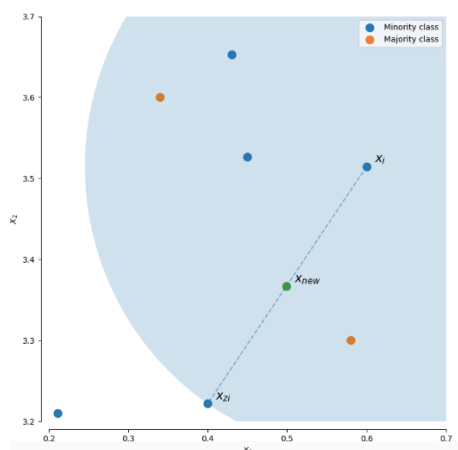


## Détection de fraudes bancaires

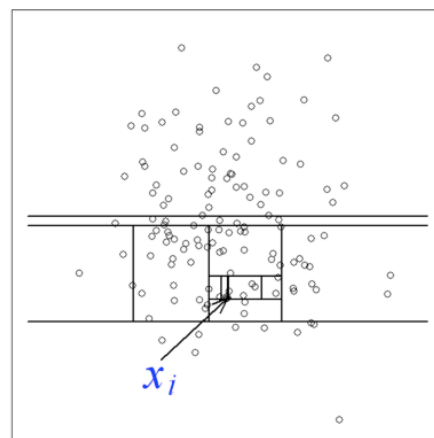


La difficulté d'un projet de fraude bancaire (problème de classification) vient du fait qu'il est très balancé (en moyenne il y a même pas 1% de fraude). Cela engendre, dans un premier temps, une problématique de métrique. En effet, en ne détectant aucune fraude, nous garderions une moyenne de 99%. Nous souhaitons dans ce cas de figure détecter les faux négatifs, c'est l'information que nous donne l'aire sous la courbe AUC (recall, precision).

Dans un second temps, avant de traiter la problématique liée à des projets très balancés comme celui-ci, nous faisons ce qu'on appelle un SMOTE qui permet de créer artificiellement des cas de fraude afin que l'algorithme puisse s'entraîner sur une population plus grande. Une autre solution est de créer un algorithme sur des transactions bancaires saines afin qu'il reconnaisse ensuite les fraudes comme des outliers (c'est le cas d'un isolation forest par exemple).



méthode SMOTE



Isolation forest