



A wide-angle photograph of a massive wind farm situated in a vast, arid mountain range. The sun is low on the horizon, casting a warm, golden glow over the hills and the numerous white wind turbines. The turbines are densely packed across the terrain, their blades catching the light. In the foreground, the tower and part of the blades of a single wind turbine are visible, looking towards the right side of the frame.

# PRÉSENTATION GROUPE 6

&

# SOMMAIRE

- 
- 1 Notre équipe
- 2 Contexte et enjeux
- 3 Analyse des données
- 4 Quelques modèles pertinents
- 5 Random-forest & difficultés
- 6 Résultats et prédictions

# L'ÉQUIPE

« Se réunir est un début, rester ensemble est un progrès, travailler ensemble est la réussite.

**Henry Ford**  
Fondateur de FORD



# L'ÉQUIPE



Tom B.



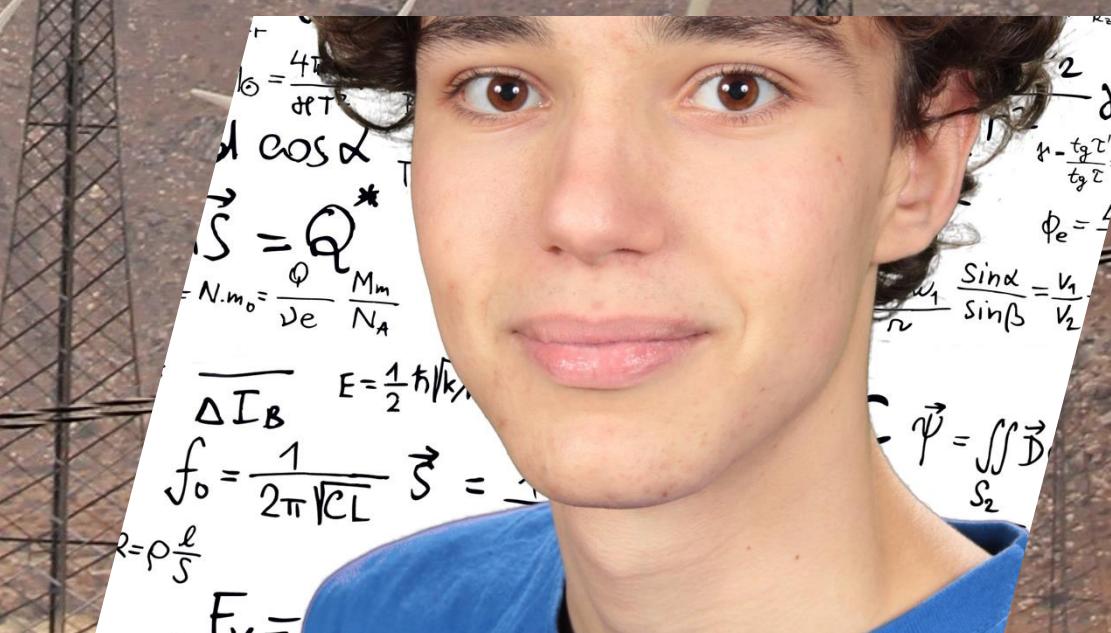
M. Adnane K.



Paul R.



Baptiste F.

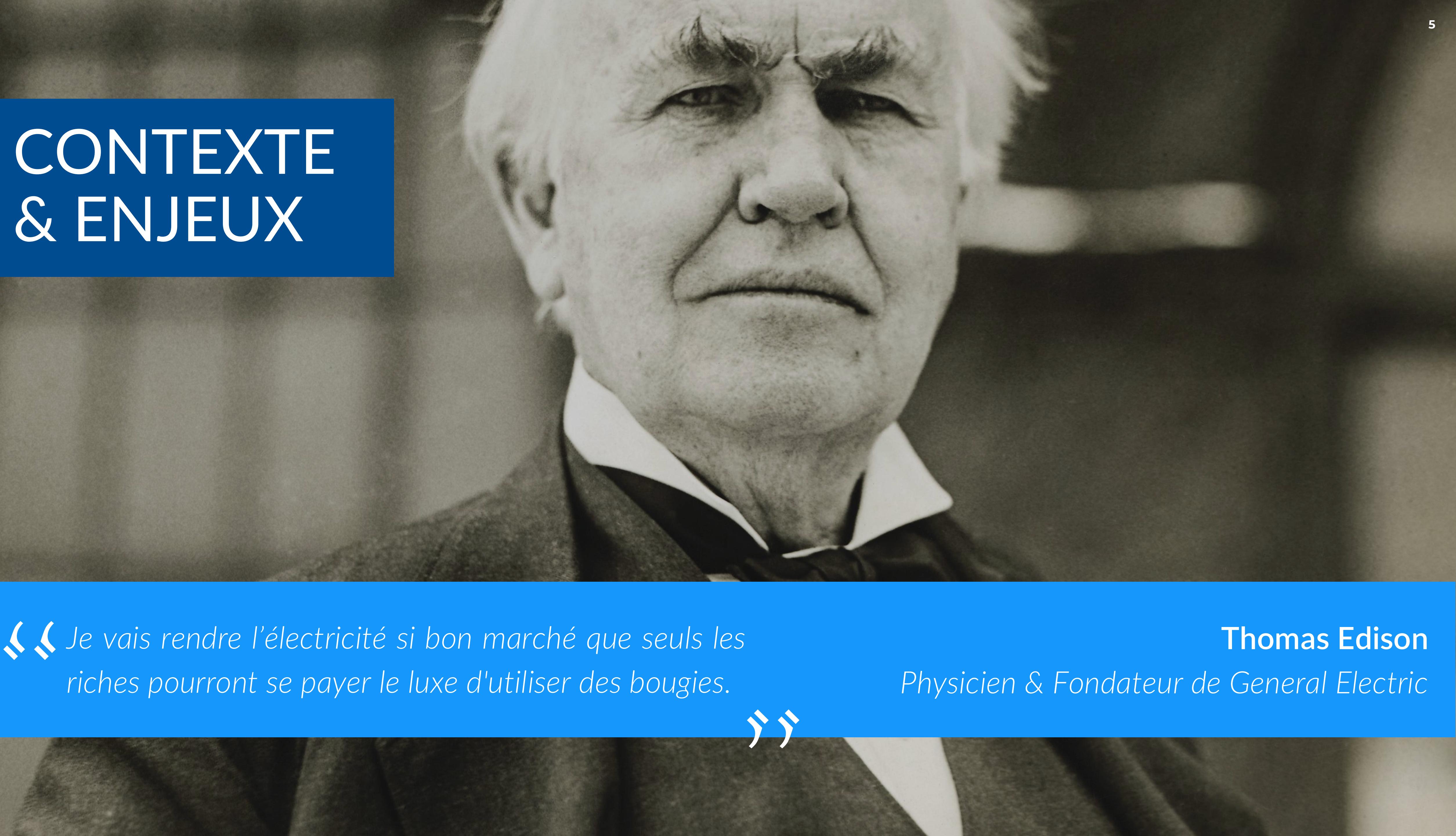


Bilal Z.

# CONTEXTE & ENJEUX

« Je vais rendre l'électricité si bon marché que seuls les riches pourront se payer le luxe d'utiliser des bougies.

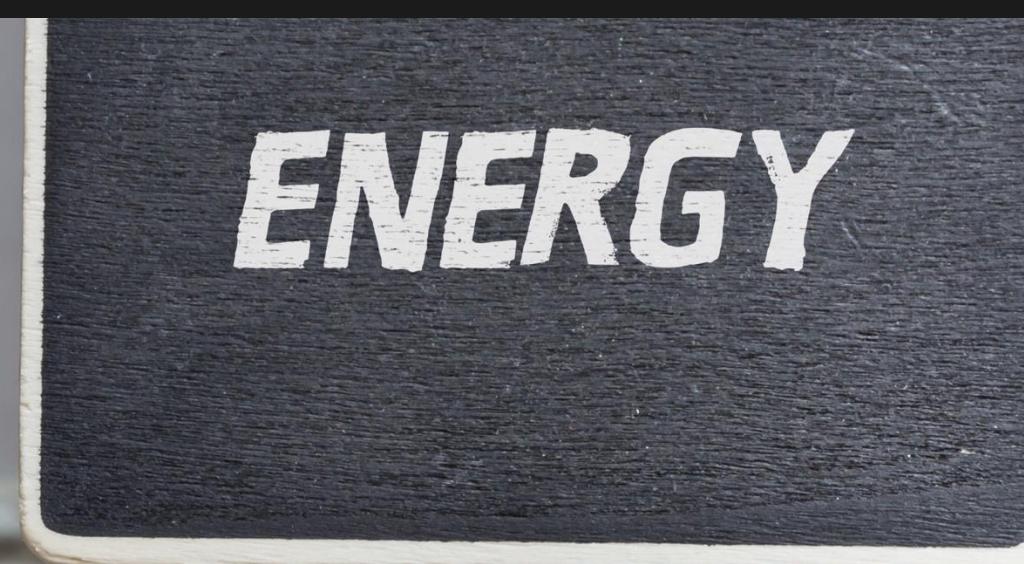
**Thomas Edison**  
Physicien & Fondateur de General Electric



# L'ÉNERGIE, LE NOUVEAU NERF DE LA GUERRE

COUPURES D'ÉLECTRICITÉ  
PAS DE PANIQUE ?

© OLYMPIEN/ANTONIO GARCIA/GETTY IMAGES



Une demande toujours plus forte ;



Des prix toujours plus élevés ;

Des énergies intermittentes.

