



*République Algérienne Démocratique et Populaire*  
*Ministère de l'enseignement supérieur et de la recherche scientifique*  
*Université Larbi Tébessi – Tébessa*  
*Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie*  
*Département de mathématique et informatique*



## **Mémoire de fin d'étude**

Pour l'obtention du diplôme de **MASTER**

**Domaine** : Mathématiques et informatique

**Filière** : Informatique

**Option** : Systèmes d'Informations

---

### **IA dans le secteur d'Énergie – Cas d'étude** **« Prédiction de consommation d'Énergie »**

---

Réalisé par : Salem Aimene

*Devant le jury :*

<b>Dr. KHEDHIRI Abderrazak</b>	<b>MCB</b>	<b>Université Larbi Tébessi</b>	<b>Président</b>
<b>Dr. SLIMI Hamda</b>	<b>MAB</b>	<b>Université Larbi Tébessi</b>	<b>Examineur</b>
<b>Pr. LAOUAR Mohammed Ridda</b>	<b>Pr</b>	<b>Université Larbi Tébessi</b>	<b>Encadreur</b>
<b>Mlle. TOUATI Hamad Zaineb</b>		<b>Université Larbi Tébessi</b>	<b>Co- Encadreur</b>

06/06/2023

# Dédicaces

*Avec une profonde gratitude et un amour sincère, je dédie humblement ce modeste travail à **mes parents** qui ont toujours été mon soutien inconditionnel, ainsi qu'à **mes frères** bien-aimés qui ont partagé chaque moment précieux de ma vie.*

*Aimene*

## **Abstract**

The effective prediction of energy consumption is crucial for efficient energy management and optimal resource allocation. In this thesis, we address the challenges related to data quality and model selection in energy consumption prediction using an artificial intelligence model, specifically utilizing time series data.

To mitigate the limitations identified in existing literature, we conducted a comprehensive data analysis process to ensure the quality and reliability of the input data. By carefully examining the data, we took measures to handle missing values and normalize the dataset to optimize model performance. We selected the LSTM (Long Short-Term Memory) model due to its superior performance and high applicability in solving sequential problems. LSTM has the ability to capture long-term dependencies in time series data, making it well-suited for energy consumption prediction tasks.

The experimental results obtained at the end of this research demonstrate the effectiveness of our approach. Overall, this thesis contributes to the field of energy consumption prediction by addressing challenges related to data quality and model selection. Our results highlight the importance of appropriate data analysis and demonstrate the relevance and efficiency of LSTM models for accurate energy consumption forecasting. The findings of this research can facilitate informed decision-making and enable better energy management strategies in various domains.

**Keywords:** Energy consumption, Energy efficiency, Prediction, Time series, Data Analysis, LSTM.

## **Abstract**

La prédiction efficace de la consommation d'énergie est cruciale pour une gestion énergétique efficace et une allocation optimale des ressources. Dans cette thèse, nous abordons les défis liés à la qualité des données et à la sélection du modèle dans la prédiction de la consommation d'énergie à l'aide d'un modèle d'intelligence artificielle, en utilisant spécifiquement des données de séries temporelles.

Pour atténuer les limitations identifiées dans la littérature existante, nous avons mené un processus complet d'analyse des données afin d'assurer la qualité et la fiabilité des données d'entrée. En examinant attentivement les données, nous avons pris des mesures pour traiter les valeurs manquantes et normaliser l'ensemble de données pour optimiser les performances du modèle, nous avons sélectionné le modèle LSTM en raison de ses performances supérieures et de sa grande applicabilité à la résolution de problèmes séquentiels. LSTM a la capacité de capturer les dépendances à long terme dans les données de séries temporelles, ce qui le rend bien adapté aux tâches de prédiction de la consommation d'énergie.

Les résultats expérimentaux obtenus à la fin de cette recherche démontrent l'efficacité de notre approche.

Dans l'ensemble, cette thèse contribue au domaine de la prédiction de la consommation d'énergie en abordant les défis liés à la qualité des données et à la sélection du modèle. Nos résultats mettent en évidence l'importance d'une analyse appropriée des données et démontrent la pertinence et l'efficacité des modèles LSTM pour une prévision précise de la consommation d'énergie. Les résultats de cette recherche peuvent faciliter la prise de décision éclairée et permettre de meilleures stratégies de gestion énergétique dans différents domaines.

**Mots clés :** Consommation d'énergie, Efficacité énergétique, Prédiction, Séries chronologiques, Data Analysis, LSTM

## ملخص

توقع استهلاك الطاقة بشكل فعال أمر حاسم لإدارة الطاقة الفعالة وتخصيص الموارد بشكل مثالي. في هذه الأطروحة، نتناول التحديات المتعلقة بجودة البيانات واختيار النموذج في توقع استهلاك الطاقة باستخدام نموذج الذكاء الاصطناعي، وذلك باستخدام بيانات السلاسل الزمنية بشكل خاص.

للتغلب على القيود المحددة في الأدبيات الحالية، قمنا بإجراء عملية تحليل بيانات شاملة لضمان جودة وموثوقية البيانات المدخلة. من خلال دراسة البيانات بعناية، اتخذنا إجراءات لمعالجة القيم المفقودة وتطبيع مجموعة البيانات لتحسين أداء النموذج. قمنا باختيار نموذج LSTM بسبب أدائه المتفوق وقابليته الكبيرة لحل المشكلات التسلسلية. يتمتع نموذج LSTM بقدرته على التقاط التبعيات ذات المدى الطويل في بيانات السلاسل الزمنية، مما يجعله مناسبًا لمهام توقع استهلاك الطاقة.

تظهر النتائج التجريبية التي تم الحصول عليها في نهاية هذا البحث فعالية نهجنا.

بشكل عام، تساهم هذه الأطروحة في مجال توقع استهلاك الطاقة من خلال معالجة التحديات المتعلقة بجودة البيانات واختيار النموذج. تسلط نتائجنا الضوء على أهمية تحليل البيانات المناسب وتوضيح ملاءمة وفعالية نماذج LSTM في التنبؤ الدقيق لاستهلاك الطاقة. يمكن أن تسهم نتائج هذا البحث في اتخاذ قرارات مستنيرة وتمكين استراتيجيات أفضل لإدارة الطاقة في مختلف المجالات.

**الكلمات المفتاحية:** استهلاك الطاقة، توقع، سلاسل زمنية، تحليل البيانات، LSTM

# Table de matières

INTRODUCTION GÉNÉRALE .....	1
Chapitre 1: L'Énergie : Concepts fondamentaux et aperçu du domaine .....	3
1.1 Introduction .....	3
1.2 L'Énergie : Définition, Formes et sources .....	3
1.2.1 Qu'est-ce que l'énergie ? .....	3
1.2.2 Formes d'énergie .....	3
1.2.3 Sources et méthodes de production d'énergie .....	4
1.3 Secteurs de consommation d'énergie .....	5
1.4 Évolution de la consommation d'énergie .....	6
1.5 L'importance de rationaliser la consommation d'énergie .....	7
1.6 Tendances actuelles dans le secteur de l'énergie .....	8
1.7 Conclusion .....	8
Chapitre 2: Etat de l'art sur l'IA et la prédiction de la consommation d'énergie .....	10
2.1 Introduction .....	10
2.2 Les types de prédiction de la consommation d'énergie .....	10
2.3 Taxonomie de modèles de prédiction .....	11
2.3.1 Modèles simples (Single models) .....	11
2.3.2 Modèles d'ensemble (Ensemble models) .....	15
2.3.3 Modèles hybrides (Hybrids models) .....	15
2.4 Les séries chronologique (Time series) .....	16
2.5 Métriques d'évaluation pour les modèles de prédiction .....	17
2.5.1 Métriques d'évaluation de régression .....	17
2.6 Datasets .....	18
2.7 Travaux connexes .....	19
2.8 Conclusion .....	22
Chapitre 3: Cas d'étude : prédiction de consommation d'Électricité .....	24
3.1 Introduction .....	24
3.2 Méthodologie .....	24
3.2.1 Dataset .....	25
3.2.2 Visualisation de données .....	25

3.2.3 Préparation de données .....	27
3.2.4 Ingénierie de caractéristiques.....	28
3.2.5 Fractionnement de Dataset .....	31
3.2.6 Sélection de modèle .....	31
3.2.7 Développement de modèle .....	32
3.2.8 Evaluation de modèle .....	33
3.3 Résultats et discussion.....	33
3.3.1 Comparaison avec les travaux connexes.....	35
3.4 Conclusion .....	37
Conclusion Générale & perspectives.....	38
Références .....	39

# Liste de figures

Figure 1-1: Consommation final total par source (TFC), Le monde 1990-2020 [7] .....	6
Figure 1-2: Consommation final total par secteur (TFC), Le monde 1990-2020 [7].....	7
Figure 2-1: Types de prédiction .....	10
Figure 2-2: Taxonomie de modèles de prédiction de la consommation d'énergie selon leur complexité.....	12
Figure 2-3: Architecture de base de ANN [25].....	14
Figure 2-4: Méthodes d'ensembles .....	15
Figure 3-1: Méthodologie proposé de travail .....	24
Figure 3-2: Consommation d'Energie en mégawatts (MW) de 2004 à 2019 .....	26
Figure 3-3: Consommation d'Energie en mégawatts (MW) de 2004 à 2019 .....	26
Figure 3-4: Décomposition de la séries temporelles (Tendance, saisonnalité et résidus) .....	28
Figure 3-5: Consommation d'Energie en mégawatts (MW) par heure .....	30
Figure 3-6: Consommation d'Energie en mégawatts (MW) par semaine .....	30
Figure 3-7: Fractionnement de Dataset .....	31
Figure 3-8: Résultats de prédiction avec LSTM sur le test set .....	33
Figure 3-9: Loss durant la phase d'entraînement.....	34
Figure 3-10: Performance de notre modèle LSTM.....	34
Figure 3-11: Comparaison avec les travaux connexes .....	36

# Liste de tables

Tableau 2-1 : Comparaison de dataset les plus utilisés dans la littérature .....	18
Tableau 2-2 : Résumé des travaux connexes .....	21
Tableau 3-1 : Dataset après l'étape d'extraction de caractéristiques .....	29
Tableau 3-2 : Comparaison avec les travaux connexes .....	35

# INTRODUCTION GÉNÉRALE

La consommation d'énergie est un problème critique qui affecte l'économie, l'environnement et la société dans son ensemble. La demande croissante d'énergie a entraîné un besoin de méthodes plus efficaces et durables de gestion de l'énergie. Les modèles d'intelligence artificielle (IA) et d'apprentissage automatique (ML) ont montré un grand potentiel à cet égard en proposant de nouvelles et innovantes solutions pour l'analyse et la prédiction de la consommation d'énergie.

Ces dernières années, l'exploitation de modèles d'IA et de ML dans le domaine de la consommation d'énergie a suscité beaucoup d'intérêt chez les chercheurs et les praticiens pour l'amélioration de **l'efficacité énergétique**.

L'évolution rapide dans la consommation d'énergie implique le développement et l'amélioration des réseaux d'énergie, cette évolution démontre aussi un défi qui implique des stratégies de gestion comme la **maîtrise de la demande en énergie** (Energy demand management).

L'utilisation de modèles d'IA et de ML a permis de réaliser des améliorations significatives à l'efficacité énergétique, permettant aux entreprises énergétiques de mieux gérer leurs ressources énergétiques et de réduire leurs déchets.

Les perspectives d'utilisation des modèles ML dans la prédiction de la consommation d'énergie sont très prometteuses. Ces modèles ont le potentiel de révolutionner la façon dont la consommation d'énergie est gérée en fournissant des solutions précises et efficaces. Avec les progrès de la technologie, les modèles ML sont devenus plus sophistiqués et capables de traiter des modèles complexes de consommation d'énergie et de prédire les tendances futures. Cela en fait des outils précieux pour l'amélioration de l'efficacité énergétique.

Dans cette thèse, nous abordons les défis liés à la qualité des données et à la sélection du modèle dans la prédiction de la consommation d'énergie à l'aide d'un modèle d'intelligence artificielle, en utilisant spécifiquement des données de séries temporelles, nous évaluerons son potentiel pour améliorer les pratiques de gestion de l'énergie et contribuer au développement d'un avenir énergétique durable. Notre objectif est de fournir une compréhension détaillée de l'utilisation d'un modèle d'apprentissage automatique (ML), et d'identifier les opportunités et les défis à relever pour une mise en œuvre réussie.

Nous présentons les concepts fondamentaux de l'énergie et les principaux défis du domaine dans le chapitre 1. Dans le chapitre 2, nous faisons un état de l'art des méthodes et des outils existants pour modéliser, analyser et optimiser l'énergie. Nous proposons ensuite une étude de cas d'un système réel dans le chapitre 3. Nous terminons par une conclusion générale et des perspectives de recherche future.

---

# **Chapitre 01 :**

## **L'Energie : concepts fondamentaux et aperçu**

---

# Chapitre 1: L'Énergie : Concepts fondamentaux et aperçu du domaine

## 1.1 Introduction

L'énergie est une ressource fondamentale pour les sociétés modernes et essentielle pour répondre à nos besoins quotidiens et assurer le développement économique. Dans ce chapitre, nous commencerons par aborder certaines notions générales relatives à l'énergie et à ses formes et sources. Ensuite, nous explorerons les différents secteurs qui le consomment afin de mieux comprendre l'évolution de la consommation, nous mettrons aussi en évidence les raisons et les facteurs qui ont contribué à cette évolution, et nous citerons les tendances actuelles dans le domaine énergétique.

## 1.2 L'Énergie : Définition, Formes et sources

### 1.2.1 Qu'est-ce que l'énergie ?

L'énergie est définie par les experts comme la capacité à accomplir du travail. Selon les théories économiques, l'énergie est considérée comme une ressource cruciale pour la production industrielle. La civilisation moderne est possible parce que les gens ont appris à transformer l'énergie d'une forme à une autre, puis à l'utiliser pour effectuer un travail. L'énergie est utilisée pour se déplacer à pied, à vélo, pour faire avancer les voitures sur les routes et les bateaux sur l'eau, pour cuisiner pour fournir de l'éclairage à nos résidences et lieux de travail, pour la fabrication de biens et pour les voyages spatiaux des astronautes. [1]

### 1.2.2 Formes d'énergie

Il existe plusieurs formes d'énergie telles que la chaleur, la lumière, le mouvement, l'électricité, la forme chimique et la gravité. Ces formes d'énergie peuvent être regroupées en deux catégories générales pour faire du travail : L'énergie potentielle (stockée) et l'énergie cinétique (active).[2]

- **L'énergie potentielle (stockée) :**

L'énergie potentielle est une forme d'énergie stockée, c'est l'énergie qui réside dans un objet en raison de sa position ou de sa configuration et qui peut être libérée et transformé en une autre forme d'énergie.

Par exemple, un objet placé en hauteur, a de l'énergie potentielle en raison de sa position dans le champ de gravité terrestre. Cette énergie peut être libérée et convertie en énergie cinétique lorsque l'objet est lâché et tombe vers le sol.

