Enefit – Predict Energy Behavior of Prosumers

Overview :

The goal of the competition is to create an energy prediction model of prosumers to reduce energy imbalance costs. This competition aims to tackle the issue of energy imbalance, a situation where the energy expected to be used doesn't line up with the actual energy used or produced. Prosumers, who both consume and generate energy, contribute a large part of the energy imbalance. Despite being only a small part of all consumers, their unpredictable energy use causes logistical and financial problems for the energy companies.

Description :

The number of prosumers is rapidly increasing, and solving the problems of energy imbalance and their rising costs is vital. If left unaddressed, this could lead to increased operational costs, potential grid instability, and inefficient use of energy resources. If this problem were effectively solved, it would significantly reduce the imbalance costs, improve the reliability of the grid, and make the integration of prosumers into the energy system more efficient and sustainable. Moreover, it could potentially incentivize more consumers to become prosumers, knowing that their energy behavior can be adequately managed, thus promoting renewable energy production and use.

Evaluation : MAE

Dataset Description :

Votre défi dans cette compétition est de prédire la quantité d'électricité produite et consommée par les clients énergétiques estoniens ayant installé des panneaux solaires. Vous aurez accès aux données météorologiques, aux prix de l'énergie pertinents et aux enregistrements de la capacité photovoltaïque installée.

Tous les ensembles de données suivent la même convention de temps. L'heure est donnée en EET/EEST. La plupart des variables sont une somme ou une moyenne sur une période d'1 heure. La colonne datetime (quel que soit son nom) donne toujours le début de la période d'1 heure. Cependant, pour les ensembles de données météorologiques, certaines variables telles que la température ou la couverture nuageuse sont données pour un temps spécifique, qui est toujours la fin de la période d'1 heure.

**train.csv**

* county : Un code ID pour le comté.
* is\_business : Booléen indiquant si le prosommateur est une entreprise ou non.
* product\_type : Code ID avec la correspondance suivante des codes aux types de contrats : {0 : "Combiné", 1 : "Fixe", 2 : "Service général", 3 : "Spot"}.
* target : La quantité de consommation ou de production pour le segment pertinent pour l'heure. Les segments sont définis par le comté, is\_business et product\_type.
* is\_consumption : Booléen indiquant si la cible de cette ligne est une consommation ou une production.
* datetime : L'heure estonienne en EET (UTC+2) / EEST (UTC+3). Elle décrit le début de la période d'1 heure pour laquelle la cible est donnée.
* data\_block\_id : Toutes les lignes partageant le même data\_block\_id seront disponibles au même moment de prévision. Cela dépend des informations disponibles lorsque les prévisions sont réellement faites, à 11h chaque matin. Par exemple, si le data\_block\_id de prévision météorologique pour les prédictions faites le 31 octobre est 100, alors le data\_block\_id météorologique historique pour le 31 octobre sera 101 car les données météorologiques historiques ne sont réellement disponibles que le lendemain.
* row\_id : Un identifiant unique pour la ligne.
* prediction\_unit\_id : Un identifiant unique pour la combinaison comté, is\_business et product\_type. De nouvelles unités de prévision peuvent apparaître ou disparaître dans l'ensemble de test.

**gas\_prices.csv**

* origin\_date : La date à laquelle les prix du jour suivant sont devenus disponibles.
* forecast\_date : La date à laquelle les prix prévus devraient être pertinents.
* [lowest/highest]\_price\_per\_mwh : Le prix le plus bas / le plus élevé du gaz naturel sur le marché du jour suivant ce jour-là, en euros par mégawattheure équivalent.
* data\_block\_id

**client.csv**

* product\_type : Type de produit
* county : Code ID pour le comté. Voir county\_id\_to\_name\_map.json pour la correspondance des codes ID avec les noms de comté.
* eic\_count : Nombre agrégé de points de consommation (EIC - Code d'identification européen).
* installed\_capacity : Capacité installée de panneaux solaires photovoltaïques en kilowatts.
* is\_business : Booléen indiquant si le prosommateur est une entreprise ou non.
* date
* data\_block\_id

