



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

Exploration des courbes de charge par modèles de fondation pour séries temporelles et données tabulaires.

Le stage :

Le département SEQUOIA (Services, Economie, Questions hUmaines, Outils innovants et IA) de la R&D d'EDF appuie la branche Commerce en proposant des approches innovantes à même d'approfondir la connaissance client.

Nous explorons notamment la mise en place de modèles capables d'analyser les courbes de charge des clients professionnels (dites « PRO »), c'est-à-dire les données temporelles de consommation électrique des entreprises. L'objectif est d'extraire des informations pertinentes dans le but d'accompagner les entreprises dans la réduction de leur empreinte carbone et la maîtrise de leur consommation.

Dans le cadre du stage proposé, nous nous intéressons à **l'expérimentation de méthodes de Deep Learning et de Machine Learning visant à identifier des usages, équipements ou comportements à partir des courbes de charge d'entreprises clientes d'EDF et d'informations collectées auprès de ces clients** (équipements électriques, système de chauffage, présence de climatisation...).

Lors de la première partie du stage, nous nous concentrerons sur les courbes de charge PRO comme source de données pour des tâches telles que la détection d'usages, le regroupement (clustering) des profils de charge ou encore la classification de comportements. L'objectif est **d'explorer des modèles basés sur les distances élastiques et leurs dérivées, notamment TimePoint, ainsi que des modèles de fondation comme TabPFN-TS, TSPulse et VQShape**. Cette première phase vise à évaluer la capacité de ces modèles à extraire l'information pertinente à partir des seules courbes de charge.

Dans un second temps, avec l'émergence de modèles de fondation adaptés aux données tabulaires, nous envisageons de fusionner les courbes de charge PRO avec les informations connues sur les clients afin de constituer un jeu de données tabulaire. L'objectif est de comparer les performances obtenues avec les seules courbes de charge à celles obtenues avec des données enrichies, et ainsi tester ce que l'on peut tirer d'une représentation agrégée des courbes dans un format tabulaire.

Ainsi, **nous étudierons des modèles de fondation adaptés aux données tabulaires, tels que TabPFNv2 et TabICL, ainsi que d'autres modèles d'apprentissage automatique comme RealMLP, TabM et XGBoost**, afin d'exploiter pleinement le potentiel de ces données enrichies.

Une phase dédiée sera consacrée à l'optimisation des performances des modèles étudiés, en recourant à des techniques telles que la recherche d'hyperparamètres, l'ensembling, et la combinaison de modèles (mélange de modèles) pour améliorer la robustesse des résultats.

Tout au long du stage, une analyse approfondie des résultats sera menée afin de comprendre les enjeux, identifier les leviers d'amélioration, et affiner les modèles explorés. Ainsi, la dernière phase du stage sera centrée sur la synthèse des résultats finaux, ainsi que sur la mise en forme des méthodes pertinentes identifiées.

Profil recherché :

Ce stage de fin d'études s'adresse aux étudiants en école d'ingénieurs ou en Master 2, spécialisés **en mathématiques appliquées, statistiques, data science, Deep Learning ou intelligence artificielle**.

Compétences attendues :

- Maîtrise du langage Python, en particulier du framework PyTorch pour le deep learning
- Connaissances de base en classification et en clustering
- Connaissances fondamentales en Deep Learning
- Compétences en analyse de données
- Aptitude à interpréter les données et à évaluer les résultats issus des méthodes appliquées

Compétences à développer ou approfondir durant le stage :

- Modèles de fondation : Transformers, MLP, in-context learning
- Deep Learning : architectures avancées
- Machine Learning : arbres de décision, mesures de distance élastique
- Analyse de séries temporelles et de données tabulaires : décomposition, clustering, classification

Une bibliographie non exhaustive des articles pertinents pour les travaux du stage est jointe à l'offre.

Compétence supplémentaire appréciée :

- Maîtrise des bonnes pratiques avec Git, GitLab, uv/poetry et SQL
- Expérience avec des outils de gestion d'expériences tels que Hydra et MLflow
- Une expérience concrète dans l'utilisation de modèles de deep learning constitue un atout majeur
- Une appétence pour la recherche appliquée dans le domaine de l'énergie sera particulièrement appréciée

Informations complémentaires :

Dates : Stage d'une durée de 6 mois. *La date de début est flexible entre février et mai 2025.*

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

Horaires : 35 h / semaine

Indemnité : en fonction des formations

Contacts :

- Philippe CHARPENTIER (Chercheur expert, EDF), mail : philippe.charpentier@edf.fr
- Diala HAWAT (Ingénierie chercheure, EDF), mail : diala.hawat@edf.fr
- Maël DESMARTIN (Ingénieur chercheur, EDF), mail : mael.desmartin@edf.fr
- Dominique VOIRIN (Chef de projet, EDF), mail : dominique.voirin@edf.fr

Pour candidater merci d'envoyer un C.V. et une lettre de motivation sur ces adresses mails

Bibliographie :

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

- Shi Bin Hoo, Samuel Müller, David Salinas, Frank Hutter. From Tables to Time: How TabPFN-v2 Outperforms Specialized Time Series Forecasting Models, (2025).
- Han-Jia Ye, Si-Yang Liu, Wei-Lun Chao. A Closer Look at TabPFN v2: Understanding Its Strengths and Extending Its Capabilities, (2025).
- Noah Hollmann, Samuel Müller, Lennart Purucker, Arjun Krishnakumar, Max Körfer, Shi Bin, Hoo, Robin Tibor Schirrmeister, and Frank Hutter. Accurate predictions on small data with a tabular foundation model. *Nature*, 637(8045):319–326, (2025).
- Nick Erickson, Lennart Purucker, Andrej Tschalzev, David Holzmüller, Prateek Mutalik Desai, David Salinas, Frank Hutter. TabArena: A Living Benchmark for Machine Learning on Tabular Data, (2025).
- Vijay Ekambaram, Subodh Kumar, Arindam Jati, Sumanta Mukherjee, Tomoya Sakai, Pankaj Dayama, Wesley M. Gifford, Jayant Kalagnanam. TSPulse: Dual Space Tiny Pre-Trained Models for Rapid Time-Series Analysis, (2025).
- Yunshi Wen, Tengfei Ma, Tsui-WeiWeng, Lam M. Nguyen, Anak Agung Julius. Abstracted Shapes as Tokens - A Generalizable and Interpretable Model for Time-series Classification, (2025).
- Si-An Chen, Chun-Liang Li, Nate Yoder, Sercan O. Arik, Tomas Pfister. TSMixer: An All-MLP Architecture for Time Series Forecasting, (2023).
- Petitjean F., Tan Chan Wei, and Webb G. Fastee. Fast ensembles of elastic distances for time series classification. *Data Mining and Knowledge Discovery*, 34 : 231–272 (2019).
- Weber, R. S., Ben Ishay, S., Lavrinenko, A., Finder, S. E., & Freifeld, O. (2025). TimePoint: Accelerated time series alignment via self-supervised keypoint and descriptor learning. In Proceedings of the 42nd International Conference on Machine Learning (ICML 2025).