prestataire et valide le budget
4. Implémentation	Validation interne du projet, contractualisation avec le fournisseur	Lance les démarches administratives (budget, appel d'offres), suit l'installation, participe à la formation sur la solution	Déploiement rapide Phase pilote maîtrisée Configuration des alertes et connectivité	Déploiement réussi sur un premier bâtiment
5. Post-implémentation	Système en fonctionnement	Suit les dashboards, vérifie les économies et le confort, échange avec le support pour ajustements, donne du feedback	Mesurer les gains réels Facilité de pilotage au quotidien Stabilité et fiabilité de la plateforme	Amélioration mesurable → Possible extension à d'autres sites

Cette trajectoire décrit comment le décideur parcourt les phases **prise de conscience, réflexion, décision, déploiement et suivi**, chaque étape reflétant ses besoins (piloter le budget, fiabilité du service, support) et ses interactions (informations reçues, contacts avec le fournisseur, formations, etc.).

Analyse du marché mondial et algérien

Marché mondial

Le marché mondial de la gestion énergétique intelligente est en forte croissance. Selon *Mordor Intelligence*, la taille du marché global des systèmes de gestion de l'énergie serait estimée à environ 55,03 milliards USD en 2024 et pourrait presque doubler d'ici 2029 (≈113,85 milliards USD, TCAC ~15,65 %). D'autres études plus spécifiques aux bâtiments signalent un marché évalué à 14,7 milliards USD en 2024 pour les BEMS (Building Energy Management Systems), avec un taux de croissance annuel projeté autour de 7,5 % sur 10 ans. Ce dynamisme s'explique par plusieurs tendances : l'essor des objets connectés (IoT) et du cloud qui facilitent la surveillance à distance, la montée rapide des prix de l'énergie poussant les entreprises à rechercher des économies, ainsi que l'intégration croissante de sources renouvelables (solaire, éolien) qui nécessite des stratégies de stockage et de réponse à la demande. Les politiques publiques jouent également un rôle clé (réglementations *LEED*, *Future Homes Standard*, plans d'économie d'énergie, etc.) en imposant des normes et en subventionnant ces technologies. En termes de géographie, l'Amérique du Nord et l'Europe dominent encore ce marché grâce à des réglementations strictes et à la présence d'acteurs industriels majeurs, tandis que l'Asie-Pacifique affiche le taux de croissance le plus élevé du fait de son urbanisation rapide et de ses programmes de villes intelligentes.

Marché algérien

En Algérie, bien que la transition énergétique soit encore à ses débuts, plusieurs signes montrent un intérêt croissant pour les solutions de gestion énergétique intelligente. Des initiatives internationales soutiennent la modernisation des systèmes de suivi énergétique (par exemple, une délégation de l'UE travaille avec le ministère de l'Énergie algérien pour moderniser le système d'information énergétique national). Un projet de la GIZ vise aussi à améliorer la gestion de l'énergie dans les communes pilotes en introduisant des outils logiciels d'évaluation des mesures d'efficacité et EnR. Parallèlement, le secteur de la construction incorpore de plus en plus des concepts de « bâtiments intelligents ». Team France Export souligne que les bâtiments intelligents se développent en Algérie grâce à l'intégration de solutions de gestion énergétique, y compris dans les infrastructures critiques (hôpitaux, stades, tunnels autoroutiers, etc.). Cette tendance laisse entrevoir une demande future pour des technologies comme Smart Grid IA, portées par la nécessité d'optimiser les installations existantes (CVC éco-efficaces, pompes à chaleur connectées) et

d'anticiper la croissance urbaine. Cependant, le marché algérien reste encore fragmenté et jeune, confronté à des barrières (coûts initiaux élevés, rareté de compétences spécialisées, incitations financières limitées). La perspective de résidences universitaires ou de parcs d'affaires verts, combinée à la volonté gouvernementale de diversifier les sources d'énergie, laisse supposer un potentiel de développement favorable dans les prochaines années.

Matrice SWOT adaptée à Smart Grid IA

Catégorie	Contenu
Forces (Strengths)	- Utilisation de technologies avancées : edge computing, IA, IoT pour une gestion énergétique en temps réel - Architecture modulaire et SaaS facilitant un déploiement rapide et flexible sans investissements lourds - Positionnement fort sur la durabilité, compatible avec les objectifs de transition énergétique - Potentiel de subventions et partenariats publics grâce à l'impact écologique - Équipe technique spécialisée en intelligence énergétique, renforçant la crédibilité du projet
Faiblesses (Weaknesses)	- Faible notoriété initiale et absence de références clients, freinant la signature des premiers contrats - Coûts élevés de R&D, maintenance technique et scalabilité de la plateforme - Risques techniques (bugs, intégration complexe avec des équipements existants) - Forte dépendance à la connectivité réseau et à la qualité des données issues des capteurs (précision, fiabilité) - Une panne ou une perte de données pourrait affecter la qualité du service rendu
Opportunités (Opportunities)	- Marché mondial en croissance rapide porté par la transition énergétique et la digitalisation des bâtiments (CAGR à deux chiffres) - Contexte algérien favorable : soutien des bailleurs (UE, GIZ), montée des exigences en matière d'efficacité énergétique - Opportunités de partenariats avec intégrateurs GTB, énergéticiens, bureaux d'étude - Potentiel d'extension géographique : Maghreb / MENA - Ajout de fonctionnalités IA avancées (machine learning prédictif, conformité réglementaire automatisée) pouvant renforcer l'attractivité du service
Menaces (Threats)	- Concurrence importante de groupes historiques : Schneider Electric, Siemens, ABB, Honeywell, Johnson Controls, etc. - Ces grands acteurs peuvent rapidement intégrer des fonctionnalités similaires avec leurs ressources et leur réputation

	- Risques économiques et politiques locaux ralentissant les projets d'efficacité énergétique (budgets publics gelés, instabilité) - Éventuels changements réglementaires défavorables ou exigences transfrontalières contraignantes - Réserves sur l'IA en entreprise (protection des données, crainte de l'automatisation) nécessitant des actions pédagogiques et de transparence
--	---

Analyse de la concurrence

Type de concurrents	Nom / Acteur	Offre / Caractéristiques	Forces	Faiblesses	Différenciation d' Smart Grid IA
Concurrents directs	BrainBox AI (Canada)	Plateforme IA autonome pour CVC (chauffage, ventilation, climatisation)	- IA temps réel - Jusqu'à 25–40 % de gains énergétiques - Ciblage tertiaire performant	- Focalisé uniquement sur le CVC- Moins modulaire - Intégration potentiellement complexe	Offre multi-équipements (CVC, éclairage, recharge, etc.) Architecture unifiée et évolutive
	AGRID Copilot (France)	Solution clé en main, sans nouveaux capteurs, installation rapide (<1 jour)	- Installation ultra-rapide- Pas de capteurs supplémentaires requis	- Données limitées - Moins personnalisable - Moins adapté à des bâtiments complexes	Plus grande richesse de donnéesModèle prédictif plus avancéPlus grande adaptabilité
	Schneider, Siemens, Legrand	Systèmes GTB numériques avancés, souvent fermés et centrés sur l'infrastructure physique	- Réputation forte - Réseau mondial d'intégration - Solutions stables	- Systèmes propriétaires - Faible composante IA prédictive - Complexité d'intégration	Ouverture, IA native, solution SaaS plus agile
Concurrents indirects	Nest, Tado, autres thermostats intelligents	Produits grand public pour pilotage local du chauffage	- Très accessibles - Installation simple - Prix abordables	- Fonctionnalité limitée - Pas de vue multi-site - Pas de machine learning ou d'analyse avancée	Plateforme IA multisite, optimisation fine, scénarios complexes
	GTB traditionnelles (legacy)	Outils de gestion énergétique classiques sans IA, souvent liés à de gros équipements	- Présents dans de nombreux bâtiments - Peu coûteux à maintenir	- Obsolètes, pas de prédiction - Faible ergonomie - Données peu exploitables	IA prédictive + tableau de bord intelligent + alertes en temps réel