**electricity\_prices.csv**

* origin\_date :
* forecast\_date : Représente le début de la période d'1 heure pendant laquelle le prix est valide.
* euros\_per\_mwh : Le prix de l'électricité sur les marchés du jour suivant en euros par mégawattheure.
* data\_block\_id

**forecast\_weather.csv =>** Prévisions météorologiques qui auraient été disponibles au moment de la prédiction. Provenant du Centre européen pour les prévisions météorologiques à moyen terme.

* [latitude/longitude] : Les coordonnées de la prévision météorologique.
* origin\_datetime : L'horodatage de la génération de la prévision.
* hours\_ahead : Le nombre d'heures entre la génération de la prévision et la météo prévue. Chaque prévision couvre au total 48 heures.
* temperature : La température de l'air à 2 mètres au-dessus du sol en degrés Celsius. Estimée pour la fin de la période d'1 heure.
* dewpoint : La température de rosée à 2 mètres au-dessus du sol en degrés Celsius. Estimée pour la fin de la période d'1 heure.
* cloudcover\_[low/mid/high/total] : Le pourcentage du ciel couvert par des nuages dans les bandes d'altitude suivantes : 0-2 km, 2-6, 6+ et total. Estimé pour la fin de la période d'1 heure.
* 10\_metre\_[u/v]\_wind\_component : La composante [est/ouest] de la vitesse du vent mesurée à 10 mètres au-dessus de la surface en mètres par seconde. Estimée pour la fin de la période d'1 heure.
* data\_block\_id
* forecast\_datetime : L'horodatage de la météo prévue. Généré à partir de origin\_datetime plus hours\_ahead. Cela représente le début de la période d'1 heure pour laquelle les données météorologiques sont prévues.
* direct\_solar\_radiation : La radiation solaire directe atteignant la surface sur un plan perpendiculaire à la direction du soleil accumulée pendant l'heure, en watt-heures par mètre carré.
* surface\_solar\_radiation\_downwards : La radiation solaire, à la fois directe et diffuse, qui atteint un plan horizontal à la surface de la Terre, accumulée pendant l'heure, en watt-heures par mètre carré.
* snowfall : Chute de neige pendant une heure en unités de mètres d'équivalent en eau.
* total\_precipitation : Le liquide accumulé, comprenant pluie et neige, qui tombe sur la surface de la Terre au cours de l'heure décrite, en unités de mètres.

**historical\_weather.csv**

* datetime : Cela représente le début de la période d'1 heure pour laquelle les données météorologiques sont mesurées.
* temperature : Mesurée à la fin de la période d'1 heure.
* dewpoint : Mesuré à la fin de la période d'1 heure.
* rain : Différent des conventions de prévision. La pluie des systèmes météorologiques à grande échelle de l'heure en millimètres.
* snowfall : Différent des conventions de prévision. Chute de neige pendant l'heure en centimètres.
* surface\_pressure : La pression atmosphérique à la surface en hectopascals.
* cloudcover\_[low/mid/high/total] : Différent des conventions de prévision. Couverture nuageuse à 0-3 km, 3-8, 8+ et totale.
* windspeed\_10m : Différent des conventions de prévision. La vitesse du vent à 10 mètres au-dessus du sol en mètres par seconde.
* winddirection\_10m : Différent des conventions de prévision. La direction du vent à 10 mètres au-dessus du sol en degrés.
* shortwave\_radiation : Différent des conventions de prévision. L'irradiation horizontale globale en watt-heures par mètre carré.
* direct\_solar\_radiation : Radiation solaire directe
* diffuse\_radiation : Différent des conventions de prévision. L'irradiation solaire diffuse en watt-heures par mètre carré.
* [latitude/longitude] : Les coordonnées de la station météorologique.
* data\_block\_id