Il existe différents types d'énergie potentielle, notamment l'énergie potentielle gravitationnelle, l'énergie potentielle élastique, l'énergie potentielle électrique,

l'énergie potentielle nucléaire, etc. Chaque type d'énergie potentielle est associé à un champ de force spécifique et peut être utilisé dans diverses applications, telles que le stockage d'énergie et la génération d'électricité. [2, 3, 4]

- **L'énergie cinétique (active) :**

C'est la capacité de l'objet à effectuer un travail en raison de sa vitesse (son mouvement). L'énergie cinétique est une forme d'énergie associée au mouvement d'un objet.

L'énergie cinétique peut être convertie en d'autres formes d'énergie, telles que l'énergie thermique ou l'énergie électrique. Elle est utilisée dans de nombreuses applications, notamment la conversion d'énergie mécanique en électricité grâce à des turbines hydrauliques ou éoliennes, etc... [2, 3, 4]

### 1.2.3 Sources et méthodes de production d'énergie

Les ressources utilisées pour générer de l'énergie afin de satisfaire les besoins humains en électricité, en chaleur, en transport et autres, sont désignées sous le terme de sources d'énergie. Les sources d'énergie se divisent principalement en deux catégories : les sources d'énergie renouvelables et les sources d'énergie non renouvelables. [3]

- **Les sources non renouvelables :**

Sont des sources d'énergie qui s'épuisent avec le temps [4]. Elles comprennent :

-**Les combustibles fossiles** : (pétrole, gaz naturel, charbon) sont le résultat de l'accumulation de matières organiques dans le sol pendant des périodes s'étendant sur des millions d'années. Ils sont brûlés pour produire de l'énergie, pour produire de l'électricité, pour le chauffage et pour les transports. C'est pour ça ils sont polluants et émettent des gaz qui contribuent au réchauffement climatique.

-**L'uranium** : Il est utilisé comme combustible dans les centrales nucléaires pour produire de l'électricité.

- **Les sources renouvelables :**

Les sources d'énergie renouvelables sont des sources d'énergie qui se renouvellent naturellement et de manière continue [5]. Elles comprennent :

-**L'énergie solaire** : L'énergie solaire est générée par le rayonnement émis par le soleil, qui peut être capté par des panneaux solaires.

-**L'énergie éolienne** : L'énergie éolienne est engendrée par le mouvement de l'air ou le vent., qui peut être capté par des éoliennes.

-**L'énergie hydraulique** : l'énergie hydraulique est produite par la force de l'eau, qui peut être utilisée pour faire tourner des turbines.

-**L'énergie géothermique** : L'énergie géothermique est engendrée par la chaleur provenant de l'intérieur de la Terre.

**-L'énergie des biomasses** : l'énergie des biomasses est produite par la combustion de matières organiques, comme le bois, les déchets agricoles, les résidus forestiers, etc.

**-L'énergie marine** : L'énergie marine est générée par les mouvements des marées, des courants océaniques et des vagues.

Les différentes sources d'énergie peuvent servir de sources primaires pour générer de la chaleur, comme pour le chauffage des bâtiments ou la production d'eau chaude sanitaire. Elles peuvent également être utilisées comme sources secondaires pour produire de l'électricité. Les centrales électriques ont la capacité d'utiliser des sources d'énergie telles que les combustibles fossiles, la biomasse, l'énergie hydraulique, l'énergie solaire ou éolienne afin de générer de l'électricité [4, 5].

Pour répondre aux besoins énergétiques, les sources d'énergie peuvent être utilisées par des manières différentes.

### **1.3 Secteurs de consommation d'énergie**

Il existe différents secteurs de consommation d'énergie, chacun ayant des besoins énergétiques spécifiques. Parmi ces secteurs de consommation d'énergie on peut citer : [2, 3]

**-Le secteur résidentiel** : il s'agit des maisons, des appartements et des bâtiments où les gens vivent. Ce secteur est responsable d'une part considérable de la consommation énergétique, en particulier pour le chauffage, la climatisation et l'éclairage.

**-Le secteur commercial** : Il englobe divers types d'établissements tels que bureaux, centres commerciaux, restaurants, hôtels, et bien d'autres. Cette industrie nécessite de l'énergie pour des besoins tels que l'éclairage, le chauffage, la climatisation, l'utilisation d'ordinateurs, d'équipements de cuisine et de systèmes de ventilation., etc.

**-Le secteur industriel** : il comprend les usines, les entreprises de fabrication, etc. Ce domaine représente la plus grande part de la consommation énergétique à l'échelle mondiale., car il nécessite beaucoup d'énergie pour faire fonctionner les machines, les équipements et les processus de production.

**-Le secteur des transports** : il comprend les voitures, les camions, les avions, les trains, les bateaux, etc. Ce secteur est également un gros consommateur d'énergie, en particulier pour les carburants fossiles.

**-Le secteur de l'agriculture** : il comprend les exploitations agricoles, les serres, les usines de transformation des aliments, etc. Ce secteur consomme de l'énergie pour l'irrigation, le chauffage, la réfrigération et les équipements de transformation des aliments.

**-Le secteur de la construction** : il comprend la construction et la rénovation de bâtiments, les travaux publics, les infrastructures, etc. Ce secteur consomme de l'énergie pour les machines, les outils, les équipements de chantier, etc.

Chaque secteur a des besoins énergétiques différents, ce qui implique que les choix que nous faisons en matière d'utilisation de l'énergie peuvent avoir un impact important sur l'environnement et le climat.

#### 1.4 Évolution de la consommation d'énergie

Au cours des dernières décennies, la consommation d'énergie a augmenté dans toutes les sources énergétiques et dans tous les secteurs économiques. [6, 7]

La hausse de la demande énergétique est principalement attribuable à la croissance démographique et à l'expansion de l'économie mondiale., ainsi qu'à l'urbanisation croissante dans de nombreux pays. [8, 9, 10]

Les combustibles fossiles, notamment le pétrole, le gaz naturel et le charbon, ont été les principales sources d'énergie utilisées pour satisfaire cette demande croissante en énergie. Néanmoins, l'utilisation des énergies renouvelables, comme l'énergie éolienne, solaire, etc., a également connu une augmentation progressive au fil du temps. (la figure 1.1).

En ce qui concerne les secteurs économiques, l'industrie est le plus grand consommateur d'énergie, suivie par les secteurs des transports et résidentiel (la figure 1.2).

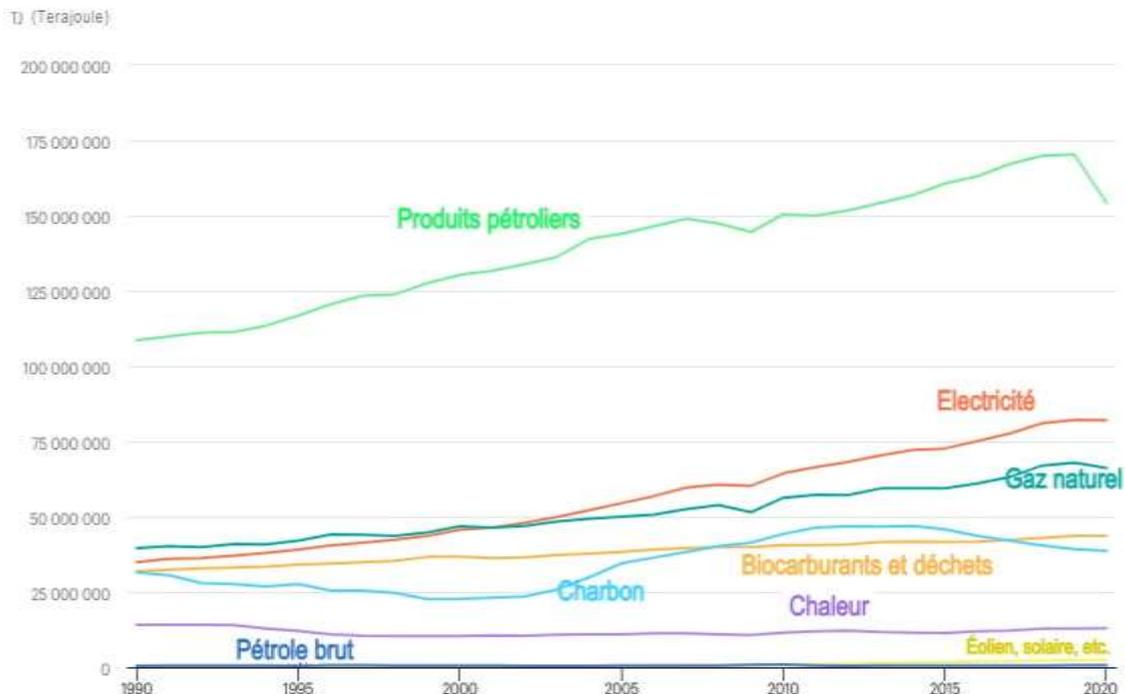


Figure 1-1: Consommation final total par source (TFC), Le monde 1990-2020 [7]

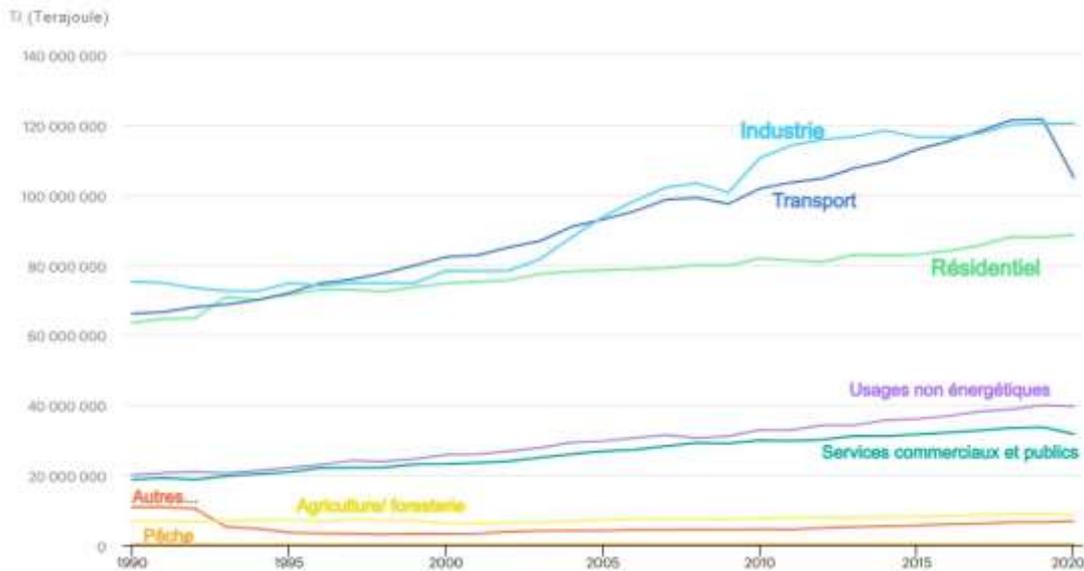


Figure 1-2: Consommation final total par secteur (TFC), Le monde 1990-2020 [7]

Il convient de souligner que cette augmentation de la consommation d'énergie a engendré des conséquences néfastes sur l'environnement, notamment en termes d'émissions de gaz à effet de serre, ainsi que de pollution de l'air et de l'eau. [9]. Il est donc essentiel de trouver des solutions visant à réduire notre consommation d'énergie, à adopter des sources d'énergie durables et moins polluantes, afin de préserver notre planète. [10]

### 1.5 L'importance de rationaliser la consommation d'énergie

La rationalisation de la consommation d'énergie est importante pour plusieurs raisons :

**-Réduire les problèmes politiques et sociales :** En diminuant notre dépendance vis-à-vis des sources d'énergie importées et en favorisant l'adoption de technologies économes en énergie ainsi que de sources d'énergie renouvelables, nous avons la possibilité d'améliorer la sécurité énergétique de notre nation, de créer de nouveaux emplois et de stimuler la croissance économique. [8]

**-Réduire les problèmes environnementaux :** La consommation excessive d'énergie a des répercussions défavorables sur l'environnement, car elle entraîne des émissions de gaz à effet de serre et contribue à la pollution de l'air et de l'eau, à la destruction des écosystèmes, etc. En rationalisant la consommation d'énergie, nous pouvons réduire cet impact. [9]

**-Réduire les problèmes économiques :** La consommation d'énergie représente souvent une part importante des dépenses des ménages et des entreprises. En rationalisant la consommation d'énergie, nous pouvons réduire nos coûts d'énergie et économiser de l'argent à long terme. [10]

### ***1.6 Tendances actuelles dans le secteur de l'énergie***

L'augmentation constante et exhaustive de la consommation d'énergie impose la nécessité de moyens avancés et intelligents pour aider à la réduction et la rationalisation de la consommation

L'une des tendances de solutions les plus prometteuses c'est l'utilisation de l'Intelligence Artificielle (IA) et l'apprentissage automatique (ML) pour résoudre ces problèmes dans de nombreux domaines, notamment les domaines énergétiques. [11]

L'Intelligence Artificielle (IA) et l'apprentissage automatique (ML) peuvent aider à prévoir la demande d'énergie de manière plus précise, permettant aux fournisseurs d'énergie d'ajuster leur offre en conséquence et d'éviter la surproduction d'énergie, cela peut entraîner une utilisation plus durable de l'énergie.

### ***1.7 Conclusion***

Dans ce chapitre nous avons mené une étude globale du domaine énergétique et de la consommation d'énergie dans les différents sources et secteurs, conduisant à la conclusion que cette consommation s'accélère à un rythme élevé, et que les technologies innovantes telles que l'IA et les modèles d'apprentissage automatique sont largement utilisées. Dans le chapitre suivant, nous exposerons certains de ces modèles et tenterons d'expliquer leurs fonctionnements.

---

## **Chapitre 2 :**

**Etat de l'art sur l'IA et la prédiction de la  
consommation d'énergie**

---

## Chapitre 2: Etat de l'art sur l'IA et la prédiction de la consommation d'énergie

### 2.1 Introduction

Les modèles d'apprentissage automatique jouent un rôle essentiel dans la planification et la prévision. L'objectif principal est de parvenir à une prédiction qui donne des informations futures en se basant sur les études antérieures. Ce chapitre examinera la littérature pour avoir une bonne connaissance sur le domaine, en identifiant les modèles de prédiction et les types de données les plus couramment utilisés, nous examinerons également les métriques d'évaluation.

### 2.2 Les types de prédiction de la consommation d'énergie

En se basant sur la période de prévision, on peut décomposer le processus de la prévision d'énergie en **trois catégories**. Il s'agit de prévisions à court terme, à moyen terme et à long terme.

Les prévisions à **court terme** (d'une heure à semaines) ; sont principalement utilisées pour optimiser l'engagement des unités de génération, la distribution d'énergie et la répartition de la charge, tout en prenant en compte le contrôle en temps réel et l'évaluation de la sécurité.

