

EDF R&D : PROPOSITION DE STAGE

Benchmark des Foundation Models pour les séries temporelles : les courbes de consommation de clients résidentielles

Le contexte :

Environnement de travail :

Le département SEQUOIA (Services, Economie, Questions hUmaines, Outils innovants et IA) de la R&D, intervient en appui de la direction Marketing de la Branche Commerce d'EDF.

Au sein de ce département, le groupe E7C est spécialisé en analyse de données de consommation, et en modélisation statistique dans le but de concevoir des nouvelles offres de fourniture ou des services.

Plus précisément, nos missions sont de :

- Mettre en place des panels de clients pour collecter des courbes de charge (globale ou sur des usages spécifiques) : des compétences en sondage, ou en dimensionnement de panels sont alors requises.
- Modéliser des profils de consommation.
- Développer des algorithmes pour faire de l'estimation/prévision de consommation.
- Estimer l'impact d'une offre ou d'un service sur la consommation des clients.

Ce stage s'inscrit dans le projet « Smart Customer Data », dont l'objectif est de recueillir et valoriser les données clients. Par ailleurs, un volet méthodologique du projet identifie et développe des traitements innovants et créateurs de valeur pour la relation client et les offres à partir des courbes de charge au pas de temps des compteurs communicants (Linky et Gazpar), mais aussi à pas de temps fin (seconde).

Contexte du stage :

Afin d'éviter les risques de blackout du réseau électrique, il est crucial d'assurer en temps réel un équilibre entre la production et la consommation d'énergie. Dans un contexte d'intégration croissante des énergies renouvelables dans le mix énergétique, il est nécessaire de mettre à disposition des solutions de flexibilité à nos clients particuliers : consommer davantage lorsque la production d'énergies renouvelables est élevée et consommer moins lors de baisses de production. Ainsi, il devient essentiel de prévoir de manière fine et individuelle la courbe de consommation électrique des clients résidentiels, au pas de temps demi-horaire.

Depuis quelques années, de nombreux foundation models performants pour le traitement du texte ont émergé. Plus récemment, certains chercheurs ont adapté ces modèles aux séries temporelles. Ce stage propose d'étudier ces foundation models de séries temporelles en réalisant un benchmark appliqué aux tâches de prédiction et de clustering, afin d'améliorer notre capacité à prévoir et gérer la consommation énergétique de nos clients.

En particulier, pour la tâche de clustering, il s'agira de générer un espace de représentation des séries temporelles via un foundation model, puis d'appliquer un modèle de clustering sur cet espace. Si l'espace de représentation permet de distinguer efficacement les différents motifs de consommation, cela peut également améliorer les performances de prédiction. Une approche intéressante serait d'utiliser le clustering comme une étape préalable à la prévision, en appliquant ensuite un modèle de prédiction spécifique à chaque cluster.

Dans ce cadre, les foundation models seront appliqués aux séries de consommation électrique du segment résidentiel. Ces séries temporelles se caractérisent par leur non-stationnarité et leur dépendance à des variables exogènes, à la fois statiques (type de logement, option tarifaire, etc.) et dynamiques (météo, prix). L'objectif est d'identifier le foundation model le plus adapté à ce type de séries temporelles et, dans un second temps, d'optimiser les prévisions grâce à des méthodes de stacking ou de transfert d'apprentissage.

Le stage :

Le stage comportera plusieurs phases :

État de l'art des méthodes envisagées et montée en compétences :

- Se familiariser avec la problématique de prédiction et de clustering d'un ensemble de séries temporelles.
- Etablir un état de l'art des nouvelles méthodes de fondation appliquées aux séries temporelles.

Bibliographie non exhaustive des modèles :

- MOMENT: A Family of Open Time-series Foundation Models: <https://arxiv.org/abs/2402.03885>
- Unified Training of Universal Time Series Forecasting Transformers: <https://arxiv.org/pdf/2402.02592>
- Chronos: Learning the Language of Time Series: <https://arxiv.org/pdf/2403.07815>

Prise en main des données :

- Jeu de données comprenant :
 - Courbe de charge au pas de temps 30 minutes.
 - Questionnaires de connaissance client associés.
 - Données météorologiques.
- Exploration des statistiques descriptives et nettoyage (valeurs aberrantes et manquantes).

Mise en œuvre :

- Implémenter les méthodes identifiées lors de la revue de littérature en Python.
- Adapter ces méthodes à notre usage et à nos données.
- Réaliser un benchmark (comparaison) des différentes méthodes.
- Mettre en place différentes méthodes de stacking de modèles/transfer learning.
- Comparer les performances avec les méthodes classiques en Machine Learning.

Capitalisation des travaux effectués :

- Rédaction d'un rapport (revue de littérature, comparaison, conclusion des résultats obtenus).
- Code Python accompagné d'une documentation.

Informations complémentaires :

La R&D propose ce stage de fin d'études, à destination d'étudiants en écoles d'ingénieurs ou Master 2, spécialisés en Statistiques / Data Science / Deep Learning / IA.

Compétences : L'étudiant(e) sera amené(e) à mettre en œuvre des compétences :

- Data management (gestion, nettoyage, exploration, SQL...).
- Mise en place de méthodes de Deep Learning sur séries temporelles.
- Le langage de programmation proposé est Python.
- Une connaissance approfondie de PyTorch serait fortement appréciée.

Dates : à partir de février/mars 2025 (flexible).

Lieu du stage : EDF Lab Paris-Saclay – Recherches et Développement, 7 Bd Gaspard Monge, 91120 Palaiseau.

Contacts :

Philippe CHARPENTIER (Ingénieur Chercheur Expert), mail : philippe.charpentier@edf.fr

Youssef KADHI (Ingénieur Chercheur), mail : youssef.kadhi@edf.fr

Claire LAMBERT (Cheffe de Projet), mail : claire-c.lambert@edf.fr

Merci d'envoyer un C.V et une lettre de motivation sur ces e-mails.

Horaires : 35 h / semaine

Indemnité : en fonction des formations