image: images/CV.png

**CROSS VALIDATION**

**If we have 1000 data points in our whole dataset we generally split that into 70-30% or 75-25% (70% train-30% test)**

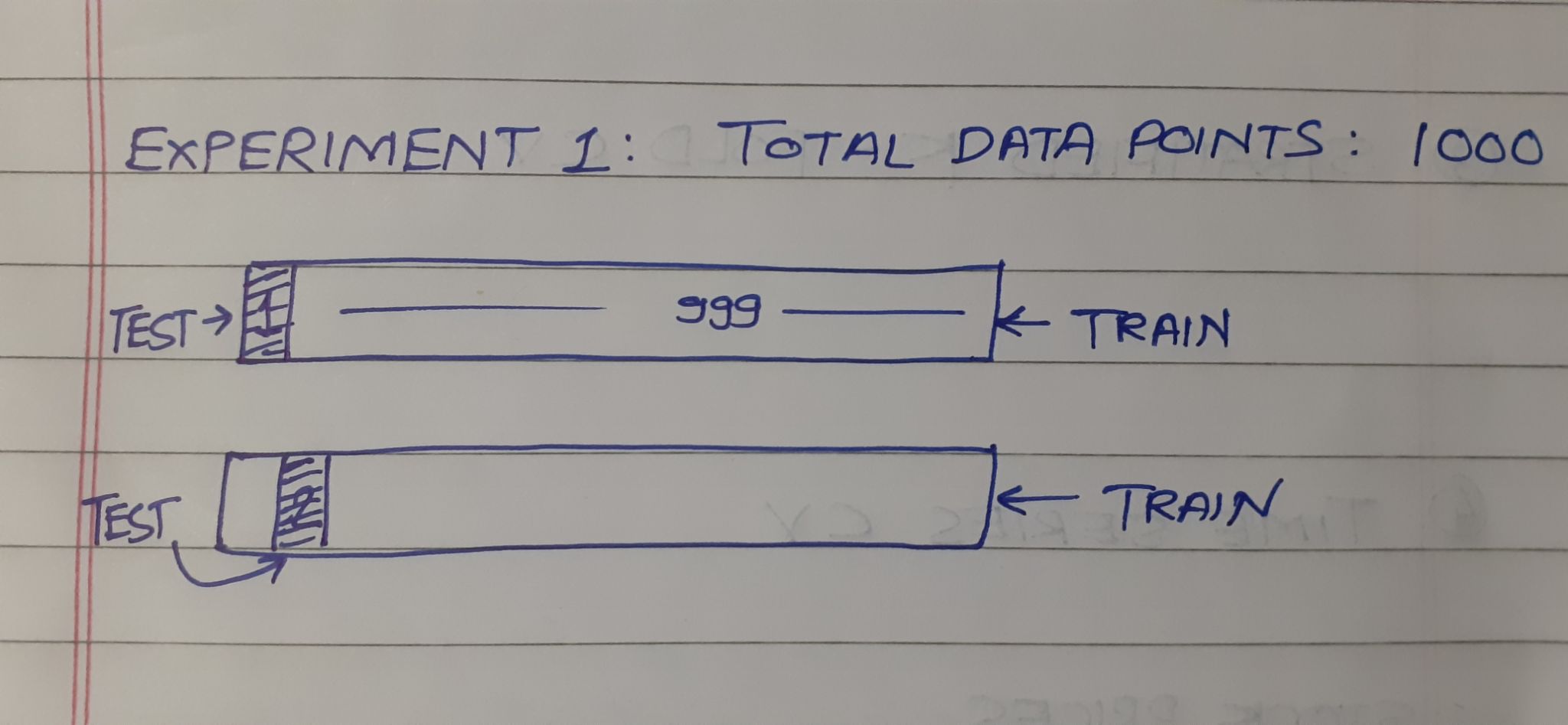
**Now we set a variable called Random state to any integer value(popular values are 42,0). For each such random state dataset is randomly split into train and test.**

**So suppose, if we get 85% accuracy on test data for Random\_state=0, then we may get 87% test accuracy for Random\_state=42**

**Types of Cross Validation-**

**1. LOO-Leave One Out CV-**

This cross validation method is very old. As the name suggests, if there are a total 1000 data points then only the 1st single data point is taken as a test data and the rest 999 are used for training the model.