Les prévisions à **moyen terme** (d'une semaine à mois) ; sont surtout appliquées à la planification des ressources et des tarifs de production d'énergie. Le but de cette prédiction donc est d'assurer un équilibre entre la demande et la génération et de maintenir le système, l'achat d'énergie et la fixation des prix.

Les prévisions à **long terme** (d'un mois à des années) ; sont surtout appliquées à la gestion des ressources et aux investissements dans le développement. L'objectif de cette prédiction est donc la planification de l'expansion du système, y compris la génération, la transmission et la distribution. Cette prédiction peut également avoir un impact sur l'achat de nouvelles unités de génération.

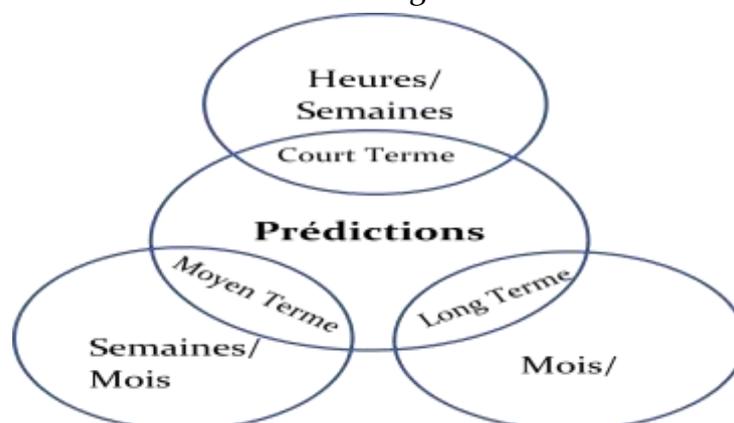


Figure 2-1: Types de prédiction

### 2.3 *Taxonomie de modèles de prédiction*

On peut nommer la catégorisation des différents types de modèles d'apprentissage automatique en fonction de leurs algorithmes sous-jacents, architectures ou de leurs méthodes d'apprentissage ; une taxonomie des modèles. [16, 17]

Différents experts, chercheurs et praticiens peuvent utiliser différentes taxonomies en fonction de leurs besoins spécifiques, de leur expérience et de leurs objectifs. Dans l'ensemble, le choix de la classification des modèles dépend du contexte et de l'objectif de l'analyse ou de l'application. [18, 19]

Selon la littérature [12–27], les modèles de ML utilisés dans la prédiction de la consommation d'énergie diffèrent en fonction de leur complexité, en modèles simples (Single models), modèles d'ensemble (Ensemble models) et les modèles hybrides (Hybrid models).

Afin de classer ces modèles, et en se basant sur des revues précédentes, la taxonomie appropriée pour notre problème de prédiction est dans (La figure 2-1).

Cette taxonomie nous permet d'englober les modèles utilisés par les experts, chercheurs et praticiens, et de comprendre comment l'utilisation de modèles avec différents niveaux de complexité peut différer.

Dans la (figure 1-3), nous classons les modèles les plus couramment utilisés selon la littérature en fonction de leur complexité. Les modèles inclus dans cette classification sont :

#### 2.3.1 **Modèles simples (Single models)**

Un modèle simple en apprentissage automatique, fait référence à un type spécifique de modèle d'apprentissage automatique conçu pour résoudre une tâche ou un problème particulier en utilisant un seul algorithme ou une seule architecture. [17]

Exemples de ces modèles :

- **Modèles statistiques**

Les modèles statistiques sont utilisés pour apprendre les formes dans les données et pour faire des prédictions sur de nouveaux points de données [16, 17, 18]. Certains des modèles statistiques couramment utilisés dans le contexte de notre problème sont :

- ARMA/ARIMA/SARIMA :

Les modèles ARMA (Autoregressive Moving Average), ARIMA (Autoregressive Integrated Moving Average) et SARIMA (Seasonal Autoregressive Integrated Moving Average) ; sont des modèles de séries chronologiques souvent utilisés en apprentissage automatique pour les tâches de prévision. Ces modèles peuvent être utilisés pour prévoir les valeurs futures d'une série chronologique en fonction de leur observation passée, ces modèles estiment les paramètres qui conviennent le mieux aux données de la série chronologique et l'utilisent pour effectuer des prévisions.

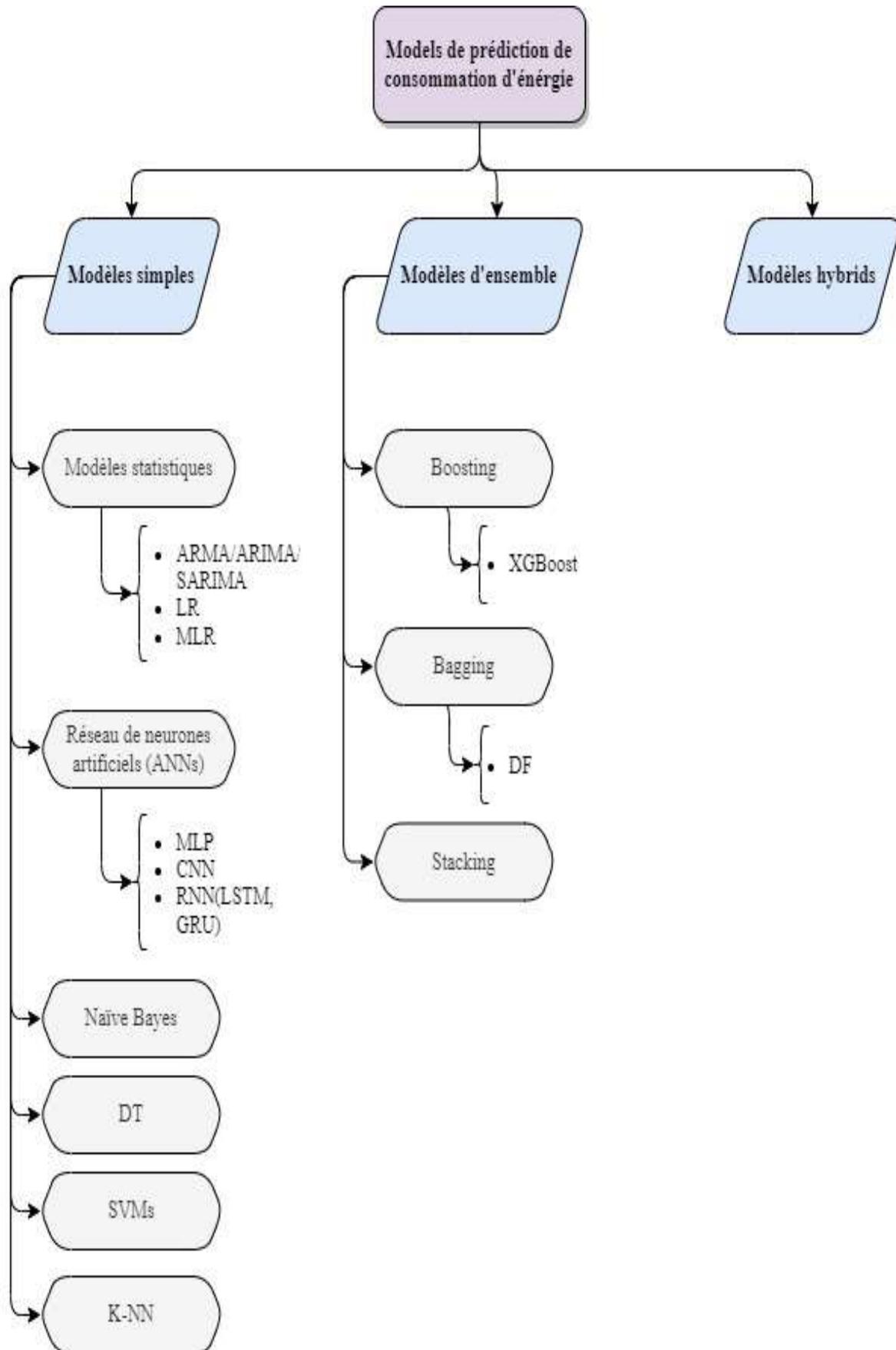


Figure 2-2: Taxonomie de modèles de prédiction de la consommation d'énergie selon leur complexité. [12-27]

- Régression linéaire :

Un modèle statistique utilisé pour prédire une variable de sortie continue en fonction d'une ou plusieurs variables d'entrée.

- Régression linéaire multiple :

C'est une méthode statistique utilisée pour modéliser la relation entre une variable dépendante et deux ou plusieurs variables indépendantes. C'est une extension du modèle de régression linéaire simple.

- **Réseaux de neurones artificiels (ANN)**

C'est un type d'algorithme d'apprentissage automatique inspiré de la structure et du fonctionnement du cerveau humain [25]. Un ANN est constitué de plusieurs couches de nœuds interconnectés (aussi appelés neurones), où chaque nœud est responsable d'une opération spécifique. Les nœuds sont reliés par des arêtes, qui représentent les poids qui déterminent la force de la connexion entre deux nœuds [25]. Certains des types d'ANN les plus courants incluent :

- Perceptron multicouche (MLP : Multi-Layer Perceptron) :

C'est un type de réseau de ANNs, qui se compose d'une ou plusieurs couches cachées de neurones entre les couches d'entrée et de sortie. Les MLP sont le type le plus basique de ANNs.

- Réseau neuronal convolutif (CNN : Convolutional Neural Networks) :

C'est une architecture de réseau de neurones profonds qui apprend directement à partir des données. Ils sont basés sur des opérations de convolution, qui permettent de capturer des formes locales dans les données d'entrée.

- Réseau de Neurones Récurrents (RNN : Recurrent Neural Networks) :

C'est un type d'algorithme de deep learning qui est capable de traiter des données séquentielles. Contrairement aux réseaux de neurones classiques, les RNN ont des connexions récurrentes entre les neurones, ce qui leur permet de conserver une mémoire à court et à long terme des informations précédemment vues dans la séquence. [36]

- LSTMs (Long Short-Term Memory Networks) :

C'est une variante des RNN, conçus pour capturer les relations à long terme dans les données séquentielles. Cette caractéristique rend les réseaux LSTM idéaux pour le traitement et la prédiction de données.

- GRUs (Gated Recurrent Unit) :

Ils sont un outil puissant pour modéliser des données séquentielles et surmonter les limites des RNN, surtout dans les cas où les dépendances à long terme sont importantes.

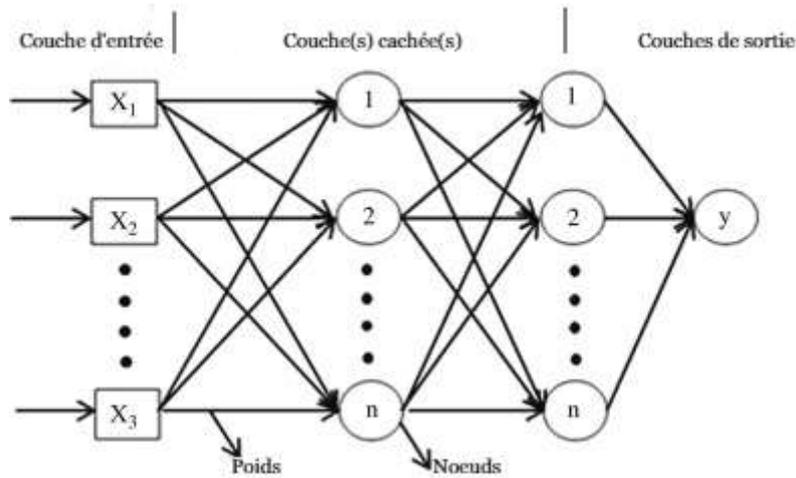


Figure 2-3: Architecture de base de ANN [25]

- **Naïve Bayes**

Cet algorithme est basé sur le théorème de Bayes et l'hypothèse d'indépendance entre les caractéristiques des données. Il est couramment utilisé dans l'apprentissage automatique pour des tâches de prédiction. En pratique, l'utilisation de Naïve Bayes comme algorithme de référence ou en combinaison avec d'autres algorithmes dans une approche ensembliste; est courante dans la littérature. [21]

- **Arbre de décision (Decision Tree)**

Un arbre de décision se compose respectivement de nœuds et de branches qui représentent les décisions et les résultats possibles de ces décisions. On peut graphiquement représenter toutes les solutions possibles d'une décision, on se basant sur certaines conditions ou caractéristiques des données.

la forêt aléatoire (Random Forest) et Arbres à gradient dopé (Gradient Boosted Trees), sont des variations de ces arbres de décision, elles visent à améliorer leur performances en combinant plusieurs arbres ou en modifiant les critères de division. [16, 17]

- **Machines à vecteurs supports (SVMs)**

Les SVMs (Support Vector Machines) sont un type d'algorithme d'apprentissage automatique supervisé utilisé pour les tâches de prédiction. Les SVMs sont basés sur l'idée de trouver un hyperplan qui sépare les données en différentes classes. L'hyperplan est choisi de manière à maximiser la marge, qui est la distance entre l'hyperplan et les points de données les plus proches de chaque classe. [18]

- **K-plus proches voisins (K-NN)**

KNN, ou k-Nearest Neighbors, fonctionne en trouvant les K points de données les plus proches dans l'ensemble d'entraînement d'un nouveau point de données d'entrée, puis en prédisant la classe ou la valeur de l'entrée en fonction de la classe ou de la valeur la plus commune ou la moyenne parmi ses K voisins. [26]

### 2.3.2 Modèles d'ensemble (Ensemble models)

Afin d'améliorer la puissance prédictive globale du modèle ; les modèles d'ensemble en apprentissage automatique ont été conçus, ils s'agit d'une collection de plusieurs modèles individuels combinés [16, 17]. L'idée derrière la modélisation en ensemble est que la combinaison des prédictions de plusieurs modèles donnera un modèle plus précis et plus robuste que n'importe quel modèle individuel. Il existe plusieurs types de modèles d'ensemble en apprentissage automatique, notamment, [19, 21] :

- **Bagging** : cela implique de former plusieurs modèles indépendants sur différents sous-ensembles des données d'entraînement, puis de combiner leurs prédictions. Un exemple populaire de bagging est **Random Forest**.
- **Boosting** : cela implique de former une séquence de modèles, où chaque modèle ultérieur essaie de corriger les erreurs commises par les modèles précédents. Un exemple populaire de boosting est **XGBoost**.
- **Stacking** : cela implique de former plusieurs modèles différents, puis d'utiliser leurs prédictions comme entrée pour un modèle final. Le modèle final apprend à combiner les prédictions des modèles individuels pour faire la prédiction finale.

