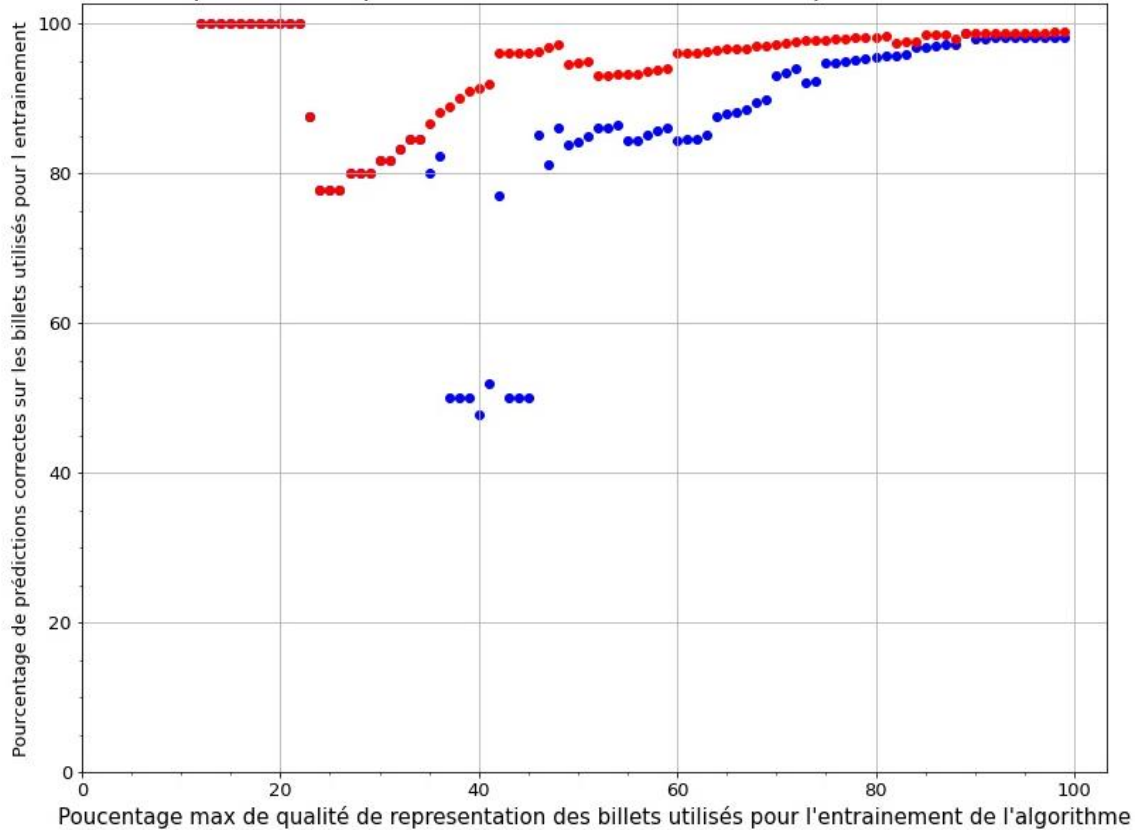


Annexe au projet 6 1/2

Pourcentage de prédictions correctes en fonction du pourcentage max de la qualité de représentation des billets utilisés pour l'entrainement



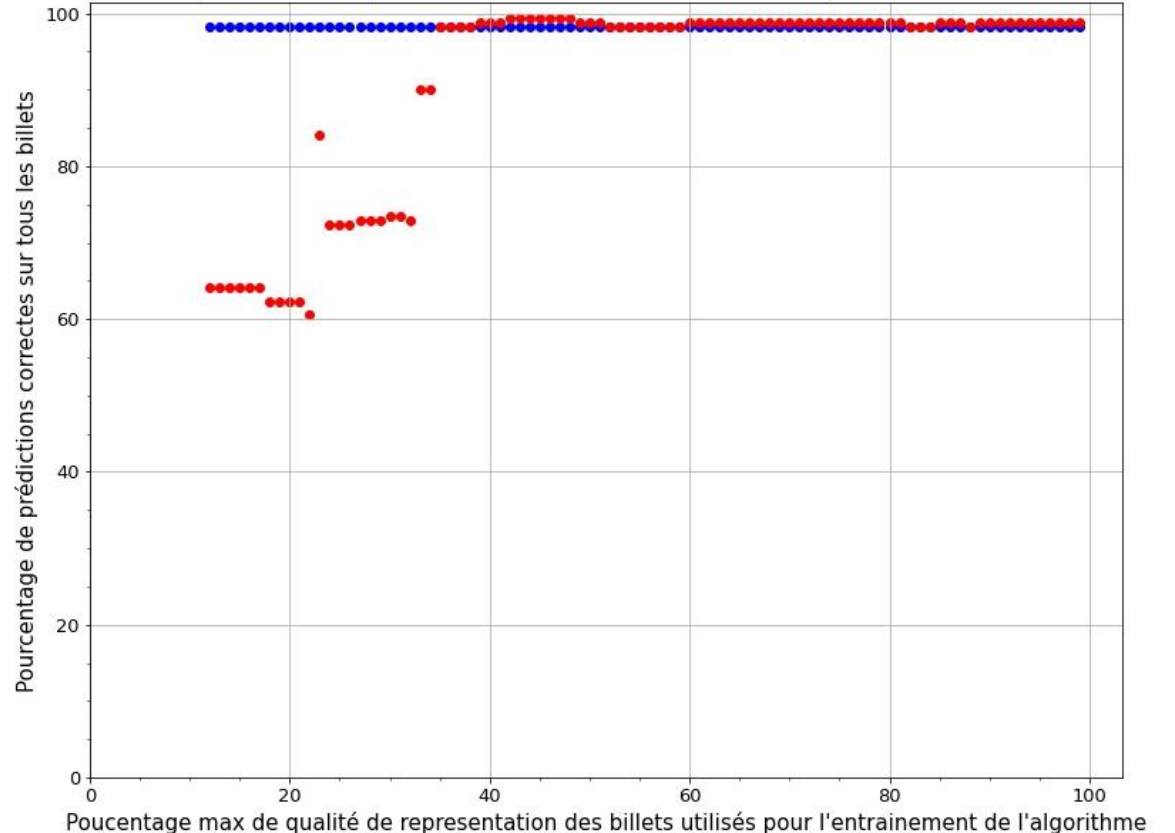
Nous remarquons que plus les données d'entraînement contiennent des billets bien représenté moins les algorithmes font d'erreurs sur les valeurs d'entraînement.

