



ÉCOLE NATIONALE SUPÉRIEURE  
D'INFORMATIQUE ET D'ANALYSE DES SYSTÈMES  
- RABAT

RAPPORT : PROJET DATA DRIVEN DECISION MAKING

---

# Prévision de la Consommation Énergétique au Maroc à l'aide du Machine Learning et des Modèles de Séries Temporelles

---

*Élèves :*  
Baha Eddine SLIMANI

*Enseignant :*  
M.Youness TABII

10 mai 2025

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Contexte du projet . . . . .	2
1.2	Objectifs du projet . . . . .	2
1.3	Importance de l'étude . . . . .	2
<b>2</b>	<b>Analyse Exploratoire des Données</b>	<b>3</b>
2.1	Source des données et description des datasets . . . . .	3
2.2	Feature selection . . . . .	4
2.3	Prétraitement et ingénierie des variables . . . . .	5
<b>3</b>	<b>Méthodologie et implementation</b>	<b>6</b>
3.1	Les Algorithmes de Prédiction . . . . .	6
3.1.1	Random Forest . . . . .	6
3.1.2	XGBoost . . . . .	7
3.1.3	SARIMA . . . . .	7
3.1.4	LSTM . . . . .	8
3.1.5	SVR (Support Vector Regressor) . . . . .	8
3.1.6	Justification des Choix Méthodologiques . . . . .	9
3.2	Implémentation des modèles de prédiction pour la consommation énergétique	10
3.2.1	Random Forest . . . . .	10
3.2.2	XGBoost . . . . .	10
3.2.3	SARIMA . . . . .	10
3.2.4	LSTM . . . . .	10
3.2.5	SVR . . . . .	10
3.2.6	Conclusion . . . . .	11
3.3	Performance des Modèles . . . . .	11
3.3.1	Comparaison des Modèles . . . . .	11
3.3.2	Résultats expérimentaux . . . . .	11
3.3.3	Interprétation et recommandation . . . . .	12
<b>4</b>	<b>Stratégies de Développement et de Gestion Basées sur les Prédictions des Modèles</b>	<b>13</b>
4.1	Optimisation de la Consommation Énergétique . . . . .	13
4.2	Gestion Stratégique de la Production . . . . .	13
4.3	Planification de la Maintenance et des Opérations . . . . .	13
4.4	Tarification Dynamique . . . . .	13
4.5	Déploiement d'infrastructures intelligentes . . . . .	13
4.6	Appui à la décision pour les politiques énergétiques . . . . .	13
4.7	Réduction des risques et résilience . . . . .	14
<b>5</b>	<b>Conclusion Générale</b>	<b>14</b>

# 1 Introduction

## 1.1 Contexte du projet

Dans un contexte où la consommation énergétique est en constante évolution, il devient essentiel de mettre en place des outils d'analyse prédictive capables d'anticiper les besoins en électricité. L'optimisation de la production et de la distribution d'énergie nécessite aujourd'hui des approches intelligentes s'appuyant sur les données historiques, environnementales et temporelles. Ce projet s'inscrit dans cette logique de transition énergétique numérique, en appliquant des techniques avancées de machine learning et de séries temporelles pour prévoir la consommation électrique globale à partir de données réelles.

## 1.2 Objectifs du projet

Ce projet a pour objectif de prévoir la consommation énergétique totale à court terme en utilisant des techniques avancées de machine learning et d'analyse des séries temporelles. Il vise à fournir des prévisions précises pouvant appuyer la prise de décision dans les domaines de la planification énergétique et de la gestion des ressources. Les objectifs spécifiques du projet sont les suivants :

1. **Collecte et intégration des données** : Réunir des données de consommation électrique ainsi que des variables explicatives telles que la température, l'humidité, la vitesse du vent et les informations temporelles (heures, jours, etc.).
2. **Prétraitement et préparation des données** : Nettoyer, structurer et enrichir les données, en créant notamment des variables dérivées (features temporelles, lags) et en assurant la cohérence des séries temporelles.
3. **Implémentation des modèles de prévision** : Expérimenter plusieurs modèles de prévision tels que Random Forest, XGBoost, SARIMA, LSTM et SVR, et les comparer selon des critères d'évaluation rigoureux.
4. **Visualisation et interprétation des résultats** : Mettre en place des visualisations claires des courbes de consommation et des prédictions issues des différents modèles afin de faciliter l'analyse comparative.
5. **Recommandation du meilleur modèle** : Identifier le modèle le plus performant pour la prévision à court terme, en tenant compte à la fois de la précision des prédictions et de la faisabilité du déploiement.