Annexe

	Cabinets de conseil énergie	Audits ponctuels, diagnostics manuels	- Expertise humaine - Recommandations personnalisées	- Non automatisé - Intervention limitée dans le temps - Pas d'analyse continue	Suivi permanent + machine learning adaptatif + visualisation automatisée des économies
	Opérateurs énergétiques (DRMS)	Offres à venir de gestion de la demande énergétique (flexibilité réseau)	- Accès aux données réseau - Intégration avec les mécanismes tarifaires	- Centrés réseau, peu adaptés à la gestion fine bâtiment - Peu personnalisables à l'échelle individuelle	Adapté aux besoins bâtimentaires + possibilité de couplage à la demande réseau

Stratégie produit/service

Le produit proposé serait une **solution clé en main** IoT + IA, livrée sous forme de box centrale plus capteurs et plateforme cloud. Ses principales caractéristiques seraient :

Élément	Description
Matériel IoT	- Capteurs sans fil : consommation électrique, température, CO₂, occupation des locaux, etc. - Reliés à une box centrale avec microcontrôleur - Boîtiers concentrateurs pour la collecte locale des données avant envoi au cloud.
Logiciel IA (Cloud/SaaS)	- Plateforme d'analyse hébergée en cloud - Ingestion des données en temps réel - Algorithmes avancés : indexation temporelle, réseaux LSTM, auto-encodeurs - Prédiction de consommation, détection d'anomalies, optimisation des équipements
Fonctionnalités clés	- Alertes en temps réel sur dérives de consommation ou pannes - Tableau de bord interactif multi-site- Rapports périodiques et recommandations personnalisées
Usage et mises à jour	- Interface intuitive, accessible sur web et mobile (bilingue FR / AR) - Maintenance et mises à jour logicielles à distance - Modularité des offres : • Pack résidentiel basique (suivi simple) • Pack tertiaire complet (multi-zones, services premium, support 24/7)

Cette stratégie produit mettrait l'accent sur l'innovation technologique (IA, big data), l'éco-conception (réduction des consommations et du carbone) et l'ergonomie

Annexe

(installation facile, interface intuitive), afin de constituer une offre différenciée et attrayante sur un marché en croissance .

Stratégie marketing et communication

La communication de Smart Grid viserait à convaincre les décideurs par **des preuves concrètes** et des messages ciblés. Les piliers stratégiques seraient :

Axe stratégique	Description
Marketing de contenu & digital	- Publications sur blogs spécialisés et infographies explicatives - Webinars B2B sur LinkedIn (valorisation de l'IA pour l'énergie) - Campagnes e-mail ciblées vers bailleurs sociaux et gestionnaires tertiaires
Promotion d'image & relations publiques	- Partenariats avec institutions publiques (ADEME, FNME, APRUE) et énergéticiens (Sonelgaz) - Participation à des projets pilotes subventionnés - Présence sur salons professionnels (énergie, bâtiments intelligents) et conférences universitaires - Communiqués de presse axés innovation verte et efficacité
Message différenciant	- Double bénéfice mis en avant : 1) Économies réelles d'énergie (20–40 %) 2) Réduction d'empreinte carbone (mesurée) - Mise en valeur des obligations réglementaires (décret tertiaire, performance énergétique) - Ton expert, étayé de preuves terrain et chiffres concrets
Ciblage multicanal	- Communication bilingue (FR/AR) pour s'adapter au marché algérien - Utilisation de réseaux locaux : universités, associations professionnelles, chambres du bâtiment - Supports pédagogiques vulgarisés : vidéos explicatives, témoignages clients, brochures illustrées - Sensibilisation large au smart building via des formats courts (vidéos, animations)

En résumé, la stratégie marketing combinerait du **digital B2B** (Inbound marketing, réseaux sociaux) et des **relations publiques sectorielles** (partenariats institutionnels, salons) pour bâtir la crédibilité d' Smart Grid IA.

Stratégie de distribution

La distribution viserait principalement un modèle **B2B2C**. Le produit serait vendu à travers plusieurs canaux complémentaires :

Canal de distribution	Description	Cible principale
-----------------------	-------------	------------------

Annexe

Contrats directs avec syndicats / entreprises	- Accords-cadres avec bailleurs sociaux, copropriétés, groupes immobiliers - Déploiement à l'échelle de plusieurs bâtiments en une seule opération - Intégration possible de clauses de performance énergétique dans les contrats	Gestionnaires de patrimoine immobilier (tertiaire / résidentiel collectif)
Appels d'offres publics et marchés tertiaires	- Réponse aux marchés publics (administrations, écoles, hôpitaux, stades, etc.) - Projets pilotes ou vitrines pour démontrer la performance de la solution	Collectivités locales, organismes publics, ministères
Partenariats technologiques	- Distribution indirecte via des partenaires locaux déjà implantés (électriciens, sociétés de maintenance, fournisseurs de capteurs IoT) - Formation et support commercial assurés par Smart Grid IA	PME du bâtiment, intégrateurs, installateurs certifiés
Plateforme SaaS en ligne	- Souscription directe via un portail web sécurisé - Accès immédiat à la plateforme cloud et à ses fonctionnalités de base - Mises à jour logicielles automatiques et support distant inclus	Petites structures autonomes, clients multisites connectés

En somme, la stratégie de distribution mêlerait vente directe (force commerciale ciblée) et indirecte (partenaires), avec un point d'entrée « clé en main » pour les clients peu technophiles (pack installation + service complet).

Stratégie publicitaire et promotionnelle

La promotion de la solution s'appuierait sur des offres et outils spécifiques pour encourager l'adoption :

Axe promotionnel	Description	Objectif visé
Pilotes gratuits ou subventionnés	Proposer des démonstrateurs cofinancés ou gratuits dans des bâtiments publics ou copropriétés pilotes (ex. écoles, hôpitaux)	Générer des retours d'expérience concrets et des cas de référence
Garanties de performance	Mettre en avant des contrats avec engagements (ex. « jusqu'à 30 % d'économies garanties ») pour sécuriser le ROI des clients	Rassurer les prospects sur la rentabilité et la fiabilité

Annexe

Publicité ciblée	Campagnes LinkedIn/Facebook orientées B2B avec témoignages clients Publications dans revues professionnelles (énergie, immobilier, GTB, etc.)	Accroître la notoriété et générer des leads qualifiés
Événements et démonstrations	Organisation de webinars techniques interactifs Présence sur des salons pro avec démonstrateurs physiques (box + dashboard IA) Possibilité de mini-ateliers en ligne ou en présentiel	Montrer la solution en action, convaincre par la preuve
Promotions de lancement	Réductions temporaires sur l'abonnement SaaS Intégration offerte aux premiers clients dans un programme pilote labellisé (bâtiment vert)	Encourager l'adoption rapide, gagner des clients ambassadeurs
Storytelling & communication humaine	Valorisation d'utilisateurs réels (gestionnaires, occupants) par des interviews vidéo Mise en avant des bénéfices concrets (confort, sécurité, fierté d'action écoresponsable)	Créer un lien émotionnel, humaniser la technologie

Cette stratégie publicitaire jouerait sur la démonstration de valeur ajoutée par la preuve empirique et sur le storytelling (parcours client) pour convaincre à la fois la raison et l'émotion des décideurs.

Stratégie de vente

Le modèle commercial envisagé serait flexible et orienté B2B/B2B2C :

Axe stratégique	Description	Objectif visé
Abonnements modulables	- Forfait de base mensuel (plateforme + box + X capteurs inclus) - Exemple : 100 €/mois + 5 €/capteur/mois pour un bâtiment tertiaire (à ajuster selon le segment) - Facturation possible par kWh géré (offre énergétique basée sur usage)	Flexibilité tarifaire selon la taille et la typologie du site
Contrats de performance énergétique	- Paiement indexé partiellement sur les économies réalisées (ex. partage des gains) - Adapté aux clients publics ou privés réticents au CAPEX initial	Réduction des freins financiers à l'entrée
Offres packagées	- Packs complets incluant matériel, installation, accès SaaS et support	Séduction de plusieurs segments avec une offre évolutive

Annexe

	- Exemple : Pack Standard (suivi simple), Pack Premium (analyses prédictives, vision multi-site) - Possibilité de modèle freemium : accès gratuit limité, options payantes avancées	
Segmentation par client	- Canaux différenciés : • Résidentiel via promoteurs, installateurs ou syndicats • Tertiaire via commerciaux, appels d'offres, institutions - Offres adaptées aux besoins spécifiques : confort résidentiel vs pilotage multi-zone tertiaire	Optimisation de la stratégie de vente par type de client
Soutien à l'aide financière	- Accompagnement dans la recherche de subventions (CEE, crédits d'impôt, dispositifs locaux) - Valorisation des avantages réglementaires (ex. conformité décret tertiaire) - Argumentaire ROI intégrant les aides dans les simulations de coût global	Lever les barrières budgétaires et renforcer l'attractivité

Cette stratégie de vente s'appuierait donc sur un tarif transparent lié à l'usage, un accompagnement au ROI, et une segmentation claire des offres selon le profil du client.

