

Dataset : PPG-Dalia
Henri de Montalembert

PROJET PYTHON FOR ANALYSIS

CONTEXT DE LA DATA SET

Objectif : Prédire la fréquence cardiaque à partir des données photopléthysmographie.

Difficultés rencontrées dans le passé :

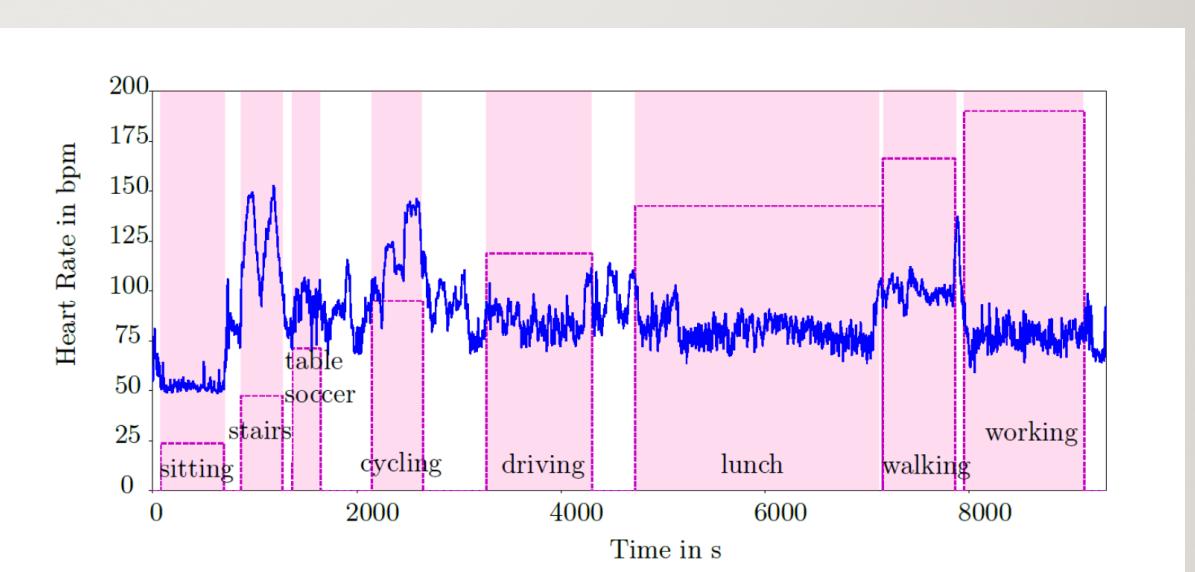
- Artefacts de mouvement
- Nombre très limité d'activité

Objectifs :

- Plus grands nombres d'activités
- Conditions plus naturelles
- Compensation des mouvements d'artefacts

SUJET DU PROJET

RÉALISER UNE CLASSIFICATION POUR
PRÉDIRE L'ACTIVITÉ D'UN INDIVIDU.



DIFFICULTÉS RENCONTRÉES

- Gros volume de données → Etude sur un seul individu pour commencer
- Format .plk → Utilisation des packages pickle et json
- Objets de tailles différentes → Moyenner les données pour arriver à la même taille que l'objet « Activity »

PREMIÈRE APPROCHE

Prédiction à partir de la fréquence cardiaque et des caractéristiques de chaque individu.

- Moins volumineux
- Plus facile à implémenter
- Moins précis

Exemple de l'input (X) :

	Gender_m	Gender_f	Age	Height	Skin	Sport	HearthRate
2114	1	0	34	182.0	3	6	79.468090
55538	0	1	26	170.0	3	4	91.686940
29506	1	0	43	179.0	3	5	68.813344
45199	1	0	43	195.0	3	5	43.863619
57819	0	1	26	170.0	3	4	96.404852

DEUXIÈME APPROCHE

Prédiction à partir des signaux provenant du poignet et de la poitrine ainsi que des caractéristiques de chaque individu.

