

# Sujet de Thèse

## Prédiction et anticipation de situations de crise dans les réseaux de distribution ruraux

Ce sujet de thèse s'inscrit dans le cadre d'un projet ADEME (Agence De l'Environnement et de la Maîtrise de l'Energie) avec ENEDIS portant sur le développement d'un démonstrateur en milieu rural de l'observabilité et de la flexibilité des réseaux de distribution. Ce démonstrateur Smart Occitania permettra de planifier les aménagements futurs en infrastructure énergétique avec l'évolution des réseaux de distribution d'énergie au service d'un territoire rural dans une approche intégrée.

### Objectifs de la thèse

A partir des données des capteurs ILD (Indicateurs Lumineux de Défaut) et capteurs météo et autres données environnementales, il est nécessaire de procéder à une :

- Mise en place d'un modèle de prédiction pour les pannes
- Mise en place d'un protocole de récupération des données pour affiner le modèle dans le temps

L'efficacité du modèle sera mesurée sur sa capacité à bien prédire les probabilités de panne sur une section. On pourra le tester à partir des données recueillies sur les pannes déjà connues. Une phase d'observation peut aussi être envisagée afin de tester le modèle sur des cas réels

### Description du sujet :

Prédire la probabilité qu'un secteur tombe en panne est un enjeu crucial à plusieurs niveaux. Dans un premier temps, cela permettrait de planifier de manière efficace l'entretien des lignes électriques en ciblant en priorité les lignes les plus susceptibles d'être en panne. Ensuite, en cas de panne, le modèle pourrait aider à identifier le type de panne le plus probable et son emplacement sur la ligne.

Le but de cette thèse est d'utiliser des méthodes d'apprentissage automatique afin de créer ces modèles prédictifs. Le principe étant d'utiliser les informations sur les précédentes pannes afin de pouvoir généraliser et construire un modèle. Dans le cas présent, différents verrous devront être levés.

Le premier verrou provient des données elles mêmes. Les facteurs pouvant influencer sur une panne étant multiple, il faut pouvoir croiser plusieurs sources de données pour obtenir un modèle fiable (informations sur la panne et son emplacement, données météorologiques, données géographiques, charge du système, état du système). Or, toutes ces données n'ont pas forcément été enregistrées lors des pannes. Il faudra donc, dans un premier temps essayer d'apprendre des modèles à partir des données disponibles (et donc incomplètes). Dans un second temps, il faudra mettre en place un protocole de description des pannes (quels types de données, quelles sources, etc ..) afin d'obtenir une base de données suffisamment complète pour notre tâche.

Le second verrou est d'ordre scientifique. En effet, il ne s'agit pas d'apprendre un modèle unique mais plusieurs modèles, dont on devra vérifier la cohérence, pour les différents aspects de la tâche. Par exemple, pour la prédiction de la probabilité d'une panne sur un secteur, le but est d'apprendre les paramètres d'une loi de probabilité (de type exponentielle par exemple) en fonction des variables d'entrée. On parle alors d'apprentissage de densité. La prédiction du type de panne est un problème de classification. Enfin, on pourra construire de manière automatique des catégories de secteurs qui seront ensuite prises en compte dans d'autres modèles. On parle alors d'apprentissage non supervisé. L'autre problème provient du fait que les données seront de types très hétérogènes (texte, image, etc ...). Il s'agit d'un problème de modélisation de données pour les algorithmes d'apprentissage qui nécessitent des données homogènes et concises. Des approches multimodales d'apprentissage non supervisée peuvent être envisagées pour ensuite modéliser les données de natures différentes via une mesure de similarité.

**Encadrants :**

Mathieu Serrurier : [Mathieu.Serrurier@irit.fr](mailto:Mathieu.Serrurier@irit.fr)

Sandrine Mouysset : [sandrine.mouysset@irit.fr](mailto:sandrine.mouysset@irit.fr)