

electricite renouvelable

Objectifs du défi/cf challenge

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.

but : • offre adapte aux profils des clients
• une ou plusieurs années de données de consommation extrapolées sont reconstituées à partir d'une seule année de données mesurées 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.

info :

--> prix varient bcq ds journee
--> 1_si conso elevee en periode creuse alors site alimentee a des couts inf
--> 2_profil de consommation d'une install ==> tarif de fourniture / frais de fonctionnement de PO
--> site > 250 kVA , analyse precise
--> conso variable selon la temperature ==> utilisation chauffage electriq
--> + 2,4GW par degre perdu
--> install indus, cas spécifique, utilisation non thermosensibles (processus chimique ou metallurgiq)

Description de l'indice de référence/cf challenge

La persistance est un repère rapide et relativement précis. Il consiste à estimer que la consommation horaire de l'année y+1 est égale à celle de l'année y, en déplaçant les données

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.

facteurs influencant :

pour chaque site, analyser les utilisations de la thermosensibilité
les saisons annuelles, hebdo, quotidienne
humidités (influence le confort thermique)
jours fériés nationaux, ponts
vacances

les miens : climatisation , vague de chaud,
fonte des neiges au printemps => augmentation de la production de 40%
autre moyen de chauffage (bois, gaz ...)
sur 1 semaine, impact du niveau de température

augmentation de T a 5 differente si soleil brille ou si il pleut
augmemntation de T de 5 differente a 10 ou 15 ou 20 degre
ecart de prix du a des promos, la dsponibilite de l'electricite,
saison
sucession de jour et de nuit
position du jourferie dans la semaine (mercredi different de vendredi)

processus nv client :

- > historiq du donnees de conso combine a des simulations de prix de l'electricite (un calcul des repartion des couts de fourniture en euros/MWh)
- > ? Un percentile donné est ensuite utilisé pour couvrir les coûts d'approvisionnement pour un large éventail de scénarios de prix

limite :

- > peu de donnees
- > periode courte, moins de 1 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 mesurées fournies par le client."

Dataset :

ID:	ID du point de donnees (site)
Timestamp:	Horodatage complet avec annee, mois, jour et heure (en heure local CET et CEST)
temp1:	Temperatures locales en °C
temp2:	Temperatures locales en °C
meannationaltemp :	moyenne des temperatures nationales en C
humidity1et 2:	Humidité relative locale (%)
loc1 et 2:	Coordonnees des sites etudies en degres decimaux et de la forme (latitude, longitude)
locsecondary 1 et 2:	Coordonnees des sites secondaires en degres decimaux et de la forme (latitude, longitude)
consumptionsecondary 1, 2 et 3 :	(kWh) donnees de consommation de 3 sites secondaires dont les correlations avec les sites etudiees peubvent etre utiles

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 secondairesi" 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.

Note de travail:

Calculer le coefficient de corrélation entre 2 variables numériques revient à chercher à résumer la liaison qui existe entre les variables à l'aide d'une droite. On parle alors d'un ajustement linéaire.

"wikipedia Correlation" [https://fr.wikipedia.org/wiki/Corr%C3%A9lation_\(statistiques\)](https://fr.wikipedia.org/wiki/Corr%C3%A9lation_(statistiques))

cf coefficient de corrélation linéaire de Bravais Pearson

cf matrice de corrélation

cf un estimateur (biaisé) du coefficient de corrélation

http://www.astro.ulg.ac.be/cours/magain/STAT/Stat_Main_Fr/Chapitre7.html

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.

1 modéliser les données avec la `meannationaltemp`

2 remplacer les moyennes par les saisons

3 calcul des facteurs d'effet météo $\text{facteur} = (\text{volume_météo} / \text{volume_normale}) - 1$

4 dépolluer l'historique de l'effet météo

Un logiciel à la manœuvre: Une fois que les jours de référence de consommation électrique similaires sont sélectionnés, un logiciel calcule les écarts de consommation entre la consommation réelle et ce qu'elle aurait été si le jour n'avait pas été férié. Le prévisionniste applique ensuite des coefficients pour pondérer ces différents écarts. La prévision pour le jour férié à venir tient compte du résultat de cette pondération. L'ensemble des données manipulées pour la prévision de consommation est alors « nettoyées » de l'impact des aléas climatiques.

Donc, plus qu'une corrélation directe du prix du gaz avec les températures, c'est la variation entre les prévisions à moyen terme et les températures réelles qui influence les prix.

kilovoltampère (kVA) mesure la puissance électrique apparente d'une installation (le kilowatt kW mesure la puissance active). Comme un voltampère est égal à un Watt ($1[\text{VA}] = 1[\text{W}]$), on exprime dans le langage courant la puissance d'un compteur électrique en kVA ou en kW.

Le kilovoltampère (kVA) mesure la puissance électrique dite apparente. Plus simplement, on peut dire qu'elle définit la valeur maximale d'une installation, c'est-à-dire la puissance maximale que pourra prendre la puissance active (kW).

GW est un symbole, qui signifie : gigawatt, unité de mesure de puissance
a unit of power equal to one billion (10^9) watts.

MWh Le kilowatt-heure . Si de l'énergie est produite ou consommée à puissance constante sur une période

donnée, l'énergie totale en kilowatts-heures est égale à la puissance en kilowatts multipliée par le temps en heures.

le rang centile d'une donnée ,dans une série statistique, indique le pourcentage des données ayant une valeur inférieure ou égale à cette donnée.

Les données de même valeur doivent se retrouver dans le même rang.

- Les données les plus faibles se retrouvent dans le rang 1.
- Les données les plus élevées se retrouvent dans le rang 100.
- Si le rang centile n'est pas un entier, on l'arrondit à l'entier supérieur.
- La série étudiée doit être ordonnée. Généralement, elle est ordonnée dans le sens croissant.

Rang centil d'une donnée (d) = $[(\text{nb de donnée de valeur inf à 'd' + la moitié du nb de données de même valeur que 'd'}) / [\text{Nb total de données}]]$

Une matrice de corrélation est utilisée pour évaluer la dépendance entre plusieurs variables en même temps. Le résultat est une table contenant les coefficients de corrélation entre chaque variable et les autres. Il existe différentes méthodes de tests de corrélation : Le test de corrélation de Pearson, la corrélation de Kendall et celle de Spearman qui sont des tests basés sur le rang.

traduction

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.