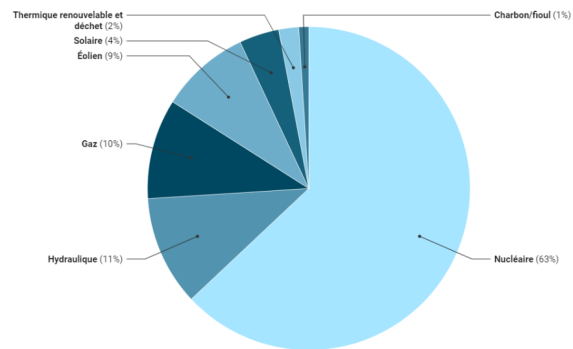


# L'optimisation du mix énergétique

Par Azhar ELBAGHAZAOUI, Nadia SYLLA, Menglin XIONG et Kehan LIU étudiants en TAF Data science.

La France dépend fortement de l'énergie nucléaire, qui représente environ 63% de sa production totale d'énergie électrique en 2022. Les 37% restants proviennent d'autres sources telles que le gaz et les énergies renouvelables. L'objectif est d'augmenter la part des énergies vertes dans les 37% de production d'électricité restantes, en particulier en utilisant au maximum la capacité de production d'énergie électrique éolienne pour réduire le besoin de production d'électricité au gaz. La production répond à la demande en consommation d'électricité, laquelle est influencée par des facteurs tels que la météo, l'heure et le jour. Il est donc intéressant d'analyser et de prévoir également la demande en électricité, pour qu'elle soit aligner à la production d'énergie électrique.

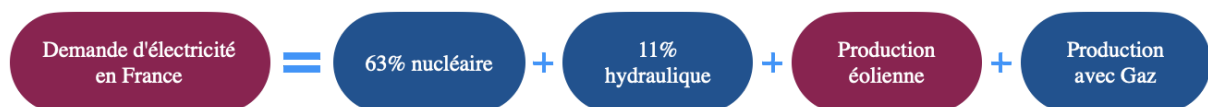


## Pourquoi alors optimiser le mix énergétique en France ?

Développer un modèle de prédiction visant à anticiper et optimiser la production éolienne en fonction de la météo contribuerait à réduire les coûts et les pertes de production. Cette capacité de prédiction revêt une importance particulière dans le cadre de l'engagement de la France et des producteurs à réduire leurs émissions de gaz à effet de serre. De même, la possibilité de prédire la consommation d'énergie électrique permettrait de cibler la réduction de la production d'énergie au strict nécessaire et d'éviter toutes importations d'énergie (car en la prévoyant, on pourrait mettre en marche les outils pour produire et répondre à la demande). Les deux prédictions associées permettent de minimiser la production basée sur le gaz, alignant ainsi le pays sur ses objectifs environnementaux et contribuant à la stabilisation des prix de l'électricité indexés sur le gaz.

## Comment optimiser le mix énergétique en France ?

Pour répondre à cette question, on a mis en place un modèle empirique se basant sur l'équation suivante :



L'équation définit la demande en électricité, avec 63% provenant de l'énergie nucléaire et 11% de l'énergie hydraulique, et le reste étant comblé par des sources telles que la

production éolienne et celle issue du gaz. L'objectif principal est de répondre de manière efficace à cette demande tout en réduisant la part des énergies fossiles, en particulier le gaz, au profit des énergies renouvelables.

Dans cette optique, la première phase du projet vise à évaluer la production d'énergie éolienne pour l'ensemble de la France. Cette évaluation se concentre initialement sur le parc éolien en mer de Saint-Nazaire en vue de généraliser les résultats obtenus à l'ensemble des parcs éoliens terrestres présents sur le territoire français. La seconde phase du projet consiste à prédire avec précision la consommation totale d'électricité pour l'ensemble du pays.

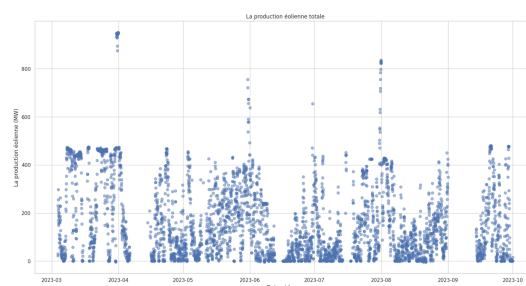
## Quelles données utiliser ?

