



EDF R&D : PROPOSITION DE STAGE DE FIN D'ETUDES (6 MOIS)

Détection d'anomalies et de dérives dans les consommations énergétiques par méthodes d'IA (Transformers & Apprentissage Auto-Supervisé)

Le contexte :

Le département SEQUOIA (Services, Economie, Questions humaines, Outils innovants et IA) de la R&D intervient en appui de la Direction Marketing des Clients Particuliers et de la Direction des Systèmes d'Informations et du Numérique de la branche Commerce d'EDF.

Un des enjeux d'EDF Commerce est l'amélioration, l'accélération du déploiement et la pérennisation de ses algorithmes d'IA, notamment ceux visant à aider les clients à mieux comprendre leur consommation, afin de mieux la maîtriser.

Depuis le déploiement du compteur Linky, selon le niveau de consentement, nous pouvons relever la consommation au pas mensuel, journalier voire demi-horaire des clients. En rejoignant notre équipe, vous utiliserez l'intelligence artificielle pour décrypter ces signaux et avoir un impact direct sur la transition énergétique.

Votre mission : développer un système capable de détecter automatiquement des anomalies dans les courbes de charge individuelles. Les « anomalies » étant de natures multiples : ponctuelles, en « patterns », en dérive, intra-séries, inter-séries, elles permettront de traiter des cas d'usage concrets : détecter les changements de comportement (adoption d'un véhicule électrique, télétravail) et les anomalies (fraude, panne de compteur, pics extrêmes dûs à la température) pour proposer des services plus intelligents, personnalisés et économies en énergie.

Ce stage viendrait également en appui d'une thèse autour de l'explicabilité des modèles multimodaux, dont l'un des uses-cases est d'expliquer les « anomalies » (au sens large) de consommation. Un des pré-requis est donc de bien les détecter. Vous travaillerez donc en étroite collaboration avec notre doctorant sur le sujet.

Le stage :

Au cours du stage, vous irez au-delà des méthodes classiques pour explorer, implémenter et évaluer des algorithmes à la pointe de la recherche mondiale. L'objectif est de créer un détecteur d'anomalies non-supervisé, robuste et précis, capable de transformer une simple courbe de charge en actions métier à forte valeur ajoutée :

- **Marketing Proactif** : Identifier l'installation d'une pompe à chaleur pour proposer une offre adaptée.
- **Service Client Augmenté** : Contacter un client suite à une hausse de consommation pour l'aider à la maîtriser.
- **Opérations Optimisées** : Prévenir une panne de compteur avant qu'elle ne survienne.

REALISATION D'UN ETAT DE L'ART

- **Lecture et recherche d'articles scientifiques pertinents sur le sujet**
 - Etude des modèles classiques (à base de distance, de densités, kNN, One Class SVM, ...)
 - Montée en compétence sur les modèles récents (Deep Learning pour la reconstruction ou la prédiction : AE, VAE, LSTM, GAN, Transformers)
- *Une bibliographie non-exhaustive est jointe à l'offre de stage.*
- **Echange avec les chercheurs et thésards au sein de la communauté scientifique d'EDF R&D**

LA FRONTIERE SCIENTIFIQUE A REPOUSSER

Les données de consommation électrique sont complexes : bruitées, multi-saisonnieres et non-stationnaires. Les modèles classiques s'épuisent vite. Votre défi sera de tester l'hypothèse selon laquelle les architectures neuronales les plus récentes peuvent non seulement "apprendre" la normalité, mais aussi être contraintes intelligemment pour ne pas considérer une anomalie durable comme un nouveau comportement normal (le problème de la "sur-généralisation").

EXPERIMENTATIONS ENVISAGEES

Après avoir établi baselines solides (Prophet, Auto-encodeurs LSTM), vous plongerez au cœur de la recherche actuelle en testant les familles de modèles les plus prometteuses :

- **Les Transformers de Nouvelle Génération** : Au-delà du célèbre *Anomaly Transformer*, vous implémenterez son successeur, **MEMTO**. Ce modèle puissant est doté d'une "mémoire" des comportements normaux, ce qui le rend exceptionnellement doué pour détecter les changements de fond sans les "oublier".
- **L'Apprentissage Auto-Supervisé (SSL) Révolutionné** : Vous explorerez **CARLA**, une approche de rupture. Au lieu de laisser le modèle deviner ce qui est abnormal, vous lui apprendrez explicitement en lui injectant des anomalies synthétiques (pannes, fraudes, etc.). C'est une méthode de pointe pour spécialiser un modèle aux enjeux métier.
- **(Bonus) Exploration des Modèles de Fondation** : En fin de stage, vous aurez l'opportunité d'évaluer le potentiel des grands modèles pré-entraînés pour les séries temporelles, comme **Moirai 2 (Salesforce)**, **TimesFM (Google)**, pour la détection d'anomalies en mode "zéro-shot" (sans ré-entraînement), une voie de recherche qui pourrait redéfinir le domaine.
- **Benchmark des modèles**
 - **Comparaison des performances** : Un défi majeur dans le domaine est la fiabilité des comparaisons expérimentales. On pourra se tourner vers des initiatives comme **TSB-AD (Time-Series Benchmark for Anomaly Detection)** qui visent à résoudre ce problème en collectant, nettoyant et organisant plus de 1000 séries temporelles provenant de diverses sources pour fournir une base d'évaluation plus fiable et standardisée. La question des **métriques** sera abordée, en adaptant leurs choix au type d'anomalie (AUC-ROC, AUC-PR, VUS – Volume Under Surface)
 - Un cas d'application concrète sur un jeu de données EDF sera mis en œuvre à partir d'un panel de mesures de clients, ayant répondu à un questionnaire décrivant leur foyer et leurs équipements électriques. Environ 450 clients ont vu leurs courbes annotées « à la main » permettant de connaître réellement des anomalies de consommation.