# ANALYSE DES DONNÉES

« « T'échappes à la police, pas aux statistiques.

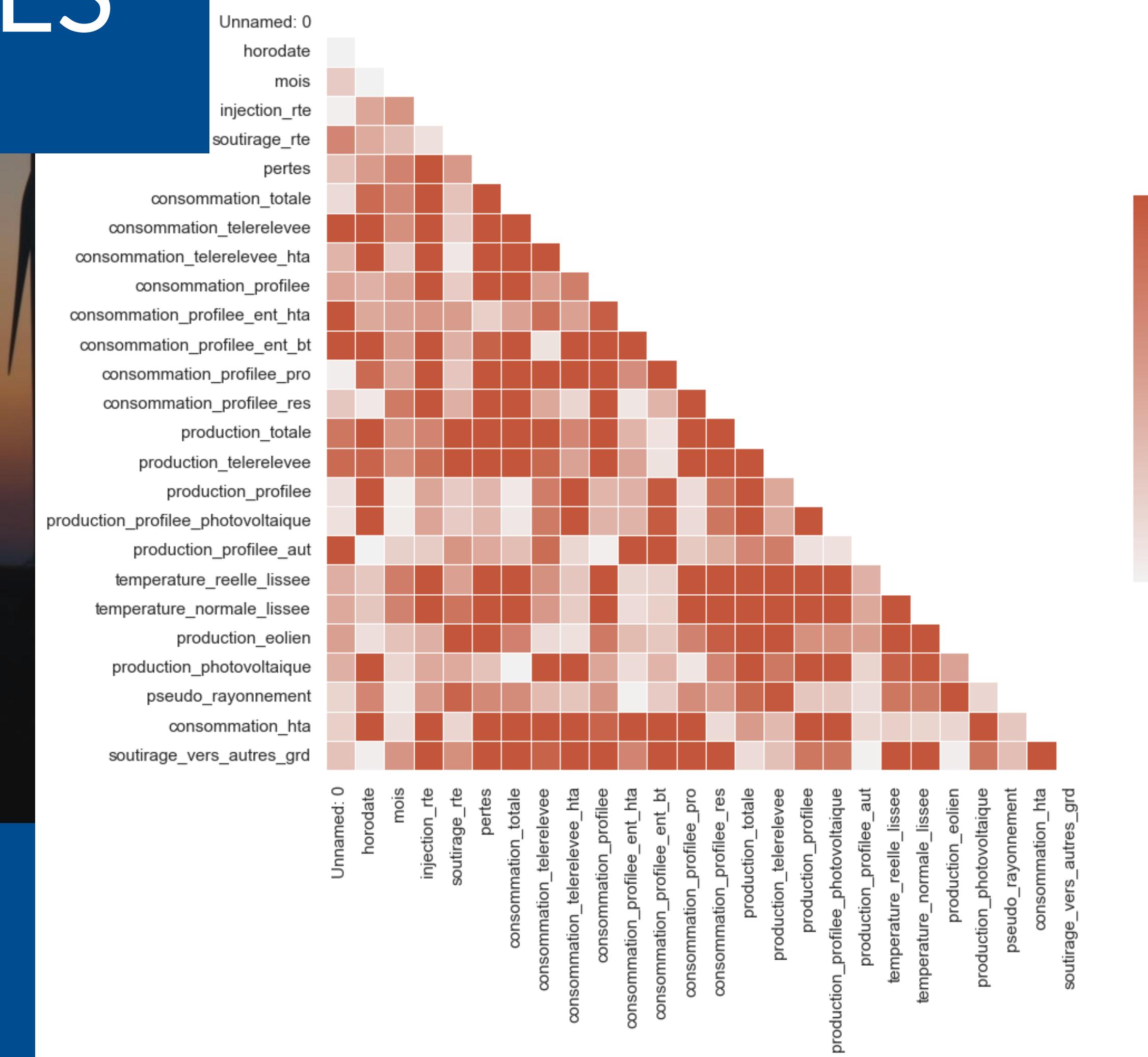


**Jean-Jacques Goldman**  
Auteur, compositeur & interprète

# CHOIX DES PARAMÈTRES

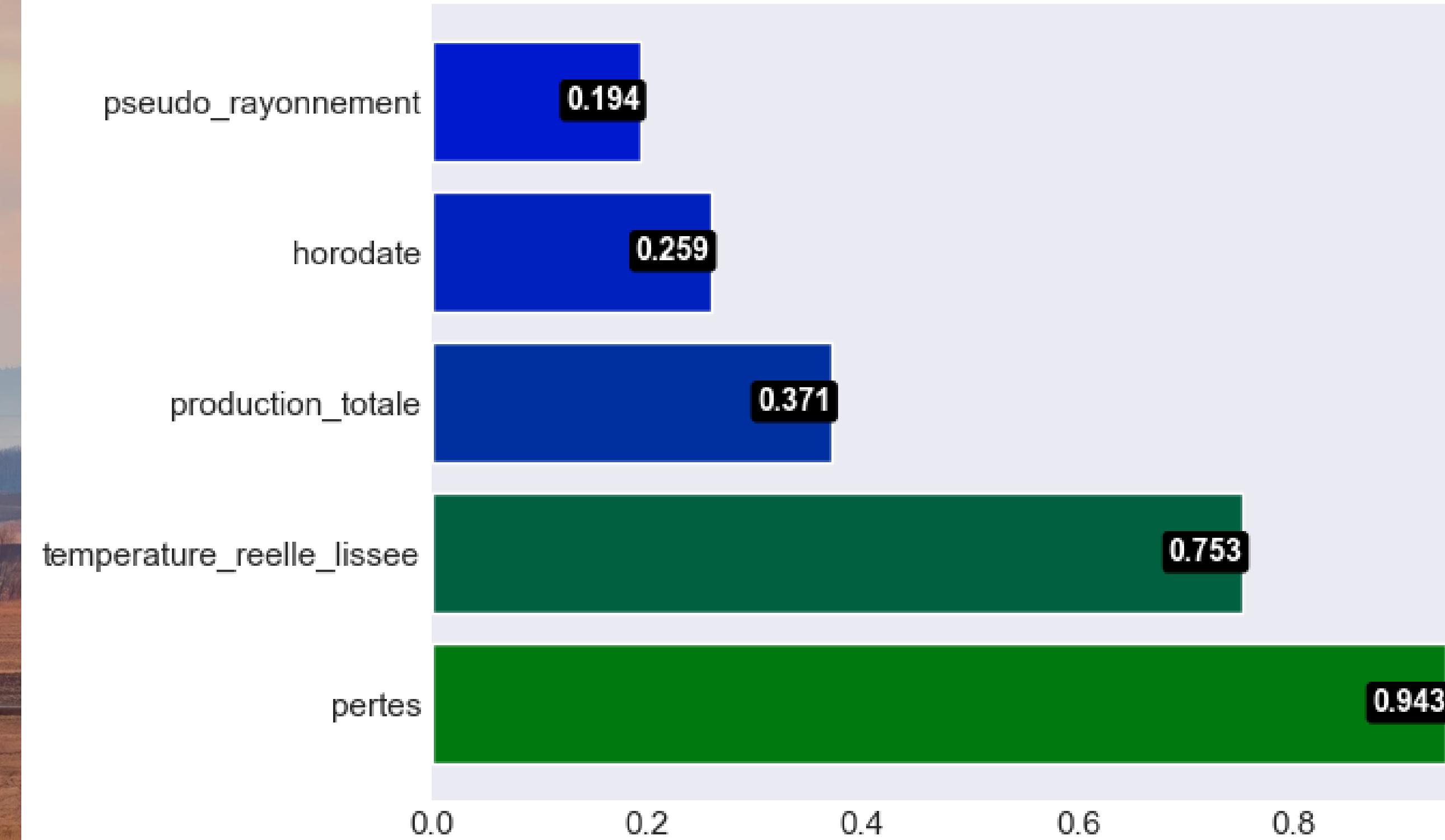
# Choix des paramètres en fonction de leurs influences

# Intercorrélation entre les différents paramètres dans la base de données en valeur absolue



# PARAMÈTRES DOMINANTS

Graphique des valeurs par paramètre



*Valeur de l'intercorrelation*

# QUELQUES MODÈLES



« Tu prends la pilule bleue, l'histoire s'arrête là. Tu te réveilles dans ton lit, et tu crois ce que tu veux. Tu prends la pilule rouge, tu restes au Pays des Merveilles et je te montre jusqu'où va le terrier.