Pour la prédiction de la production de l'énergie éolienne, nous avons utilisé trois bases de données différentes :

- Les [données météorologiques](#) comprennent des détails heure par heure entre le 1er mars 2023 et le 30 septembre 2023. Ces données incluent des paramètres tels que la direction et la vitesse du vent, les rafales de vent, la température, les précipitations, l'humidité et le pourcentage de nuages.
- Les [données de production de l'énergie éolienne au parc de Saint-Nazaire](#) heure par heure pour la période du 1er mars 2023 au 30 septembre 2023. Ces données comprennent la quantité d'énergie produite en mégawatts (MW) pour chaque heure durant cette période pour les deux groupes du parc.
- Les [données des parcs éoliens terrestres en France](#), incluant le nom de chaque parc, ainsi que la puissance maximale en mégawatts (MW) qu'ils peuvent produire.

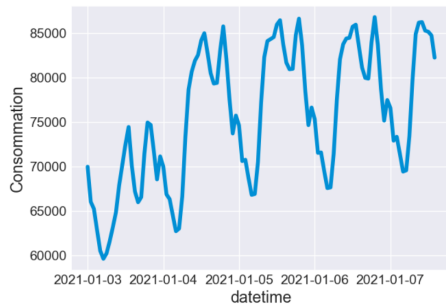
Pour la prédiction de la consommation de l'électricité en France, on a utilisé un [dataset contenant les données de consommation électrique](#) couvrant une période horaire allant du 1er janvier 2021 au 1er novembre 2023.

## Description de dataset

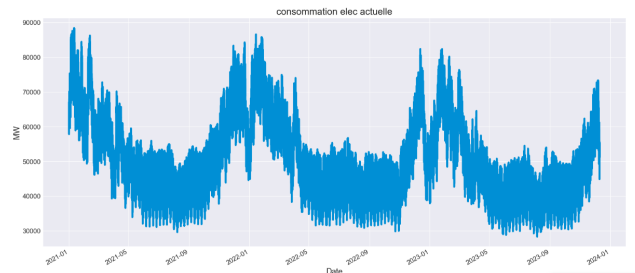


Production d'électricité du parc de St Nazaire sur 6 mois

Après une étude statistique descriptive de nos données, nous constatons que la production d'électricité du parc éolien de St Nazaire est comprise entre 0 et 450 MW, en dehors des valeurs aberrantes.



Consommation d'électricité en France sur 4 jours.



Consommation d'électricité en France sur 3 ans

De même, nous constatons que la consommation d'électricité en France présente une périodicité journalière et annuelle.

## Préparation et analyse des données

Avant d'intégrer ces données dans nos modèles de prédiction de la production d'énergie éolienne et de la consommation d'électricité en France, nous avons appliqué plusieurs opérations. Les opérations effectuées comprennent la sélection, le nettoyage, la construction, l'intégration, et le formatage des données. Plus précisément, la sélection des données implique la suppression de colonnes non pertinentes, tandis que le nettoyage adresse les valeurs manquantes et les anomalies. L'intégration des données combine les informations sur la production éolienne, les conditions météorologiques, et la consommation en un nouveau Data Frame. Enfin, la préparation des données pour l'apprentissage de l'algorithme inclut la structuration des ensembles d'entraînement et de test, ainsi que le formatage des données pour représenter l'énergie totale produite dans l'ensemble du parc éolien. La préparation des données de consommation implique également la décomposition de la colonne 'Date et heure' pour faciliter l'utilisation d'un modèle Random Forest, permettant au modèle de comprendre les relations temporelles dans la consommation d'électricité en France.

## Hypothèses

En raison de la complexité de notre projet et de la nécessité de prendre en compte de nombreux facteurs pour optimiser la structure énergétique, nous avons formulé de nombreuses hypothèses tout au long du processus du projet afin de simplifier l'analyse:

- Il n'y a que 4 sources d'énergétique : nucléaire, hydraulique, éolien, gaz.
- La proportion de la production d'électricité provenant de l'énergie nucléaire et hydraulique par rapport à la production totale d'électricité reste constante selon les pourcentages dans le rapport de RTE en 2022.
- Les données météorologiques des stations et des parcs sont pareilles.

## Algorithmes utilisés ?

Pour prédire la production d'énergie éolienne, nous faisons face à un problème de régression sur un apprentissage supervisé. Pour ce type de problème, plusieurs algorithmes de machine learning peuvent être utilisés. Toutefois nous avons choisi 5 algorithmes plus courants : *Linear Regression*, *Random Forest*, *KNN Régression*, *Gradient Boosting Regressing* et *Support Vector Regression*.

Pour la prédiction de la consommation électrique, 2 modèles ont été choisis permettant de capturer des relations complexes, temporelles et sur du long terme dans des séries temporelles : *Random Forest* et *LSTM* (Long Short-Term Memory).

