Au début, mon algorithme ne me renvoyait aucune erreur, je me suis ainsi rendu compte de l'utilité de séparer les données. En effet, je prenais k = 1, et les données testées étaient en mémoire dans mon ordinateur. Ainsi, l'algorithme n'avait qu'à dire que le plus proche voisin d'un individu était lui-même, d'où l'absence d'erreur.
Ensuite, même avec les données séparées il me donnaient encore aucune erreur pour k = 1. C'était parce que je m'étais trompé dans le choix de mes données de vérification.
70% des données sont en mémoire dans mon ordinateur et 30% sert de vérification.

Puis, tout marchait relativement bien, j'ai donc essayé d'optimiser mon algorithme. J'ai réduit les données mais les résultats pour les $k \in [|1,10|]$ étaient tous moins bon que pour les données non réduites. J'ai donc abandonné cet idée et mis la réduction en commentaire.
Pour trouver la meilleur de k, j'ai exécuter mon algorithme pour $k \in [|1,20|]$ et le meilleur k est 1.
J'ai fais de même avec le nouveau set de données contenu dans « preTest.csv », donnant les même conclusions.

J'ai ainsi environs 12% d'erreur avec mon algorithme knn, ce qui est a mon sens plutôt satisfaisant.