- Plus volumineux
- Moins rapide
- Plus précis

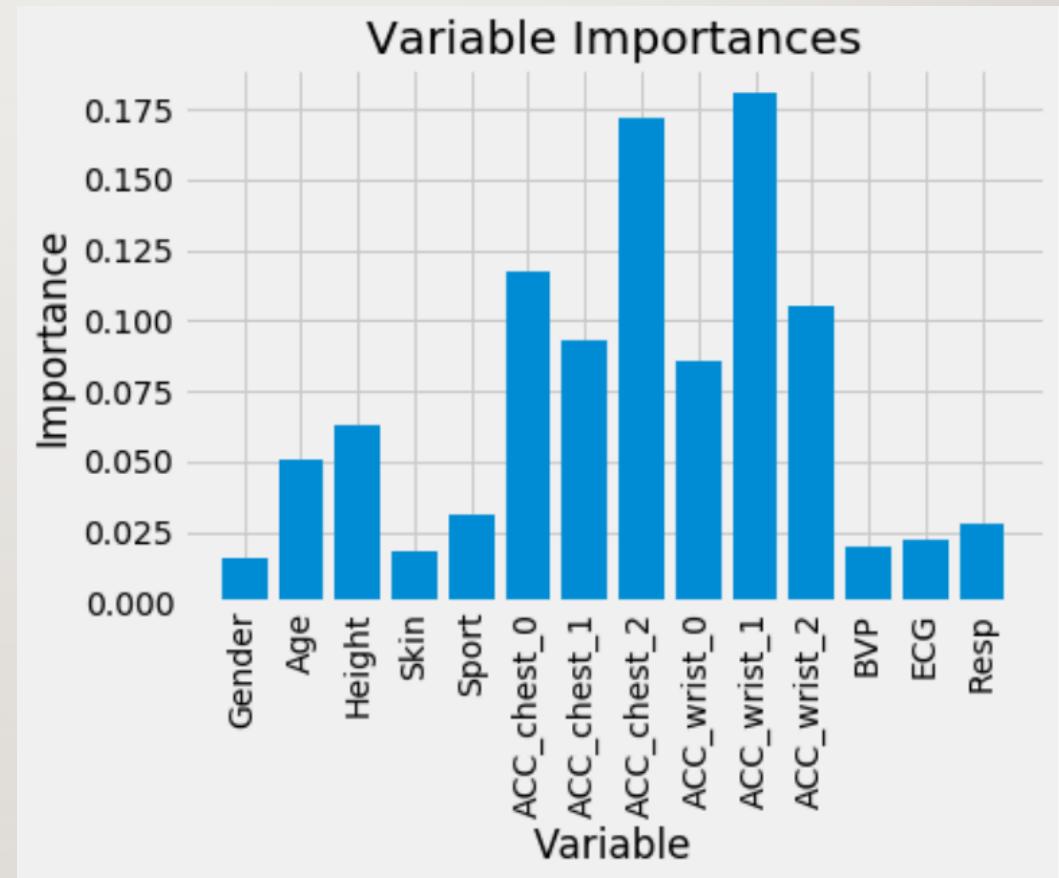
On gardera cette approche pour la suite du projet

Gender_m	Gender_f	Age	Height	Skin	Sport	ACC_chest_0	ACC_chest_1	ACC_chest_2
1	0	25	168.0	4	5	0.776443	-0.027226	0.460662
0	1	28	167.0	4	5	0.891751	-0.058672	-0.234487
0	1	21	170.0	3	6	0.749979	0.099183	0.034139
1	0	34	182.0	3	6	0.873191	-0.042329	-0.320115
1	0	25	168.0	4	5	0.862061	0.039984	-0.265171

ACC_wrist_0	ACC_wrist_1	ACC_wrist_2	BVP	ECG	Resp	Labels
-0.193359	-0.910156	0.218750	12.024375	0.085356	-8.468131	130.437595
-0.830078	-0.054688	0.611328	12.207500	0.033108	15.674072	111.305273
-0.277344	-0.906250	0.199219	24.407500	-0.037639	-2.773246	118.815090
-0.378906	-1.177734	0.173828	-349.126875	-0.027772	5.532881	124.936098
-0.382812	1.218750	0.183594	-1.549375	0.132616	-3.135908	93.981005

ANALYSE DES VARIABLES

- Les données ACC provenant du poignet et de la poitrine sont les plus importantes
- La taille et l'âge sont importants aussi pour la prédiction



MODÈLES UTILISÉS

- **Arbres de décision : 0,85 de précision**

Meilleurs paramètres (grid search) : `'criterion': 'entropy', 'max_features': None, 'min_samples_split': 10]`

- **Random forest : 0,92 de précision**

Meilleurs paramètres : (grid search) :

```
{'bootstrap': False,  
 'max_depth': 40,  
 'max_features': 'auto',  
 'min_samples_leaf': 1,  
 'min_samples_split': 2,  
 'n_estimators': 140}
```