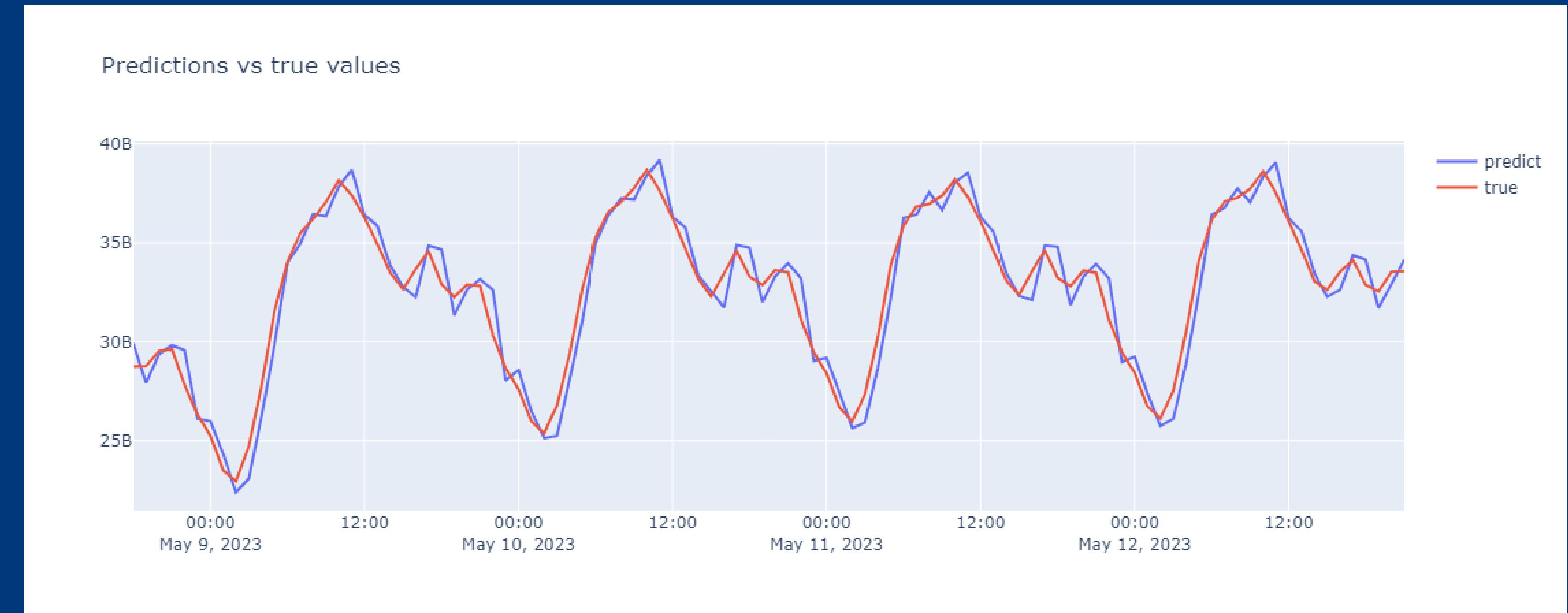
»

**Morpheus**  
*The Matrix*



# ARIMA

- Modélisation temporelle Auto Regressive Integration Mean Average de la consommation totale ;
- Optimisation en brute force ;
- Pas de features prises en compte ;
- Best  $(p,d,q) = (1,1,2)$ , MSE=0,9

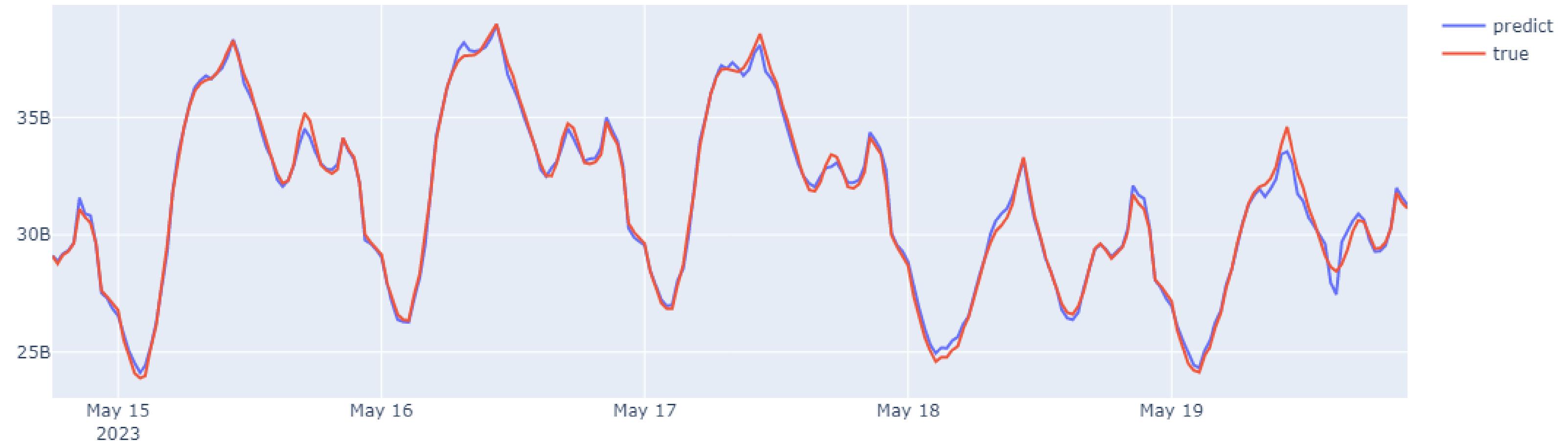




# XGBOOST

- Apprentissage supervisé sous Gradient Boost et arbres avec features ;
- Prise en compte de 17 features ;
- Problème au forecasting : inputs indisponibles.

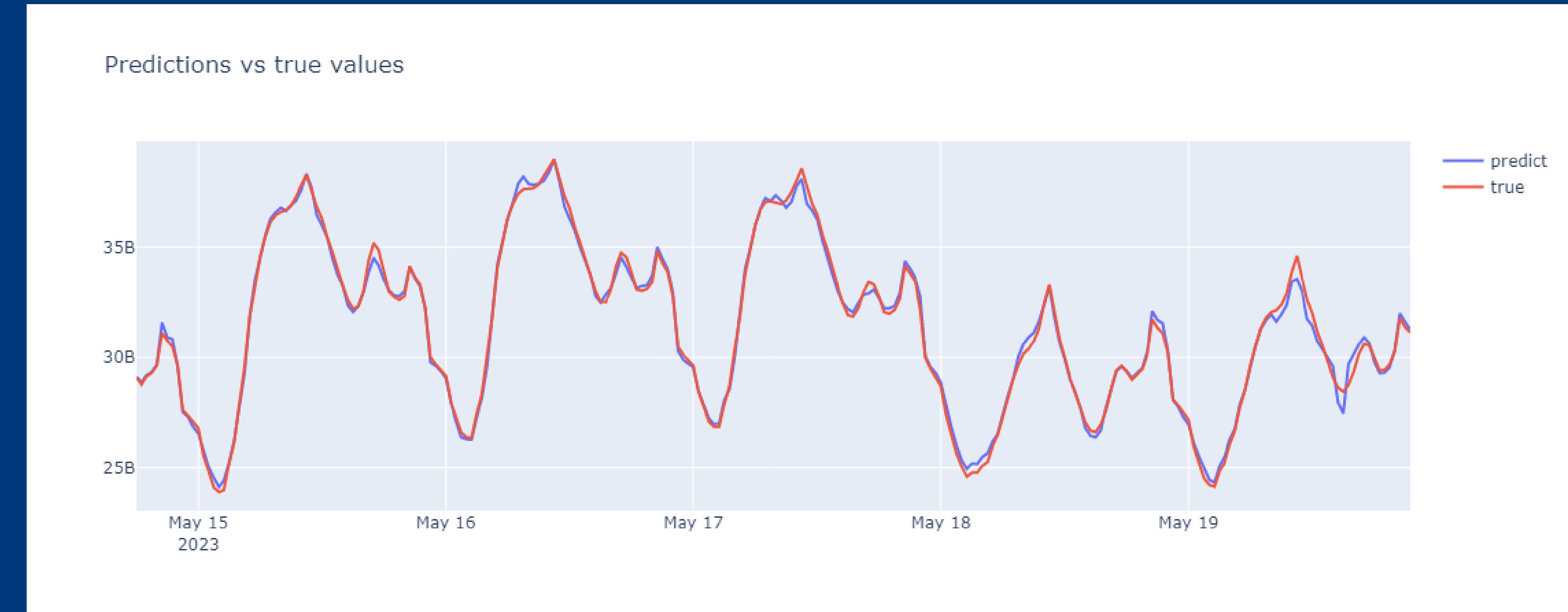
Predictions vs true values

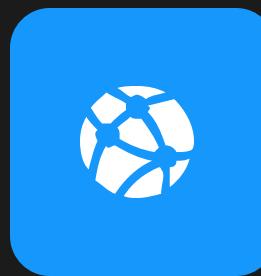




# RANDOMFOREST

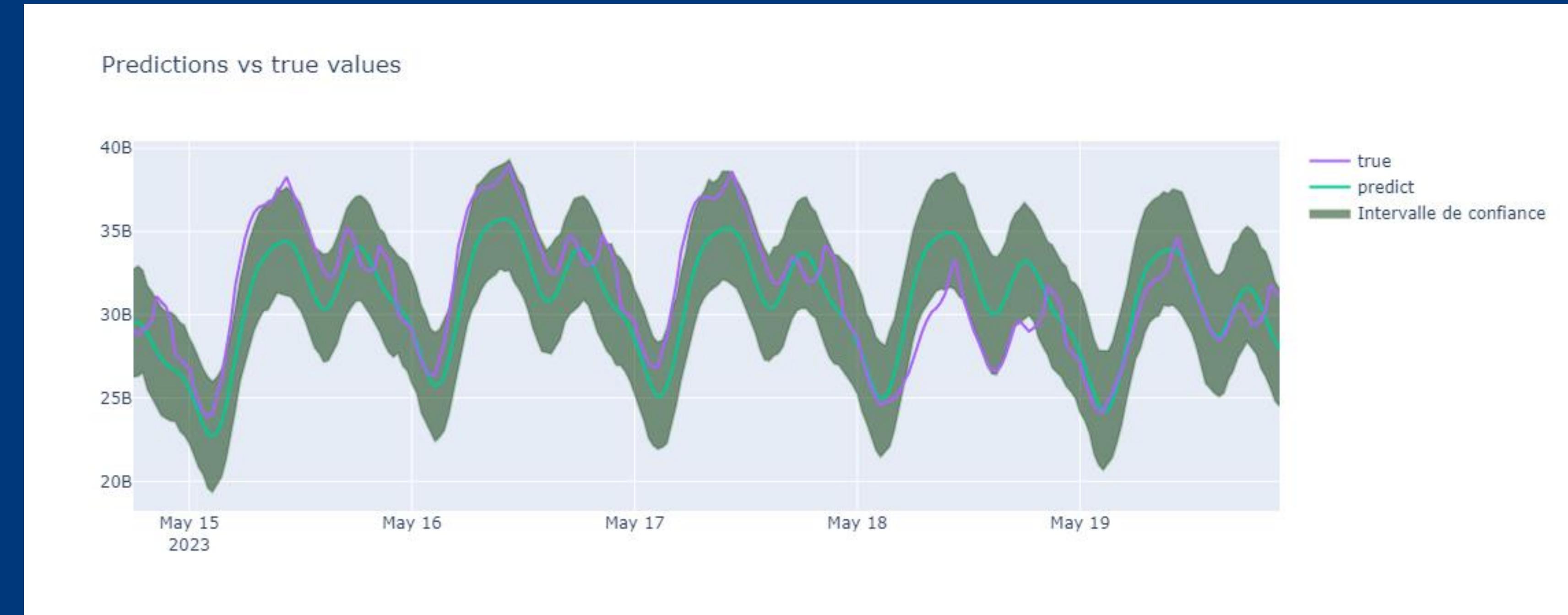
- Apprentissage supervisé sous arbres avec features
- Prise en compte de 17 features
- Problème au forecasting : inputs indisponibles