Plan de production et d'organisation

La mise en œuvre du projet suivrait un calendrier progressif, avec un équilibre entre R&D, tests et montée en charge :

Phase	Période	Objectifs clés	Activités principales
1. Prototypage	Mois 0 à 6	Concevoir une première version fonctionnelle du système	- Développement logiciel de base (collecte des données, prévision simple) - Installation de capteurs sur 2 sites pilotes (1 résidentiel, 1 tertiaire) - Mise en place des infrastructures cloud - Test technique de la chaîne de données
2. Version 1.0	Mois 6 à 12	Ajouter les fonctionnalités avancées et améliorer l'expérience utilisateur	- Intégration des modules de : • Budget énergétique • Détection d'anomalies • Interface utilisateur optimisée UX/UI- Collecte

Annexe

			de retours utilisateurs pilotes
3. Pilote terrain étendu	Mois 12 à 18	Tester la solution en conditions réelles dans plusieurs contextes locaux	- Déploiement sur 2 villes cibles (ex. zones climatiques ou réglementaires différentes) - Suivi des indicateurs de performance (économies réalisées, confort, robustesse du réseau IoT) - Réglages selon feedback
4. Industrialisation	À partir de M18	Passer à l'échelle et structurer l'entreprise pour le déploiement massif	- Recrutement (équipes commerciales, support technique, production) - Mise en place des processus (SLA, logistique capteurs, documentation client) - Renforcement des outils de gestion

Sur le plan organisationnel, l'équipe projet serait pluridisciplinaire :

Domaine	Rôles clés	Fonctions / Missions
Développement algorithmique	- Data scientists- Ingénieurs IA	- Conception du moteur prédictif (LSTM, auto-encodeurs) - Détection d'anomalies et optimisation énergétique
Conception hardware (IoT)	- Ingénieurs IoT / systèmes embarqués- Électroniciens	- Développement de la box et des capteurs intégrés - Validation technique et certification matérielle
Développement logiciel	- Développeurs back-end / cloud - Développeurs front-end UX/UI- Designer UX	- Déploiement SaaS de la plateforme - Création du tableau de bord utilisateur interactif et multilingue
Relation client & support	- Techniciens support / installation - Chargés de relation client	- Accompagnement à l'installation - Suivi technique et formation utilisateur
Commercial & marketing	- Commerciaux B2B - Responsable marketing digital	- Vente de la solution auprès des bailleurs et collectivités - Animation des canaux LinkedIn, salons, RP
Recherche & innovation	- Partenariats avec laboratoires de recherche	- Renforcement de la R&D appliquée

Annexe

	- Coopérations avec incubateurs	- Veille technologique et expérimentation sur le terrain
Fournisseurs & logistique	- Accord avec fabricants de capteurs - Intégrateurs GTB / installateurs	- Sécurisation de la chaîne d'approvisionnement - Déploiement rapide sur site client
Infrastructure technique	- Externalisation via AWS / cloud équivalent	- Stockage des données, traitement cloud scalable, haute disponibilité
Gestion de projet	- Méthodologie agile (Sprints, outils collaboratifs) - Indicateurs mensuels de performance projet (KPIs techniques, commerciaux, satisfaction)	- Organisation itérative pour optimiser les fonctionnalités - Ajustement continu du calendrier et des priorités

Prévision financière

Le plan financier prévisionnel détaille les coûts initiaux, les revenus attendus et la trajectoire de rentabilité.

Hypothèse : parc pilote de 5 bâtiments publics (écoles, mairies, hôpitaux)

1. Coûts d'équipement initial par bâtiment

Poste	Estimation unitaire	Coût total (5 bâtiments)
Capteurs IoT (50 unités)	~2 500 DA / capteur	~625 000 DA
Boîtier central + passerelle	~75 000 DA	~375 000 DA
Installation locale	~1 000 DA / capteur	~250 000 DA
Total / bâtiment	~250 000 – 300 000 DA	~1,5 million DA

2. Développement logiciel & IA (MVP)

Élément	Montant estimatif
Dév. logiciel IA + interface web	~1,5 million DA
Hébergement cloud basique (6 mois)	~150 000 DA
ests + support technique initial	~200 000 DA
Total	~1,85 million DA

Annexe

3. Coûts de fonctionnement annuels (OPEX)

Poste	Estimation annuelle
Hébergement / serveurs	~300 000 DA
Support client / hotline	~150 000 DA
Communication (site, plaquettes)	~200 000 DA
Total fonctionnement	~650 000 DA/an

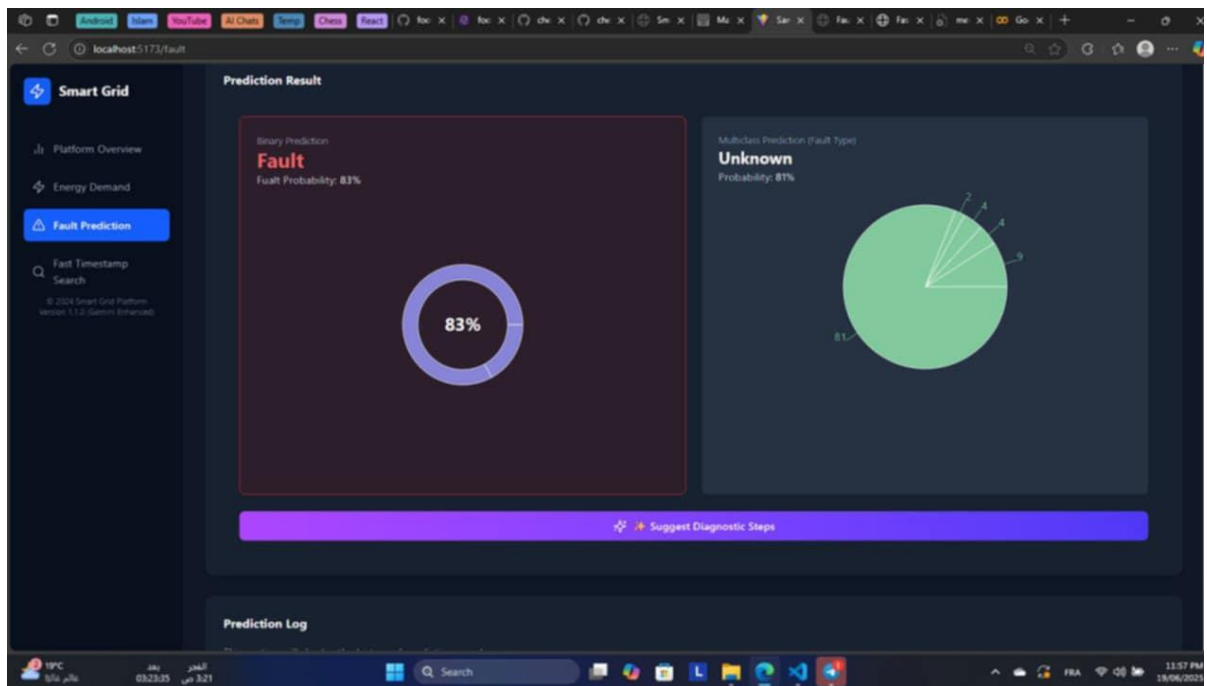
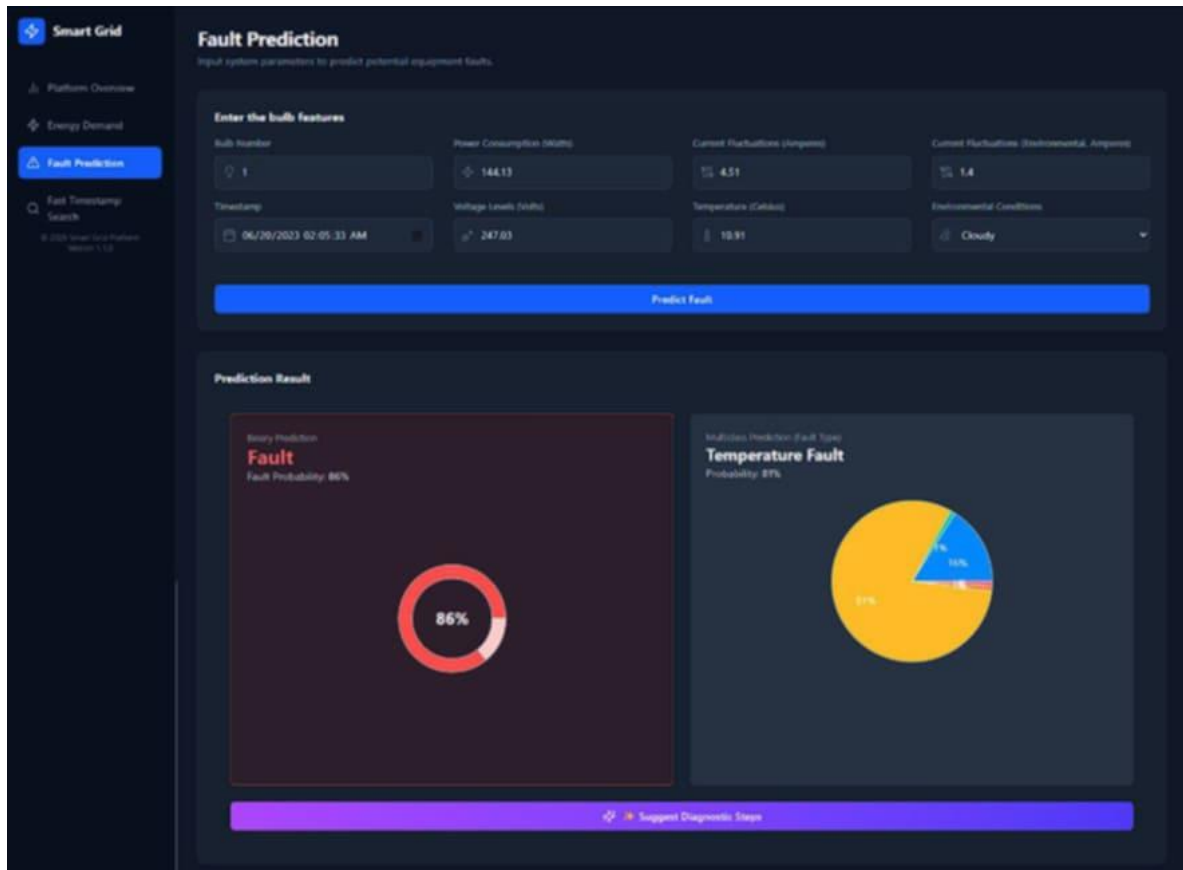
4. Revenus attendus (année 1)

