Cross Validation: Cross validation is used to evaluate model performance that means which model is best for my dataset.

There are three options to know which model is best for our dataset.

Option 1: Use all available data for training and test on same dataset.

Suppose I have 100 math questions. I train a kid how to solve this 100 math questions. When the kid goes for exam I ask exact same questions. Then I try to measure the mathematical skill of that kid based on the score. This is not a good way measuring someone math skill because he has already seen those questions.

Option 2: Split available dataset into training and test sets.

Again suppose we have 100 math questions to train a kid. This time I take 70 questions for training and 30 questions to test that kid. This option is used almost all of our supervised learning using train_test_split from sklearn. When the kid goes for the test he is not seen this 30 questions which is good to measure the math skill of that kid. But there is a problem, in the 70 question, they are all geometry and remaining 30 questions is from calculus. Those questions, he has not seen before or he doesn't have any knowledge about calculus. This method is good sometimes but is not perfect.

Option 3: K fold cross validation.

In this technic, we divide our 100 samples into folds. Suppose we divide our 100 samples into 5 folds. Each fold containing 20 samples and then we ran multiple iterations.

In the first iteration we use fold number 2 to 5 for training and first fold for testing and note down the score.

Second iteration we use fold number 1 and 3 to 5 for training and second fold for testing and note down the score.

This process will run until the last fold is for testing. After that we average our score. This is very good because we giving the variety of samples to train the model.

StratifiedKFold is similar to **KFold.** It is little better in a way that when you are separating your folds it will divide each of the classification categories in a uniform way. These could be very helpful. Suppose we are creating 3 folds in iris dataset. Two folds have two types of flower and third fold has different type of flower. Then it might creat problems. That's why **StratifiedKFold** is better.