Ce qui est également valable pour la régression logistique lors d'un test sur tous les billets, contrairement à Kmeans qui dans ce deuxième cas est efficace dès le début.

Un point est dit bien représenté sur un axe ou un plan factoriel si il est proche de sa projection sur l'axe ou le plan. S'il est éloigné, on dit qu'il est mal représenté.

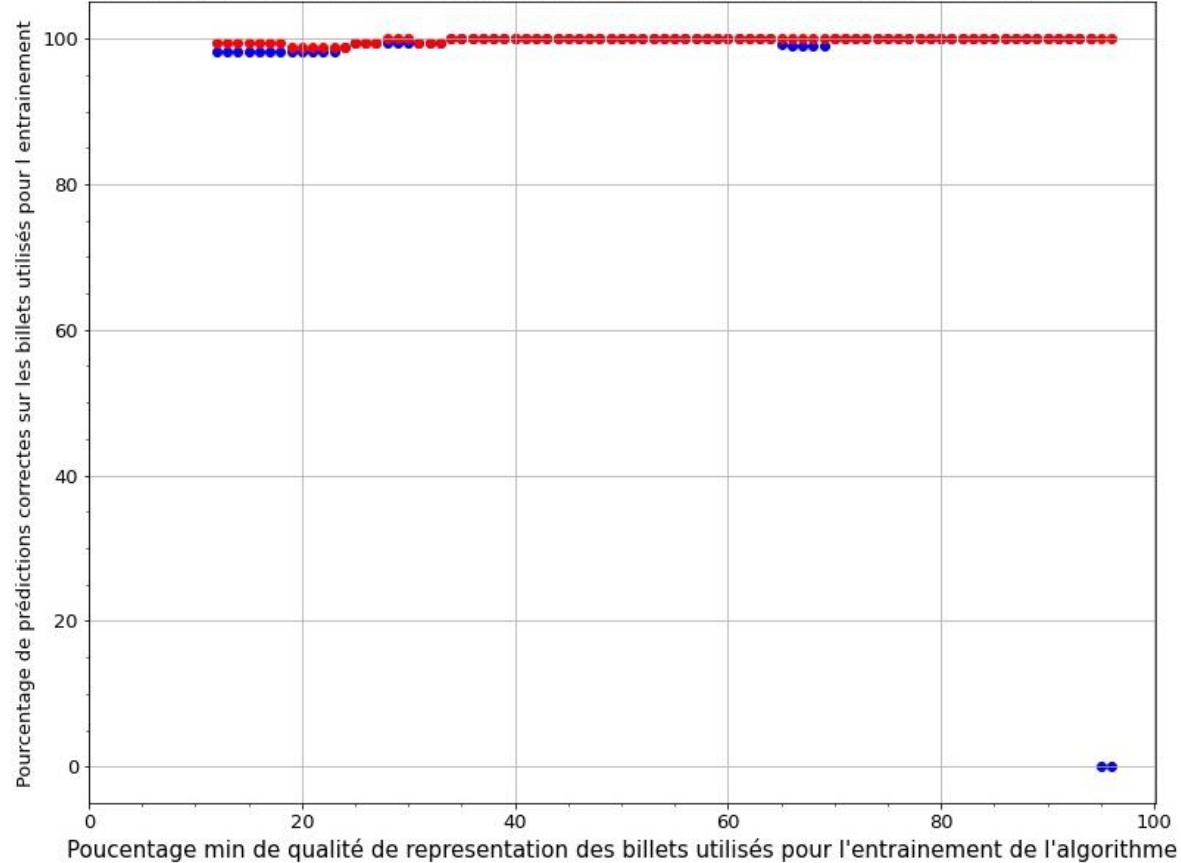
Vérifions quelle est l'influence de la qualité de représentation des billets utilisés avec deux algorithmes aux caractéristiques différentes, Kmeans en bleu et la régression logistique en rouge sur nos 2 graphiques.

Pourcentage de prédictions correctes en fonction du pourcentage max de la qualité de représentation des billets utilisés pour l'entrainement



Annexe au projet 6 2/2

Pourcentage de prédictions correctes en fonction du pourcentage min de la qualité de representation des billets utilisés pour l'entrainement



Les billets qui ont la meilleure qualité de représentation sont les plus proches de leur projection(et correspondent plus a la réalité), ils contribuent pour beaucoup à l'efficacité des algorithmes utilisés.

Ces deux graphiques dont chaque pourcentage de prédiction est calculé à partir des billets dont les qualités de représentation sont les plus importantes, nous démontrent, que nos deux algorithmes (Kmeans en bleu et la régression logistique en rouge), sont plus efficaces lorsqu'ils utilisent les billets dont la qualité de représentation sont les plus importants.

Pourcentage de prédictions correctes en fonction du pourcentage min de la qualité de representation des billets utilisés pour l'entrainement