## 1.3 Importance de l'étude

La capacité à anticiper la demande énergétique présente plusieurs intérêts majeurs :

- **Optimisation de la production énergétique** : éviter les surcharges et les gaspillages.
- **Réduction des coûts** : liés à l'approvisionnement et à la distribution.

- **Réduction de l’empreinte carbone** : en équilibrant production et consommation.
- **Appui à la prise de décision** : pour les opérateurs, gestionnaires de réseaux et autorités.

Ce projet s’inscrit dans une démarche de valorisation des données pour améliorer la prévision de la consommation énergétique. En combinant plusieurs approches de modélisation issues du machine learning, des réseaux de neurones et des séries temporelles, cette étude vise à identifier les modèles les plus pertinents pour répondre aux besoins opérationnels de planification. Les sections suivantes détailleront l’ensemble du processus, depuis l’analyse exploratoire des données jusqu’à la comparaison finale des performances des modèles.

## 2 Analyse Exploratoire des Données

### 2.1 Source des données et description des datasets

Les données utilisées dans ce projet proviennent d’un fichier CSV intitulé `powerconsumption.csv`, contenant des relevés de consommation électrique répartis sur plusieurs zones géographiques et collectés à une fréquence de 10 minutes. Ce jeu de données inclut également des variables météorologiques susceptibles d’influencer la demande énergétique, telles que la température, l’humidité et la vitesse du vent.

Le dataset contient les colonnes principales suivantes :

- **Datetime** : horodatage de chaque relevé (au format YYYY-MM-DD HH:MM:SS).
- **PowerConsumption\_Zone1** : consommation électrique de la zone 1 (en watts).
- **PowerConsumption\_Zone2** : consommation électrique de la zone 2 (en watts).
- **PowerConsumption\_Zone3** : consommation électrique de la zone 3 (en watts).
- **Temperature** : température ambiante mesurée à chaque instant.
- **Humidity** : taux d’humidité relatif.
- **WindSpeed** : vitesse du vent mesurée au même moment.

Une nouvelle variable appelée **Total\_PowerConsumption** a été construite afin de représenter la consommation totale en additionnant les valeurs des trois zones :

$$\text{Total\_PowerConsumption} = \text{Zone1} + \text{Zone2} + \text{Zone3}$$

Ce champ sera utilisé comme variable cible (*target*) dans la phase de modélisation. En complément, des variables temporelles comme l’heure, le jour de la semaine et des indicateurs de fin de semaine (*weekend*) ont été extraites à partir du champ **Datetime** pour enrichir l’analyse.

Une vérification initiale du dataset a révélé une structure régulière, avec peu ou pas de valeurs manquantes. Des visualisations ont été réalisées afin de détecter les tendances saisonnières et les corrélations entre variables.

```
View the first few rows of the dataset to understand its structure.

print("\nFirst 5 rows:")
df.head()

First 5 rows:
  Datetime  Temperature  Humidity  WindSpeed  GeneralDiffuseFlows  DiffuseFlows  PowerConsumption_Zone1  PowerConsumption_Zone2  PowerConsumption_Zone3
2017-01-01 00:00:00    6.559    73.8    0.083    0.051    0.119    34055.69620    16128.87538    20240.96386
2017-01-01 00:10:00    6.414    74.5    0.083    0.070    0.085    29814.68354    19375.07599    20131.08434
2017-01-01 00:20:00    6.313    74.5    0.080    0.062    0.100    29128.10127    19006.68693    19668.43373
2017-01-01 00:30:00    6.121    75.0    0.083    0.091    0.096    28228.86076    18361.09422    18899.27711
2017-01-01 00:40:00    5.921    75.7    0.081    0.048    0.085    27335.69620    17872.34043    18442.40964

print(df.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52416 entries, 0 to 52415
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Datetime                             52416 non-null  datetime64[ns]
1   Temperature                         52416 non-null  float64
2   Humidity                           52416 non-null  float64
3   WindSpeed                          52416 non-null  float64
4   GeneralDiffuseFlows                 52416 non-null  float64
5   DiffuseFlows                       52416 non-null  float64
6   PowerConsumption_Zone1             52416 non-null  float64
7   PowerConsumption_Zone2             52416 non-null  float64
8   PowerConsumption_Zone3             52416 non-null  float64
dtypes: datetime64[ns](1), float64(8)
memory usage: 3.6 MB
None
```