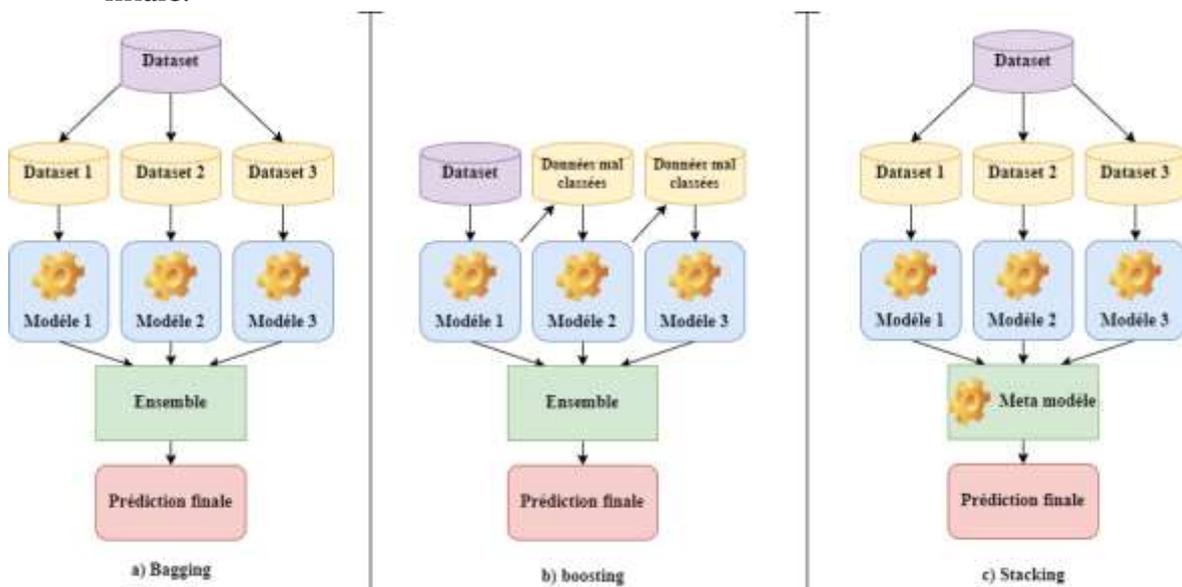


Figure 2-4: Méthodes d'ensembles

Si un modèle simple ne peut pas fournir des prédictions précises en raison de la complexité du problème ou de la taille limitée des données, on peut utiliser les modèles d'ensemble dans ces situations. En combinant des plusieurs prédictions, les modèles d'ensemble peuvent renforcer la précision et la robustesse. Cependant, les modèles d'ensemble peuvent également être coûteux en termes de calcul et nécessitent de plus de ressources par rapport aux modèles individuels.

### 2.3.3 Modèles hybrides (Hybrid models)

La combinaison de différents types de modèles et techniques est un moyen de concevoir et créer un nouveau modèle avec des performances améliorées ; ils sont les modèles hybrides.

Par exemple, un modèle hybride peut combiner un réseau de neurones profond avec l'apprentissage par renforcement, ou un réseau neuronal convolutif avec un réseau neuronal récurrent. [13]

Souvent, ont fait face à des problèmes complexes qui nécessitent plusieurs techniques pour être résolus, ici, les modèles hybrides peuvent être utilisés pour les résoudre.

La différence entre les modèles hybrides et les modèles d'ensemble, c'est que les modèles hybrides combinent de différents types de modèles ou techniques pour créer un modèle entièrement nouveau avec des performances améliorées ; tandis que les modèles d'ensemble combinent de plusieurs modèles individuels pour créer un modèle plus robuste et précis. En se basant sur le problème spécifique et les données analysées, ces modèles peuvent être efficaces pour améliorer les performances des modèles d'apprentissage automatique. [17]

#### **2.4 Les séries chronologique (Time series)**

Les données de séries temporelles, également connues sous le nom de données horodatées, sont un ensemble de points de données disposés dans l'ordre chronologique, ils sont faits séquentiellement au fil du temps. Ces points de données représentent généralement des mesures successives prises à partir de la même source sur une période de temps fixe et sont utilisés pour surveiller les changements au fil du temps [28]. Par conséquent, les données chronologiques se prêtent naturellement à la modélisation comme problème séquentiel, où l'ordre des observations est important et influence les observations subséquentes.

Les données de séries temporelles suivent le comportement d'un sujet ou d'une entité unique à différents intervalles de temps, ces comportements peuvent être également espacés, comme dans le cas des métriques ; ou irrégulièrement espacés, comme dans le cas des événements. [29]

Par exemple, le comportement de la charge électrique à Tébessa collectée à la première heure de chaque jour est un exemple de données de séries temporelles. L'importance du temps en tant qu'axe est ce qui distingue les données de séries temporelles des autres types de données.[30]

La prédiction des observations futures est l'un des objectifs principaux de l'étude des séries temporelles, qui trouve ses applications dans différents domaines tels que l'économie, la finance, la médecine... [33]

Nous pouvons diviser les données de séries temporelles en : séries temporelles **univariées** et **multivariées**. [28]

Lorsque nous sommes concernés par une seule variable dans notre étude, on parle de données de séries temporelles univariées, c'est une variable est observée sur une période de temps. En d'autres termes, l'ensemble de données de séries temporelles univariées ne comporte qu'une colonne de données, qui représente une séquence de valeurs observées à différents moments dans le temps. C'est une prédiction univariée, elle utilise uniquement les valeurs présentes et passées de la série pour prédire les

valeurs futures. D'autre part, les données de séries temporelles multivariées impliquent deux variables ou plus qui sont observées sur une période de temps. Les variables peuvent être liées ou non liées. L'objectif ici est d'étudier la relation entre elles. On l'appelle une prévision multivariée, elle utilise au moins une variable supplémentaire pour prévoir les valeurs futures.

Une série temporelle comprend généralement trois composantes principales ; sont la **tendance**, la **saisonnalité** et les **résidus**. La tendance décrit le mouvement à long ou court terme de la série temporelle. La saisonnalité représente un phénomène périodique, tandis que les résidus représentent des fluctuations aléatoires. Pour comprendre le comportement global de la série temporelle, il est nécessaire d'analyser et de séparer toutes les composantes.[28]

## 2.5 Métriques d'évaluation pour les modèles de prédiction

Les métriques d'évaluation des modèles en apprentissage automatique sont des mesures quantitatives utilisées pour évaluer les performances des modèles de prédiction. Ces mesures permettent de faire une comparaison entre les performances de différents modèles et aussi à l'optimisation des paramètres du modèle pour obtenir de meilleures prédictions. On peut exprimer chaque métrique d'évaluation par une formule mathématique correspondante, qui calcule la mesure quantitative de performance du modèle.

Il existe différentes métriques d'évaluation selon le type de problème de prédiction ; régression ou classification [13]. Dans notre travail, nous nous concentrons sur les métriques de régression.

### 2.5.1 Métriques d'évaluation de régression

- **Erreur moyenne absolue (Mean Absolute Error - MAE) :**

L'erreur moyenne absolue est la mesure de la différence entre les valeurs prédites et les valeurs réelles. Elle est calculée en prenant la moyenne de la valeur absolue de ces différences. [14, 15]

Formule : 
$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

$y_i$  = prédiction,  $x_i$  = valeur réelle,  $n$  = nombre total de points de données

- **Erreur quadratique moyenne (Mean Squared Error - MSE) :**

L'erreur quadratique moyenne est la mesure de la différence entre les valeurs prédites et les valeurs réelles. Elle est calculée en prenant la moyenne des carrés de ces différences. [16, 18]

Formule : 
$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$n$  = nombre de points de données,  $Y_i$  = valeurs observées,  $\hat{Y}_i$  = valeurs prédites

- **Erreur quadratique moyenne racine (Root Mean Squared Error - RMSE):**

L'erreur quadratique moyenne racine est la racine carrée de l'erreur quadratique moyenne. Elle est une mesure plus intuitive de l'erreur que le MSE. [19, 20]

Formule :  $RMSE = \sqrt{MSE}$

- **Coefficient de détermination (R-squared) :**

Le coefficient de détermination est une mesure de la proportion de variance dans les données expliquée par le modèle. Il est calculé en divisant la variance expliquée par la variance totale. [21, 22]

Formule :  $R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$

- **Erreur absolue moyenne pondérée (Mean Absolute Percentage Error - MAPE) :**

L'erreur absolue moyenne pondérée est une mesure de l'erreur moyenne en pourcentage entre les valeurs prédites et les valeurs réelles. Elle est calculée en prenant la moyenne de la valeur absolue de ces différences pondérées par les valeurs réelles. [23, 24]

Formule :  $M = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$

M= MAPE, n = nombre de fois que l'itération de sommation se produit,  $A_t$  = valeur réelle,  $F_t$  = valeur prévue

## 2.6 Datasets

Tableau 2-1 : Comparaison de dataset les plus utilisés dans la littérature

Dataset	PJM	IHEPC
<b>Description</b>	Contient des données de consommation d'énergie horaire de PJM Interconnection LLC, une organisation régionale de transmission aux États-Unis. Couvre plus de 15 ans. Fournit des informations précieuses sur les tendances de consommation d'énergie dans le secteur résidentiel.	Mesures de la consommation d'énergie électrique sur presque 4 ans ; dans le secteur résidentiel.
<b>Source de Données</b>	Données accessibles au public depuis le site web de PJM sous la licence CC0 : domaine public	EDF R&D, Clamart, France
<b>Période Temporelle</b>	Décembre 2004 – août 2018	Décembre 2006 – Novembre 2010

<b>Nombre d'Instances</b>	121272 (Heures)	2075259 (Minutes) 34587 (Heures)
<b>Type</b>	Univarié, Séries chronologiques	Multivarié, Séries chronologiques
<b>Valeurs Manquantes</b>	Non	Oui
<b>Tâches Associées</b>	Régression, Clustering	Régression, Clustering

## 2.7 Travaux connexes

Dans cette section, nous passons en revue la littérature pertinente sur la prévision de la demande d'énergie, en mettant l'accent sur les méthodes et modèles utilisés, données utilisées ainsi que sur leurs résultats.

**Hadjout et al [31]**, ont proposé un modèle ensembliste basé sur des techniques d'apprentissage profond pour prédire la consommation mensuelle d'électricité pour le secteur économique du marché algérien. Le modèle proposé a utilisé une combinaison de LSTM, GRU et CNN. Le modèle a été évalué sur une série chronologique de données provenant de 2000 clients et de 14 ans de consommation mensuelle d'électricité de Bejaia, en Algérie. Les résultats ont montré que le modèle proposé atteint une MAPE et un RMSE respectifs de 3,04% et 60,66.

**Shakouri et Sahed [32]**, ont proposé un modèle de prévision de la consommation d'électricité basé sur des réseaux neuronaux artificiels ANN pour prédire la consommation annuelle pour le marché algérien. Le modèle proposé a utilisé une architecture de réseau neuronal à propagation avant avec un algorithme de rétropropagation. Le modèle a été évalué sur un ensemble de données multivariées provenant du système électrique algérien, et les résultats ont montré que le modèle proposé a atteint une MAPE de 0,033.

**Bezzar et al [33]**, ont proposé une approche de prévision de séries chronologiques basée sur l'analyse de données pour la gestion de la consommation d'électricité domestique. Ils ont proposé un modèle XGBoost qui prend en compte les lacunes mentionnées telles que le mauvais choix du modèle approprié, les ensembles de données de séries chronologiques ont été utilisés directement sans être sérieusement analysés. Les résultats expérimentaux menés sur l'ensemble de données de consommation d'électricité individuelle (IHEPC) ont montré une nette supériorité de leur modèle proposé sur la plupart des modèles basés sur ST et/ou IA proposés dans la littérature, avec respectivement 0,229 et 0,026 pour RMSE et MAPE.

**Kim et Cho [34]**, ont proposé un modèle de réseau neuronal CNN-LSTM pour prédire la consommation d'énergie résidentielle. Le modèle a été entraîné et testé sur l'ensemble de données disponible publiquement "individual household electricity power consumption" (IHEPC). Le modèle proposé a atteint une MAPE de 32,83 ; MSE 0,3549 ; RMSE 0,5957 et MAE 0,3317.

**Han et al [35]**, ont proposé un cadre efficace d'apprentissage profond pour la gestion intelligente de l'énergie dans les réseaux IoT. Le cadre proposé a utilisé une combinaison de modèles LSTM et CNN pour la prévision de la consommation d'énergie. Le cadre a été évalué en utilisant l'ensemble de données PJM. Les résultats ont démontré que le cadre proposé atteint respectivement 0,15 et 3,77 pour MSE et RMSE.

**Khan et al [36]**, ont proposé un modèle efficace de prévision de charge électrique à court terme pour une gestion efficace de l'énergie. Le modèle proposé a utilisé une approche hybride de techniques d'apprentissage profond. Le modèle a été évalué sur l'ensemble de données PJM disponible publiquement, et les résultats ont démontré que le modèle proposé a atteint une bonne performance, avec un RMSE de 3,4.

**Mujeeb et Javaid [37]**, ont proposé deux nouvelles approches, ESAENARX et DE-RELM, pour l'analyse prédictive de gros volumes de données de charge et de prix d'électricité. Ces approches ont utilisé des techniques d'analyse de données telles que l'analyse de régression, le clustering et la classification pour prévoir la consommation d'énergie et les prix. Les approches proposées ont été évaluées sur l'ensemble de données PJM disponible publiquement, et les résultats ont démontré que les approches proposées surpassaient les modèles de référence en termes de MAPE et de RMSE avec 1,09 et 5,24 respectivement pour DE-RELM ; et 1,08, 3,86 respectivement pour ESAENARX.

**Gao et al [38]**, ont proposé un moteur de prévision basé sur plusieurs blocs pour la prédiction de prix et de charge. Le moteur proposé a utilisé une approche hybride qui a combiné différentes combinaisons de réseaux neuronaux testés dans les mêmes conditions de prévision pour montrer leurs capacités. Le moteur a été évalué sur l'ensemble de données PJM disponible publiquement. Les résultats ont montré que le moteur proposé a atteint une performance élevée, avec un RMSE de 1,14 et une MAPE de 0,49.

Tableau 2-2 : Résumé des travaux connexes

Auteurs	Approche proposée	Données utilisées	Résultats
<b>Hadjout et al. [31]</b>	Apprentissage en profondeur (LSTM, GRU, CNN) ensembliste	Séries chronologiques de 2000 clients et 14 ans, Bejaia, Algérie (Dataset Privé)	MAPE= 3,04% RMSE= 60,66
<b>Chekouri et Sahed [32]</b>	Réseaux de neurones artificiels (ANN)	Séries chronologiques de l'Algérie de 1980 à 2016	MAPE=0,033
<b>Bezzar et al [33]</b>	XGBoost	IHEPC Dataset (séries chronologiques)	RMSE= 0,229 MAPE=0,026
<b>Kim et Cho [34]</b>	CNN-LSTM	IHEPC Dataset	MAPE= 32,83 ; MSE= 0,3549 ; RMSE= 0,5957 ; MAE= 0,3317.
<b>Han et al [35]</b>	Apprentissage ensembliste/LSTM/CNN	PJM Dataset (séries chronologiques)	0,15 et 3,77 pour MSE et RMSE.
<b>Khan et al [36]</b>	Modèle hybride	PJM Dataset	Un RMSE de 3,4
<b>Mujeeb et Javaid [37]</b>	Nouvelles approches, ESAENARX et DE-RELM	PJM Dataset	MAPE= 1,09 RMSE= 5,24 pour DE-RELM ; et 1,08, 3,86 respectivement pour ESAENARX.
<b>Gao et al [38]</b>	ANN	PJM Dataset	Un RMSE de 1,14 et une MAPE de 0,49.