# PROPHET

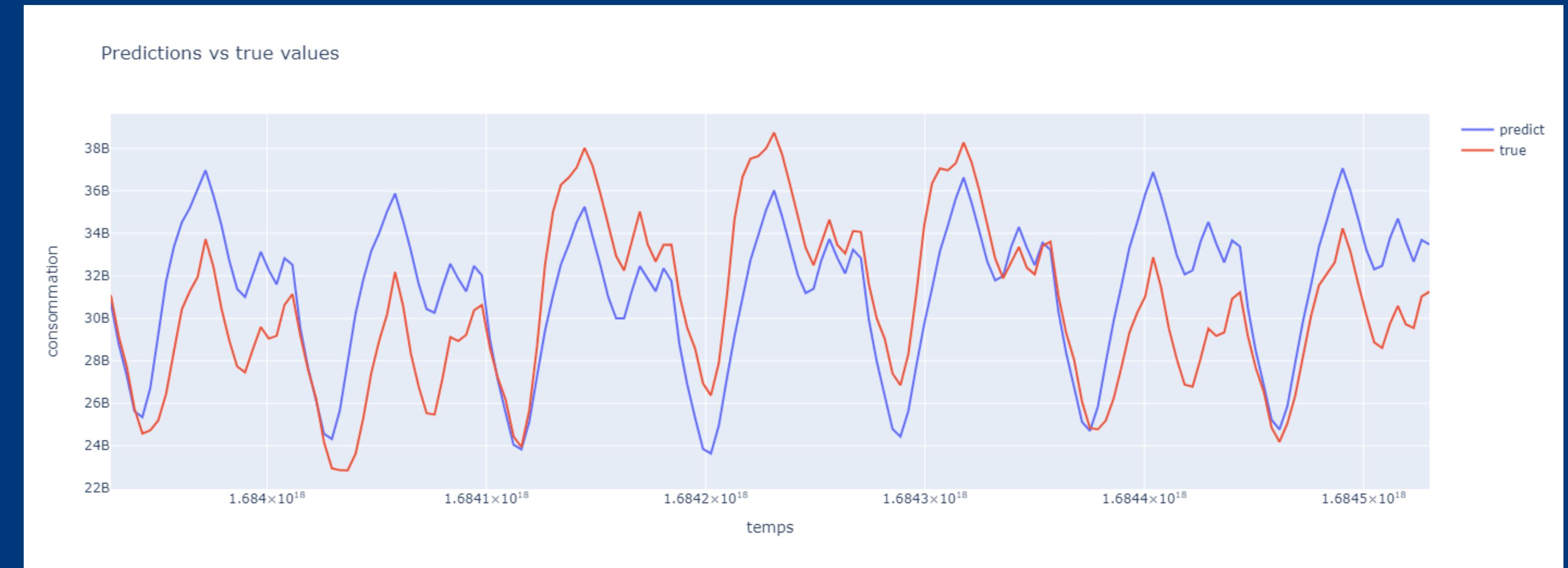
- Apprentissage supervisé sans features ;
- Algorithme Facebook ;
- Adapté big data ;
- Intervalle de confiance extrêmement élevé.





# RNN

- RNN Simple, LSTM et GRU ;
- Peuvent prendre en compte des dépendances temporelles entre entrée et sortie ;
- Peuvent traiter des séquences de longueurs variables ;
- Parfois difficile d'apprendre des dépendances long terme → LSTM ;
- Problème de vanishing/exploding gradient.



# CHALLENGES

Résultats *parfois* frustrants.

*Choix des hyperparamètres et de la méthode de prédiction.*

*Prétraitement des données*

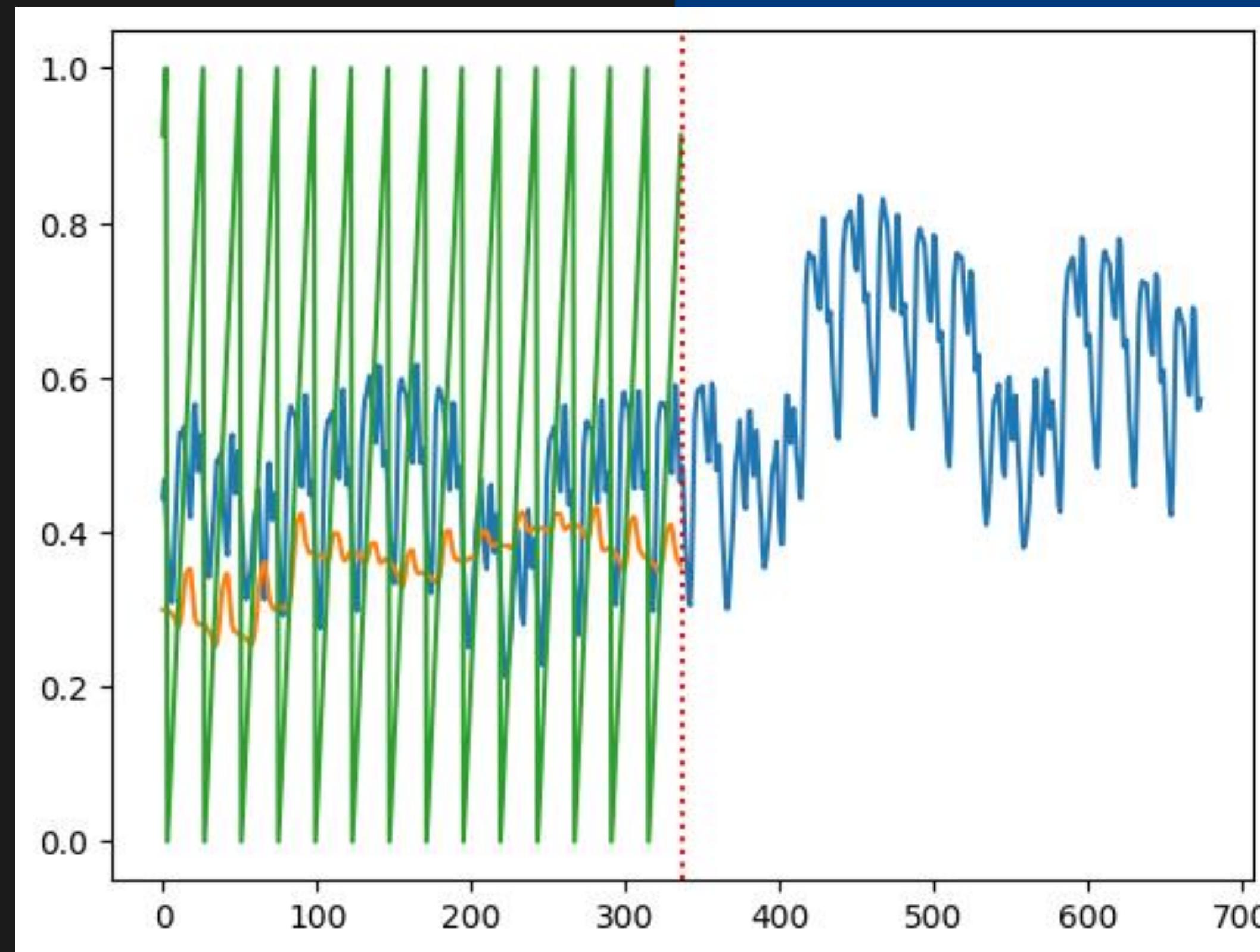
Choix des variables d'intérêt,  
normalisation, train/test split,  
formalisme de Pytorch.