## Évaluation ?

Lors de l'évaluation des différents modèles, il est essentiel d'utiliser diverses mesures pour garantir la qualité des prédictions et les comparer entre elles. Les résultats de MSE, RMSE et MAE reflètent les erreurs entre les valeurs prédites et les valeurs réelles, où des valeurs plus basses indiquent de meilleures prédictions. Le NMAE prend en compte l'échelle des données qui permet une comparaison équitable entre différents datasets. Le  $R^2$  quantifie dans quelle mesure les variations de la variable prédite peuvent être expliquées par les variations de la variable réelle, avec une valeur proche de 1 indiquant une prédiction de qualité.

Tableau 1 - L'évaluation des modèles de la production

Algorithme	MSE	RMSE	MAE	R <sup>2</sup>	NMAE
Linear Regression	12101.40	110.01	80.10	0.56	0.084
Random Forest	5010.11	70.78	47.09	0.82	0.049
KNN Régression	6061.14	77.85	51.44	0.78	0.054
Gradient Boosting Regressing	6117.95	78.22	55.10	0.78	0.058
Support Vector Regression	17151.52	130.96	94.02	0.38	0.099

Nous constatons donc que le Random Forest est l'algorithme qui apprend le mieux et prédit le mieux la production d'énergie éolienne.

Tableau 2 - L'évaluation des modèles de la consommation

Algorithme	MSE	RMSE	MAE
Random Forest	10224116.48	3197.52	2406.78
LSTM	540994.35	735.52	571.55

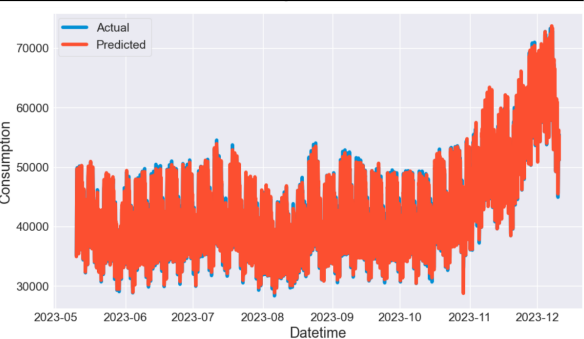
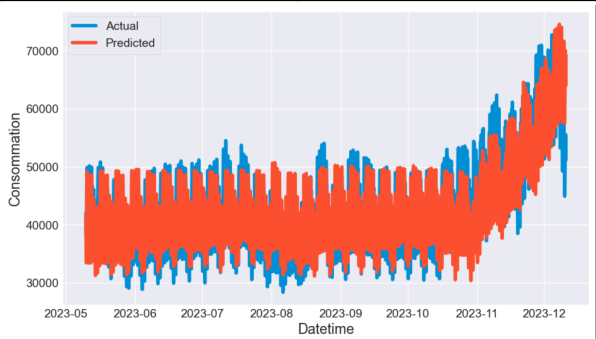
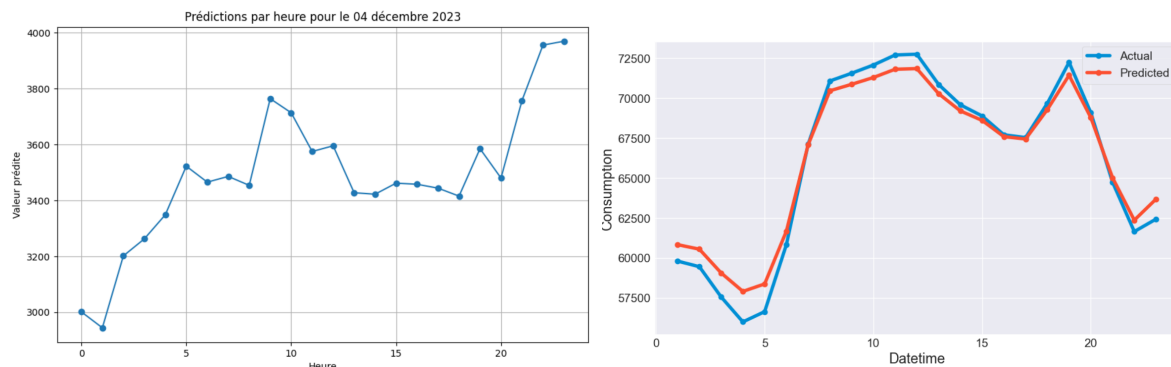
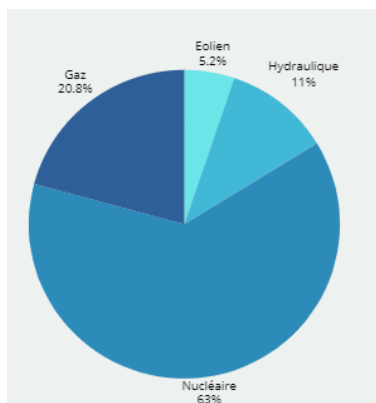
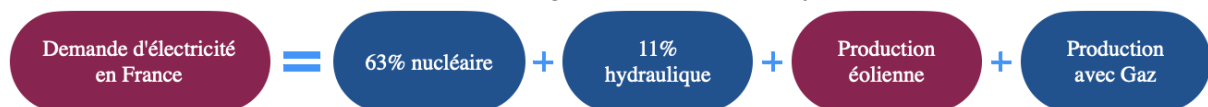