DEROULEMENT SUGGERÉ

- **Mois 1-2** : Maîtrise des fondamentaux, bibliographie ciblée et construction d'un framework d'évaluation robuste.
- **Mois 3-4** : Plongée dans le Deep Learning. Implémentation, entraînement et optimisation des modèles de pointe (en priorité **CARLA** et **MEMTO**).
- **Mois 5** : Analyse comparative rigoureuse et exploration des modèles de fondation.
- **Mois 6** : Synthèse, rédaction du rapport et recommandations pour une future industrialisation.

CE QUE CE STAGE VOUS APPORTERA :

- **Une expertise technique de pointe** : Vous développerez une compétence rare sur les architectures d'IA les plus avancées (Transformers, apprentissage auto-supervisé) pour l'analyse des séries temporelles.
- **La maîtrise d'un projet de R&D complet** : Vous piloterez toutes les étapes du projet, de l'analyse de la recherche mondiale à la validation sur des données industrielles réelles et la formulation de recommandations.
- **Un impact direct et valorisant** : Votre travail contribuera concrètement à des enjeux de transition énergétique et d'amélioration du service client.
- **Une immersion dans un écosystème d'innovation** : Vous serez pleinement intégré(e) à notre communauté scientifique, en collaboration étroite avec des chercheurs et un doctorant sur le sujet.

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. Une appétence pour la recherche sera appréciée.

Profil recherché :

- Excellentes compétences en Python et maîtrise du framework de Deep Learning PyTorch. Une expérience démontrée avec des modèles de Deep Learning et ces outils serait particulièrement appréciée.
- Solide compréhension des architectures neuronales (RNN, Transformers) et de l'apprentissage non-supervisé.
- Curiosité, rigueur scientifique et un fort intérêt pour le secteur de l'énergie et son impact sociétal.
- Lecture et synthèses d'articles de recherche
- De bonnes pratiques sur Git seront valorisées.

Dates : Stage d'une durée de 6 mois. La date de début est flexible entre février et mai 2026.

Lieu du stage : EDF Lab Paris-Saclay – Recherche et Développement, 7 Bd Gaspard Monge, 91120 Palaiseau. Le stagiaire pourra bénéficier de mesures de télétravail en fonction du niveau d'autonomie.

Contacts :

Laurent BOZZI (Ingénieur Chercheur Senior), mail : laurent.bozzi@edf.fr
Adrien PETRALIA, PhD, Ingénieur-Chercheur, mail : adrien.petralia@edf.fr
Alexandre LEGROSSE, Doctorant, EDF / Lyon2 mail : alexandre.legrosse@edf.fr
Merci d'envoyer un C.V et une lettre de motivation sur ces trois e-mails.

Horaires : 35 h / semaine.

Indemnité : en fonction des formations.

Bibliographie :

Les articles de recherche suivants sont pertinents pour les travaux du stage :

- Wang, F., Jiang, Y., Zhang, R., Wei, A., Xie, J., & Pang, X. (2025). A Survey of Deep Anomaly Detection in Multivariate Time Series: Taxonomy, Applications, and Directions. *Sensors*, 25(1), 190.
<https://doi.org/10.3390/s25010190>
- Song, J., Kim, K., Oh, J., & Cho, S. (2023). **Memto**: Memory-guided transformer for multivariate time series anomaly detection. *Advances in Neural Information Processing Systems*, 36, 57947-57963.
- Darban, Z. Z., Webb, G. I., Pan, S., Aggarwal, C. C., & Salehi, M. (2025). **CARLA**: Self-supervised contrastive representation learning for time series anomaly detection. *Pattern Recognition*, 157, 110874.
- Boniol, P., Linardi, M., Roncallo, F., & Palpanas, T. (2020, April). Automated anomaly detection in large sequences. In *2020 IEEE 36th international conference on data engineering (ICDE)* (pp. 1834-1837). IEEE.
- Boniol, P., Linardi, M., Roncallo, F., Palpanas, T., Meftah, M., & Remy, E. (2021). Unsupervised and scalable subsequence anomaly detection in large data series. *The VLDB Journal*, 30(6), 909-931.
- Hundman et al. (2018), Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding
- Malhotra et al. (2015), Long Short Term Memory Networks for Anomaly Detection in Time Series
- Park et al. (2018), A Multimodal Anomaly Detector for Robot-Assisted Feeding using an LSTM-based Variational Autoencoder
- Su et al. (2022), Time Series Anomaly Detection Using LSTM-Based Variational Autoencoder-Generative Adversarial Networks
- Tuli et al. (2022), TranAD: Deep Transformer Networks for Anomaly Detection in Time Series Data
- Deng & Hooi (2021), Graph Neural Network-Based Anomaly Detection in Multivariate Time Series