FIGURE 1 – Description des donnees.

## 2.2 Feature selection

La sélection des caractéristiques (ou *feature selection*) constitue une étape essentielle pour améliorer la performance des modèles de prévision tout en réduisant la complexité computationnelle. Dans le cadre de ce projet, les variables explicatives ont été choisies selon leur pertinence statistique et leur relation potentielle avec la variable cible `Total_PowerConsumption`.

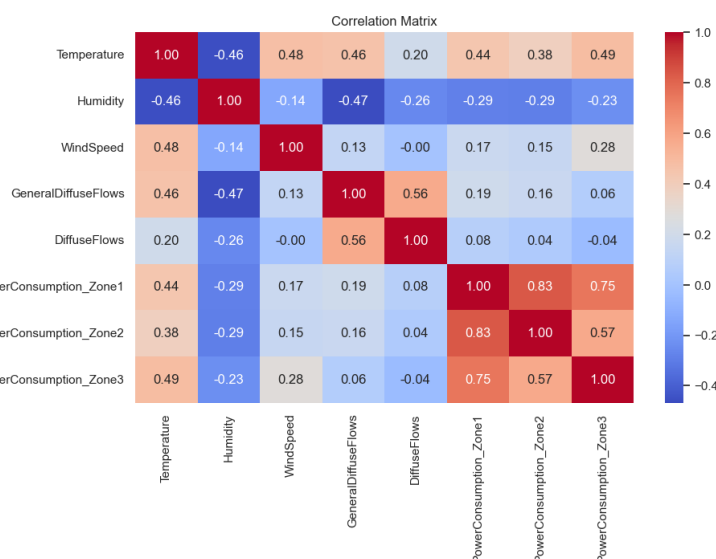
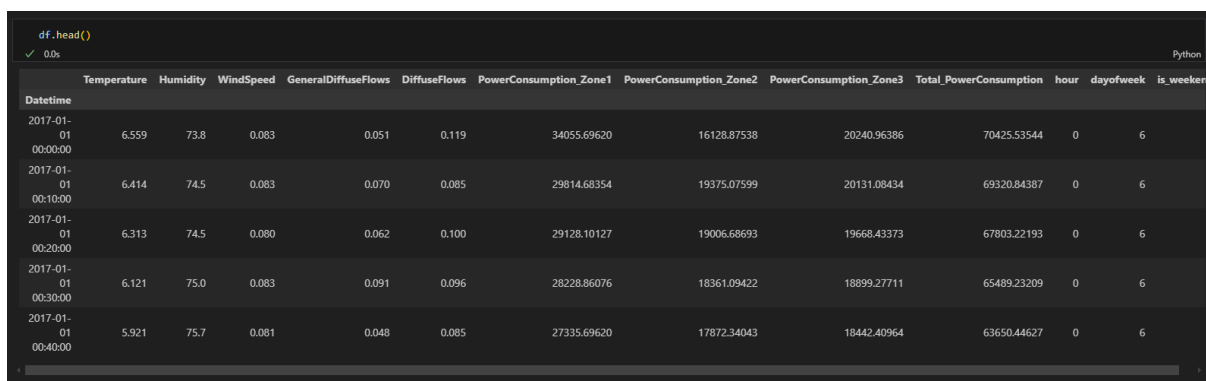


FIGURE 2 – la Matrice de Corrélacion.