Figure 1 - Performance du modèle Random Forest  
LSTM a une meilleure précision et une meilleure performance dans le cas de la prédiction de la consommation d'électricité, c'est pourquoi nous avons choisi cet algorithme pour la prédiction de la consommation de l'électricité.

## Déploiement du modèle



Avec l'aide de nos deux modèles sélectionnés, nous avons effectué des prédictions pour le 4 décembre 2023 concernant la consommation d'électricité et la production d'énergie éolienne en France. Une fois ces deux éléments trouvés, nous avons ensuite calculé les autres éléments de l'équation du mix énergétique selon nos hypothèses.



- Demande d'électricité = 71 298.33 MW
- Production éolienne = 3 713.47 MW (Pourcentage éolien prédite = 5.21%)
- Production nucléaire =  $71\,298.33 \times 63\% = 44\,917.95$  MW
- Production hydrauliques =  $71\,298.33 \times 11\% = 7\,842.82$  MW
- Consommation de gaz =  $71\,298.33 - 44\,917.95 - 7\,842.82 - 3\,713.47 = 14\,824.09$  MW

## Et par la suite ?

Dans cette étude, la faisabilité de la prédiction de l'équation du mix énergétique en France a été démontrée. Bien que l'équation soit empirique avec des pourcentages fixes des d'autres sources d'énergie en dehors des éoliens, il est envisageable d'élargir le projet en incluant d'autres énergies et en tenant compte de la flexibilité de leurs pourcentages tout au long de la journée. La collaboration avec des fournisseurs de données météorologiques est recommandée pour améliorer la généralisation et le calcul de la production éolienne totale en France. De plus, l'exploration de l'utilisation d'algorithmes d'optimisation, intégrés au machine learning, pourrait permettre d'optimiser les pourcentages des sources d'énergie de l'équation du mix énergétique en faveur des énergies vertes pour répondre à la demande.

## Bibliographie

- [1] <https://prix-elec.com/energie/production/mix-energetique>
- [2] [https://www.rte-france.com/actualites/bilan-electrique-2022?fbclid=IwAR158\\_C84WXimQtYBCu8ughPvk9rUVVkv3WoMI9zvK4zJ5rKA5rN4cjYNig](https://www.rte-france.com/actualites/bilan-electrique-2022?fbclid=IwAR158_C84WXimQtYBCu8ughPvk9rUVVkv3WoMI9zvK4zJ5rKA5rN4cjYNig) (Bilan Mix 2022)
- [3] [https://www.lemonde.fr/les-decodeurs/article/2022/10/06/pourquoi-le-prix-de-l-electricite-d-epend-de-celui-du-gaz-et-autres-questions-sur-les-factures-a-venir\\_6140985\\_4355771.html](https://www.lemonde.fr/les-decodeurs/article/2022/10/06/pourquoi-le-prix-de-l-electricite-d-epend-de-celui-du-gaz-et-autres-questions-sur-les-factures-a-venir_6140985_4355771.html)
- [4] <https://www.sciencedirect.com/science/article/pii/S0960148117302550#sec6> (Rapport scientifique sur l'utilisation du Random Forest pour l'estimation de production éolienne)
- [5] <https://www.rte-france.com/eco2mix/telecharger-les-indicateurs> (Données consommations)
- [6] <https://www.visualcrossing.com/weather/weather-data-services> (Données meteorologiques)
- [7] <https://www.services-rte.com/fr/visualisez-les-donnees-publiees-par-rte/production-realisee-par-groupe.html> (Données production électrique parc Saint Nazaire)
- [8] <https://www.georisques.gouv.fr/donnees/bases-de-donnees/eolien-terrestre> (liste des parcs éoliens terrestres)