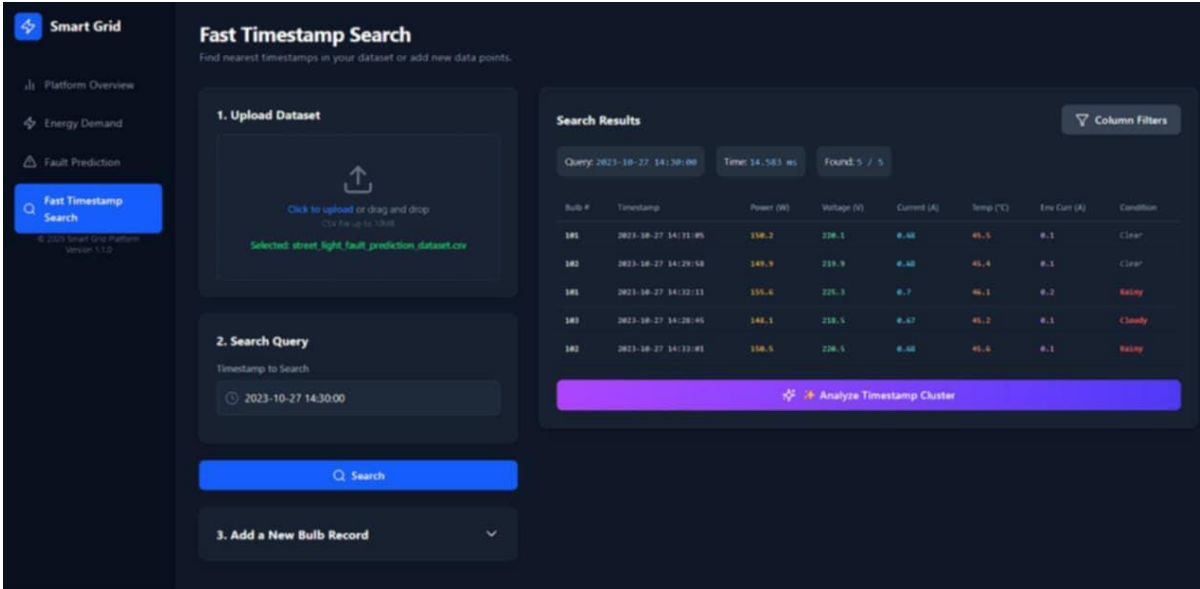
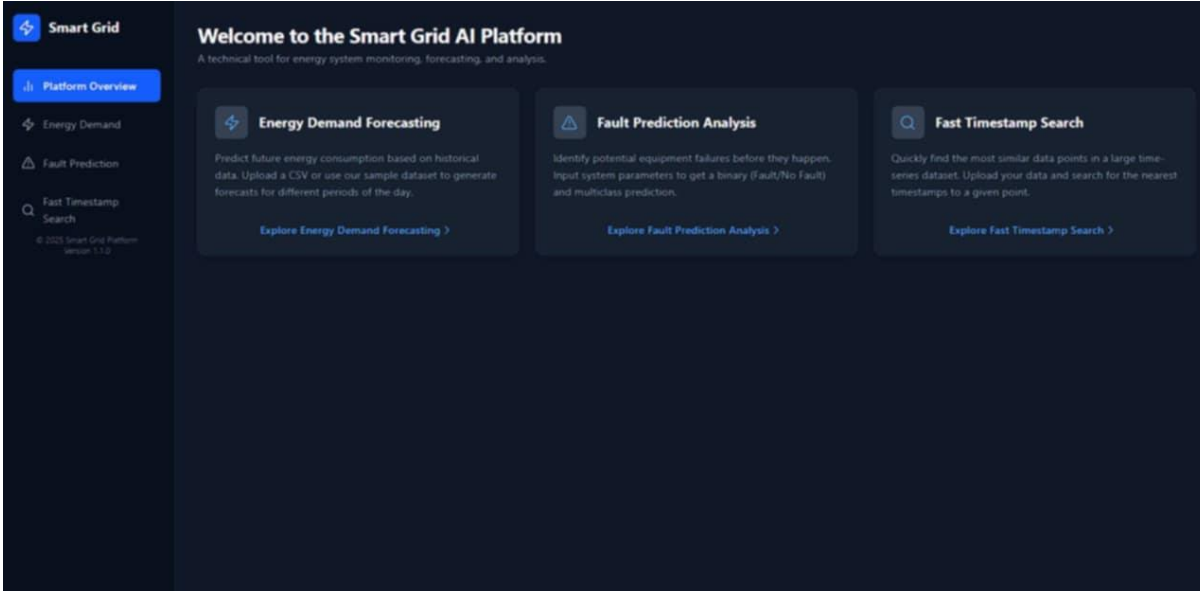
Poste	Tarif unitaire	Revenu estimé (5 bâtiments)
Abonnement par bâtiment	~15 000 DA / mois	900 000 DA / an
Complément par capteur (50 x 5)	~400 DA / capteur / mois	1,2 million DA / an
Total annuel brut	~2,1 millions DA	