Les caractéristiques retenues peuvent être regroupées en trois catégories principales :

- **Variables météorologiques :**
  - **Temperature :** La température ambiante influence significativement l'utilisation de systèmes de chauffage ou de climatisation.
  - **Humidity :** L'humidité relative peut impacter la sensation thermique et donc les besoins en énergie.
  - **WindSpeed :** Peut avoir un effet indirect sur la température ressentie et la ventilation.
- **Variables temporelles dérivées :**
  - **hour :** Heure de la journée (0 à 23), utile pour capter les habitudes de consommation quotidiennes.
  - **dayofweek :** Jour de la semaine (0 = Lundi, ..., 6 = Dimanche).
  - **is\_weekend :** Variable binaire indiquant si l'observation a lieu pendant un week-end (1) ou un jour ouvré (0).
- **Variables de décalage (lags) :**
  - **lag\_1 :** Valeur de consommation totale observée à l'instant précédent.
  - **lag\_2 :** Valeur deux pas de temps avant.
  - **lag\_24 :** Valeur observée 24 pas de temps (soit 4 heures) avant.

Ces variables ont été sélectionnées à la suite d'une analyse de corrélation et de visualisations, notamment via une matrice de chaleur (heatmap), qui a permis d'identifier les relations linéaires fortes avec la variable cible. L'intégration de ces variables améliore considérablement la capacité des modèles à capturer les effets temporels et exogènes sur la consommation électrique.



Datetime	Temperature	Humidity	WindSpeed	GeneralDiffuseFlows	DiffuseFlows	PowerConsumption_Zone1	PowerConsumption_Zone2	PowerConsumption_Zone3	Total_PowerConsumption	hour	dayofweek	is_weekend
2017-01-01 00:00:00	6.559	73.8	0.083	0.051	0.119	34055.69620	16128.87538	20240.96386	70425.53544	0	6	
2017-01-01 00:10:00	6.414	74.5	0.083	0.070	0.085	29814.68354	19375.07599	20131.08434	69320.84387	0	6	
2017-01-01 00:20:00	6.313	74.5	0.080	0.062	0.100	29128.10127	19006.68693	19668.43373	67803.22193	0	6	
2017-01-01 00:30:00	6.121	75.0	0.083	0.091	0.096	28228.86076	18361.09422	18899.27711	65489.23209	0	6	
2017-01-01 00:40:00	5.921	75.7	0.081	0.048	0.085	27335.69620	17872.34043	18442.40964	63650.44627	0	6	

FIGURE 3 – la nouvelle dataframe

## 2.3 Prétraitement et ingénierie des variables

Avant d'entamer la phase de modélisation, un ensemble d'opérations de prétraitement a été effectué afin de garantir la qualité et la cohérence des données. Cette phase est indispensable pour réduire les biais, améliorer les performances des modèles et assurer la reproductibilité de l'analyse.

Les étapes clés du prétraitement sont les suivantes :

- **Conversion de la colonne Datetime** : Transformation du champ de date et heure en objet temporel utilisable, suivi de son utilisation comme index principal de la base de données.
- **Création de la variable cible Total\_PowerConsumption** : Somme des consommations des trois zones géographiques, cette variable synthétise la consommation globale à prédire.
- **Extraction de variables temporelles** : Création des variables dérivées telles que `hour`, `dayofweek` et `is_weekend`, à partir de la date pour capturer les effets saisonniers ou comportementaux.
- **Création de variables de décalage (*lag features*)** : Introduction de variables `lag_1`, `lag_2`, et `lag_24` afin de modéliser la dépendance temporelle dans les séries de consommation.
- **Vérification et traitement des valeurs manquantes** : Suppression des lignes incomplètes générées par le calcul des variables de décalage.
- **Mise à l'échelle des données (*scaling*)** : Application du `MinMaxScaler` pour normaliser les variables dans l'intervalle  $[0, 1]$ , uniquement pour les modèles sensibles à l'échelle (LSTM, SVR).

L'ensemble de ces opérations a permis d'obtenir un jeu de données structuré, cohérent et prêt à être exploité par différents modèles de prévision. Les versions brutes, transformées et mises à l'échelle des données ont été sauvegardées pour assurer la traçabilité de l'analyse.