89 %

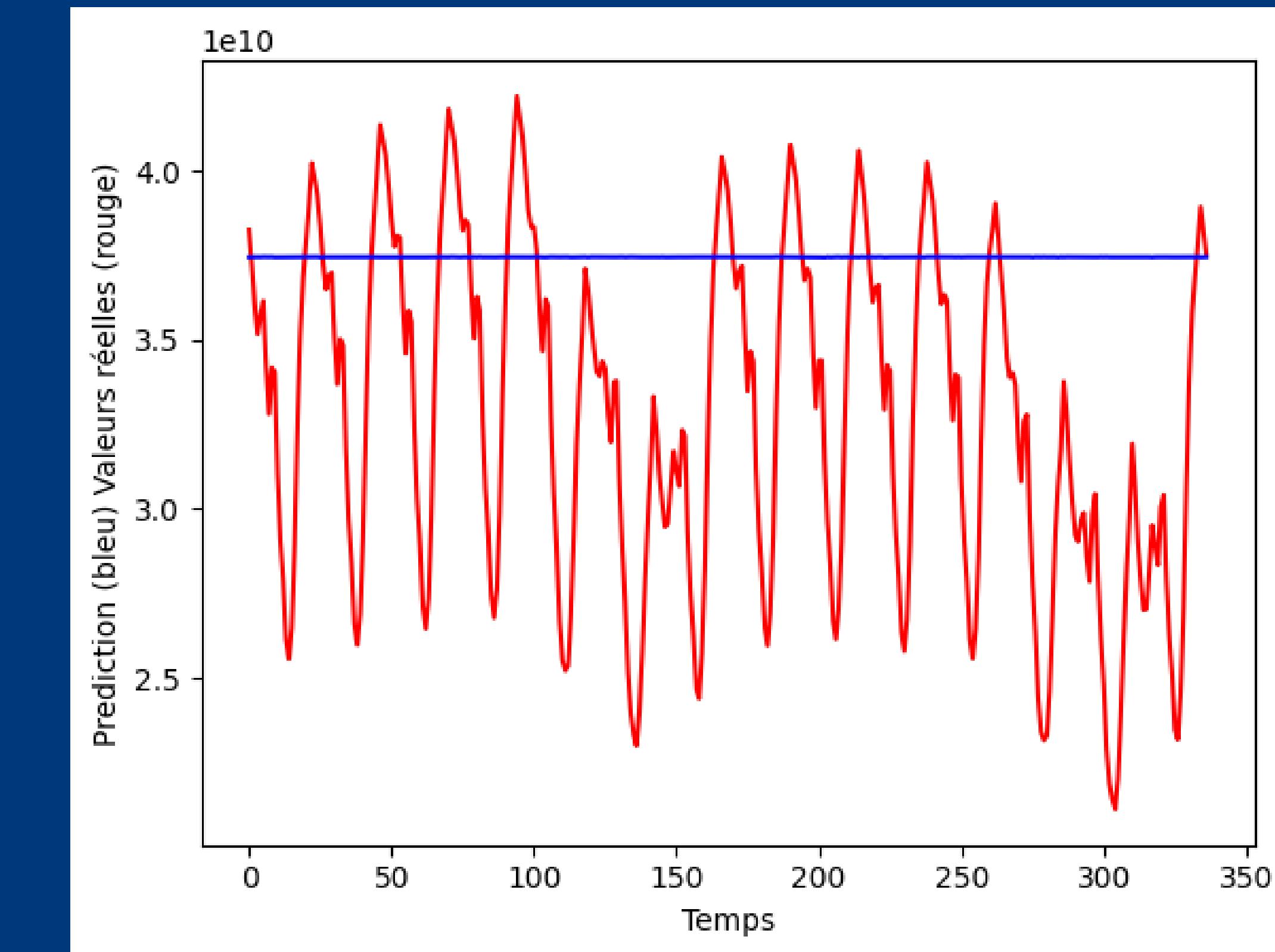
Processing

Grande puissance nécessaire :  
Google Colab gratuit ne suffit pas.

## Objectif



## Résultats...



# RANDOM-FOREST ET DIFFICULTÉS



«

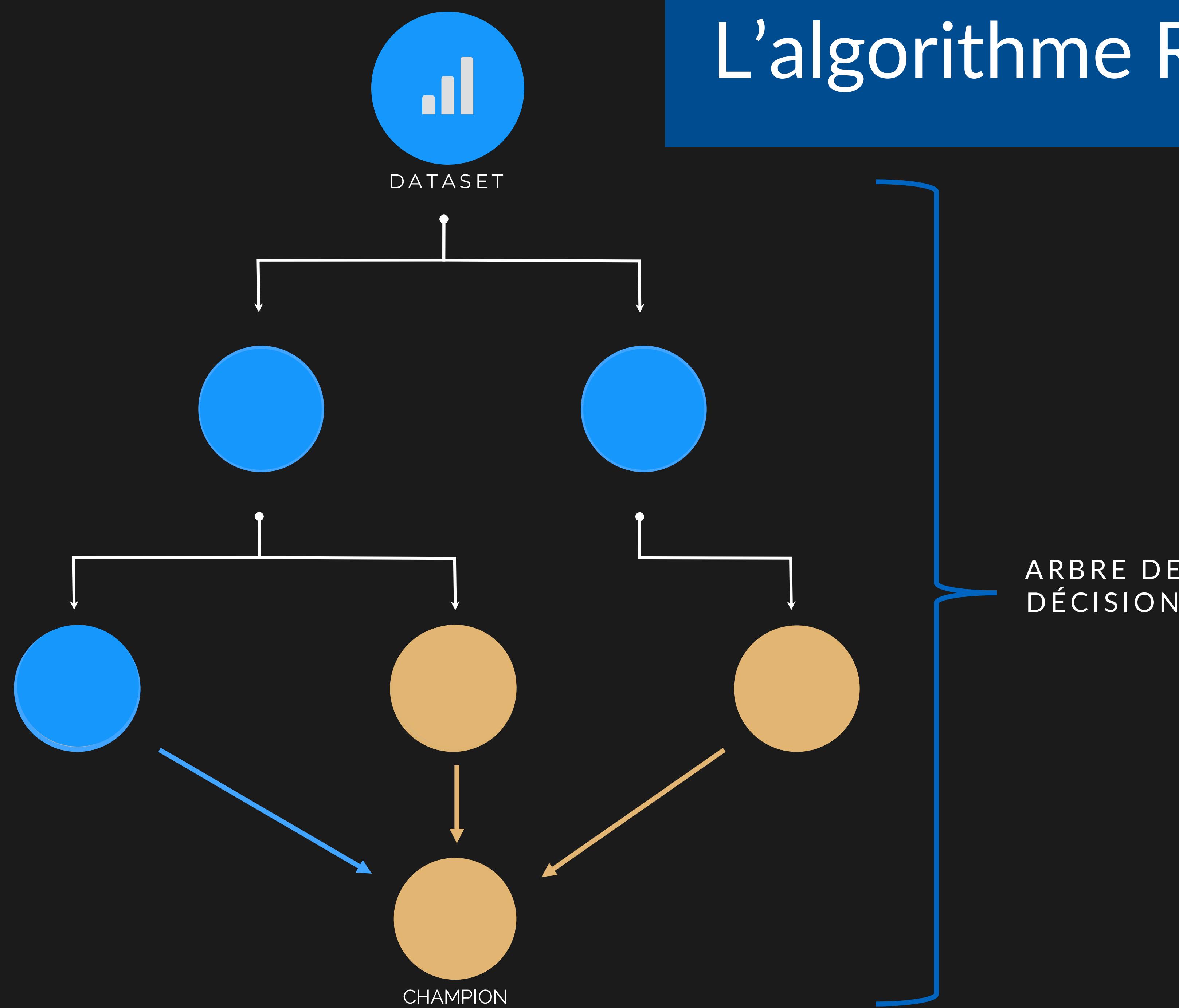
*Les emmerdes, ça vole toujours en escadrille.*

»

**Jacques Chirac**

*Ancien Président de la République*

# L'algorithme RandomForest



## PRINCIPE

Un unique arbre de décision n'est pas performant ;

Mise en parallèle de plusieurs arbre de décision ;

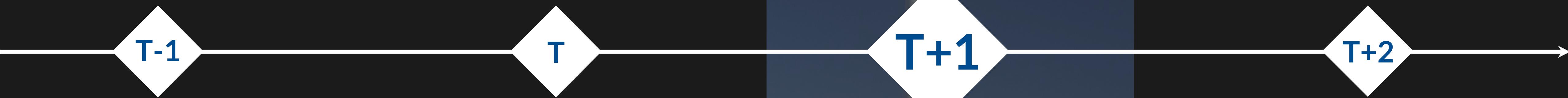
Chaque arbre de décision se forme sur différents samples ;

Ils élisent chacun un champion : victoire par démocratie.



