

**Projet sur les énergies renouvelables**

*Cahier des charges / fiche projet*



**Formation FITEC « Analyste Big Data 10/2019 »**

Equipe / participants :

Hamza ISMAILI [ismaili.hamza00@gmail.com](mailto:ismaili.hamza00@gmail.com)

Sabrina NEMEUR [cherdiouisabrina@live.fr](mailto:cherdiouisabrina@live.fr)

Flora DEMGNE [floradems@yahoo.fr](mailto:floradems@yahoo.fr)

Olivier BRUNET [o.brunet@outlook.com](mailto:o.brunet@outlook.com)

**Contexte**

La transformation de l’énergie est un enjeu stratégique et toujours d’actualité : Comment passer d’une énergie fossile à une énergie dite renouvelable, d’une production centralisée à un système décentralisé afin de répondre à la nécessité d’un développement durable et responsable d’un point de vue écologique ?

Les énergies renouvelables proviennent de phénomènes naturels et sont des [sources d'énergie](https://fr.wikipedia.org/wiki/Source_d%27%C3%A9nergie) dont le renouvellement naturel est assez rapide pour qu'elles puissent être considérées comme inépuisables à l'échelle du temps humain.

Au cours de ce projet, nous allons nous intéresser plus particulièrement à l’éolien et au solaire : dans quelle mesure le big data peut-il permettre de donner un éclaircissement sur cette problématique énergétique ?

**Données exploitées :**

* Pour l’énergie solaire :
  + Intitulé : *«*[*30 Years of European Solar Generation*](https://www.kaggle.com/sohier/30-years-of-european-solar-generation)*»*
  + Description sommaire : estimations heure par heure de différents sites à travers l’Europe de 1985 à 2015
  + Source Kaggle :

<https://www.kaggle.com/sohier/30-years-of-european-solar-generation>

* + Licence : Creative Commons [CC0 : Public Domain](https://creativecommons.org/publicdomain/zero/1.0/)
* Pour l’énergie éolienne :
  + Intitulé : *«*[*30 Years of European Wind Generation*](https://www.kaggle.com/sohier/30-years-of-european-wind-generation)*»*
  + Description sommaire : estimations heure par heure de différents sites à travers l’Europe de 1985 à 2015
  + Source Kaggle :

<https://www.kaggle.com/sohier/30-years-of-european-wind-generation>

* + Licence : Creative Commons [CC0 : Public Domain](https://creativecommons.org/publicdomain/zero/1.0/)
* Climat en Europe :
  + Source : *« the European Climate Assessment & Dataset project »*
  + Lien : <https://www.ecad.eu//dailydata/index.php>
  + Description sommaire : les informations contenues dans les différents jeux de données présentent les changements de climat et météorologiques des mesures extrêmes ainsi que des informations journalières pour analyser et monitorer ces extrêmes.
  + Licence : *« strictly for use in non-commercial research and education projects only »*
* Informations démographiques et données sur les matériaux constitutifs des éoliennes et panneaux solaire (**en cours de recherche**)

**Objectifs / Use cases :**

La finalité de note P.O.C est multiple :

* Analyser les zones géographiques en Europe les plus adaptées à la production d'énergie solaire et éolienne : croisement entre données météo et celles issues des capteurs des stations ainsi que des informations démographiques.
* Quelle périodicité retrouve-t-on d’une année sur l’autre dans les performances des stations éoliennes et solaires (jour / nuit, saisons ?)
* Peut-on mettre en évidence un dérèglement climatique sur la base des données de ces 30 dernières années (par exemple : les moyennes thermales peuvent ne pas évoluer d’une année sur l’autre, mais sur de courtes périodes, est-ce que des variations de plus en plus importantes peuvent être relevées ? en terme d’ensoleillement, pluviométrie, pression, vitesse de vent, températures etc…)
* Peut-on prédire dans le temps la performance énergétique des installations existantes ou futures. Quelle est l’indice de confiance ou la précision de ces prédictions ? Ces paramètres peuvent ils être suffisamment fiables si l’on devait envisager la création de nouvelles éoliennes ou de nouveaux panneaux solaires ?
* Les métaux utilisés pour la construction des panneaux solaires et des éoliennes sont-ils réellement écologiques ? (durée de vie des installations et recyclage des matériaux versus apports d’énergie)

**Déroulé du projet / grandes phases :**

* Recherche des données (cf. sources) : à ce jour manquent les informations démographiques et sur les matériaux de construction.
* Préparation des données :
  + Nettoyage (valeurs aberrantes, manquantes)
  + Corrélation entre features
  + Analyse exploratoire
* Stockage de la data
* Modèle prédictif de machine learning avec Spark & analyse approfondie
* Présentation des résultats :
  + comparaison des prédictions aux valeurs réelles
  + Graphiques de visualisations des différents objectifs
* Perspectives (optionnel) :
  + ingestion d’un flux de données sur la base de data plus récentes pour les futures prédictions
  + traitement de ce flux

**Planning prévisionnel**

Objectif : remise des livrables : 15/12/2019

* Du 04/11/2019 au 10/11/2019 : recherches des données manquantes et première exploration, paramétrage de l’infra -> 1 semaine.
* Du 11/11/2019 au 17/11/2019 : préparation des données, stockage de la data

et fin de paramétrage de l’infra-> 1 semaine

* Du 18/11/2019 au 01/12/2019 : analyse et modèle prédictif de machine learning -> 2 semaines
* Du 02/12/2019 au 08/12/2019 : analyse critique des résultats -> 1 semaine
* Du 09/12/2019 au 15/12/2019 : soit partie streaming optionnelle soit amélioration des étapes précédentes -> 1 semaine.