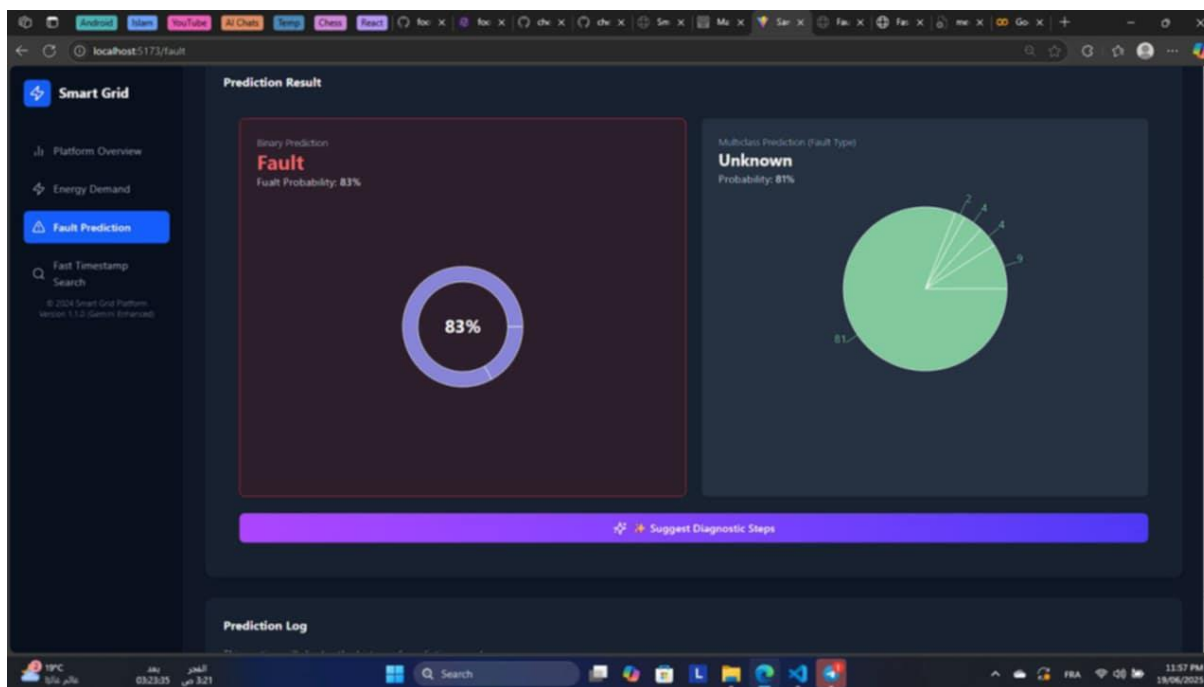
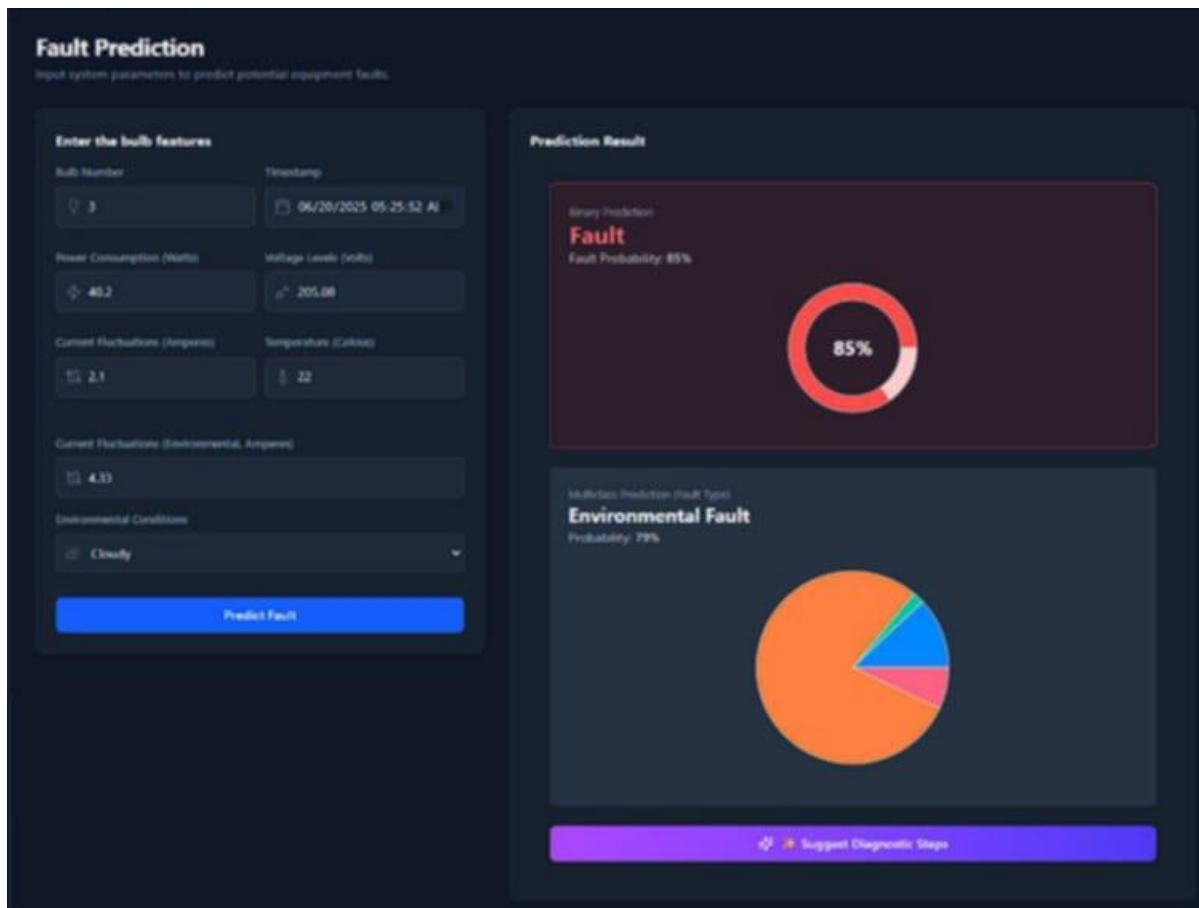
5. Résultat financier prévisionnel

Année	Revenus	Coûts totaux	Résultat
Année 1	~2,1 M DA	~4,0 M DA	Perte de lancement (~1,9 M DA)
Année 2	~4,0 M DA	~2,5 M DA	Équilibre / légère marge
Année 3	~6,0 M DA	~3,0 M DA	Rentabilité nette (~3 M DA)

Annexe II : Interface Utilisateur







Search Results

Query: 2023-10-27 14:30:00 Time: 14:30:00

Bulb #	Timestamp	Power (W)
101	2023-10-27 14:31:05	150.0
102	2023-10-27 14:29:58	149.0
101	2023-10-27 14:32:11	155.0
103	2023-10-27 14:28:45	148.0
102	2023-10-27 14:33:01	150.0

Column Filters

Bulb Number

Contains...

Environmental Condition

Contains...

Power (W)

Min

Max

Voltage (V)

Min

Max

Temperature (°C)

Min

Max

Reset

Apply Filters

Fault Prediction

Input system parameters to predict potential equipment faults.

Enter the bulb features

Bulb Number

101

Timestamp

06/28/2025 05:25:52 AM

Power Consumption (Watts)

120

Voltage Levels (Volts)

230

Current Fluctuations (Ampere)

55

Temperature (Celsius)

25

Current Fluctuations (Environmental, Anger)

3

Environmental Conditions

Rainy

Predict Fault

Prediction Result

Binary Prediction

Fault

Fault Probability: 88%

88%

Multiclass Prediction (Fault Type)