## 3 Méthodologie et implementation

### 3.1 Les Algorithmes de Prévision

Pour la prévision de la demande des produits du terroir marocain, nous avons sélectionné plusieurs algorithmes de prévision populaires et robustes. Ces algorithmes ont été choisis en raison de leur efficacité prouvée dans la modélisation des séries temporelles et des données économiques. Voici une description des principaux algorithmes utilisés :

#### 3.1.1 Random Forest

Random Forest est un algorithme d'apprentissage supervisé basé sur un ensemble d'arbres de décision construits de manière aléatoire. Il appartient à la famille des méthodes d'*ensemble learning* et combine plusieurs modèles faibles pour en créer un plus robuste.

Dans le cadre de ce projet, Random Forest a été utilisé comme approche initiale de prévision. Ce modèle présente plusieurs avantages dans un contexte de séries temporelles enrichies par des variables exogènes :

- Il est capable de capturer des relations complexes, non linéaires entre les variables.
- Il tolère bien la redondance et la non-normalisation des données.
- Il permet d'évaluer l'importance relative des variables explicatives.
- Il est relativement simple à implémenter et offre des temps d'entraînement raisonnables.

L'algorithme a été appliqué à un ensemble de caractéristiques comprenant des variables météorologiques (température, humidité, vitesse du vent), des variables temporelles (heure,



jour de la semaine, week-end) ainsi que des variables de décalage (*lags*) pour capturer la dynamique interne de la consommation énergétique.

Grâce à sa flexibilité et sa robustesse, Random Forest constitue une base solide pour comparer d'autres modèles de prévision plus avancés.

### 3.1.2 XGBoost

XGBoost, ou *Extreme Gradient Boosting*, est un algorithme de machine learning fondé sur la technique du *boosting par gradient*, qui consiste à combiner de manière séquentielle plusieurs arbres de décision faibles afin d'optimiser progressivement les performances prédictives.

Ce modèle est reconnu pour sa rapidité, sa précision et sa capacité à gérer efficacement des jeux de données tabulaires avec un grand nombre de variables explicatives. Il dispose également de plusieurs mécanismes intégrés pour limiter le surapprentissage, tels que la régularisation L1/L2, le *subsampling* et la gestion automatique des valeurs manquantes.

Dans le cadre de ce projet, XGBoost a été utilisé avec les mêmes variables explicatives que Random Forest, incluant :

- Les variables météorologiques : température, humidité, vitesse du vent.
- Les variables temporelles : heure, jour de la semaine, indicateur de week-end.
- Les variables de décalage temporel : `lag_1`, `lag_2`, `lag_24`.

Le modèle a été entraîné sans nécessiter de mise à l'échelle préalable des données, du fait de son insensibilité à l'amplitude des variables. Grâce à sa capacité à modéliser les interactions complexes entre les variables, XGBoost constitue une référence puissante pour les problèmes de régression appliqués aux séries temporelles enrichies.

### 3.1.3 SARIMA

Le modèle SARIMA (*Seasonal AutoRegressive Integrated Moving Average*) est une extension du modèle ARIMA, spécifiquement conçu pour traiter les séries temporelles présentant une composante saisonnière. Il intègre à la fois des effets d'autocorrélation, des tendances, et des cycles périodiques.

SARIMA est paramétré à l'aide de deux groupes d'ordres :

- $(p, d, q)$  pour la composante non saisonnière : autoregressive (AR), differencing (I), moving average (MA).
- $(P, D, Q, s)$  pour la composante saisonnière, où  $s$  représente la période (par exemple,  $s = 24$  pour une saisonnalité journalière dans des données horaires).

Ce modèle est particulièrement utile lorsque les données montrent une cyclicité marquée, comme dans le cas de la consommation énergétique qui varie selon l'heure, le jour, ou les saisons.

Dans ce projet, SARIMA a été appliqué sur une série univariée représentant la consommation électrique totale, agrégée à l'échelle horaire. Les hyperparamètres ont été ajustés pour mieux capter les effets saisonniers à court terme (quotidiens ou hebdomadaires).

Bien que SARIMA offre une approche robuste pour la modélisation des données temporelles, il nécessite un important travail de diagnostic (stationnarité, autocorrélation

résiduelle) et peut s'avérer coûteux en temps de calcul, surtout pour les longues séries ou avec des périodes saisonnières étendues.