D'après la littérature, Il est évident que la majorité des études de recherche reposent sur des données de séries chronologiques.

Nous pouvons également constater à partir de la littérature que la tendance générale est l'utilisation croissante de modèles d'apprentissage automatique (ML) et d'apprentissage profond (DL) pour la prévision de la consommation d'électricité. Les modèles ML et DL ont des performances supérieures à celles d'autres types de modèles dans ce domaine.

Aussi, les datasets de données chronologiques ont été utilisés directement sans être entièrement analysés. Cela peut avoir un impact sur l'exactitude des modèles. Il y a peu de recherches sur l'impact de la qualité des données sur la performance des modèles, ce qui limite la généralisation des résultats.

Il est donc important d'identifier l'impact de la sélection du modèle sur l'exactitude de la prévision. Bien que certains modèles soient plus performants que d'autres, le choix du modèle le plus approprié dépend de plusieurs facteurs, tels que le type de données et la complexité du modèle.

## **2.8 Conclusion**

Dans ce chapitre, nous avons présenté un état de l'art complet sur les types de prédiction et les modèles de prédiction, en soulignant les différentes techniques et types de modèles qui sont couramment utilisés en pratique. Nous avons également abordé un aperçu sur les séries chronologiques et les méthodes d'évaluation des modèles de prédiction, afin de mesurer leur précision et leur fiabilité dans la prédiction de valeurs futures. Dans les suivantes chapitres, nous allons appliquer les connaissances acquises pour construire un modèle de prédiction de la consommation d'énergie. Pour ce faire, nous allons suivre une approche méthodique en plusieurs étapes.

---

## **Chapitre 03 :**

**Cas d'étude : prédiction de consommation d'électricité**

---

## Chapitre 3: Cas d'étude : prédiction de consommation

### d'Électricité

#### 3.1 Introduction

Ce chapitre présente la méthodologie et les résultats en utilisant le modèle LSTM. La section méthodologie décrit les étapes impliquées dans la visualisation des données, la sélection du modèle, l'entraînement du modèle et l'évaluation du modèle. La section des résultats et discussions analyse les performances du modèle proposé et les compare avec d'autres modèles existants.

#### 3.2 Méthodologie

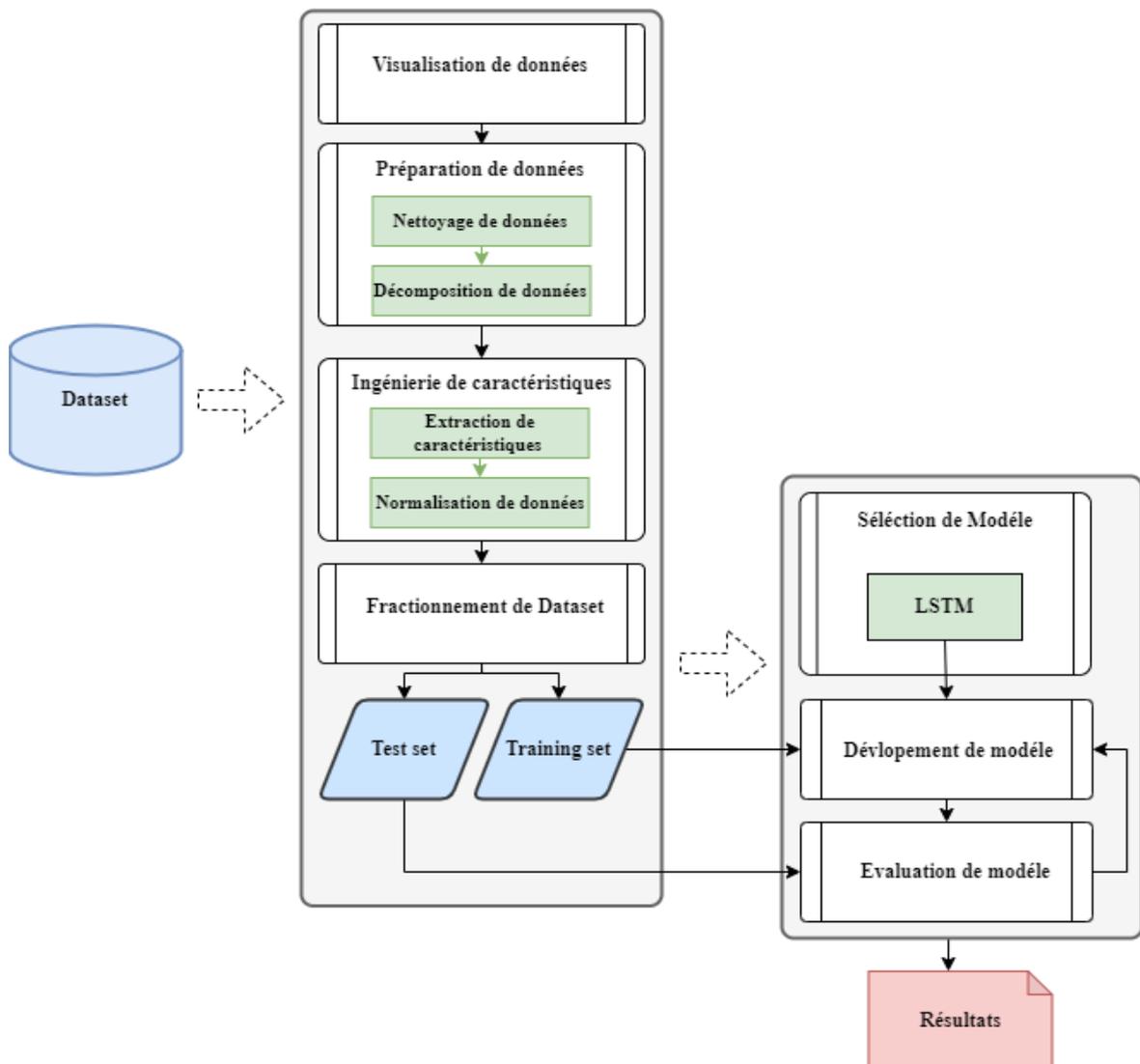


Figure 3-1: Méthodologie proposé de travail

### 3.2.1 Dataset

Les données utilisées dans cette étude sont un dataset contenant des données de consommation d'énergie horaire de « PJM Interconnection LLC (PJM) », une organisation régionale de transmission aux États-Unis qui exploite un système de transmission électrique résidentiel. Cet ensemble de données s'étend sur plus de 15 ans et fournit des informations précieuses sur les tendances de consommation d'énergie.

Les données de consommation d'énergie électrique horaire proviennent du site Web de PJM et sont accessibles au public sous la licence "CC0 : domaine public", ce qui les rend accessibles à des fins de recherche. L'ensemble de données est extrait de **Kaggle** et peut être trouvé sur le lien [39 ,40].

Il s'agit d'un ensemble de données de séries chronologiques univariées qui contient des données de consommation d'énergie horaire en mégawatts (MW). Nous avons choisi cette dataset pour notre travail pour les raisons suivantes :

- **Pertinence** : Le dataset est pertinent pour notre recherche, il fournit une ressource précieuse pour l'utilisation afin de prévoir ; et pour les techniques de prétraitement et d'analyse de données.
- **Taille** : dataset est suffisamment grand pour être un défi intéressant à travailler avec, et pour l'utiliser afin de former et de tester notre modèle.
- **Disponibilité** : dataset est disponible publiquement ce qui permet d'une utilisation académique, il est souvent utilisé dans les travaux universitaires et a été utilisé dans des études de recherche précédentes, ce qui peut fournir un point de référence utile ; il est connu aussi pour sa grande qualité.

### 3.2.2 Visualisation de données

La visualisation des données est le processus de création de représentations visuelles de séries chronologiques afin de comprendre les distributions de données, d'obtenir des informations et de communiquer les résultats. Afin de pouvoir analyser les composantes clés de séries chronologiques, telles que le tendance, saisonnalité et la composante résiduelle ; nous visualisons ces données qui ne sont peut-être pas évidentes en observant simplement les chiffres bruts.

L'observation et l'analyse visuelle de ces graphiques est souvent utile pour modéliser la série chronologique et avoir une compréhension des différentes composantes qui la constituent.

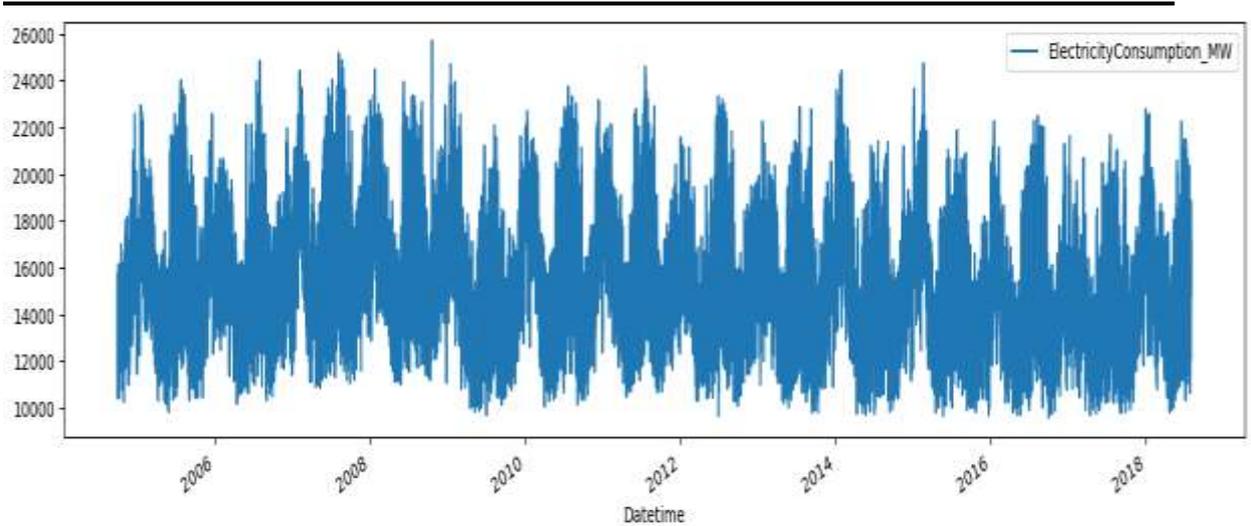


Figure 3-2: Consommation d'Énergie en mégawatts (MW) de 2004 à 2019

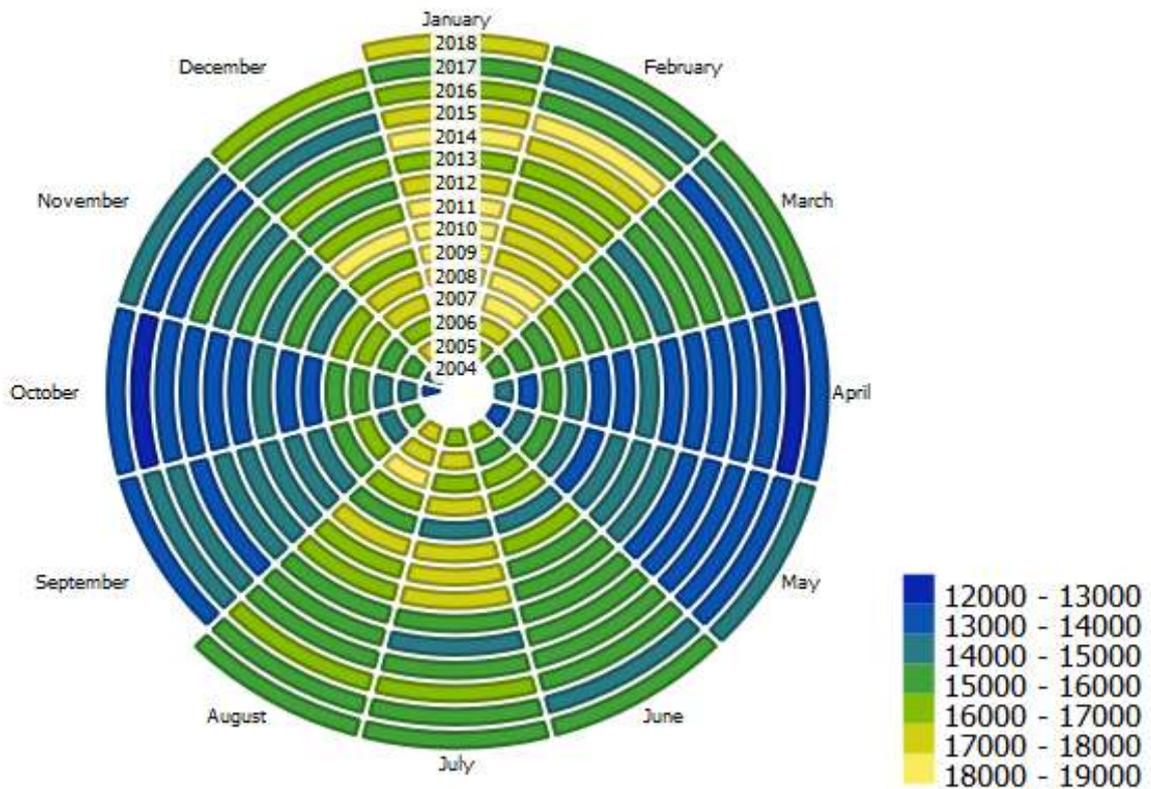


Figure 3-3: Consommation d'Énergie en mégawatts (MW) de 2004 à 2019

La figure 2-3 représente une visualisation en **spiralogramme** des années de 2004 à 2018, telle que présentée dans la figure précédente. Chaque tour de la spirale représente une année entière décomposée en mois. La progression du temps se fait de l'intérieur vers l'anneau extérieur. En observant la variation des couleurs des couches, nous pouvons prédire si des variations se trouvent dans les données. Nous pouvons constater que la consommation d'énergie est élevée pendant les mois d'hiver et d'été, et qu'elle est modérée à faible pendant les autres mois.

---

### 3.2.3 Préparation de données

#### i. Nettoyage de données

Comme première étape de notre travail, nous examinons le dataset pour détecter tout occurrences de valeurs manquantes, de nuls, de doublons ou de valeurs aberrantes. Notre dataset ne contient pas de nuls ou de doublons à traiter. En identifiant et en résolvant tout problème potentiel, nous pouvons améliorer la précision et la fiabilité de notre analyse.

#### ii. Décomposition de données

La décomposition de séries chronologiques est une méthode utilisée pour décomposer une série chronologique en ses composantes, telles que la tendance, la saisonnalité et les résidus, et pour vérifier la stationnarité de la série chronologique. L'analyse des données de la série chronologique est une étape nécessaire, afin d'estimer et de distinguer toutes les composantes présentes, pour but de souligner l'impact de chaque composante sur le comportement global de la série chronologique. La figure 3-3 montre les résultats obtenus.

