

J'ai corrigé les erreurs de distribution entre similaire et différents dans les scripts, je vous n'ai pas demandé de le programmer. Mais vous devrez être capable d'en parler et expliquer pourquoi on fait ça. Le reste même vrai avec tout en soi.

Le Jupyter notebook

Si votre data.csv suit la même structure que le miens, vous pouvez tout considérer comme une boîte noire.

La première résultats va sélectionner les meilleurs attributs (ceux qui sont le plus influents) dans la détermination de la similarité des processus.

Ensuite je considère seulement ceux qui apparaissent dans les 3 modèles.

Ensuite je vais estimer l'erreur rate et la sensibilité pour chaque classe pour chaque modèle → à part la régression (car il est forcément binaire (dans ce cas j'ai donc groupé tous les différents niveaux de similaire ensemble)).

Script R

Comme j'exécute les script R à travers Python → les erreurs en R ne vont pas affichées. Donc il faudra que vous faites bien attention que tous les packages R soient installés sur vos machines.

Normalement les seules choses que vous allez devoir modifier dans les scripts R sont les noms de vos annotations → les miens sont D, C, S et VS → je vous laisse faire les modifications en fonction de vos noms d'annotations.

Faite attention à l'adresse du data.csv dans les scripts R