Electrical Fault

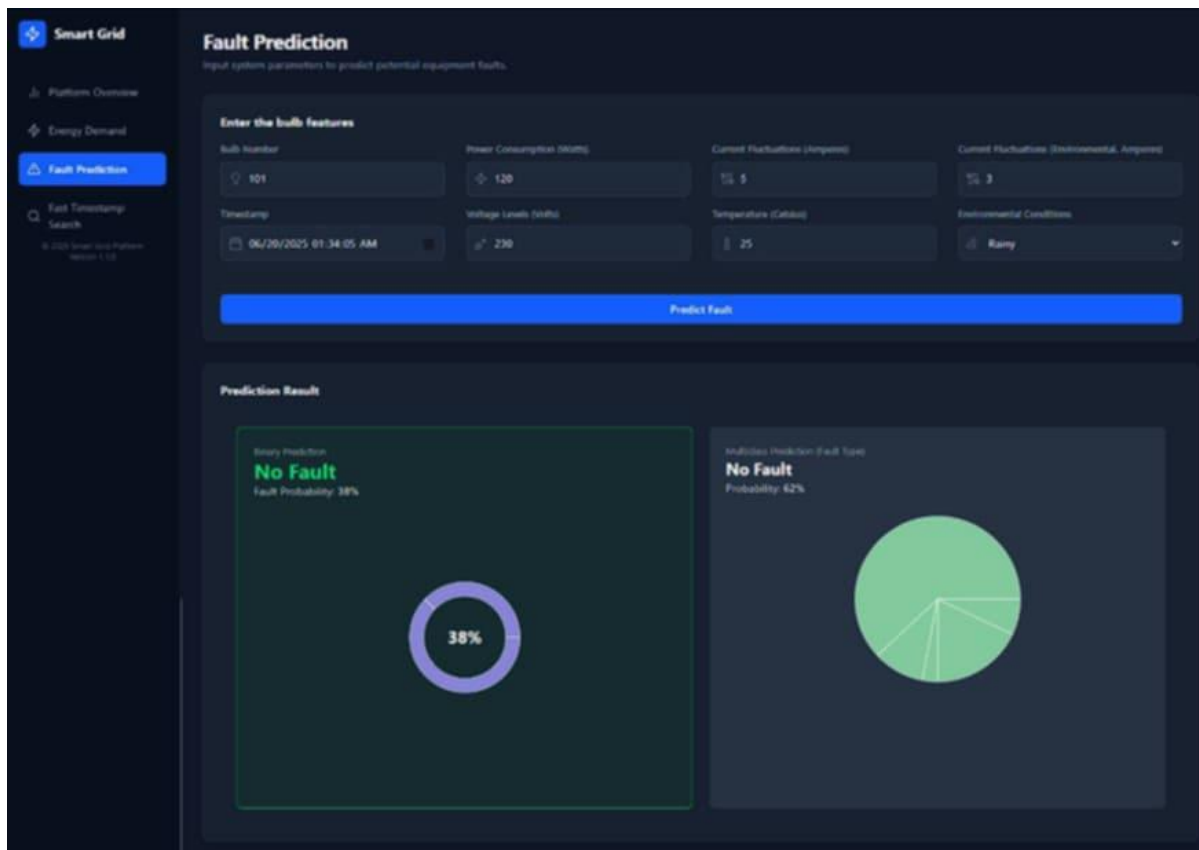
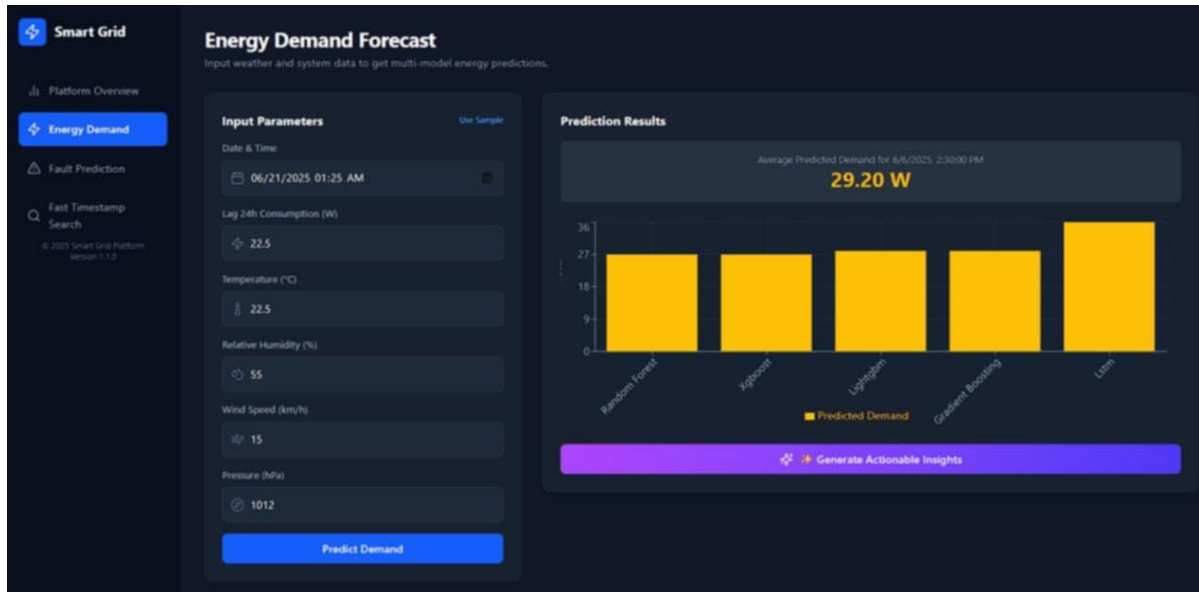
Probability: 88%