Le graphique de tendance de notre ensemble de données, est présenté dans la figure 3-3(a). Le composant de tendance représente le mouvement à long terme des données. On peut voir du graphique qu'il y a une augmentation de la consommation d'énergie pendant les périodes chaudes et froides, et une légère diminution pendant les autres périodes.

Aussi dans la figure 3-3(a), le schéma de changement de la consommation diffère en fonction de la durée considérée. Sur une longue période (comme plusieurs années), la tendance de la consommation n'est pas une ligne droite mais un courbe, indiquant une relation non linéaire entre le temps et la consommation. Ce comportement signifie que le taux de changement de la consommation et n'est pas constant et varie au fil du temps. Cependant, sur une période plus courte, la tendance de la consommation semble être une ligne droite, indiquant une relation linéaire entre le temps et la consommation. Cela signifie que le taux de changement de la consommation est constant et prévisible sur cette période plus courte.

La figure 3-3(b) illustre le composant de saisonnalité. Le composant de saisonnalité représente le motif régulier des données qui se répète sur une période fixe, telle que quotidienne, hebdomadaire, mensuelle ou annuelle. La saisonnalité peut avoir une valeur positive ou négative, indiquant si les données ont tendance à être plus élevées ou plus basses pendant certaines périodes.

Cette information peut être utile pour comprendre et prédire les valeurs futures de consommation d'électricité.

La figure 3-3(c) montre les résidus. Les résidus sont les fluctuations aléatoires et imprévisibles des données. En d'autre termes, c'est la variation des données qui ne peut être expliquée par les composants de tendance ou de saisonnalité.

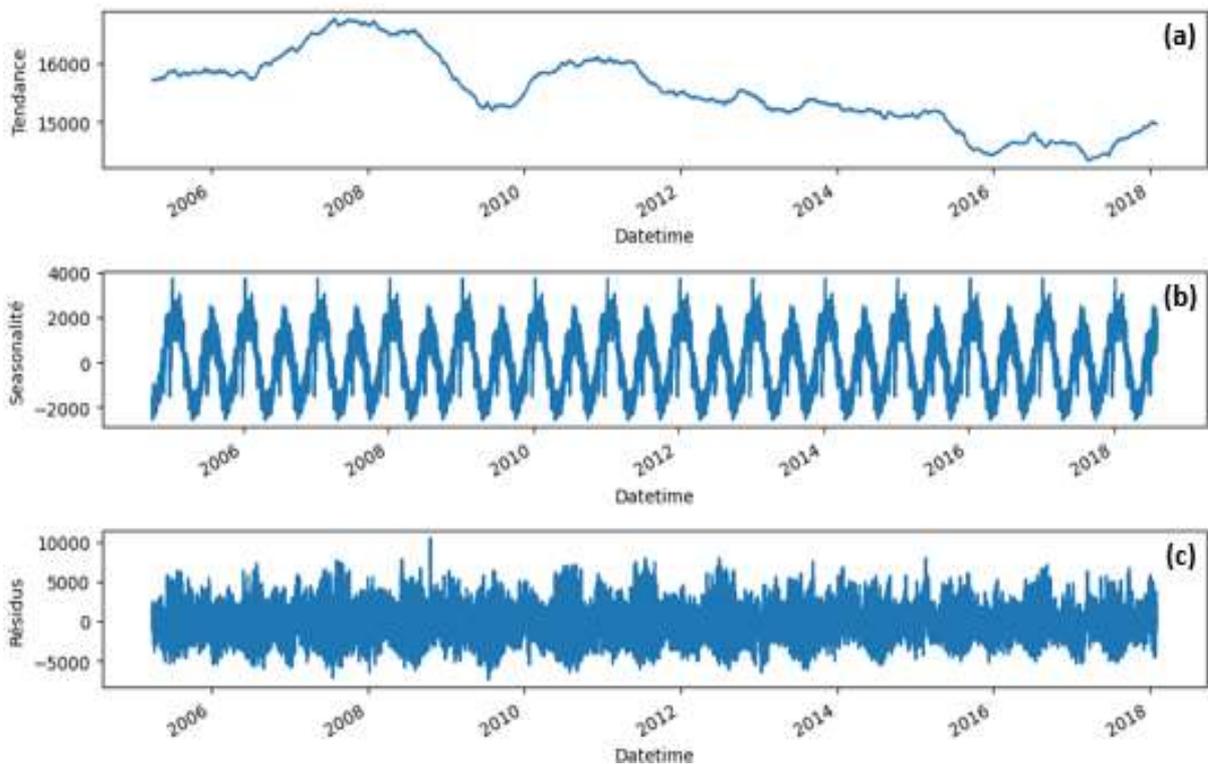


Figure 3-4: Décomposition de la séries temporelles (Tendance, saisonnalité et résidus)

### 3.2.4 Ingénierie de caractéristiques

#### i. Extraction de caractéristiques

Le processus d'extraction d'informations pertinentes ou de caractéristiques à partir de données de séries temporelles, s'appelle l'extraction de caractéristiques de séries temporelles.

L'extraction de caractéristiques a comme objectif d'identifier et de capturer des motifs, des tendances ou d'autres caractéristiques dans la série temporelle qui peuvent être utilisés dans la tâche de prédiction. La possibilité d'améliorer l'exactitude et l'efficacité de ces modèles reposent sur les techniques d'extraction de caractéristiques efficaces.

Ce processus revêt une importance significative dans tout projet de science des données impliquant l'apprentissage automatique. Les caractéristiques choisies servent d'entrée pour le modèle d'apprentissage utilisé.

Nous avons extrait de notre données les caractéristiques suivantes :

- **hour** : valeurs de l'heure de la journée (00:00=0, 23:00=24) pour chaque donnée. L'heure de la journée peut être utile pour identifier les tendances quotidiennes dans les données, telles que les pics d'activité à certains moments de la journée.
- **dayofweek** : semaine: valeurs du jour de la semaine (lundi=0, dimanche=6) pour chaque donnée. Le jour de la semaine peut aider à capturer les tendances

hebdomadaires dans les données, telles que les changements de consommation en semaine versus week-end.

- **quarter** : valeurs du trimestre de l'année (1-4) pour chaque donnée. Le trimestre de l'année peut être utile pour identifier les tendances saisonnières dans les données, telles qu'une consommation accrue à différents moments de l'année.
- **month** : valeurs du mois de l'année (1-12) pour chaque donnée. Le mois de l'année peut également être utile pour identifier les tendances saisonnières dans les données.
- **year** : valeurs de l'année (2004-2018) pour chaque donnée. L'année peut être utile pour identifier les tendances à plus long terme dans les données, telles que les changements graduels de comportement des consommateurs au fil du temps.
- **dayofyear** : valeurs du jour de l'année (1-365 ou 1-366) pour chaque donnée. Le jour de l'année peut être utile pour identifier les tendances cycliques dans les données, telles que les changements de consommation d'énergie liés aux saisons changeantes.

Tableau 3-1 : Dataset après l'étape d'extraction de caractéristiques

Datetime	EnergyConsumption_MW	hour	dayofweek	quarter	month	year	dayofyear
2004-10-01 01:00:00	12379.0	1	4	4	10	2004	275
2004-10-01 02:00:00	11935.0	2	4	4	10	2004	275
2004-10-01 03:00:00	11692.0	3	4	4	10	2004	275
2004-10-01 04:00:00	11597.0	4	4	4	10	2004	275
2004-10-01 05:00:00	11681.0	5	4	4	10	2004	275
...	...	...	...	...	...	...	...
2018-08-02 20:00:00	17673.0	20	3	3	8	2018	214
2018-08-02 21:00:00	17303.0	21	3	3	8	2018	214
2018-08-02 22:00:00	17001.0	22	3	3	8	2018	214
2018-08-02 23:00:00	15964.0	23	3	3	8	2018	214
2018-08-03 00:00:00	14809.0	0	4	3	8	2018	215

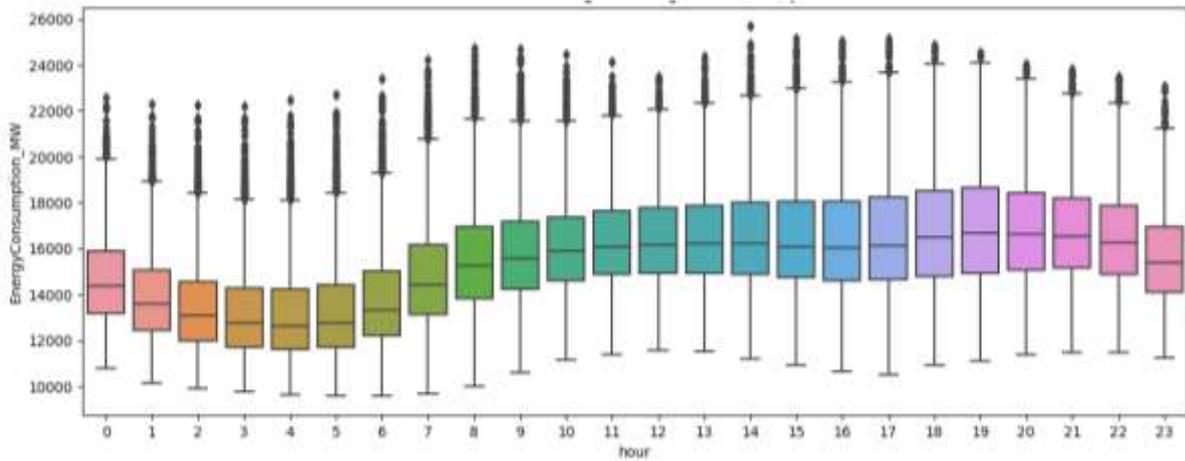


Figure 3-5: Consommation d'Énergie en mégawatts (MW) par heure

La Figure 3-5 montre que la consommation atteint généralement son point le plus bas à 04 :00, et continue à augmenter jusqu'à atteindre son maximum à 19 :00.

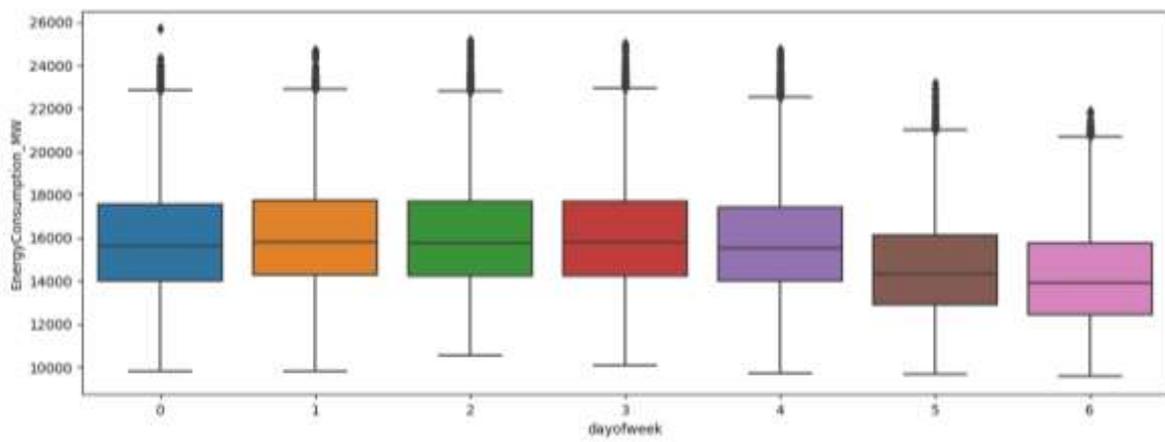


Figure 3-6: Consommation d'Énergie en mégawatts (MW) par semaine

Figure 3-6 montre que pendant les jours de semaine, la consommation est à un niveau élevé, mais elle diminue jusqu'à atteindre son point le plus bas pendant le week-end.

## ii. Normalisation de données

Pour garantir une échelle similaire entre différentes variables dans une série chronologique et réduire l'effet des valeurs aberrantes ; c' est essentielle d'appliquer la normalisation des données. La mise à l'échelle Min-Max est une méthode courante, ramenant les caractéristiques à un intervalle fixe [0, 1] [33]. La normalisation Min-Max est effectuée en utilisant la formule suivante :

$$\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Dans la formule, x représente la valeur de la caractéristique que nous essayons de normaliser, tandis que  $x_{min}$  et  $x_{max}$  représentent respectivement la plus petite et la plus grande valeur observée pour cette caractéristique. [33]

### 3.2.5 Fractionnement de Dataset

Les données ont été divisées en ensembles d'entraînement et de test lors de l'étude. La figure 3-7 illustre la division des données en ensembles d'entraînement et de test, Le ratio de division des ensembles d'entraînement et de test était d'environ 70-30%. la division était basée sur la série temporelle du 01-01-2015.

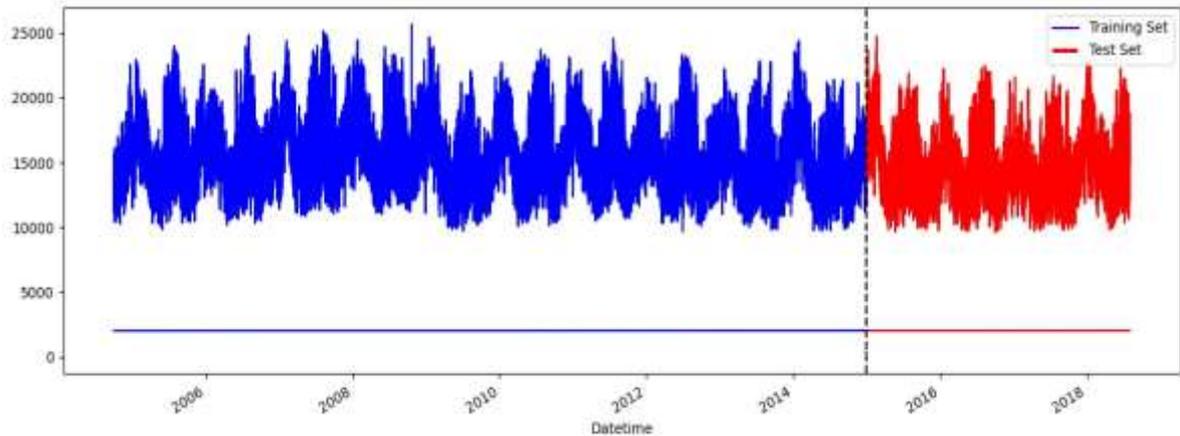


Figure 3-7: Fractionnement de Dataset

### 3.2.6 Sélection de modèle

Nous utiliserons un type de réseau de neurones, en particulier Long Short-Term Memory (LSTM), dans notre recherche pour les raisons suivantes :

- Les séries chronologiques peuvent être modélisées sous forme de séquences comme nous avons noté dans le chapitre 2 et notre recherche est basée sur des données de séries chronologiques. Par conséquent, cela est considéré comme un problème séquentiel, et pour le résoudre, nous devons utiliser un modèle d'apprentissage automatique capable de comprendre et d'apprendre de la nature séquentielle des données.
- LSTM est l'un de ces modèles conçus pour traiter des données séquentielles en utilisant des informations des étapes précédentes pour faire des prédictions à chaque nouvelle étape. C'est un type de réseau de neurones récurrents (RNN) particulièrement utile pour traiter les dépendances à long terme dans les données séquentielles.
- La tendance actuelle dans la littérature est d'utiliser des modèles d'apprentissage automatique (ML) et d'apprentissage profond (DL), comme nous avons également observé ultérieurement, ce qui soutient notre choix d'utiliser LSTM.
- En outre, pour obtenir les meilleurs résultats, LSTM dépend de plusieurs facteurs, notamment de la taille des données, qui est dans notre cas une dataset large et importante ; et de la qualité des données d'entrée, qui a été assurée par le processus complet d'analyse de données qui nous avons établie.

