



UNIVERSITÉ RENNES 2

PROJET DATA SCIENCE - MASTER 2 MAS PARCOURS SCIENCES DES
DONNEES

RAPPORT DE L'ÉTUDE

Prédictions des températures de 7 stations différentes à un horizon de 36 heures

Défi Grosses Data 2018



BENJAMIN ALLEAU
ABDESSAMAD AZNAGUE
GUILLAUME LE FLOCH

Année 2017-2018

Organisateur : INSA DE TOULOUSE
*Professeur encadrant : M. ROMAIN
TAVENARD (UNIVERSITÉ DE RENNES 2)*

05 Octobre 2017 — 08 Janvier 2018

Plan

| | | |
|----------|---|----------|
| 1 | Présentation du projet et problématique | 2 |
| 2 | Remerciements | 2 |
| 3 | Les différentes étapes du projet | 2 |
| 3.1 | Les différentes stratégies envisagées | 2 |
| 3.2 | La gestion et le remplacement des valeurs manquantes | 2 |
| 3.3 | Le choix et l'application d'algorithmes d'apprentissage supervisé | 2 |
| 3.3.1 | Premiers algorithmes : la régression linéaire et ses variantes | 2 |
| 3.3.2 | La forêt aléatoire | 2 |
| 3.3.3 | Gradient Boosting : l'algorithme XGBoost | 2 |
| 3.3.4 | Deep Learning : l'utilisation de réseaux de neurones | 2 |
| 4 | Bilan | 2 |
| 5 | Bibliographie | 2 |

- 1 Présentation du projet et problématique**
- 2 Remerciements**
- 3 Les différentes étapes du projet**
 - 3.1 Les différentes stratégies envisagées**
 - 3.2 La gestion et le remplacement des valeurs manquantes**
 - 3.3 Le choix et l'application d'algorithmes d'apprentissage supervisé**
 - 3.3.1 Premiers algorithmes : la régression linéaire et ses variantes**
 - 3.3.2 La forêt aléatoire**
 - 3.3.3 Gradient Boosting : l'algorithme XGBoost**
 - 3.3.4 Deep Learning : l'utilisation de réseaux de neurones**
- 4 Bilan**
- 5 Bibliographie**