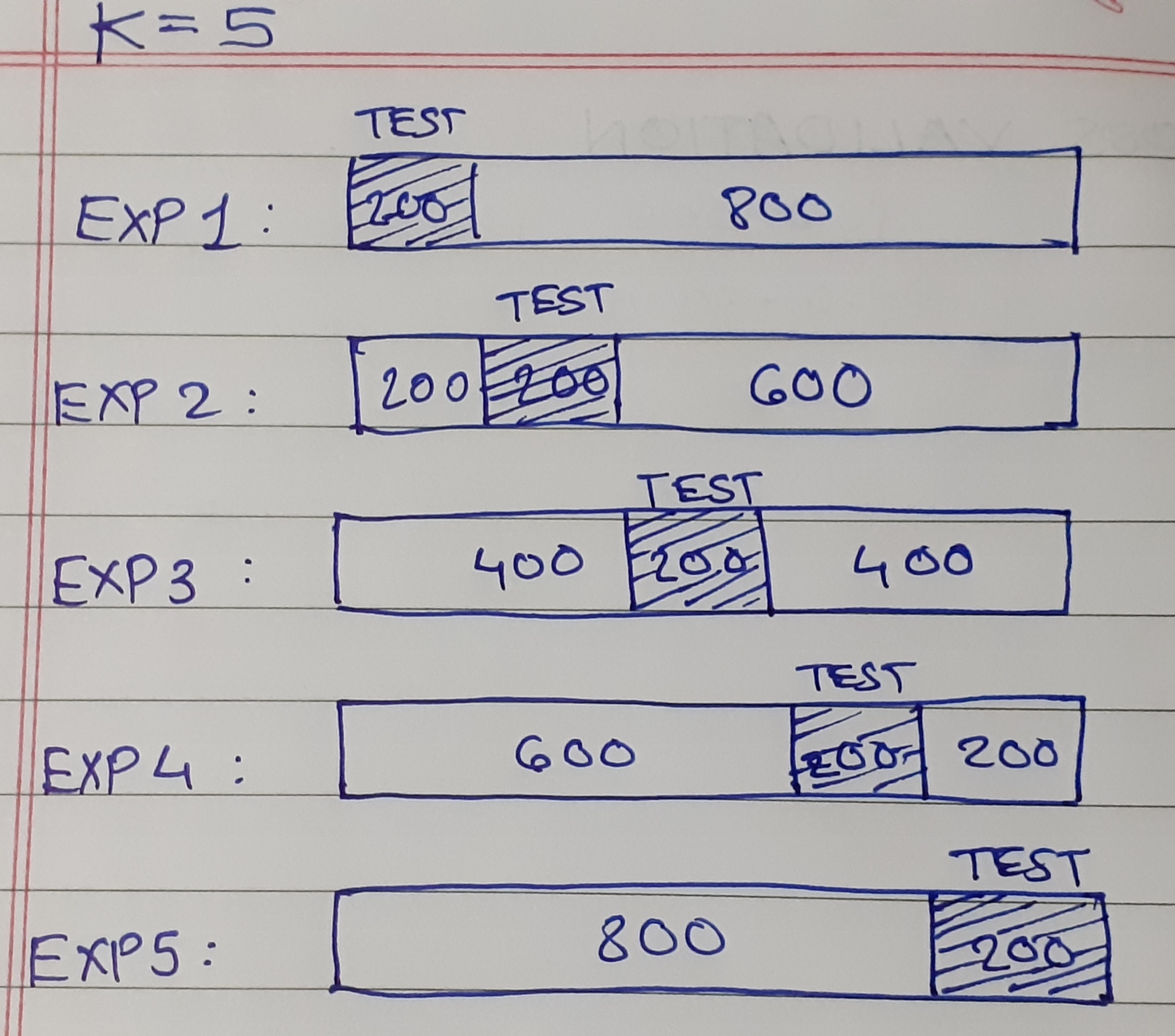


Obvious disadvantages-

1. A lot of computation power require
2. As there is only single test data point Bias is very Low

**2. K-Fold Cross Validation-**

In K-Fold CV, we consider some small value of K lets say K=5. So now if we have a dataset containing a total 1000 data points we will perform 5 experiments where 1000/5=200, 200 data points randomly get selected as test data for each experiment.