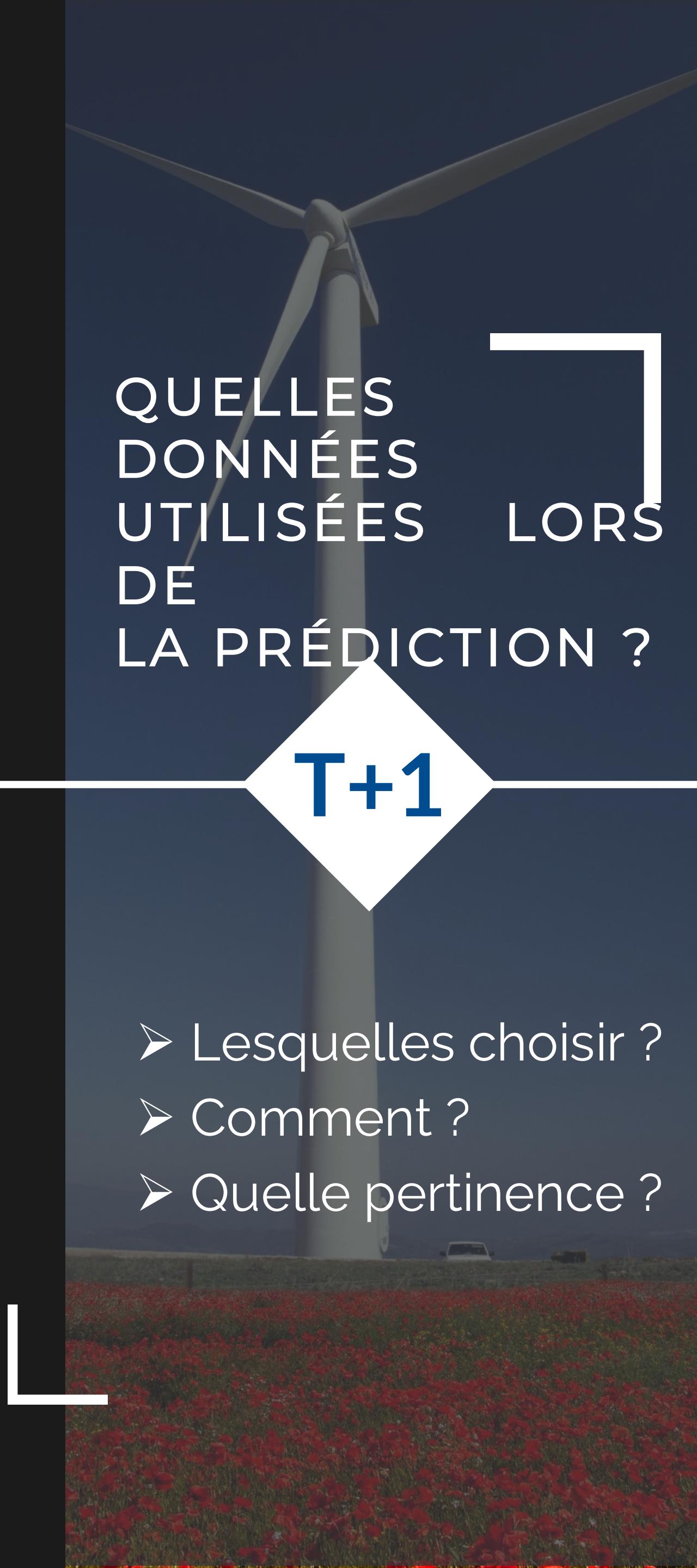
Évite l'overfitting

# LES PROBLÈMES



L'entraînement se base sur les données passées que l'on possède effectivement.

Cependant, pour la prédiction, ces paramètres d'entrée sont alors **inconnus**.



- Lesquelles choisir ?
- Comment ?
- Quelle pertinence ?

## NOTRE SOLUTION

Il faut donc trouver un moyen de se procurer des données pertinentes pour prédire la consommation d'électricité.

Pour cela, on a choisi des les prédire à l'aide :

- D'autres modèles
- D'outils à notre disposition



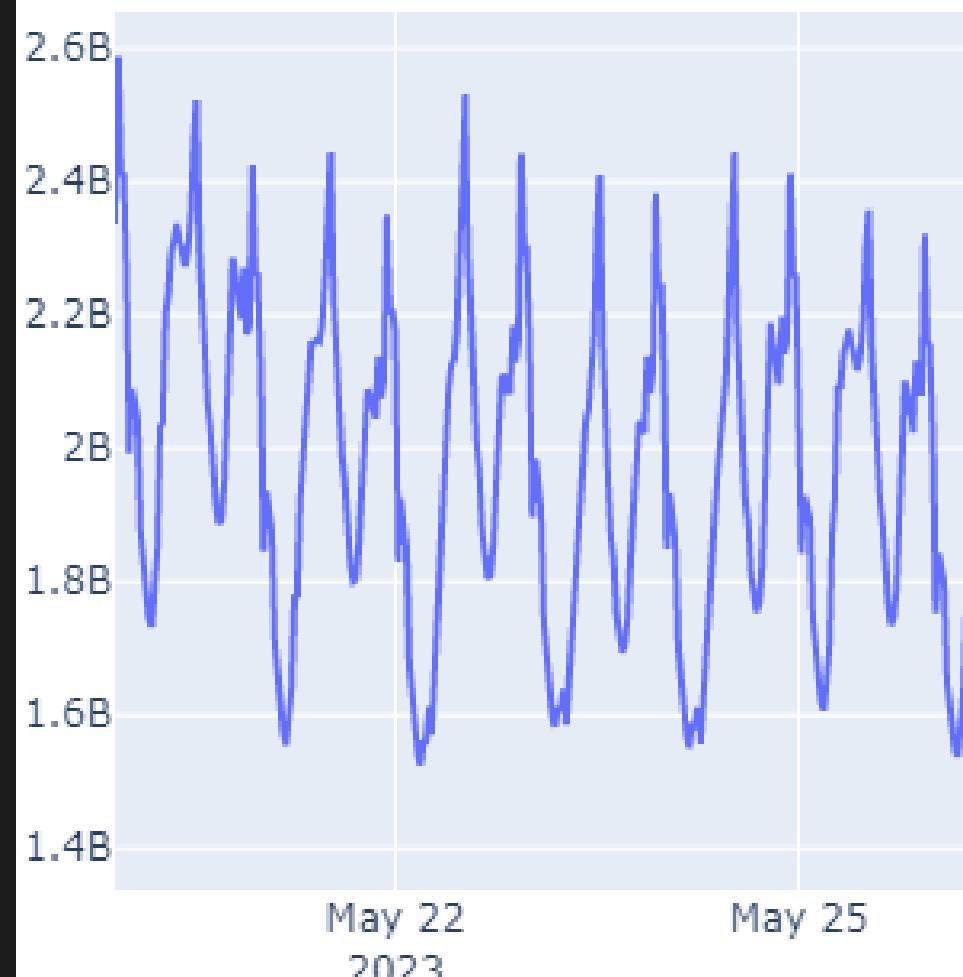
# CHOIX DES FEATURES

Certains de nos modèles sous apprentissage supervisé nécessitaient **plusieurs features d'entraînement et de test**. Or ces dernières ne sont pas forcément disponibles.

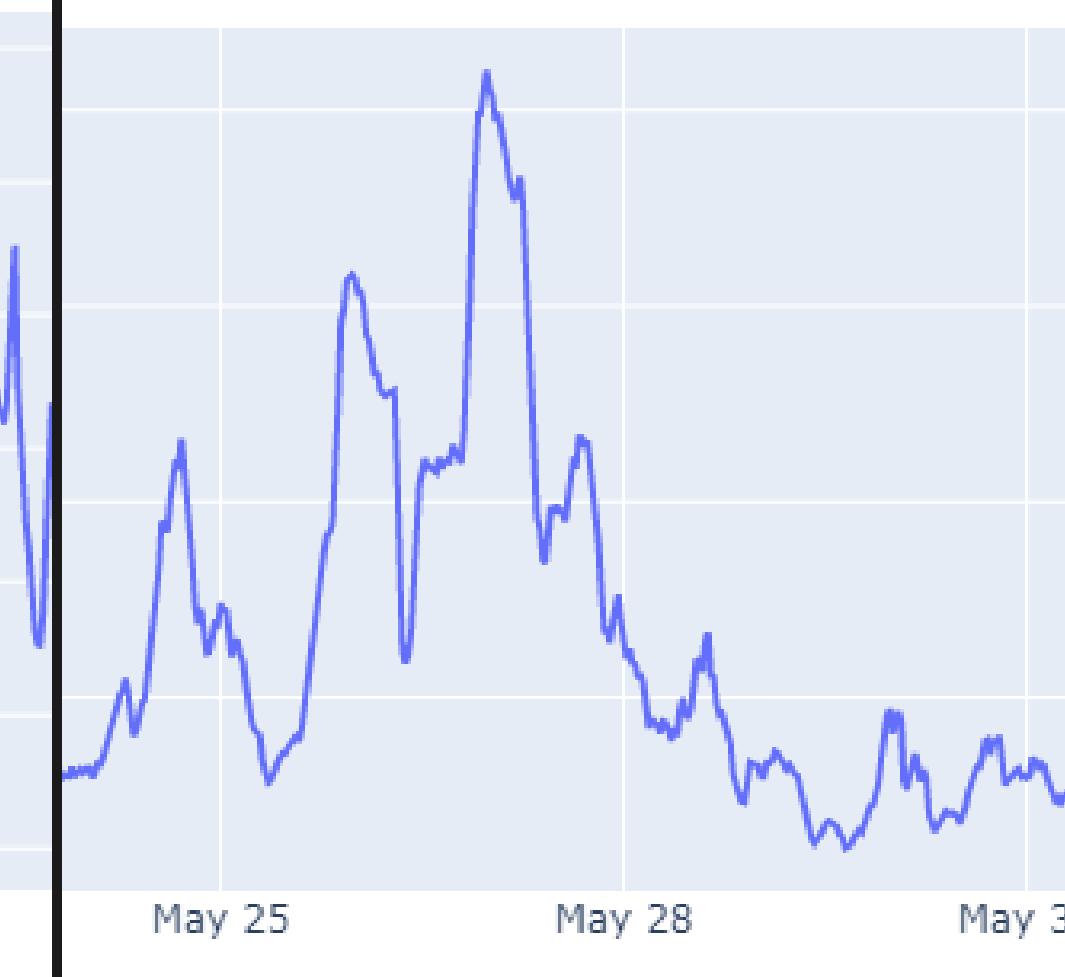
Nous avons décidé de les modéliser en **ARIMA** :

- Température normale : (2,1,2), MSE = 1e-2
- Pertes : (2,1,1), MSE = 1e-2
- Production totale : (2,2,2), MSE = 1e-2
- Production éolienne : (1,1,2), MSE = 1e-2
- Pseudo Rayonnement : (2,0,2), MSE = 1e-2