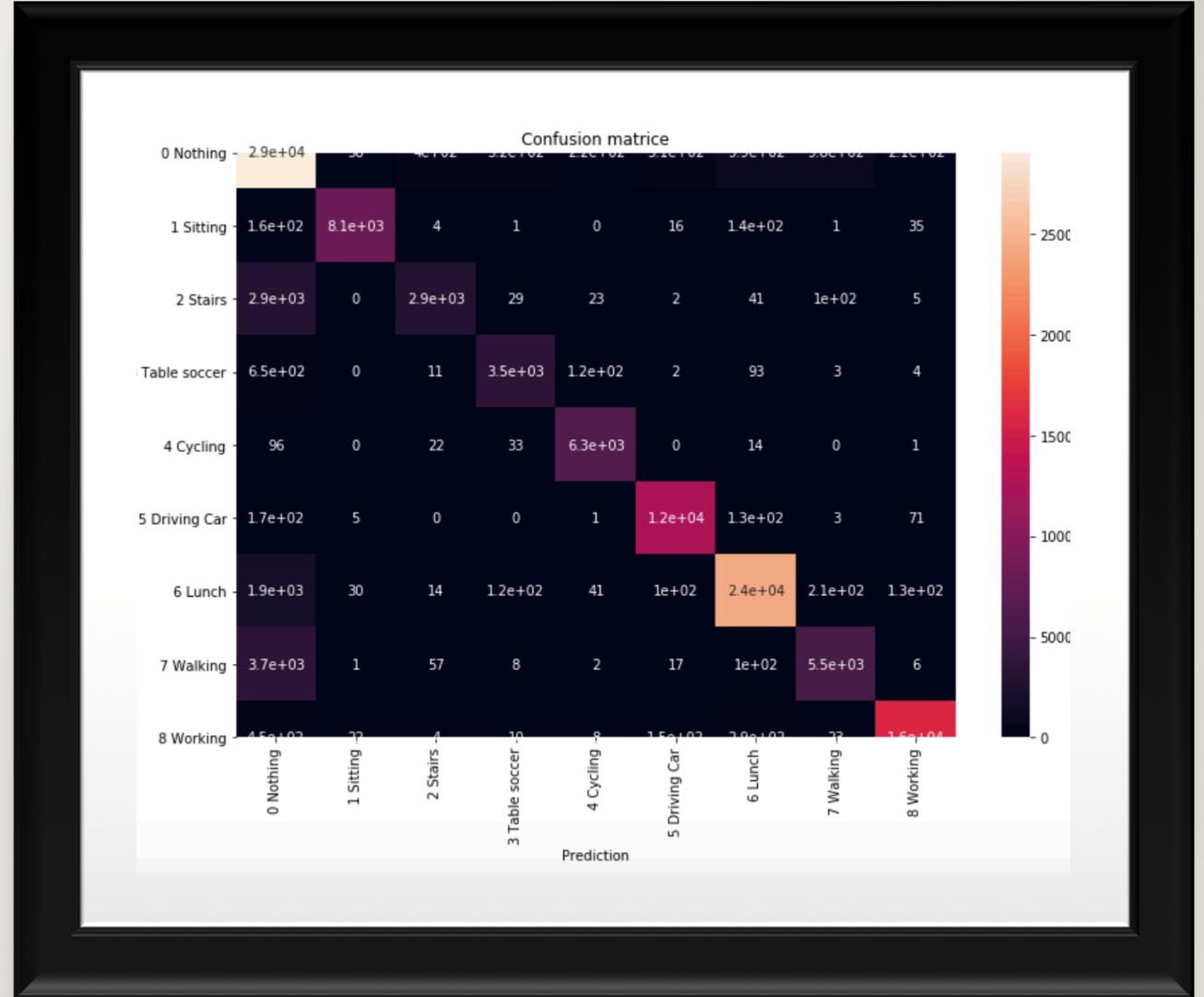


On retiendra ce modèle
pour la suite du projet

- **Gradient Boosting : 0,85 de précision**

MATRICE DE CONFUSION

Nous remarquons que le modèle se trompe généralement avec « Activity = 0 », c'est à dire quand l'individu ne fait pas d'activité particulière. En effet, la première colonne à des valeurs élevées.



CLASSIFICATION REPORT

		precision	recall	f1-score	support
0	Nothing	0.82	0.92	0.87	32519
1	Sitting	1.00	0.99	0.99	8472
2	Stairs	0.89	0.70	0.79	6075
3	Table soccer	0.93	0.88	0.90	4368
4	Cycling	0.98	0.99	0.98	6439
5	Driving Car	0.97	0.98	0.97	12775
6	Lunch	0.96	0.93	0.95	27054
7	Walking	0.85	0.73	0.78	9440
8	Working	0.98	0.96	0.97	17008
	micro avg	0.92	0.92	0.92	124150
	macro avg	0.93	0.90	0.91	124150
	weighted avg	0.92	0.92	0.92	124150