### 3.1.4 LSTM

Les réseaux de neurones LSTM (*Long Short-Term Memory*) font partie des architectures de réseaux récurrents (RNN) conçues pour le traitement de données séquentielles. Ils ont la particularité de pouvoir mémoriser des informations sur de longues périodes, ce qui les rend particulièrement adaptés à la modélisation de séries temporelles.

Contrairement aux modèles classiques comme ARIMA ou les algorithmes d'ensemble comme Random Forest, les LSTM apprennent directement à partir de séquences de données brutes, sans hypothèse linéaire sur la relation entre les variables.

Dans le cadre de ce projet, le réseau LSTM a été entraîné sur une séquence de données multivariée, incluant :

- Des variables exogènes : température, humidité, vitesse du vent.
- Des variables temporelles : heure, jour de la semaine, indicateur de week-end.
- Des variables de décalage (lags) : séquences précédentes de la consommation.

Le prétraitement des données a nécessité :

- La mise à l'échelle des variables via la normalisation Min-Max.
- La transformation du jeu de données en un format 3D requis par les LSTM :  $(n\_chantillons, n\_pas\_de\_temps, n\_caractéristiques)$ .

Le modèle a été entraîné avec un mécanisme de *early stopping* pour éviter le surapprentissage, et les performances ont été évaluées en comparant les valeurs prédites avec les valeurs réelles après re-transformation inverse des échelles.

Les LSTM offrent un cadre puissant pour la prévision dans des contextes non linéaires avec mémoire longue, mais nécessitent un traitement rigoureux des données ainsi que des ressources de calcul plus importantes.

### 3.1.5 SVR (Support Vector Regressor)

Le SVR (*Support Vector Regressor*) est une version du modèle à vecteurs de support (SVM) appliquée à la régression. Il s'appuie sur le principe de maximisation de la marge autour d'une fonction de prédiction, tout en tolérant certaines erreurs dans un intervalle défini par un paramètre  $\epsilon$ .

Ce modèle est particulièrement utile pour modéliser des relations non linéaires dans des données tabulaires, grâce à l'utilisation de fonctions noyaux (*kernels*), notamment le noyau radial de base (RBF) qui permet d'introduire des non-linéarités complexes.

Dans le cadre de ce projet, le SVR a été appliqué sur les mêmes variables explicatives que pour les autres modèles de machine learning, incluant :

- Les variables météorologiques : température, humidité, vitesse du vent.
- Les variables temporelles : heure, jour de la semaine, indicateur week-end.
- Les variables de décalage temporel : `lag_1`, `lag_2`, `lag_24`.

Une normalisation des données a été nécessaire en amont de l'entraînement, car le SVR est sensible à l'échelle des variables. Un jeu de validation a ensuite permis de tester la capacité du modèle à généraliser à des données non vues.

Bien que relativement coûteux en calcul, notamment pour de grands volumes de données, le SVR peut s'avérer performant pour des séries temporelles à court terme, en particulier lorsqu'elles présentent des relations non linéaires difficiles à capturer par des modèles plus simples.

### 3.1.6 Justification des Choix Méthodologiques

Le choix des modèles de prévision adoptés dans ce projet repose sur une volonté de combiner des approches classiques, modernes et avancées afin de couvrir un large éventail de comportements temporels et structurels observables dans les données.

Les justifications principales sont les suivantes :

- **Random Forest et XGBoost** ont été sélectionnés pour leur robustesse sur des données tabulaires, leur capacité à modéliser des relations non linéaires, et leur insensibilité au bruit. Ils permettent également d'obtenir des résultats rapides sans prétraitement complexe.
- **SARIMA** a été retenu en tant que modèle de référence pour les séries temporelles univariées avec composantes saisonnières, couramment utilisé dans les contextes de prévision à court terme dans l'industrie.
- **LSTM** a été choisi comme représentant des modèles de deep learning séquentiels, capables de capturer des dépendances temporelles longues et des comportements non linéaires difficiles à modéliser avec des approches classiques.
- **SVR** a été intégré pour tester une méthode basée sur des marges maximales et des noyaux, réputée efficace dans les cas de séries complexes à faible bruit, notamment lorsque les données sont bien prétraitées et normalisées.