Feature Predictions of pertes

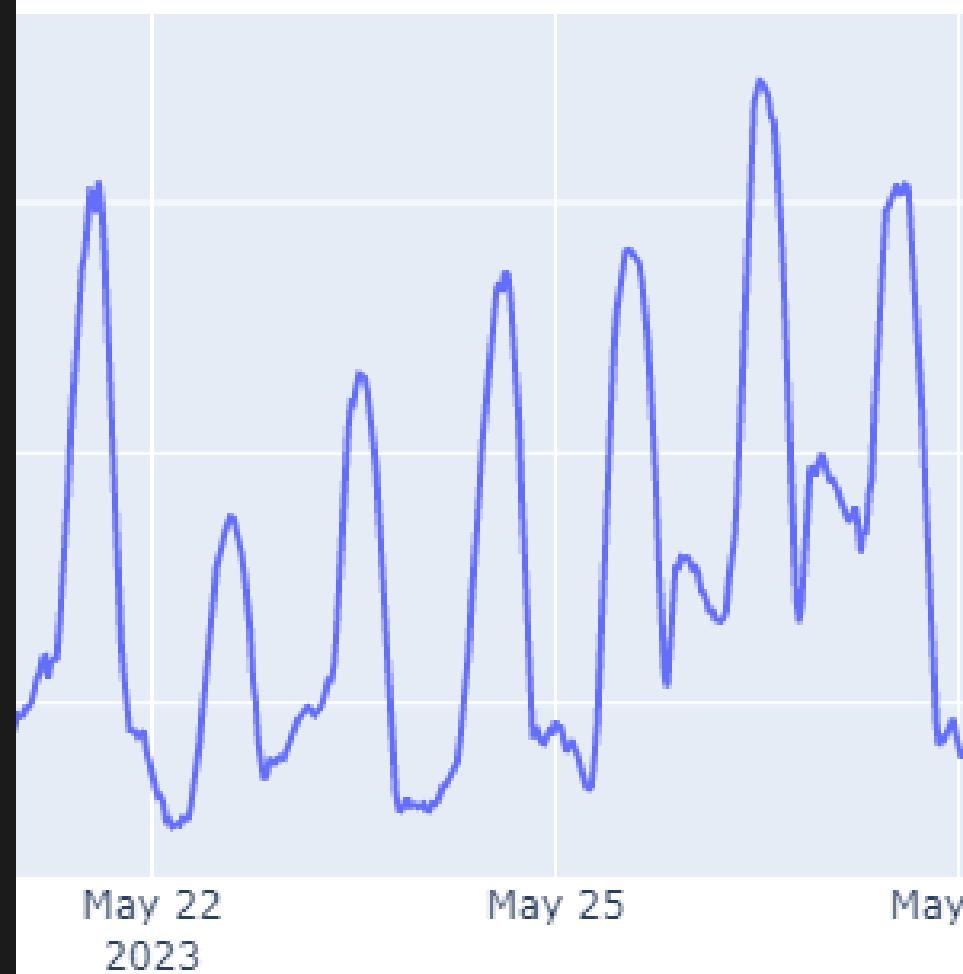


production éolienne

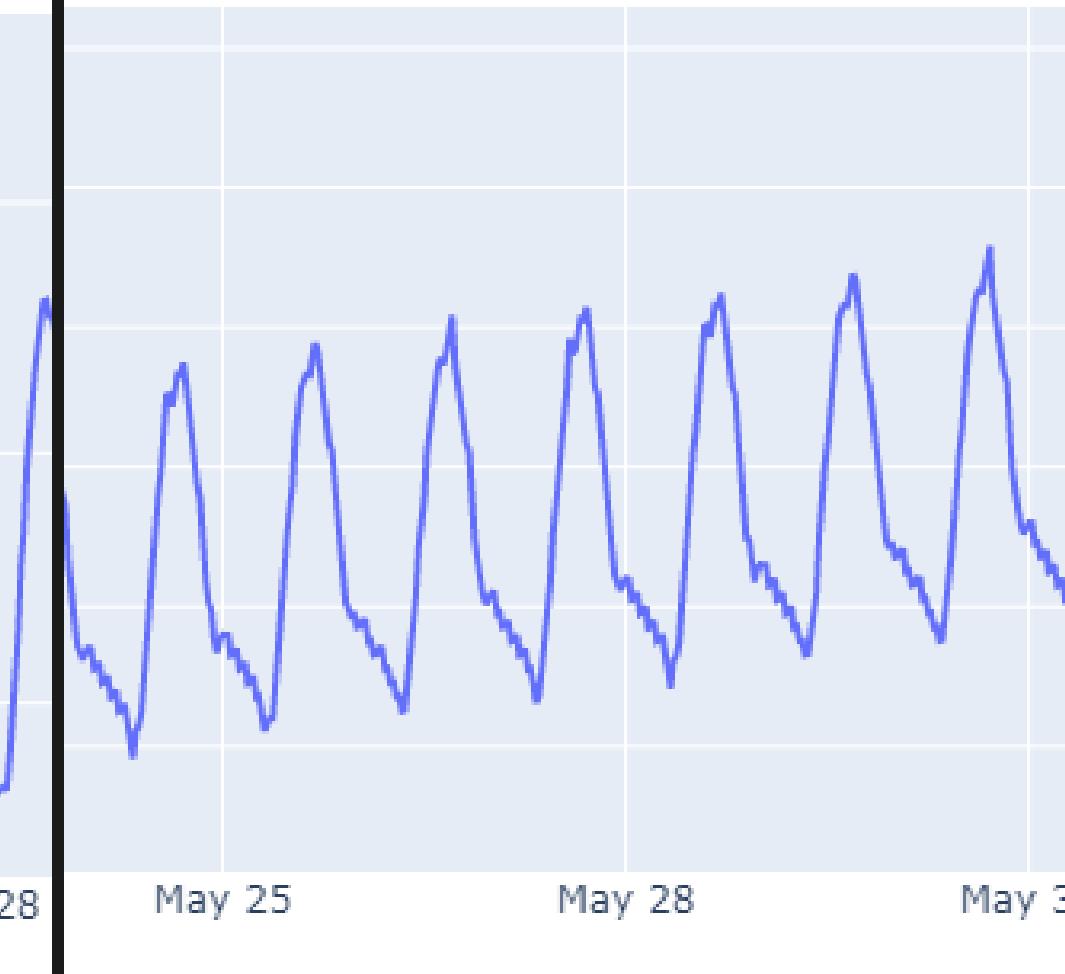


Autre solution non explorée ici : importer les prédictions des différents acteurs (Météo France etc.)

