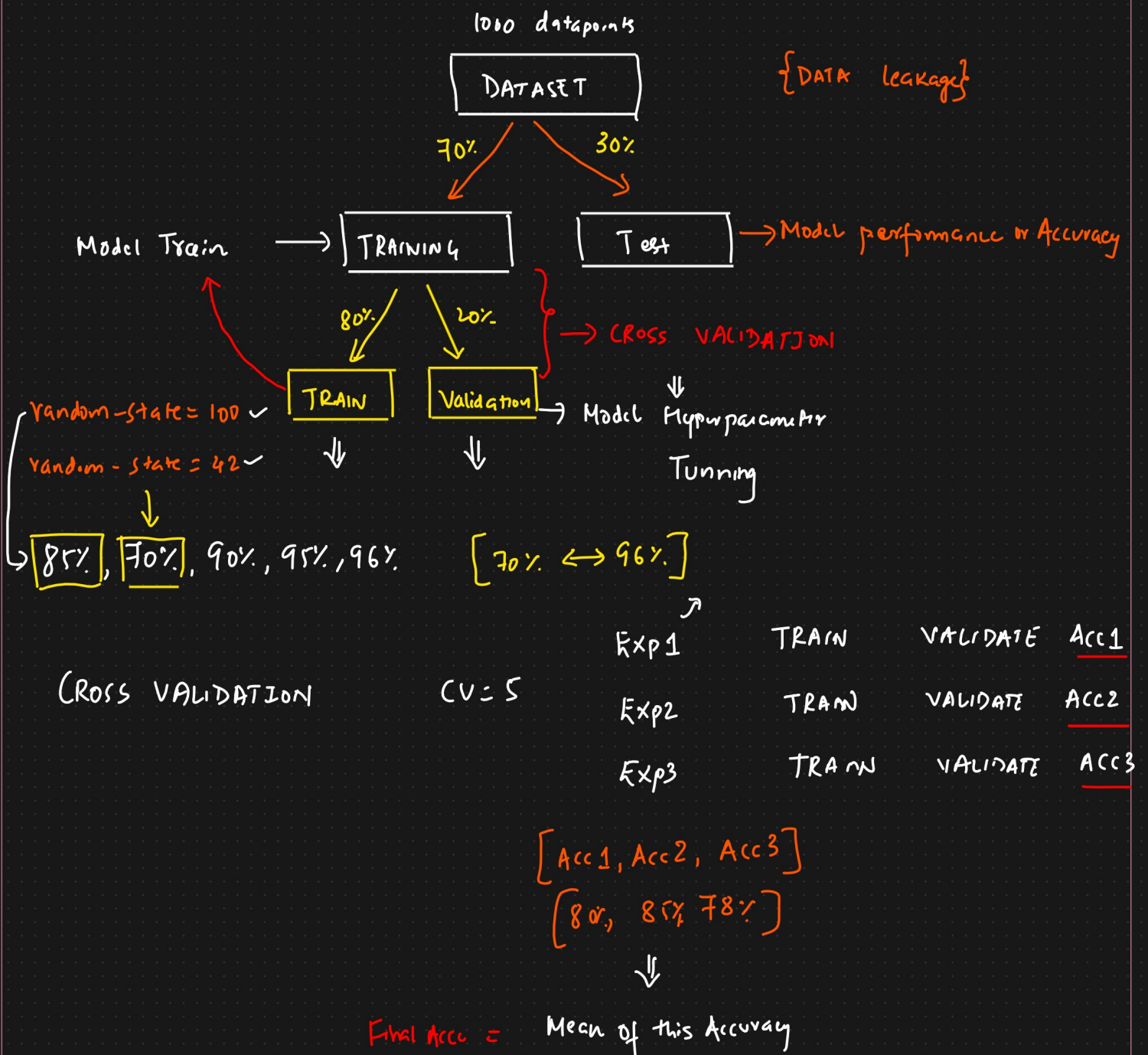


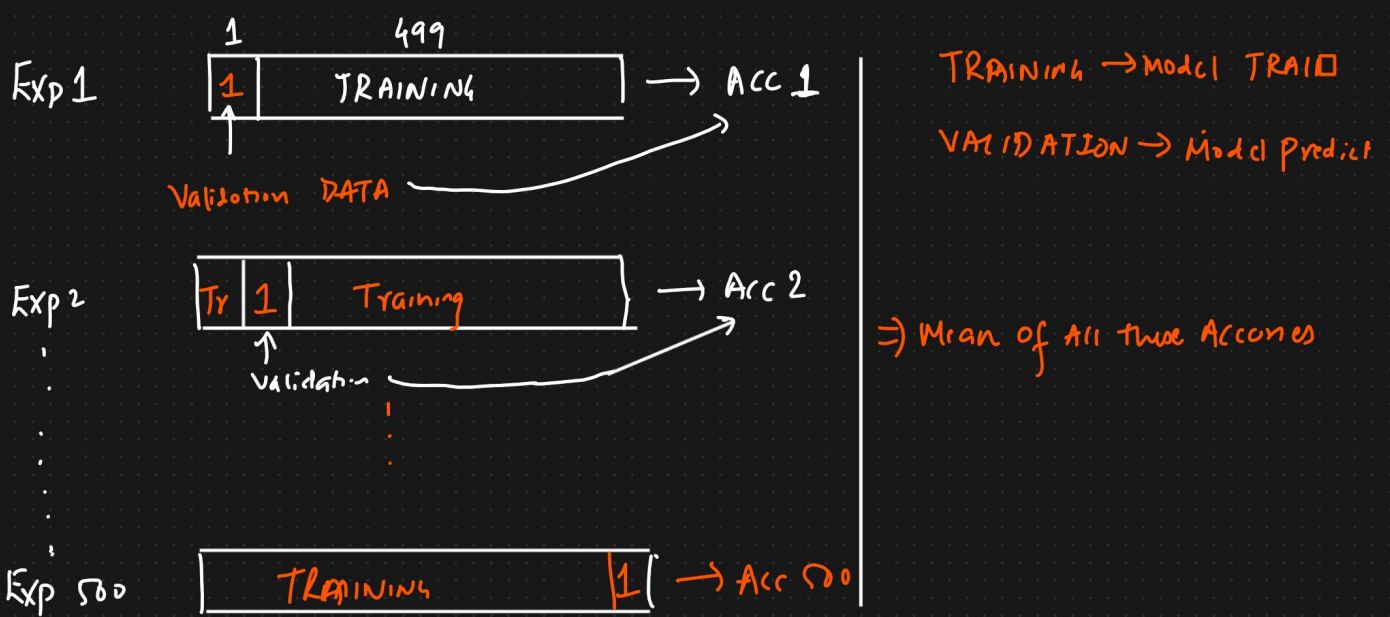
# CROSS VALIDATION AND ITS TYPES



## Types of Cross Validation

### ① Leave One Out Cross Validation (LOOCV)

TRAINING DATA → 500 Records



### Disadvantage

- ① Time Complexity is huge for training Big datasets
- ② Model Overfit → TRAINING Acc ↑↑  
New data → Validation Acc ↓↓

### ② Leave p out CV

$p = 10, 20, 30, 40 \rightarrow$  Hyperparameter

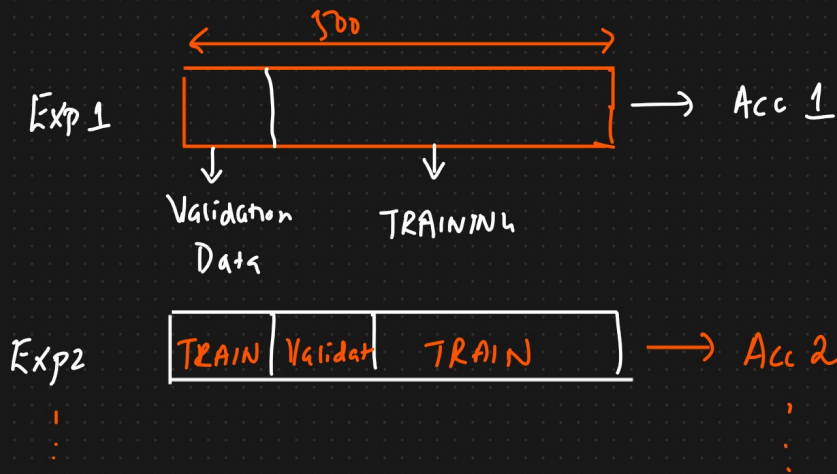
### ③ K Fold Cross Validation

$K = 5$

$n = 500$

↓  
 $CV = 5$

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



Average of all the Accuracy.

TRAIN | Validation → Acc5

$$\begin{array}{r|l} f_1 & f_2 & y \\ \hline & 0 & \\ & 0 & \\ & 0 & \\ & 0 & \\ \hline & 1 & \\ & 1 & \\ & 1 & \end{array}$$

④ Stratified K Fold CV

## { Imbalanced DATASET }

$$\left\{ \begin{array}{l} \text{50's 1} \\ \text{50's 0} \end{array} \right\}$$

$K=5$   
No. of 0's & 1's  $\sim \Sigma \text{ans}$

$$n = 500$$

Validation = 100

## Exp 1

A diagram illustrating data partitioning. A large rectangle is divided into two horizontal sections. The top section is labeled "TRAINING DATA" and contains a smaller box with the number "100" inside. An arrow points down from the "100" box to the word "Validation" written below the bottom section of the rectangle.

$$n = 100$$
$$\left\{ \begin{array}{l} 350 \rightarrow 1 \\ 150 \rightarrow 0 \end{array} \right\}$$

100

$$\begin{Bmatrix} 50 & 5 & 0 \\ 50 & 5 & 1 \end{Bmatrix}$$

## Exp2

A diagram showing a horizontal bar representing a memory segment. Inside the bar, there is a smaller vertical rectangle containing the number '100'. An arrow points down from the top of this rectangle to the word 'validation' written below the bar.

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