

Python - Data Analysis

...

Kilian Weydert

Data

15 individus

Subject ID	Gender	Age [years]	Height [cm]	Weight [kg]	Skin Type	Fitness
S1	m	34	182	78	3	6
S2	m	28	189	80	3	5
S3	m	25	170	60	3	5
S4	m	25	168	57	4	5
S5	f	21	180	70	3	4
S6	f	37	176	70	3	1
S7	f	21	168	58	3	2
S8	m	43	179	70	3	5
S9	f	28	167	60	4	5
S10	f	55	164	56	4	5
S11	f	24	168	62	3	5
S12	m	43	195	105	3	5
S13	f	21	170	63	3	6
S14	f	26	170	67	3	4
S15	m	28	183	79	2	5
		30.60± 9.59	175.27± 8.78	69.00± 12.35		

8 activités

Activity	Duration [min]
Sitting still	10
Ascending/Descending stairs	5
Table Soccer	5
Cycling	8
Driving car	15
Lunch break	30
Walking	10
Working	20

Data

1 fichier pickle par individu avec les données synchronisées

Fréquences cardiaque captées par :

- RespiBAN Professional
- Empatica

Objectif

Prédire l'activité de la personne en fonction de ses données cardiaques

Étape 1 : Import et preprocessing

Récupération du fichier avec wget

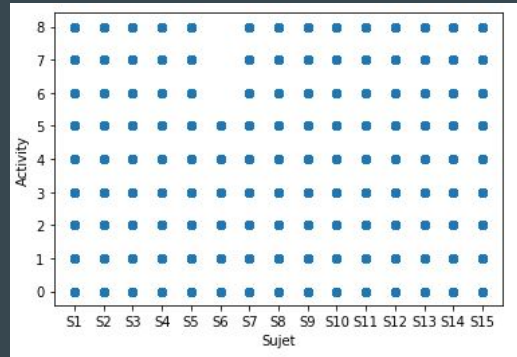
Unzip

Exploration data sujet 1

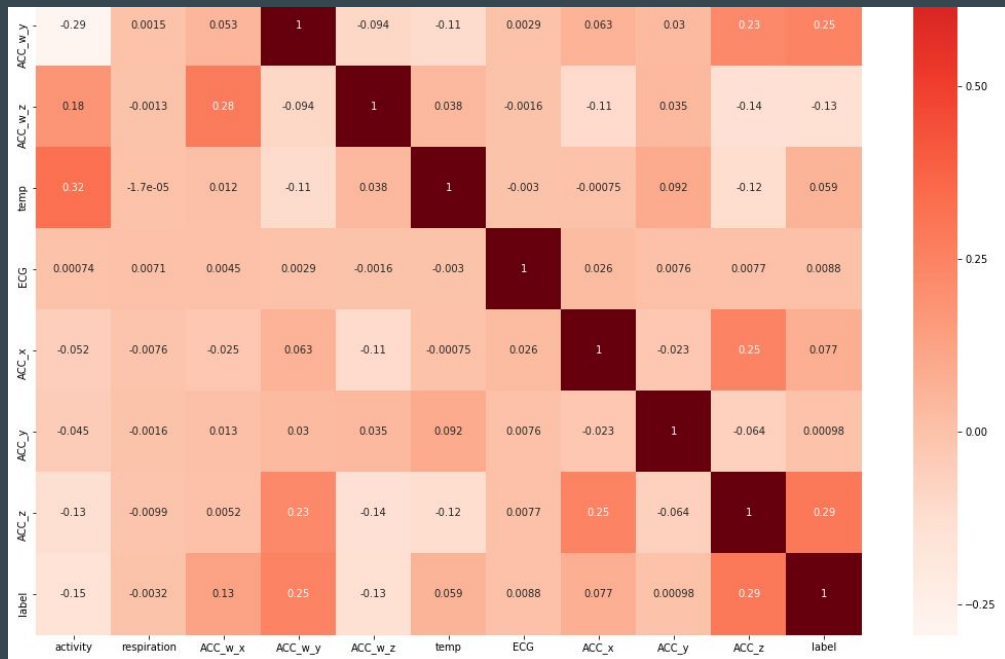
Réunir toutes les données dans un dataframe final

Étape 2 : Visualisation

On remarque le manque de données pour les activités 6, 7 et 8 du patient 6 (comme annoncé dans le readme du dataset).



Étape 2 : Visualisation



Étape 3 : Modélisation

Random Forest

Regression Logistic