Predictions of production totale



Température Normale Lisee



pseudo rayonnement

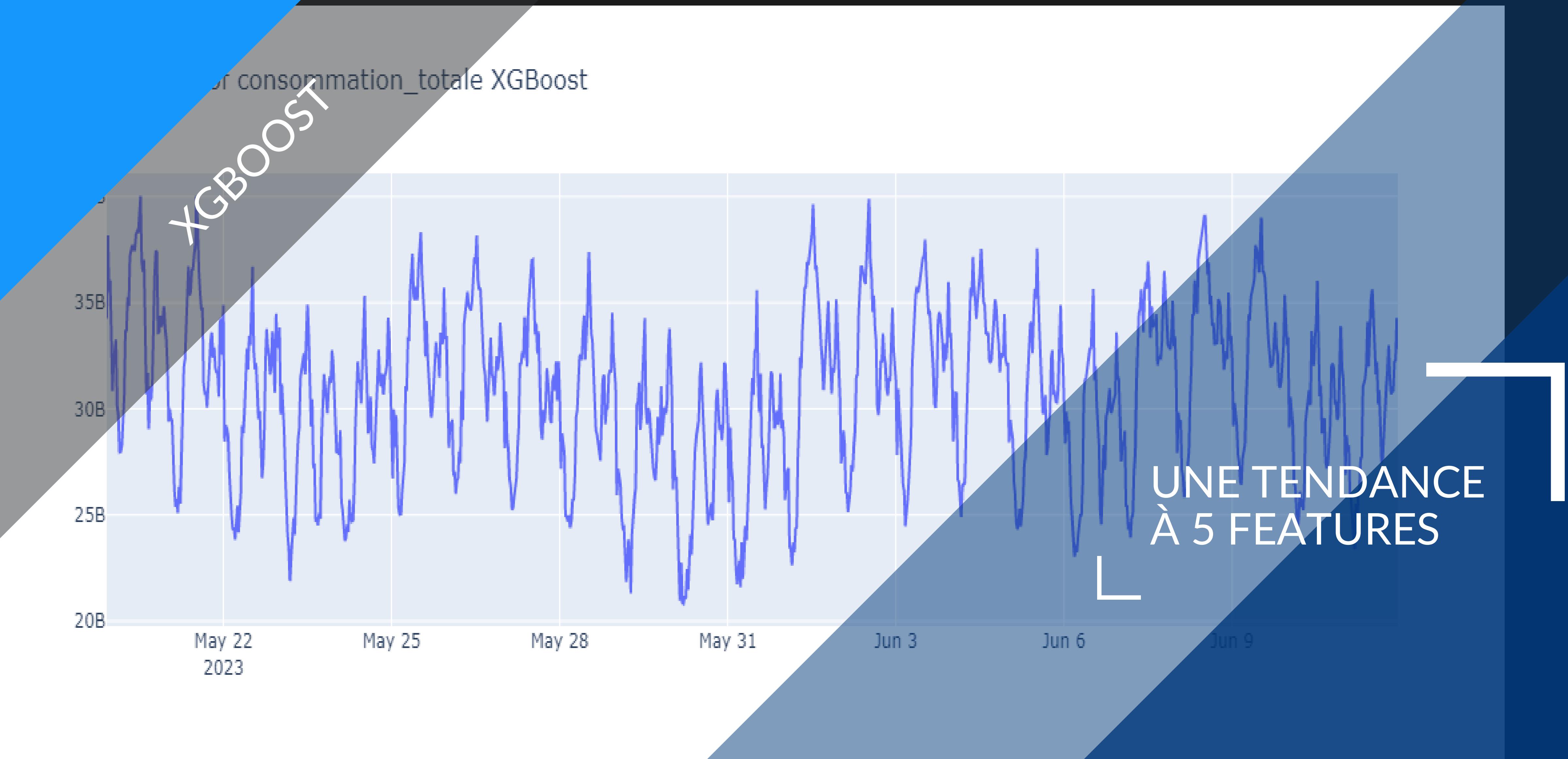


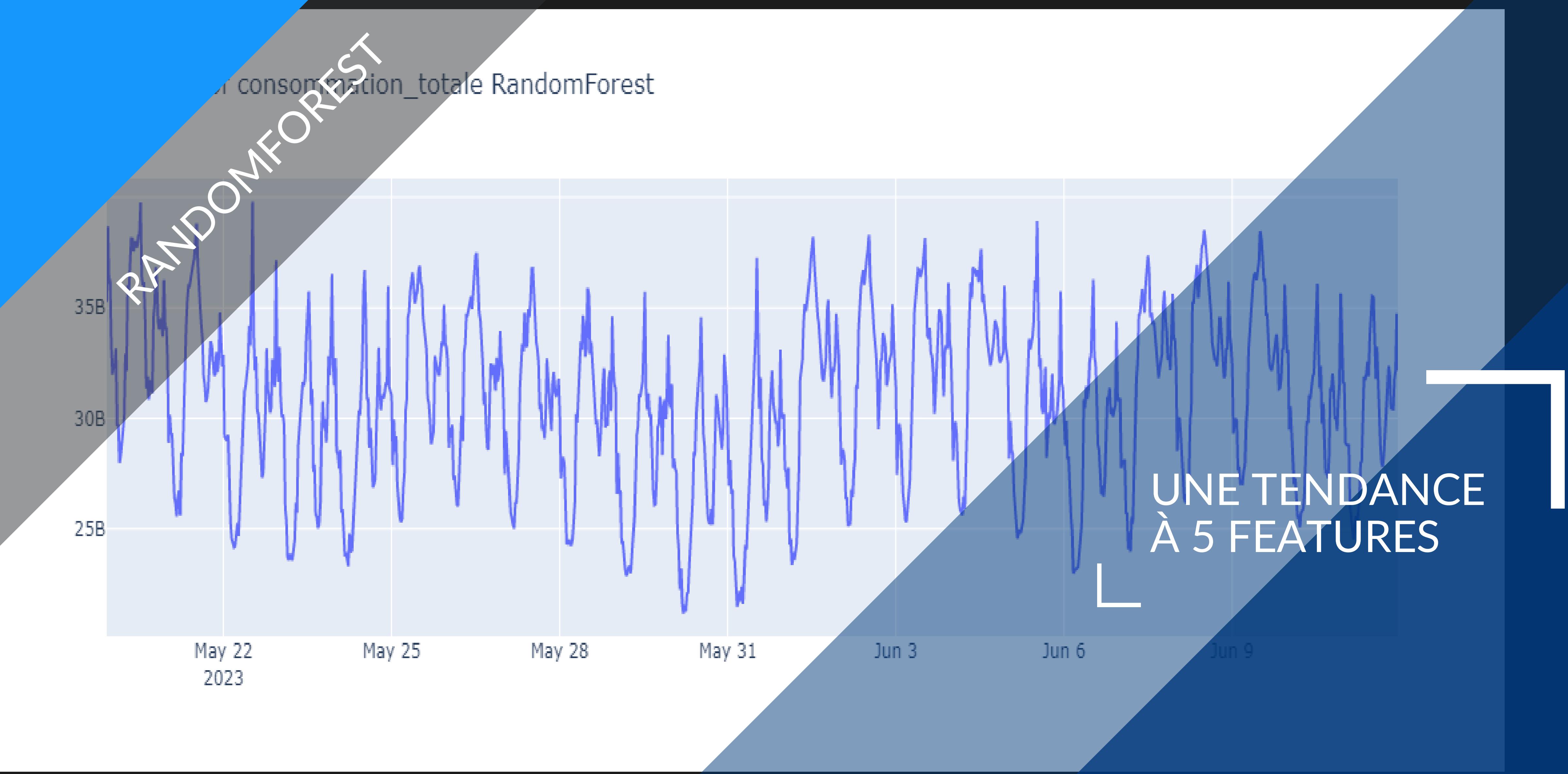
# RÉSULTATS ET PRÉDICTIONS

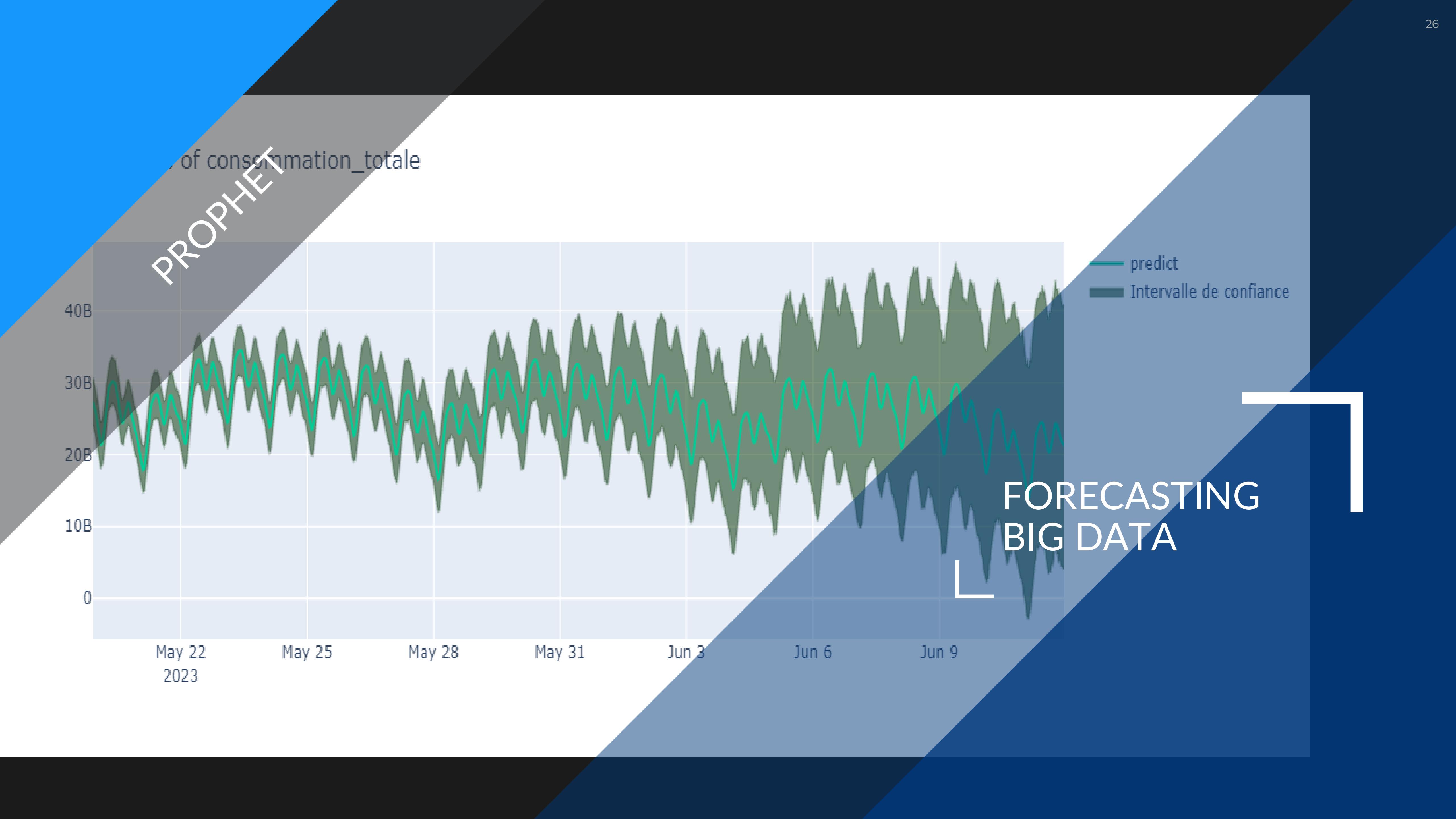


« Plus vous saurez regarder loin dans le passé,  
plus vous verrez loin dans le futur. »

**Winston Churchill**  
Ancien Premier Ministre britannique







# COMPARAISON MODÈLES

Comparaison des modèles de forecasting sur des valeurs de train et test connues.

	 COMPLEXITÉ	 INPUTS	 TECHNICITÉ	 MARGE	 MSE
ARIMA	✓	—	—	—	$9 * 10^{-1}$
XGBOOST	✓	✓	—	—	$8 * 10^{-4}$
RANDOMFOREST	✓	✓	—	—	$8 * 10^{-4}$
PROPHET	—	—	—	✓	$2 * 10^{-2}$
RNN	✓	✓	✓	—	2
LSTM	✓	✓	✓	—	$2 * 10^{-2}$

# ET ENSUITE ?

