Projet meteo

Travaux réalisé : Durant cette première période de projet, nous avons réalisé les tâches suivantes :

* Création d’un fichier « train » regroupant dans un même fichier tous les fichiers train disponible
* Suppression des individus ayant un nombre de valeurs nulle > 35%
* Utilisation d’une régression linéaire de la manière suivante :
  + 1 ère tentative :
* Utilisation du fichier « brute » (sans suppression des blocs de valeurs nulles)
* Les villes et l’échéance n’ont pas été traitées dans le modèle
* Score obtenu : 1.53
* 2 ème tentative :
  + Utilisation du fichier nettoyé
  + Utilisation des villes et de l’échéance en facteurs dans le modèle
  + Score obtenu : 1.41

Séance du 27 octobre 2017 :

Durant cette séance, nous avons réalisé les tâches suivantes :

* Création d’un modèle linéaire par ville
* Création de variable décrivant en fonction de l’échéance, les écarts entre TH2 et TH2 obs
* Imputation des valeurs manquantes (méthode « PMM » dans un premier temps), création d’une variable binaire par ville
* Implémentation de la méthode « Random Forest » sur les données brutes, pas normalisées et imputation des valeurs manquantes par la médiane
* Remise en forme du code + normalisation + modele par ville (FLO)
* Amélioration du fichier train + courbe modélisant les écarts de pred en fonction de l’échéance