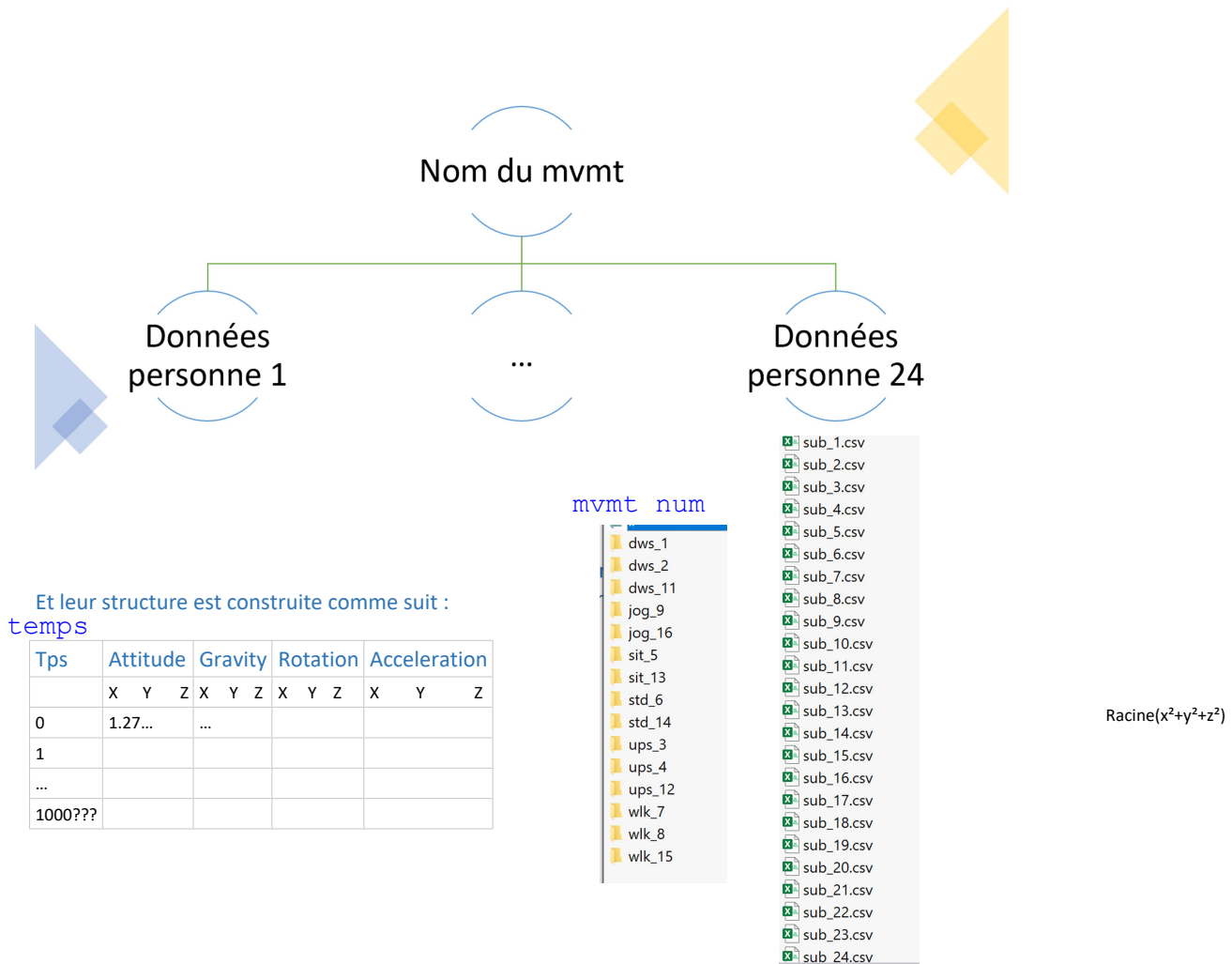


Phase1 : comprendre les données à disposition.

vendredi 12 mars 2021 08:08

STRUCTURE DES FICHIERS

360 fichiers (15*24) qui reprennent les données provenant du smartphone lorsqu'une des 24 personnes testées fait un mouvement. Ces fichiers sont répertoriés comme suit :



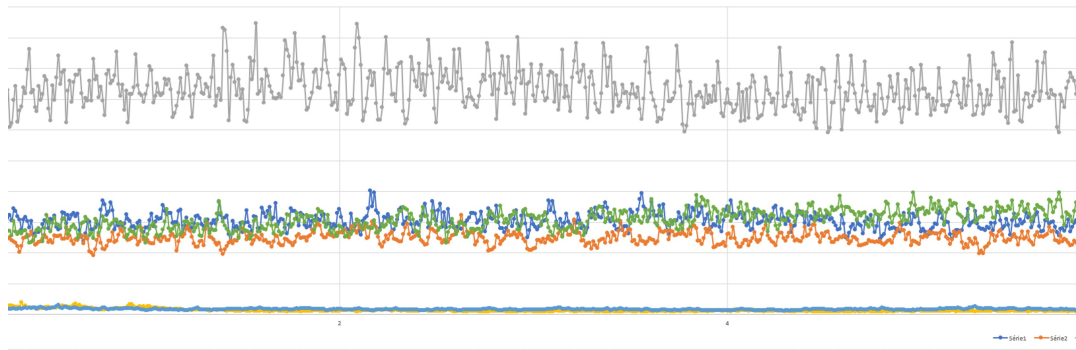
OBJECTIF PHASE 1 :

2 fichiers qui reprennent chacun une partie des données comme suit :

Ident	Var	Var	Var	...	Mouvement	Genre
Ident

trainSet.csv
et
testSet.csv

Un des fichiers va servir de training set pour trouver des patterns selon le mouvement.



Bleu foncé: downstairs
 Orange : upstairs
 Gris : jogging
 Jaune : sitting
 Bleu clair : standing up
 Vert : walking

L'autre fichier sera le test set pour tester les patterns et générer `realClasses` et `estimateClasses`.



Il est très important de scinder le fichier à partir duquel les patterns sont générés du fichier sur lequel ces mêmes patterns sont testés. En effet, il est interdit de tester un modèle sur les valeurs qui ont servi à le créer. De même, dans la phase de mise au point du modèle, on est amené à évaluer plusieurs fois le modèle. Ces tests ne peuvent pas non plus se faire tout le temps sur le même fichier test sinon le risque est d'avoir un modèle qui « colle » au train set et au test set mais pas à des données nouvelles. C'est pourquoi, dans la pratique, la cross validation est souvent utilisée. Cependant, pour faire simple dans ce projet, nous utiliserons uniquement train set et test set.