### 3.2.7 Développement de modèle

Nous avons commencé à partir du début pour construire et mettre en œuvre notre modèle d'apprentissage automatique LSTM, afin de prévoir la consommation d'énergie. Pour le mettre en œuvre, nous avons utilisé les bibliothèques Python suivantes : **pandas**, **NumPy**, **statsmodels**, **matplotlib**, **TensorFlow**, **keras** et **sklearn**, importées dans "Colab".

Notre modèle LSTM se compose de quatre couches au total :

La première couche est une couche LSTM avec 40 cellules de mémoire. Elle sert de couche d'entrée initiale, elle attend des données séquentielles avec une longueur de séquence prédéfinie et sept caractéristiques. Le rôle de cette couche est de traiter la séquence d'entrée, de capturer les dépendances entre les pas de temps et de générer des séquences en sortie.

Afin d'éviter le surajustement (overfitting), une couche de désactivation (dropout) est insérée après la première couche LSTM. Elle a comme rôle de définir de manière aléatoire une fraction (10 %) des unités d'entrée à zéro pendant l'entraînement. Cette technique de régularisation aide le modèle à généraliser mieux les données non vues.

La deuxième couche est une autre couche LSTM avec 40 cellules de mémoire. Elle a la même architecture que la première couche LSTM et a pour but de traiter davantage l'information séquentielle et de générer des séquences en sortie. Elle est suivie d'une autre couche de désactivation pour atténuer le surajustement.

La troisième couche est la dernière couche LSTM du modèle, elle est aussi composée de 40 cellules de mémoire, mais contrairement aux deux couches LSTM précédentes, ne renvoie aucune séquence. Elle résume plutôt l'information apprise à partir de la séquence et produit une seule sortie. Une couche de désactivation est insérée après cette couche pour régulariser le modèle.

La dernière couche est une couche dense (dense layer) avec une seule unité. Elle effectue une transformation linéaire sur l'entrée et produit une seule valeur de sortie. Cette couche agit comme la couche de sortie du modèle et est responsable de la génération de prédictions basées sur les motifs appris à partir des données séquentielles.

En résumé, le modèle comprend trois couches LSTM qui traitent les données séquentielles, capturant les dépendances entre les pas de temps et produisant des séquences en sortie. Les couches de désactivation aident à prévenir le surajustement et la dernière couche dense fournit la prédiction du modèle.

### 3.2.8 Evaluation de modèle

Il est important d'évaluer la performance des modèles en utilisant des mesures qui peuvent capturer efficacement l'exactitude et la précision des prévisions ; parce que dans le contexte de la prévision de la consommation d'énergie, il est important d'avoir un modèle avec une précision élevée.

En utilisant une combinaison de RMSE, MAE, MAPE et  $R^2$  nous pouvons fournir une évaluation complète de la performance du modèle et aider à identifier les domaines spécifiques où le modèle peut être en sous-performance. RMSE est utile lorsque les grandes erreurs ont un impact plus important sur la performance globale que les erreurs plus petites. MAE est utile lorsque toutes les erreurs, grandes ou petites, sont également importantes pour la performance globale. MAPE est utile lorsqu'il est important d'évaluer la performance du modèle en termes d'erreur en pourcentage.  $R^2$  à quel point le modèle LSTM s'adapte bien au problème de prédiction.

Pour avoir une vision plus globale des avantages et des limites du modèle, et utiliser ces informations pour améliorer sa performance ; nous avons utilisé plusieurs mesures d'évaluation.

### 3.3 Résultats et discussion

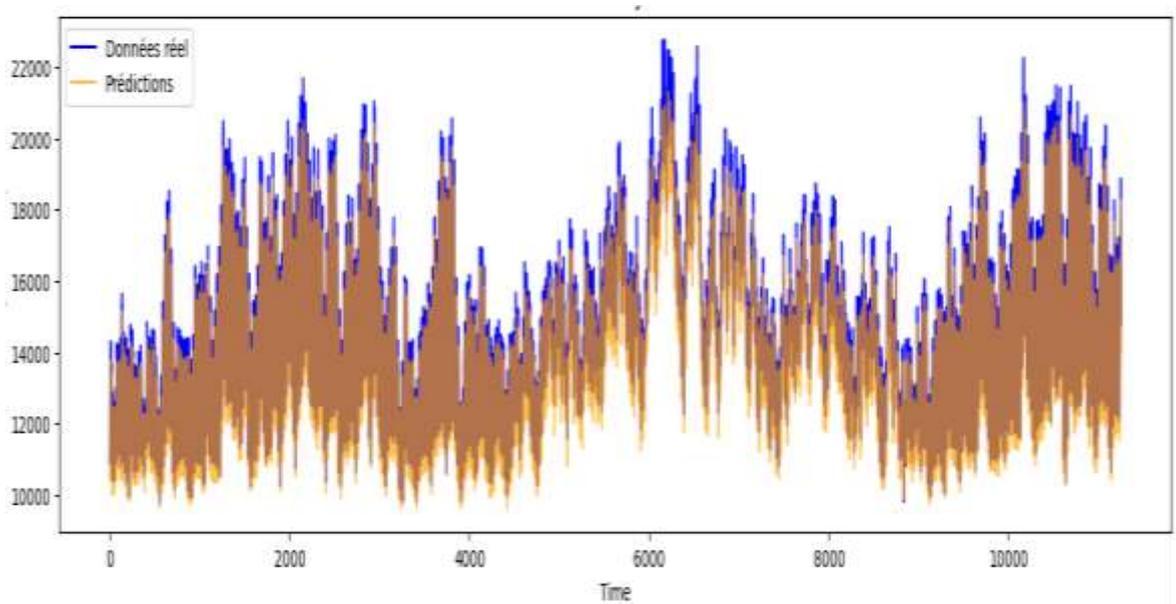


Figure 3-8: Résultats de prédiction avec LSTM sur le test set

En analysant les figures 3-8 et 3-9, il est clairement observable que notre modèle LSTM démontre une capacité remarquable à effectuer des prédictions précises, suivant de près les points de données réels.

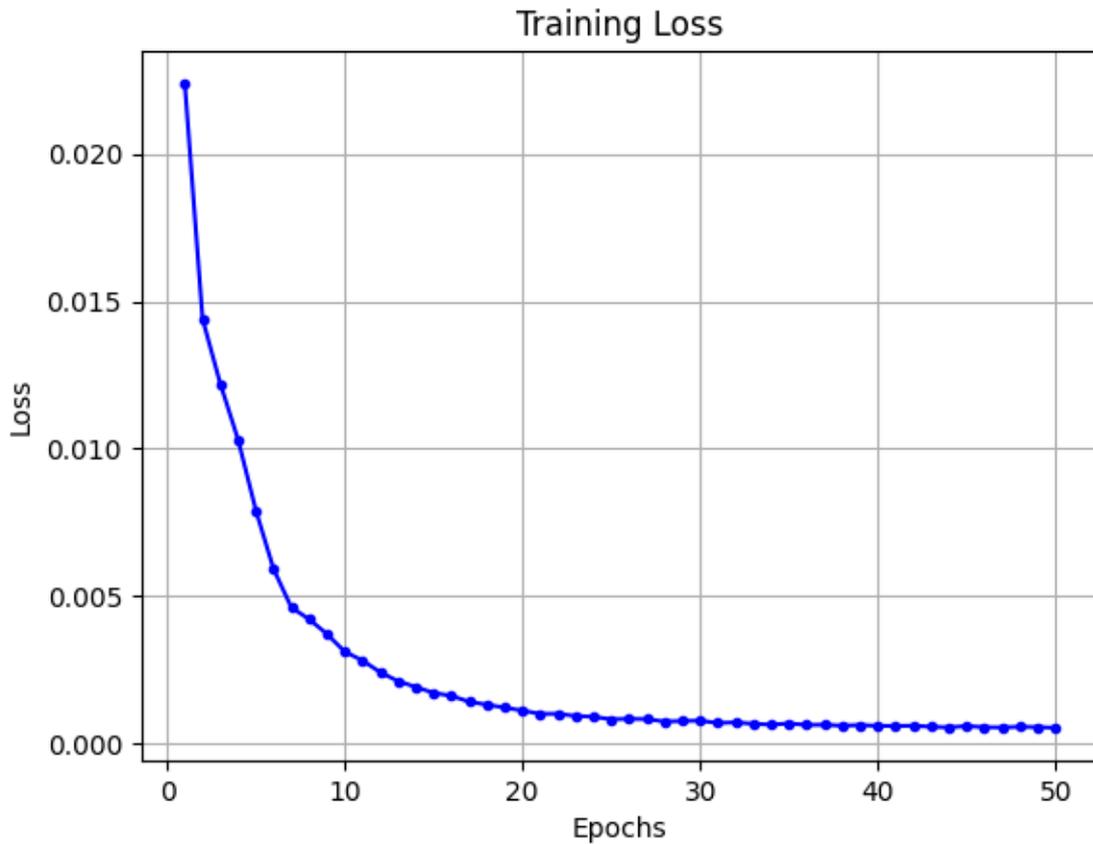


Figure 3-9: Loss durant la phase d'entraînement

Les performances du modèle sur l'ensemble d'entraînement peuvent être évaluées en se basant sur diverses métriques : un RMSE de 0,03, un MAPE de 0,08, un MAE de 0,02 et un coefficient de détermination ( $R^2$ ) de 0,96 ; comme le montre la figure 3-10.

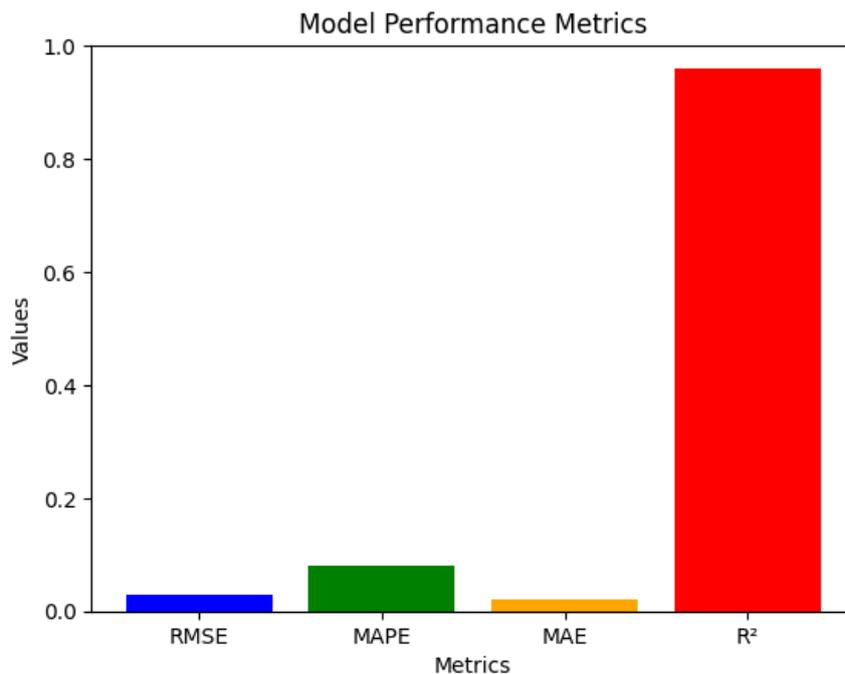


Figure 3-10: Performance de notre modèle LSTM

Ces résultats témoignent de la robustesse et de la fiabilité de notre modèle LSTM dans la prédiction de la consommation d'électricité. La faible valeur du RMSE indique que les prédictions sont proches des valeurs réelles, tandis que le MAPE et le MAE révèlent une précision élevée avec des erreurs minimales. De plus, le coefficient de détermination élevé ( $R^2$ ) de 0,96 indique que notre modèle est capable d'expliquer 96% de la variance observée dans les données.

Notre modèle LSTM démontre observations positives ce qui nous donnent une grande confiance dans son utilisation pour avoir des prévisions précises de la consommation d'électricité, ce qui peut être extrêmement bénéfique pour la planification et l'optimisation des ressources énergétiques..

### 3.3.1 Comparaison avec les travaux connexes

Pour garantir une comparaison ciblée et concise, nous évaluerons exclusivement les performances de notre modèle LSTM par rapport aux études qui ont utilisé le même dataset que le nôtre.

Tableau 3-2 : Comparaison avec les travaux connexes

Travail	RMSE	MAPE	MAE	$R^2$
LSTM-CNN [35]	3,77	–	–	–
Hybride [36]	3,4	–	–	–
Hybride [37]	5,24	1,09	–	–
ANN [38]	1,14	0,49	–	–
<b>Proposé LSTM</b>	<b>0,03</b>	<b>0,08</b>	<b>0,02</b>	<b>0,96</b>

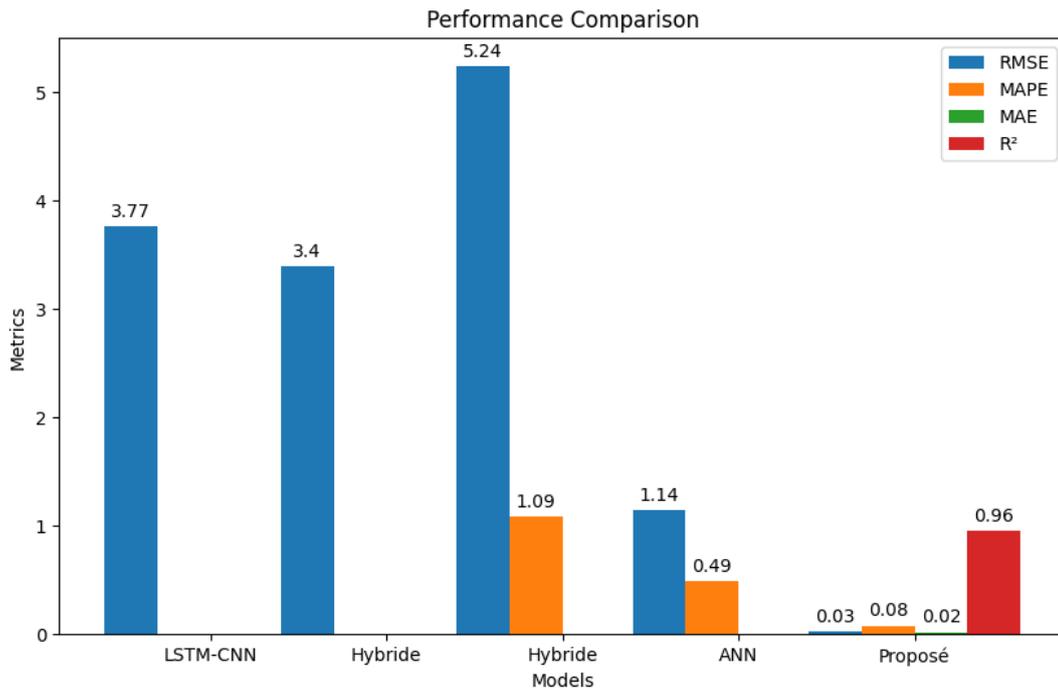


Figure 3-11: Comparaison avec les travaux connexes

Le tableau 3-2 et la figure 3-11, met en évidence la comparaison des performances de notre modèle avec les travaux connexes. Les mesures de performance utilisées sont le RMSE le MAPE le MAE et R<sup>2</sup>.