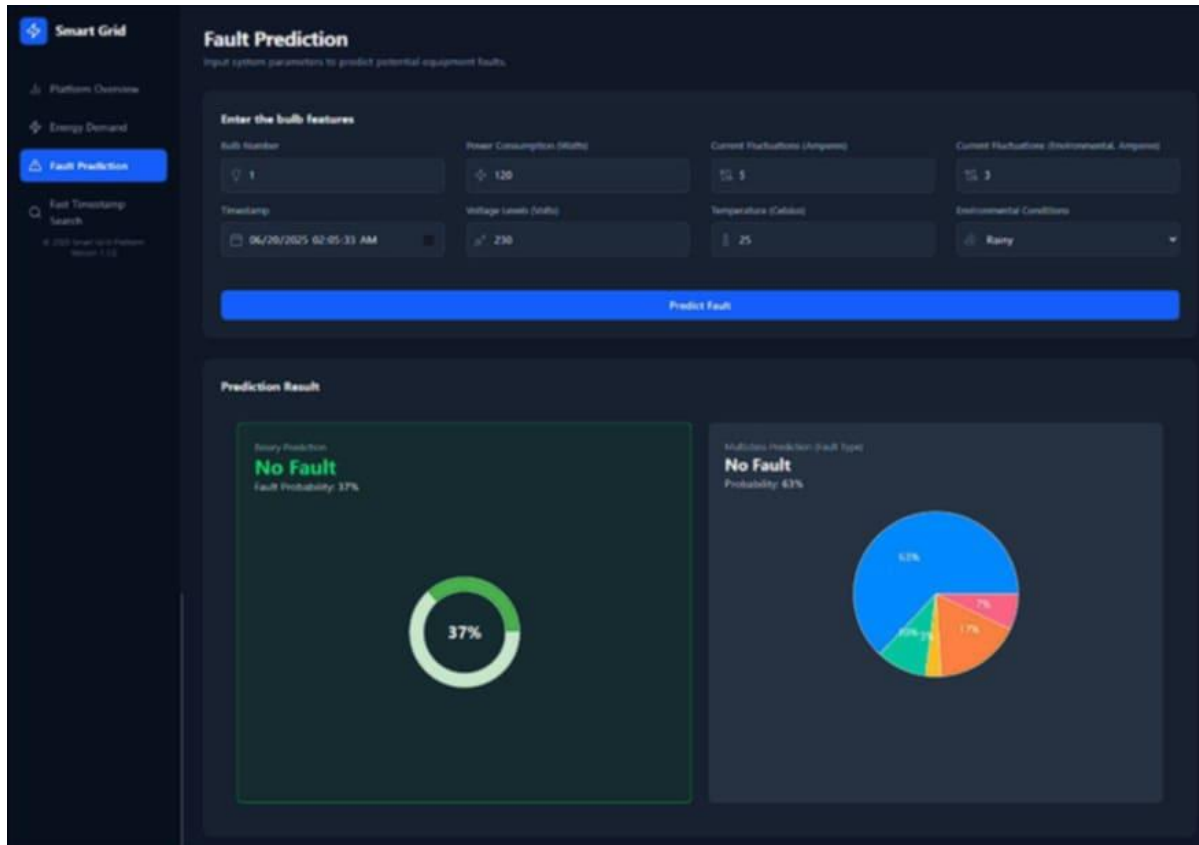
88%

2%

10%

Suggest Diagnostic Steps





Références

- [1] J. POWELL, A. MCCAFFERTY-LEROUX, W. HILAL, AND S. A. GADSDEN, "SMART GRIDS: A COMPREHENSIVE SURVEY OF CHALLENGES, INDUSTRY APPLICATIONS, AND FUTURE TRENDS," ENERGY REPORTS, TO BE PUBLISHED, 2024.
- [2] NIST, "SMART GRID: A BEGINNER'S GUIDE," NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY.
- [3] STELLARIX, "SMART GRID: ARCHITECTURE, DEVELOPMENTS AND USE CASES."
- [4] A. AUTHOR ET AL., "INTEGRATION OF SMART GRID WITH RENEWABLE ENERGY SOURCES," SCIENCE DIRECT, 2023.
- [5] U.S. DEPARTMENT OF ENERGY, "HOW THE SMART GRID PROMOTES A GREENER FUTURE."
- [6] X. FANG ET AL., "SMART GRID – THE NEW AND IMPROVED POWER GRID: A SURVEY," IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 14, NO. 4, PP. 944–980, 2012.
- [6] X. FANG, S. MISRA, G. XUE, AND D. YANG, "SMART GRID—THE NEW AND IMPROVED POWER GRID: A SURVEY," IEEE COMMUN. SURVEYS TUTORIALS, VOL. 14, NO. 4, PP. 944–980, 2012.
- [7] V. C. GUNGOR ET AL., "A SURVEY ON SMART GRID POTENTIAL APPLICATIONS AND COMMUNICATION REQUIREMENTS," IEEE TRANS. IND. INFORMAT., VOL. 9, NO. 1, PP. 28–42, 2013.
- [8] M. S. MOURSI, A. S. AL-SUMAITI, AND H. H. ZEINELDIN, "SMART GRID—A COMPREHENSIVE STUDY OF OPPORTUNITIES, CHALLENGES, AND TECHNOLOGIES," RENEW. SUSTAIN. ENERGY REV., VOL. 67, PP. 1090–1103, 2017.
- [9] Y. ZHANG, L. WANG, AND W. WANG, "REVIEW ON PROBABILISTIC FORECASTING OF WIND POWER GENERATION," RENEW. SUSTAIN. ENERGY REV., VOL. 13, NO. 9, PP. 2758–2766, 2011.
- [10] S. CHEN, H. WANG, AND H. QIN, "UNSUPERVISED LEARNING-BASED EVENT DETECTION FOR SMART GRIDS USING DEEP AUTOENCODERS," IEEE ACCESS, VOL. 8, PP. 224050–224060, 2020.
- [11] T. AHMAD AND H. CHEN, "SHORT AND MEDIUM-TERM FORECASTING OF PHOTOVOLTAIC ENERGY PRODUCTION USING MACHINE LEARNING TECHNIQUES," ENERGY, VOL. 166, PP. 257–267, 2019.
- [12] R. J. BESSA, M. A. MATOS, AND V. MIRANDA, "ELECTRIC VEHICLE INTEGRATION IN THE ELECTRIC POWER SYSTEM WITH RENEWABLE ENERGY SOURCES," IN PROC. IEEE POWER ENERGY SOC. GEN. MEETING, 2012.
- [13] A. S. MUSLEH, M. R. YUCE, AND A. ABBOSH, "REAL-TIME DETECTION OF FALSE DATA INJECTION ATTACKS IN SMART GRID USING DEEP LEARNING," IEEE SYST. J., VOL. 15, NO. 1, PP. 1195–1203, 2020.

Références

- [14] P. SAMADI ET AL., "OPTIMAL REAL-TIME PRICING ALGORITHM BASED ON UTILITY MAXIMIZATION FOR SMART GRID," IN PROC. 1ST IEEE INT. CONF. SMART GRID COMMUN., 2010.
- [15] A. P. MELIOPOULOS, G. J. COKKINIDES, AND F. XIA, "WIDE AREA MONITORING AND CONTROL FOR SMART GRIDS," IN PROC. IEEE PES GEN. MEETING, 2010.
- [16] X. LIU, Z. YU, AND Y. WANG, "REINFORCEMENT LEARNING-BASED ENERGY MANAGEMENT IN MICROGRIDS: A REVIEW," IEEE SYST. J., VOL. 15, NO. 3, PP. 4012–4023, 2021.
- [17] I. GOODFELLOW, Y. BENGIO, AND A. COURVILLE, DEEP LEARNING. CAMBRIDGE, MA, USA: MIT PRESS, 2016.
- [18] "LINEAR REGRESSION – MACHINE LEARNING," GEEKSFORGEEKS.
- [19] "ADVANTAGES AND DISADVANTAGES OF LOGISTIC REGRESSION," GEEKSFORGEEKS.
- [20] "DECISION TREE – ADVANTAGES AND DISADVANTAGES," STEM ACADEMY.
- [21] "ADVANTAGES AND DISADVANTAGES OF RANDOM FOREST," GEEKSFORGEEKS.
- [22] "ADVANTAGES AND DISADVANTAGES OF SUPPORT VECTOR MACHINES," FLYRANK.
- [23] J. WANG, "ADVANTAGES AND DISADVANTAGES OF K-NN," MEDIUM.
- [24] "ARTIFICIAL NEURAL NETWORKS AND ITS APPLICATIONS," GEEKSFORGEEKS.
- [25] J. H. FRIEDMAN, "GREEDY FUNCTION APPROXIMATION: A GRADIENT BOOSTING MACHINE," ANN. STATIST., VOL. 29, NO. 5, PP. 1189–1232, 2001.
- [26] "K-MEANS CLUSTERING," GEEKSFORGEEKS.
- [27] "K-MEANS STRENGTHS & WEAKNESSES," TOWARDS DATA SCIENCE.
- [28] "DBSCAN ALGORITHM," GEEKSFORGEEKS.
- [29] "DBSCAN CLUSTERING," SCIKIT-LEARN DOCUMENTATION.
- [30] "HIERARCHICAL CLUSTERING," GEEKSFORGEEKS.
- [31] "UNDERSTANDING HIERARCHICAL CLUSTERING," ANALYTICS VIDHYA.
- [32] "PCA EXPLAINED," GEEKSFORGEEKS.
- [33] "AUTOENCODER OVERVIEW," DEEPAI.
- [34] "AUTOENCODERS IN ANOMALY DETECTION," IBM.
- [86] "SELF-ORGANIZING MAPS EXPLAINED," ANALYTICS VIDHYA.
- [35] "UNDERSTANDING SOM," MEDIUM.
- [36] O. CHAPELLE, B. SCHÖLKOPF, AND A. ZIEN, SEMI-SUPERVISED LEARNING, MIT PRESS, 2006.

