لما تيجي تستخدم train\_test\_split و تقسم ال داتا ل train و test و و تعمل feed ل model و تحسب ال score بتاعك

ف كل مرة هـ تعمل Run ل train\_test\_split هتلاقي ال accuracy بتتغير معاك slightly diff. عشان كل مرة بتختار عدد من ال records تقسم بيهم ال train , test

في حالة ال cross validation هي بتعمل كدا automatic بعدد ال folds اللي بتحدده و بعدين بتحسب ال score لكل split و بتديك on average estimate لل score

You cant rely on the score is given by train\_test\_split cuz the score is changing based on samples generated from train\_test\_split method for this u use CrossValidation (folds & average scores) this is good approach cuz u are going across all the samples

GridSearchCV for hyper parameter training

Issue that happens with gridsearchcv is the computation cost if you have millions of data points and for parameter you have so many values range for example

To tackle this problem we can use

RandomizedSeachcv

It wont try every single permutation and combination of parameters but It will try random combination of those values , if u don’t have too much computation power