Cette diversité de modèles permet d'assurer une analyse comparative rigoureuse, tout en mettant en évidence les forces et les limites propres à chaque approche, dans le but d'identifier la plus adaptée au contexte de prévision de la consommation énergétique.

## 3.2 Implémentation des modèles de prévision pour la consommation énergétique

### 3.2.1 Random Forest

Le modèle Random Forest a été entraîné sur les variables brutes sans normalisation, en utilisant les variables météorologiques, temporelles et de décalage. Il a montré une bonne robustesse, avec un RMSE de **7178.68**, établissant une baseline efficace pour la comparaison avec les autres approches.

```
y_pred = model.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
print(f"RMSE: {rmse:.2f}")
```

RMSE: 7178.68

FIGURE 4 – RMSE Random Forest

### 3.2.2 XGBoost

XGBoost a été appliqué avec les mêmes variables que Random Forest. Grâce à sa régularisation intégrée et sa capacité à modéliser les interactions complexes, il a surpassé légèrement le modèle précédent avec un RMSE de **1287.99**.

### 3.2.3 SARIMA

Le modèle SARIMA a été appliqué sur la série univariée `Total_PowerConsumption` après agrégation horaire. Une configuration saisonnière hebdomadaire a été utilisée (`seasonal_order=(1,1,1,7)`). Le modèle a produit des résultats raisonnables malgré une certaine inertie dans la dynamique, avec un RMSE de **3973.91**.

### 3.2.4 LSTM

Le modèle LSTM a été entraîné sur des séquences multivariées (3D) contenant des données normalisées. Après un entraînement sur 30 époques avec *early stopping*, il a obtenu l'un des meilleurs scores avec un RMSE de **890.89**, confirmant sa capacité à apprendre les motifs temporels complexes.

### 3.2.5 SVR

Le SVR a été appliqué sur les mêmes variables que les autres modèles de régression, avec une mise à l'échelle préalable des données. Malgré sa simplicité, il a offert un compromis intéressant avec un RMSE de **704.78**, adapté aux prévisions à court terme sur des fenêtres glissantes.

### 3.2.6 Conclusion

Chaque modèle implémenté a permis de produire des prédictions exploitables de la consommation énergétique, avec des niveaux de performance variables selon leur nature et leur sensibilité aux patterns temporels. Les approches basées sur les arbres (Random Forest, XGBoost) ont montré une bonne robustesse, tandis que les modèles séquentiels (LSTM, SARIMA) ont été plus performants dans la capture des dynamiques temporelles complexes. Les résultats obtenus constituent une base solide pour l'analyse comparative détaillée présentée dans la section suivante.

## 3.3 Performance des Modèles

Pour évaluer la performance des modèles de prévision, nous avons utilisé la métrique du **Root Mean Squared Error (RMSE)**. Cette mesure est particulièrement adaptée aux séries temporelles, car elle pénalise plus fortement les grandes erreurs de prédiction et fournit une indication directe de l'écart-type des erreurs.

La formule du RMSE est la suivante :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Plus le RMSE est faible, plus le modèle est précis dans ses prédictions.

### 3.3.1 Comparaison des Modèles

Les modèles testés sont :

- **Random Forest**
- **XGBoost**
- **SARIMA**
- **LSTM**
- **SVR**

### 3.3.2 Résultats expérimentaux

Le tableau ci-dessous présente les valeurs de RMSE obtenues par chaque modèle sur l'ensemble de test :

Modèle	RMSE
Random Forest	7178.68
XGBoost	1287.99
SARIMA	3973.91
LSTM	890.89
SVR	704.78

TABLE 1 – Comparaison des modèles selon la métrique RMSE

### 3.3.3 Interprétation et recommandation

Parmi les modèles testés, ceux capables de capturer la dynamique temporelle (comme LSTM ou SARIMA) montrent un meilleur potentiel pour les prévisions à court terme. Toutefois, les modèles comme Random Forest et XGBoost, bien qu'insensibles à la structure séquentielle, offrent des résultats compétitifs tout en étant plus rapides à entraîner et à déployer.