En analysant les résultats, on peut observer que le travail [37] présente un RMSE élevé de 5,24, indiquant une erreur de prédiction considérablement élevée par rapport aux autres travaux. En revanche, notre modèle propose un RMSE remarquablement faible de 0,03, ce qui suggère une grande précision dans la prédiction de la consommation d'énergie.

En termes de MAPE, notre modèle affiche un résultat de 0,08, ce qui est nettement inférieur à la plupart des travaux connexes. Le travail [37] présente le MAPE le plus élevé de 1,09, ce qui suggère une plus grande déviation par rapport aux autres modèles.

En ce qui concerne le MAE, notre modèle LSTM obtient un MAE très faible de 0,02, démontrant ainsi une précision élevée dans les prédictions de consommation d'énergie.

Enfin, le coefficient de détermination (R<sup>2</sup>) évalue la proportion de la variance des données expliquée par le modèle de prédiction. Notre modèle affiche un coefficient R<sup>2</sup> exceptionnel de 0,96, ce qui suggère une excellente capacité à expliquer la variation de la consommation d'énergie par rapport aux autres modèles.

Les résultats obtenus indiquent une précision élevée, une faible erreur et une capacité élevée à expliquer la variation de la consommation d'énergie. Ce qui démontre de

manière convaincante les performances supérieures de notre modèle par rapport aux autres travaux connexes. Ces performances renforcent la crédibilité et l'efficacité de notre modèle LSTM pour les tâches de prédiction de la consommation d'énergie dans le contexte de notre thèse.

### **3.4 Conclusion**

Dans ce chapitre, nous avons abordé les principaux aspects de notre étude portant sur l'utilisation des modèles d'apprentissage automatique pour la prédiction de la consommation d'énergie. Au cours de cette étude, nous avons pris soin d'explorer les différentes étapes nécessaires à une analyse approfondie, nous avons exposé en détail la méthodologie que nous avons suivie, depuis la visualisation des données jusqu'à l'évaluation de modèle et l'obtention des résultats ainsi qu'une comparaison avec les travaux précédents.

---

## Conclusion Générale & perspectives

Dans cette thèse, nous avons exploré l'application d'un modèle d'IA pour la prédiction de la consommation d'énergie, en utilisant spécifiquement des données de séries temporelles et le modèle Long Short-Term Memory (LSTM). Notre objectif était de répondre aux limitations présentes dans la littérature concernant la qualité des données et la sélection du modèle. Nous avons mené un processus complet d'analyse des données pour garantir la qualité et la fiabilité des données d'entrée, et nous avons choisi le modèle LSTM en raison de ses capacités de pointe dans la résolution de problèmes séquentiels.

Grâce à nos recherches, nous avons réalisé des progrès significatifs dans la prédiction précise de la consommation d'énergie. Les mesures d'évaluation, y compris l'erreur quadratique moyenne (RMSE), l'erreur absolue moyenne en pourcentage (MAPE), l'erreur absolue moyenne (MAE) et le coefficient de détermination ( $R^2$ ), reflètent les performances élevées et la fiabilité de notre modèle prédictif. En particulier, nous avons obtenu des résultats exceptionnels avec  $RMSE = 0.03$ ,  $MAPE = 0.08$ ,  $MAE = 0.02$  et  $R^2 = 0.96$ , démontrant ainsi l'efficacité de notre approche.

En surmontant les défis liés à la qualité des données et à la sélection du modèle, nous avons contribué au domaine de la prédiction de la consommation d'énergie. Nos recherches soulignent la nécessité d'une analyse minutieuse des données et mettent en évidence la pertinence et l'efficacité des modèles LSTM pour parvenir à une prévision précise de la consommation d'énergie. Les résultats de cette étude ont des conséquences notables pour la gestion de l'énergie, l'allocation des ressources et la prise de décision dans différents domaines.

Dans les futurs travaux, nous concentrerons sur l'amélioration de la précision des modèles, l'exploration de leur application dans les pays en développement et l'analyse de l'impact de la qualité des données sur leur performance. Nous sommes convaincus que ces avancées contribueront à une meilleure compréhension et utilisation des modèles de ML dans le domaine de la prévision de la demande d'énergie.

---

## Références

1. EIA: What is energy? [en ligne]. Disponible sur <https://www.eia.gov/energyexplained/what-is-energy/> . [Accessed: Feb. 7, 2023]
2. -W. H. Wisner, Energy Resources: Occurrence, Production, Conversion, Use, 1st ed. New York, NY: Springer, 1999, xv, 377 pages. doi: 10.1007/978-1-4612-1226-3.
3. -Y. Demirel, Energy: Production, Conversion, Storage, Conservation, and Coupling, 3rd ed. Cham: Springer, 2021, xxii, 650 pages. doi: 10.1007/978-3-030-56164-2.
4. -V. Smil, Energy: A Beginner's Guide, Oneworld Publications, Oxford, UK, 2006. ISBN: 1-85168-452-2.
5. -A. Ghatak, B. Roy, B. Malakar, D. Chakraborty, S. Dutta, D. Roy, and A. Pandey, "A Literature Review on Different Renewable Energy Resources," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 11, pp. 356-361, Nov. 2021. ISSN: 2278-0181. doi : 10.17577/IJERTCONV9IS11045.
6. -H. Ritchie, M. Roser, and P. Rosado, "Energy," OurWorldInData.org, 2022. [Online]. Available: <https://ourworldindata.org/energy> . [Accessed: Mar. 7, 2023].
7. IEA (2022), Energy Statistics Data Browser, IEA, Paris <https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser>
8. Understanding the Global Energy Crisis (Purdue Studies in Public Policy) by Richard A. Simmons and Eugene D. Coyle | Mar 15, 2014, chapter 2.
9. <https://www.who.int/news/item/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution> (consulté le 07/02/2023)
10. Dr. Ahmed Qarareh Yusef Kabbas, Department of Curriculum and Instruction, Tafila Technical University, Jordan, THE DEGREE OF RATIONALIZATION OF ENERGY CONSUMPTION CONCEPTS INCLUDED IN THE SCIENCE TEXTBOOKS OF THE BASIC STAGE IN JORDAN AND THE STUDENTS' ATTITUDES TOWARDS THEM, European Scientific Journal, October 2015, edition vol.11.
11. -World Economic Forum, "Harnessing Artificial Intelligence to Accelerate the Energy Transition," Geneva, Switzerland, 2019. [Online]. Available: <https://www.weforum.org/whitepapers/harnessing-artificial-intelligence-to-accelerate-the-energy-transition/> . [Accessed: Mar. 7, 2023].
12. Kazemi A, Hosseinzadeh M. A multi-level fuzzy linear regression model for forecasting industry energy demand of Iran. Procedia-Social and Behavioral Sciences. 2012 Dec 31; 41:342-8.
13. Ghalekhondabi, Iman & Ardjmand, Ehsan & Weckman, Gary & Young, William. (2017). An overview of energy demand forecasting methods published in 2005–2015. Energy Systems. 8. 10.1007/s12667-016-0203-y.
14. Kavaklioglu K, Ceylan H, Ozturk HK, Canyurt OE. Modeling and prediction of Turkey's electricity consumption using artificial neural networks. Energy Conversion and Management. 2009 Nov 30 ;50(11) :2719- 27.
15. Antara Mahanta Barua, Pradyut Kumar Goswami, A Survey on Electric Power Consumption Prediction Techniques, International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 10 (2020), pp. 2568-2575 © International Research Publication House. <https://dx.doi.org/10.37624/IJERT/13.10.2020>, 2568-2575,
16. Mosavi, Amir & Bahmani, Abdullah. (2019). Energy Consumption Prediction Using Machine Learning; A Review.
17. Jui-Sheng Chou, Duc-Son Tran, Forecasting Energy Consumption Time Series using Machine Learning Techniques based on Usage Patterns of Residential Householders, Energy (2018), doi: 10.1016/j.energy.2018.09.144

18. Yixuan Wei, Xingxing Zhang, Yong Shi, Liang Xia, Song Pan, Jinshun Wu, Mengjie Han, Xiaoyun Zhao, A review of data-driven approaches for prediction and classification of building energy consumption, *Renewable and Sustainable Energy Reviews*, Volume 82, Part 1, 2018, Pages 1027-1047, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2017.09.108>.
19. Ying Sun, Fariborz Haghighat, Benjamin C.M. Fung, A review of the-state-of-the-art in data-driven approaches for building energy prediction, *Energy and Buildings*, Volume 221, 2020, 110022, ISSN 0378-7788, <https://doi.org/10.1016/j.enbuild.2020.110022>.
20. Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, Kwok Wei Shah, A review on time series forecasting techniques for building energy consumption, *Renewable and Sustainable Energy Reviews*, Volume 74, 2017, Pages 902-924, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2017.02.085>.
21. Zeyu Wang, Ravi S. Srinivasan, A review of artificial intelligence-based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models, *Renewable and Sustainable Energy Reviews*, Volume 75, 2017, Pages 796-808, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2016.10.079>.
22. L. Suganthi, Anand A. Samuel, Energy models for demand forecasting—A review, *Renewable and Sustainable Energy Reviews*, Volume 16, Issue 2, 2012, Pages 1223-1240, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2011.08.014>.
23. Hammad, Mahmoud A. & Jereb, Borut & Rosi, Bojan & Dragan, Dejan. (2020). Methods and Models for Electric Load Forecasting: A Comprehensive Review. *Logistics & Sustainable Transport*. 11. 51-76. <http://dx.doi.org/10.2478/jlst-2020-0004>.
24. Vujovic, Zeljko. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*. Volume 12. 599-606. 10.14569/IJACSA.2021.0120670.
25. N.J. Sairamya, L. Susmitha, S. Thomas George, M.S.P. Subathra, Chapter 12 - Hybrid Approach for Classification of Electroencephalographic Signals Using Time-Frequency Images With Wavelets and Texture Features, Editor(s): D. Jude Hemanth, Deepak Gupta, Valentina Emilia Balas, In *Intelligent Data-Centric Systems, Intelligent Data Analysis for Biomedical Applications*, Academic Press, 2019, Pages 253-273, ISBN 9780128155530, <https://doi.org/10.1016/B978-0-12-815553-0.00013-6>.
26. Saleh Seyedzadeh, Farzad Pour Rahimian, Ivan Gleski and Marc Roper, Machine learning for estimation of building energy consumption and performance: a review, *Visualization in Engineering* (2018) 6:5. <https://doi.org/s40327-018-0064-7>.
27. Cheng, B., & Titterton, D. M. (1994). Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, 9(1), 2-30. <http://www.jstor.org/stable/2246275>
28. A. Lagnoux, "Cours de Series Chronologiques," ISMAG, MASTER 1 - MI00141X, Université de Toulouse le Mirail. [Online]. Available: [https://www.math.univ-toulouse.fr/~lagnoux/Poly\\_SC.pdf](https://www.math.univ-toulouse.fr/~lagnoux/Poly_SC.pdf).
29. B. Abhilash Singh, "A Literature Review on Time Series Forecasting Methods," 2022.
30. T. C. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164-181, Jan. 2011, doi: 10.1016/j.engappai.2010.09.007.
31. D. Hadjout, J.F. Torres, A. Troncoso, A. Sebaa, F. Martínez-Álvarez, Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market, *Energy*, Volume 243, 2022, 123060, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2021.123060>.
32. شكوري س. م. & ساهد ع. ا. (2022). Forecasting Electricity Consumption in Algeria Using Artificial Neural Networks. 261-247, (1)12, مجلة رؤى اقتصادية. <https://www.asjp.cerist.dz/en/article/205040>
33. N. E. Bezzar, L. Laimeche, A. Meraoumia, and L. Houam, "Data analysis-based time series forecast for managing household electricity consumption," *Demonstratio Mathematica*, vol. 55, no. 1, pp. 900-921, 2022. doi: 10.1515/dema-2022-0176.

- 
34. Tae-Young Kim, Sung-Bae Cho, Predicting residential energy consumption using CNN-LSTM neural networks, *Energy*, Volume 182, 2019, Pages 72-81, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2019.05.230>.
  35. T. Han, K. Muhammad, T. Hussain, J. Lloret and S. W. Baik, "An Efficient Deep Learning Framework for Intelligent Energy Management in IoT Networks," in *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3170-3179, 1 March, 2021, doi: 10.1109/JIOT.2020.3013306.
  36. Zulfiqar Ahmad Khan, Amin Ullah, Ijaz Ul Haq, Mohamed Hamdy, Gerardo Maria Mauro, Khan Muhammad, Mohammad Hijji, Sung Wook Baik, Efficient Short-Term Electricity Load Forecasting for Effective Energy Management, *Sustainable Energy Technologies and Assessments*, Volume 53, Part A, 2022, 102337, ISSN 2213-1388, <https://doi.org/10.1016/j.seta.2022.102337>.
  37. Sana Mujeeb, Nadeem Javaid, ESAENARX and DE-RELM: Novel schemes for big data predictive analytics of electricity load and price, *Sustainable Cities and Society*, Volume 51, 2019, 101642, ISSN 2210-6707, <https://doi.org/10.1016/j.scs.2019.101642>.
  38. Wei Gao, Ayda Darvishan, Mohammad Toghiani, Mohsen Mohammadi, Oveis Abedinia, Noradin Ghadimi, Different states of multi-block based forecast engine for price and load prediction, *International Journal of Electrical Power & Energy Systems*, Volume 104, 2019, Pages 423-435, ISSN 0142-0615, <https://doi.org/10.1016/j.ijepes.2018.07.014>.
  39. "PJM dataset," Kaggle, 2021. [Online]. Available: <https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption>. [Accessed: May 11, 2023].
  40. "PJM dataset," PJM Interconnection, 2021. [Online]. Available: <https://www.pjm.com/>. [Accessed: May 11, 2023].