

Date de la demande : 24/01/2023

Nature du contrat (Stage ou Alternance) : Stage

Lieu d'exécution de la mission : Engie Digital Paris, Paris La Défense

## Le STAGE

**Intitulé précis du poste :**

Stage en data science et analyse de données appliqué aux énergies renouvelables : analyses des sous-performances de parcs éoliens et solaires

**Equipe / Plateforme :**

(présentation succincte de l'équipe / de la plateforme et de votre mission chez ED)

DARWIN est la plateforme numérique dédiée aux énergies renouvelables, totalement hébergée dans le Cloud et déployée dans une vingtaine de pays.

DARWIN est connectée à l'ensemble des centrales de production d'énergies renouvelables du Groupe à travers le monde ; ainsi, les données opérationnelles de ces centrales sont collectées massivement et sont utilisées pour améliorer les performances des parcs éoliens et solaires grâce à des algorithmes data science et des produits logiciels. Avec cette plateforme, ENGIE souhaite faciliter la prise de décision des exploitants, permettre une maintenance prédictive et améliorer la rentabilité des installations.

**Durée :**

6 mois

**Date de début souhaitée :**

A partir de début 2023

## CONTEXTE

**Mission :**

(Phrase clé qui définit la mission)

Le/La stagiaire s'intégrera à l'équipe Data Science de la plateforme DARWIN, qui depuis plusieurs années analyse les données des actifs de production d'énergies renouvelables, conçoit des algorithmes, et développe des solutions logicielles pour déployer à l'échelle ces algorithmes en production.

### Activités / Missions détaillées :

L'équipe data science de Darwin a développé et déployé un ensemble de modèles de machine learning visant à détecter des sous-performances / sous-productions sur les parcs éoliens et solaires du groupe Engie.

Objet du stage :

- Amélioration des méthodes de calcul de météo de référence en place (vent de référence, irradiance de référence), en utilisant différentes sources de données (données mesurées sur les parcs éoliens et solaires, données météo externes...); cette étape nécessitera une bonne analyse des données et de leur qualité ;
- Amélioration de la méthode d'identification des périodes d'indisponibilité des équipements (éoliennes, onduleurs PV) ;
- Utilisation de cette météo de référence et de cette méthode de détection des indisponibilités pour développer de nouveaux modèles de machine learning permettant de détecter des sous-performances (sous-productions) sur ces parcs.

En fonction de l'avancement, d'autres applications / utilisations des étapes précédentes pourront être discutées.

Les résultats attendus sont :

- une étude bibliographique ;
- un rapport incluant méthodologies et résultats, et / ou une présentation complète faisant office de rapport ;
- le code développé pour répondre à la problématique.

### Données

Au sein de la plateforme DARWIN, ENGIE dispose de données très détaillées telles que :

- des données structurées de type séries temporelles au pas de temps 10 minutes enregistrées au niveau des parc éoliens et solaires ;
- des données statiques décrivant un ensemble de caractéristiques des parcs ;
- des données météorologiques externes de différentes natures (plusieurs sources disponibles).

## PROFIL Souhaité

### Savoirs :

(Formation requise, Ecole ciblée, Outils, Langues)

- français
- anglais

### Savoir-faire :

- analyses statistiques et visualisations
- modélisation statistique / data science / machine learning / séries temporelles
- langages R et Python

# ENGIE

digital

- écosystème Microsoft Azure
- développement logiciel
- GitHub

**Savoir-être :**  
(comportements attendus)

- fort intérêt pour les énergies renouvelables
- autonomie
- créativité / force de proposition
- bonnes capacités rédactionnelles