Références

- [37] “SELF-TRAINING FOR SEMI-SUPERVISED LEARNING,” SCIKIT-LEARN.
- [38] “SELF-TRAINING IN SEMI-SUPERVISED LEARNING,” GEEKSFORGEEKS.
- [39] V. SINDHWANI, “A REVIEW OF RESEARCH ON CO-TRAINING,” RESEARCHGATE.
- [40] “CO-TRAINING UNVEILED: HARNESSING MULTIFACETED LEARNING,” TOWARDS DATA SCIENCE.
- [41] X. ZHU AND Z. GHAHRAMANI, “LEARNING FROM LABELED AND UNLABELED DATA WITH LABEL PROPAGATION,” TECH. REP. CMU-CALD-02-107, CARNEGIE MELLON UNIVERSITY, 2002.
- [42] “LABEL PROPAGATION FOR CLASSIFICATION,” SCIKIT-LEARN.
- [43] “GRAPH-BASED SEMI-SUPERVISED LEARNING: A COMPREHENSIVE REVIEW,” ARXIV.
- [44] X. ZHU, GRAPH-BASED SEMI-SUPERVISED LEARNING, SPRINGER, 2005.
- [45] “A SEMI-SUPERVISED AUTOENCODER FRAMEWORK FOR JOINT GENERATION AND CLASSIFICATION,” ARXIV.
- [46] “AUTOENCODERS AND THEIR APPLICATIONS IN MACHINE LEARNING: A SURVEY,” SCIENCE DIRECT.
- [47] “A REVIEW OF SMART GRID ANOMALY DETECTION APPROACHES PERTAINING TO ARTIFICIAL INTELLIGENCE,” SCIENCE DIRECT.
- [48] Y. XU ET AL., “A SEMI-SUPERVISED LEARNING APPROACH FOR SMART GRID DATA INTEGRITY ATTACKS DETECTION,” IEEE TRANS. IND. INFORMAT., VOL. 15, NO. 5, PP. 2730–2739, 2019.
- [49] R. HE, T. WU, AND H. GUO, “AN ENHANCED DBSCAN CLUSTERING METHOD BASED ON DENSITY DIFFERENCE,” IEEE ACCESS, VOL. 9, PP. 120292–120305, 2021.
- [50] Z. LI, Q. YANG, W. FAN, AND J. HAN, “MINING LATENT SEMANTICS FOR RECOMMENDATION,” IN PROC. ACM SIGKDD INT. CONF. KNOWL. DISCOV. DATA MIN., 2011, PP. 569–577.
- [51] L. BREIMAN, “RANDOM FORESTS,” MACH. LEARN., VOL. 45, NO. 1, PP. 5–32, 2001.
- [52] Y. BENGIO, A. COURVILLE, AND P. VINCENT, “REPRESENTATION LEARNING: A REVIEW AND NEW PERSPECTIVES,” IEEE TRANS. PATTERN ANAL. MACH. INTELL., VOL. 35, NO. 8, PP. 1798–1828, 2013.
- [53] A. KRIZHEVSKY, I. SUTSKEVER, AND G. HINTON, “IMAGENET CLASSIFICATION WITH DEEP CONVOLUTIONAL NEURAL NETWORKS,” IN ADV. NEURAL INF. PROCESS. SYST., 2012, PP. 1097–1105.
- [54] I. SUTSKEVER, O. VINYALS, AND Q. V. LE, “SEQUENCE TO SEQUENCE LEARNING WITH NEURAL NETWORKS,” IN ADV. NEURAL INF. PROCESS. SYST., 2014.

Références

- [55] R. SUTTON AND A. BARTO, REINFORCEMENT LEARNING: AN INTRODUCTION, 2ND ED. MIT PRESS, 2018.
- [56] M. J. WAINWRIGHT, HIGH-DIMENSIONAL STATISTICS: A NON-ASYMPTOTIC VIEWPOINT, CAMBRIDGE UNIVERSITY PRESS, 2019.
- [57] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN, THE ELEMENTS OF STATISTICAL LEARNING, 2ND ED., SPRINGER, 2009.
- [58] S. RUSSELL AND P. NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH, 4TH ED., PEARSON, 2020.
- [59] INTERNATIONAL ENERGY AGENCY (IEA), SMART GRIDS: TRACKING REPORT, 2022.
- [60] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST), FRAMEWORK AND ROADMAP FOR SMART GRID INTEROPERABILITY STANDARDS.
- [61] IEEE SMART GRID, "BIG DATA ISSUES IN SMART GRID SYSTEMS."
- [62] WORLD ECONOMIC FORUM, "DATA-DRIVEN ENERGY SYSTEMS."
- [63] M. A. ET AL., "INDEXING TECHNIQUES FOR BIG DATA: A SURVEY," SPRINGER.
- [64] S. J. ET AL., "BIG DATA ANALYTICS IN SMART GRID: A REVIEW," ELSEVIER.
- [65] T. R. ET AL., "TIME SERIES DATA MANAGEMENT IN SMART GRIDS," MDPI.
- [66] H. D. ET AL., "ADAPTIVE INDEXING FOR TIME SERIES IN SMART GRIDS," ACM.
- [67] S. H. ET AL., "SMART GRID RESILIENCE AND INDEXING STRATEGIES," IEEE.
- [68] GEEKSFORGEEKS, "B+ TREE INTRODUCTION."
- [69] ORACLE CORPORATION, "BITMAP INDEXES."
- [70] POSTGRESQL, "INDEX TYPES."
- [71] GEEKSFORGEEKS, "TRIE DATA STRUCTURE."
- [72] APACHE LUCENE, "INDEXING AND SEARCHING."
- [73] GEEKSFORGEEKS, "R-TREE INDEXING."
- [74] T. KRASKA ET AL., "THE CASE FOR LEARNED INDEX STRUCTURES," IN PROC. SIGMOD, 2018.
- [75] S. IDREOS, M. L. KERSTEN, AND S. MANEGOLD, "DATABASE CRACKING: TOWARDS AUTO-TUNING DATABASE KERNELS," IN PROC. CIDR, 2007.
- [76] J. LIN, E. KEOGH, L. WEI, AND S. LONARDI, "EXPERIENCING SAX: A NOVEL SYMBOLIC REPRESENTATION OF TIME SERIES," DATA MIN. KNOWL. DISCOV., VOL. 15, NO. 2, PP. 107–144, 2007.
- [77] F. FUAD, A. A. GANI, N. HASHEM, AND O. ZAKARIA, "TS-INDEX: A BIG DATA INDEXING STRATEGY FOR TIME-SERIES DATA," IEEE ACCESS, VOL. 8, PP. 77650–77663, 2020.

Références

- [78] APACHE PARQUET, "COLUMNAR STORAGE."
- [79] P. INDYK AND R. MOTWANI, "APPROXIMATE NEAREST NEIGHBORS: TOWARDS REMOVING THE CURSE OF DIMENSIONALITY," IN PROC. STOC, 1998.
- [80] FACEBOOK AI, "FAISS LIBRARY."
- [81] G. E. HINTON AND R. R. SALAKHUTDINOV, "REDUCING THE DIMENSIONALITY OF DATA WITH NEURAL NETWORKS," SCIENCE, VOL. 313, NO. 5786, PP. 504–507, 2006.
- [82] R. AGRAWAL, C. FALOUTSOS, AND A. SWAMI, "EFFICIENT SIMILARITY SEARCH IN SEQUENCE DATABASES," IN PROC. FODO, 1993.
- [83] C. FALOUTSOS, M. RANGANATHAN, AND Y. MANOLOPOULOS, "FAST SUBSEQUENCE MATCHING IN TIME-SERIES DATABASES," IN PROC. SIGMOD, 1994.
- [84] J. SHIEH AND E. KEOGH, "ISAX: INDEXING AND MINING TERABYTE SIZED TIME SERIES," IN PROC. SIGKDD, 2009.
- [85] A. CAMERRA, T. PALPANAS, J. SHIEH, AND E. KEOGH, "ISAX 2.0: INDEXING AND MINING ONE BILLION TIME SERIES," IN PROC. ICDM, 2010.
- [86] E. KEOGH, K. CHAKRABARTI, M. PAZZANI, AND S. MEHROTRA, "DIMENSIONALITY REDUCTION FOR FAST SIMILARITY SEARCH IN LARGE TIME SERIES DATABASES," KNOWL. INF. SYST., VOL. 3, NO. 3, PP. 263–286, 2001.
- [87] K.-P. CHAN AND A. W.-C. FU, "EFFICIENT TIME SERIES MATCHING BY WAVELETS," IN PROC. ICDE, 1999.
- [88] P. MALHOTRA ET AL., "LSTM-BASED ENCODER-DECODER FOR MULTI-SENSOR ANOMALY DETECTION," IN PROC. ICML WORKSHOP, 2016.
- [89] C. ZHOU AND R. PAFFENROTH, "ANOMALY DETECTION WITH ROBUST DEEP AUTOENCODERS," IN PROC. KDD, 2017.
- [90] Y. DING ET AL., "ALEX: AN UPDATABLE ADAPTIVE LEARNED INDEX," MICROSOFT RESEARCH, 2020.

Webographies:

- [W1] SOLUTIONS DE RESEAU INTELLIGENT A INDIANAPOLIS, IN
- [W2] SMART GRIDS : LES RESEAUX ELECTRIQUES ET LE RESEAU EN EVOLUTION
- [W3] ITERATIONS DU DEVELOPPEMENT DES RESEAUX ELECTRIQUES : ÉVOLUTION DE L'INTELLIGENCE... | TELECHARGER LE DIAGRAMME SCIENTIFIQUE
- [W4] OTECH.UAETH.EDU.MX/NOTI/INDEX.PHP/MACHINE-LEARNING/MACHINE-LEARNING-QUE-ES-Y-CUAL-ES-SU-RELACION-CON-LA-IA/