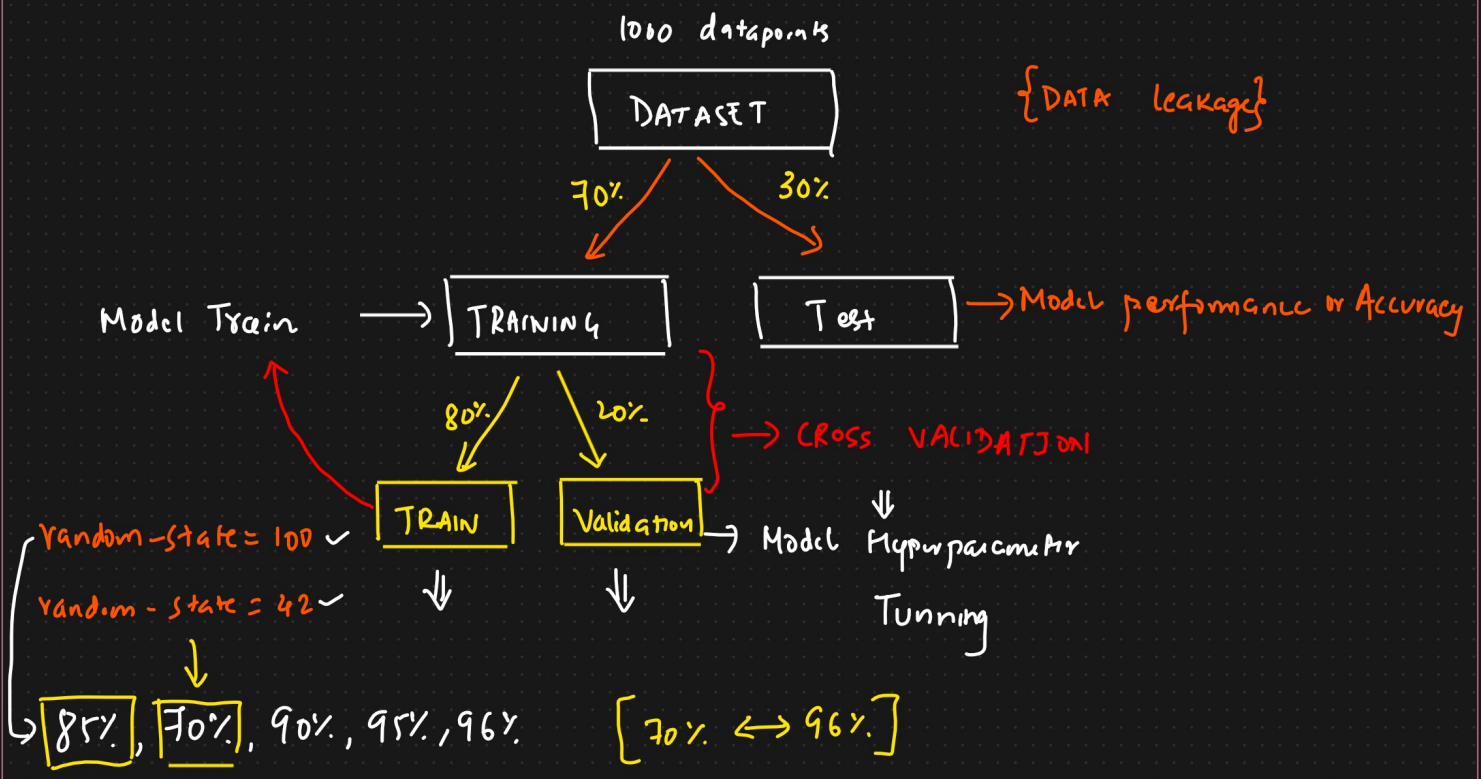
CROSS VALIDATION AND ITS TYPES



CROSS VALIDATION CV = 5

Exp1	TRAIN	VALIDATE	Acc1
Exp2	TRAIN	VALIDATE	Acc2
Exp3	TRAIN	VALIDATE	Acc3

[Acc1, Acc2, Acc3]