Advantages-

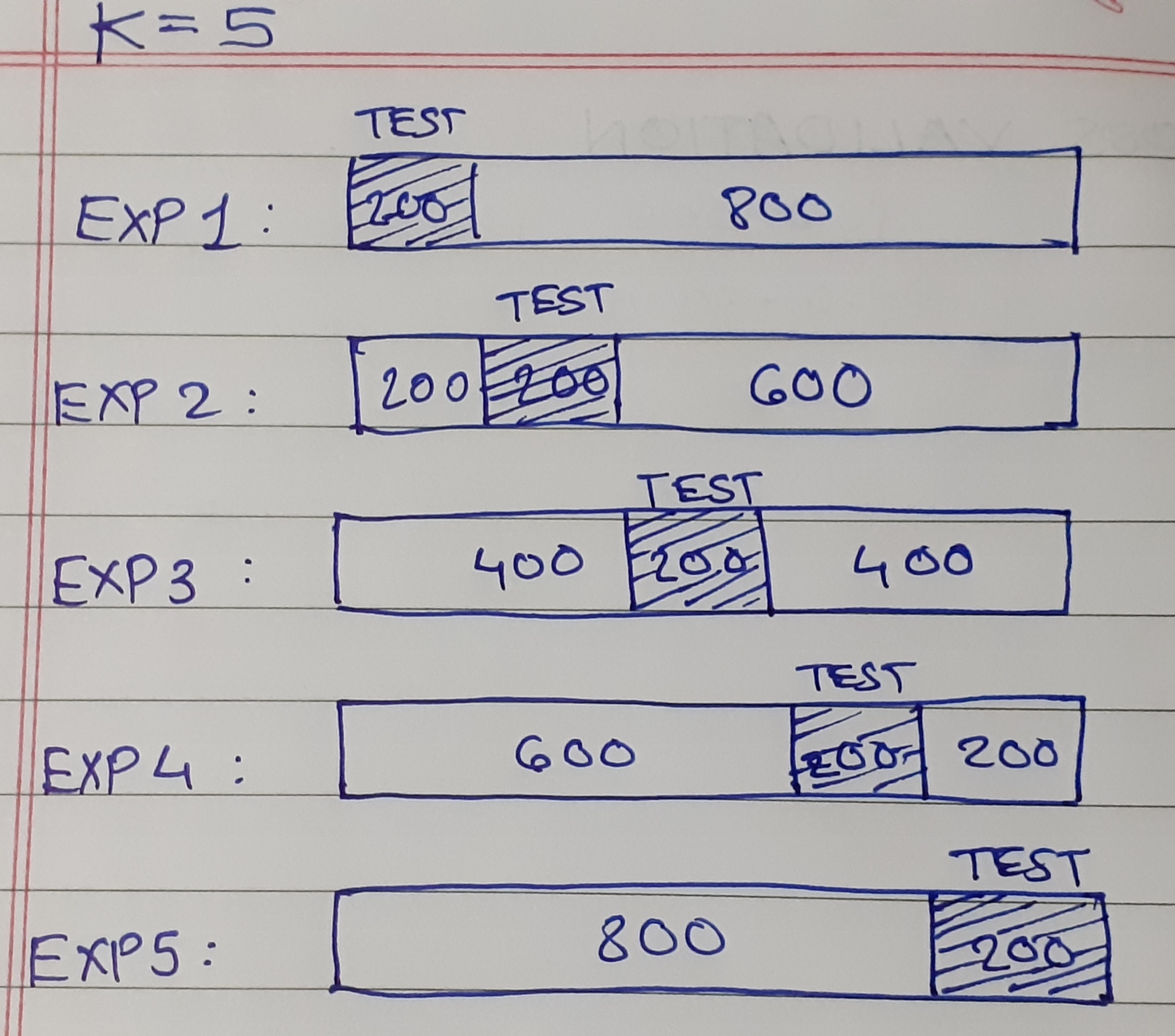
1. It overcomes Low Bias issue of first method

Disadvantages-

1. If there is a Binary Classification problem of 1's and 0's so I can get my test data for some experiment containing only 1's or maximum 0's or 1's in my training data. This will result in Low Bias and bad performance of model

**3. Stratified K-Fold Cross Validation-**

This is probably the most used CV method now. This is similar to normal K-Fold CV but it overcomes the issue in the above method. In this, the algorithm makes sure that the number of instances of each class gets split into train and test data in a proper ratio.



Advantages-

1. Overcomes the Low Bias problem of K-Fold CV

**4. Time Series Cross Validation-**

This method is used for Time Series problems as you just CAN NOT use the above method for TS data as random splitting is useless. So we cross validate something like this-

