

## Contexte du défi

Planète OUI se doit d'offrir un approvisionnement en électricité verte à des prix adaptés aux profils de consommation de ses clients. L'électricité est en effet achetée sur les marchés de l'électricité, où les prix sont très variables au cours d'une journée. Un site dont la part de consommation est élevée lorsque les prix sont bas, par exemple au milieu de l'après-midi, sera approvisionné à des coûts inférieurs à ceux d'un site consommant de l'électricité principalement pendant les périodes de pointe. Le profil de consommation d'une installation doit alors être évalué pour calculer la meilleure estimation des tarifs de fourniture, permettant à Planète OUI de couvrir ses coûts d'exploitation tout en fournissant son client au meilleur prix. Les sites dont la puissance de comptage est supérieure à 250 kVA font l'objet d'une méthodologie d'analyse précise. La plupart des sites se caractérisent par une consommation variant fortement avec la température en raison des systèmes de chauffage électrique. La consommation nationale d'électricité française est la plus sensible à la température en Europe, avec une augmentation de 2,4 GW par degré perdu, soit la production de 2 ou 3 réacteurs nucléaires. Néanmoins, les installations industrielles sont des cas particuliers, car leur consommation peut être fortement liée à des utilisations non thermosensibles, par exemple des procédés chimiques ou métallurgiques. Pour chaque site, l'objectif est notamment d'analyser les utilisations de la thermosensibilité. Toutefois, cela ne suffit pas pour déterminer précisément les profils de consommation, d'autres facteurs affectant la consommation, par exemple les saisonnalités annuelles, hebdomadaires et journalières. Lorsque Planète OUI prépare son offre de fourniture d'électricité, elle reçoit des données historiques de consommation des clients potentiels. Ces profils sont combinés à des simulations de prix de l'électricité pour calculer une distribution des coûts d'approvisionnement en €/MWh. Un percentile donné est ensuite utilisé pour couvrir les coûts d'approvisionnement pour un large éventail de scénarios de prix. Cependant, les données du client sont souvent incomplètes et s'étalent sur une période relativement courte, rarement supérieure à un an. Afin d'obtenir une tarification plus précise de l'offre, l'objectif de Planète OUI est de développer un modèle d'apprentissage machine où une ou plusieurs années de données de consommation extrapolées sont reconstituées à partir d'une seule année de données de mesure fournies par le client. Ces profils extrapolés seront également combinés avec les prix de l'électricité afin d'obtenir un ensemble de données d'analyse plus large.

## Objectifs du défi

L'objectif du défi est de prédire, à partir de l'analyse de la corrélation entre une année de consommation et les données de formation météorologique, la consommation électrique de deux sites donnés pour une année d'essai. Dans des

conditions opérationnelles, les nouveaux profils de consommation seraient intégrés à l'analyse des prix de la fourniture d'électricité.

#### Description des données

##### Description des données

Les données d'entrée du modèle à développer se composent des colonnes suivantes, les données étant enregistrées sur une base horaire :

"PIÈCE D'IDENTITÉ" : ID du point de données ;

"Horodatage" : Horodatage complet avec année, mois, jour et heure, en heure locale (CET et CEST) ;

"temp1", "temp2", "meannationaltemp" : Températures locales et moyennes nationales (°C) ;

"humidité1", "humidité2" : Humidité relative locale (%) ;

"loc1", "loc2", "locsecondary1", "locsecondary2", "locsecondary3" : les coordonnées des sites étudiés et secondaires, en degrés décimaux et de la forme (latitude, longitude).

"consommation secondaire1", "consommation secondaire2", "consommation secondaire3" : les données de consommation de trois sites secondaires, dont les corrélations avec les sites étudiés peuvent être utiles (kWh). En effet, les deux sites étudiés et les trois sites secondaires sont utilisés aux mêmes fins ; les données de sortie du modèle à développer prennent la forme suivante :

"PIÈCE D'IDENTITÉ" : ID du point de données ;

"consommation1", "consommation2" : les données de consommation des deux sites étudiés (kWh).

Les humidités relatives sont fournies avec les données de température car elles représentent des variables importantes pour la consommation d'électricité :

l'humidité influence en effet fortement le confort thermique. Pour reproduire les conditions de fonctionnement, certains points de données de température et d'humidité seront manquants. La méthode d'imputation doit être soigneusement examinée. Les variables "consommations secondaires" sont les données de consommation de plusieurs sites dont la puissance de comptage est supérieure à 250 kVA du portefeuille Planète OUI. Cette relation entre les différentes consommations des sites doit être étudiée pour compléter ou interpoler avec précision les données. Les horodatages peuvent être exprimés en mois ou jour de l'année, jour de la semaine et heures, pour étudier l'impact des saisonnalités annuelles, hebdomadaires et quotidiennes. Une attention particulière devrait être accordée au traitement des jours fériés nationaux.

#### Description de l'indice de référence

La persistance est un repère rapide et relativement précis. Elle consiste à estimer que la consommation horaire de l'année y+1 est égale à celle de l'année y, en déplaçant les données relatives aux week-ends et aux jours fériés ; par exemple, la

méthode de référence considérerait que la consommation horaire du samedi 4 août 2018 est égale à celle du samedi 5 août 2017.