[80%, 85% 78%]

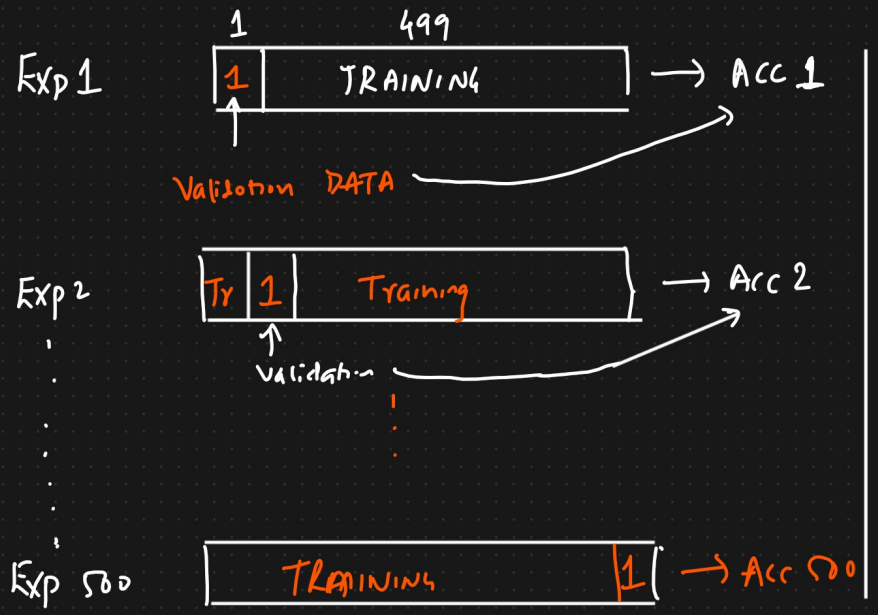
⇓

Final Accu = Mean of this Accuracy

Types of Cross Validation

① Leave One Out Cross Validation (100 CV)