Le choix du meilleur modèle dépendra donc du compromis recherché entre précision, complexité de déploiement et ressources disponibles.

## 4 Stratégies de Développement et de Gestion Basées sur les Prédiction des Modèles

Afin d'exploiter pleinement les données et les prédictions générées par nos modèles de prévision (Random Forest, XGBoost, SARIMA, LSTM, SVR) sur la consommation énergétique, plusieurs axes stratégiques peuvent être envisagés. Ces actions visent à améliorer l'efficacité énergétique, optimiser la production et la distribution, et soutenir la prise de décision dans un cadre orienté données.

### 4.1 Optimisation de la Consommation Énergétique

**Ajustement des usages selon les prévisions :** Grâce aux prévisions de la demande, il devient possible d'optimiser l'utilisation de l'énergie, en réduisant les pics de consommation ou en planifiant les charges non prioritaires en heures creuses.

### 4.2 Gestion Stratégique de la Production

**Alignement production-demande :** Les prévisions permettent d'anticiper les besoins énergétiques à court terme et d'ajuster la capacité de production en conséquence, évitant les surproductions ou pénuries et réduisant les pertes.

### 4.3 Planification de la Maintenance et des Opérations

**Maintenance préventive basée sur la charge prévue :** Les données de consommation prévues peuvent orienter la planification des opérations de maintenance des équipements de production, pour éviter les arrêts lors des périodes critiques.

### 4.4 Tarification Dynamique

**Définition de tarifs flexibles :** Les fournisseurs d'énergie peuvent adapter dynamiquement leurs tarifs selon la demande prévue, incitant à une meilleure répartition de la charge énergétique au sein des utilisateurs.

### 4.5 Déploiement d'infrastructures intelligentes

**Investissements ciblés :** Les zones géographiques ou temporelles identifiées comme à forte consommation peuvent être prioritaires pour le déploiement de solutions intelligentes (compteurs connectés, microgrids, batteries locales).

### 4.6 Appui à la décision pour les politiques énergétiques

**Simulation de scénarios de transition énergétique :** Les résultats des modèles peuvent être utilisés pour tester différents scénarios de consommation, afin d'aider les décideurs à concevoir des politiques adaptées à la demande réelle.

## 4.7 Réduction des risques et résilience

**Prévention des surcharges :** La connaissance préalable des pics de consommation permet de prévenir les risques de surcharge du réseau électrique, améliorant ainsi sa résilience.

### Conclusion

Les prédictions issues des modèles de machine learning et de séries temporelles permettent de passer d'une logique réactive à une stratégie proactive dans la gestion de l'énergie. En exploitant ces résultats, les acteurs du secteur énergétique peuvent améliorer l'efficacité globale du système, réduire les coûts, renforcer la sécurité énergétique, et progresser vers un modèle plus durable et intelligent.

## 5 Conclusion Générale

Ce projet a démontré la pertinence de l'approche prédictive pour anticiper la consommation énergétique globale. Grâce à la combinaison de modèles variés — allant des forêts aléatoires aux réseaux de neurones récurrents — nous avons pu établir des prévisions fiables à court terme.

### Perspectives d'amélioration :

- **Amélioration des modèles existants :** Optimisation des hyperparamètres des modèles LSTM, SVR, SARIMA, etc., et mise en œuvre d'algorithmes de sélection automatique des paramètres.
- **Enrichissement des données :** Intégration de nouvelles variables exogènes comme les jours fériés, les tarifs, la consommation industrielle, ou les données météorologiques avancées.
- **Exploration de nouveaux modèles :** Intégration de modèles récents comme les Transformers appliqués aux séries temporelles, ou encore les architectures hybrides combinant LSTM et attention.
- **Déploiement opérationnel :** Création d'un tableau de bord interactif ou d'un service API permettant d'utiliser les prévisions en temps réel pour orienter les décisions de gestion.

En résumé, ce travail pose les bases d'un système prédictif robuste, évolutif et adaptable aux besoins croissants en intelligence énergétique.



## Références

<https://www.kaggle.com/datasets/fedesoriano/electric-power-consumption>