TRAINING DATA → 500 Records



TRAINING → Model TRAIN

VALIDATION → Model Predict

⇒ Mean of All these Accuracies

Disadvantage

① Time Complexity is huge for training Big datasets

② Model Overfit ⟶ TRAINING Acc ↑↑
New data → Validation Acc ↓↓

② Leave p out CV

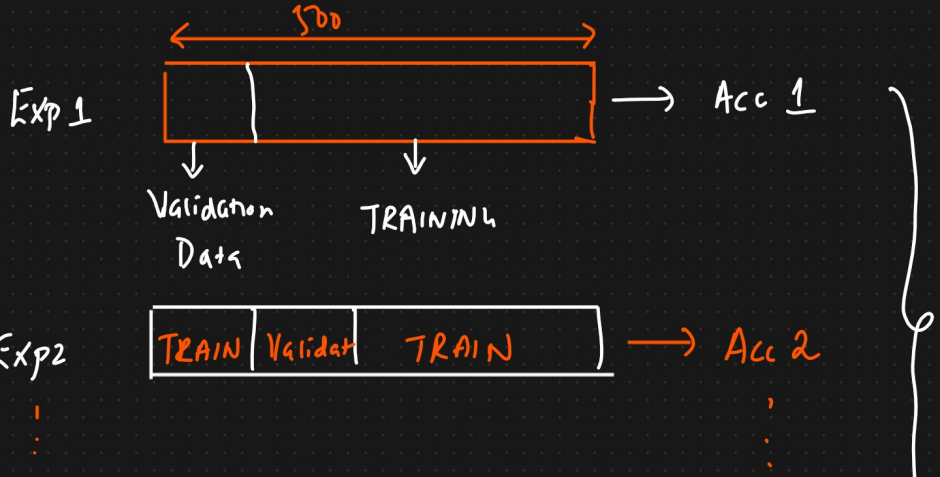
p = 10, 20, 30, 40 → Hyperparameter

③ K Fold Cross Validation

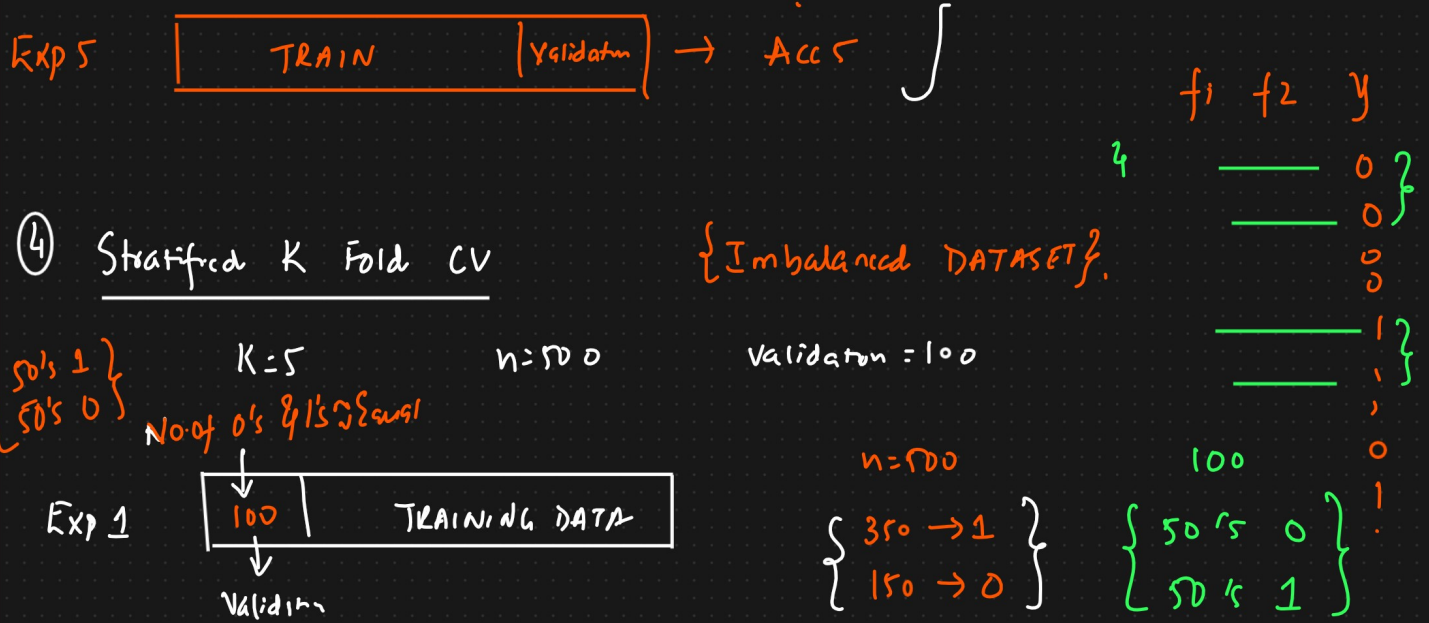
CV = 5

Validation size = $\frac{500}{5}$ = 100

K = 5 n = 500



Average of all the Accuracy.


Exp 5
TRAIN
Validation
Acc 5
f1 f2 y
4
0
0
0
1
1
0
1
{Imbalanced DATASET}.
④ Stratified K Fold CV
K=5
n=500
Validation = 100
50's 1
50's 0
No of 0's & 1's are Equal
Exp 1
100
TRAINING DATA
Validation
n=500
350 → 1
150 → 0
100
50 's 0
50 's 1
Exp 2
100
validation
⑤ Time Series CV
Amazon review Sentiment Analysis
Product A
JAN → DEC
TRAINING
VALIDATION
Day 1
Day 2
Day 3
Day 4
Day N
Time Series Application