CROSS VALIDATION AND ITS TYPES
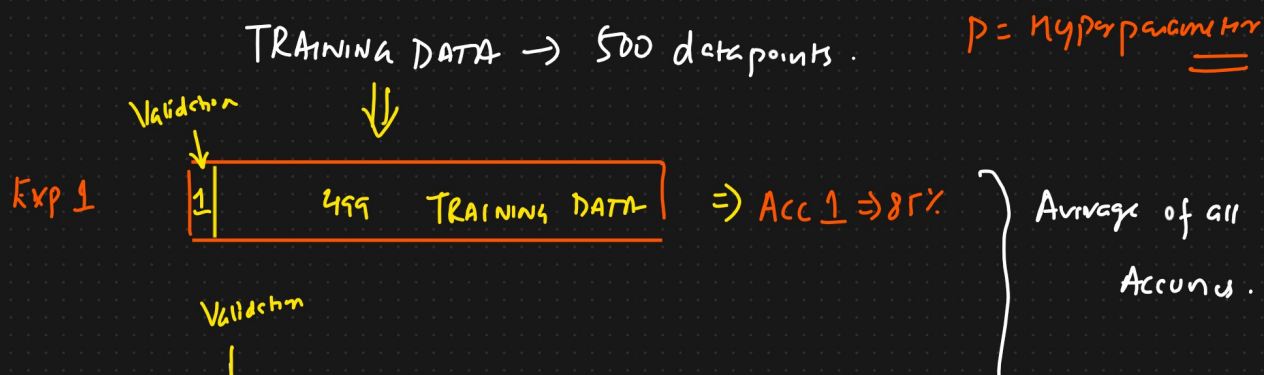
1000 datapoint

```
        ┌──────────────┐
        │   DATASET    │
        └──────────────┘
         70%        30%
        ┌──────────┐      ┌────────┐
Model → │ TRAINING │↰     │  TEST  │ → Model performance
Training└──────────┘      └────────┘
        80%      20%
     ┌───────┐  ┌────────────┐
     │ TRAIN │  │ Validation │  } CROSS VALIDATION → Hyperparameter
     └───────┘  └────────────┘        ⇑              Tuning
```

random-state = 100 }
          = 42 }

→ [85%, 75%, 70%, 68%, 95%] ⇒ Average Accuracy
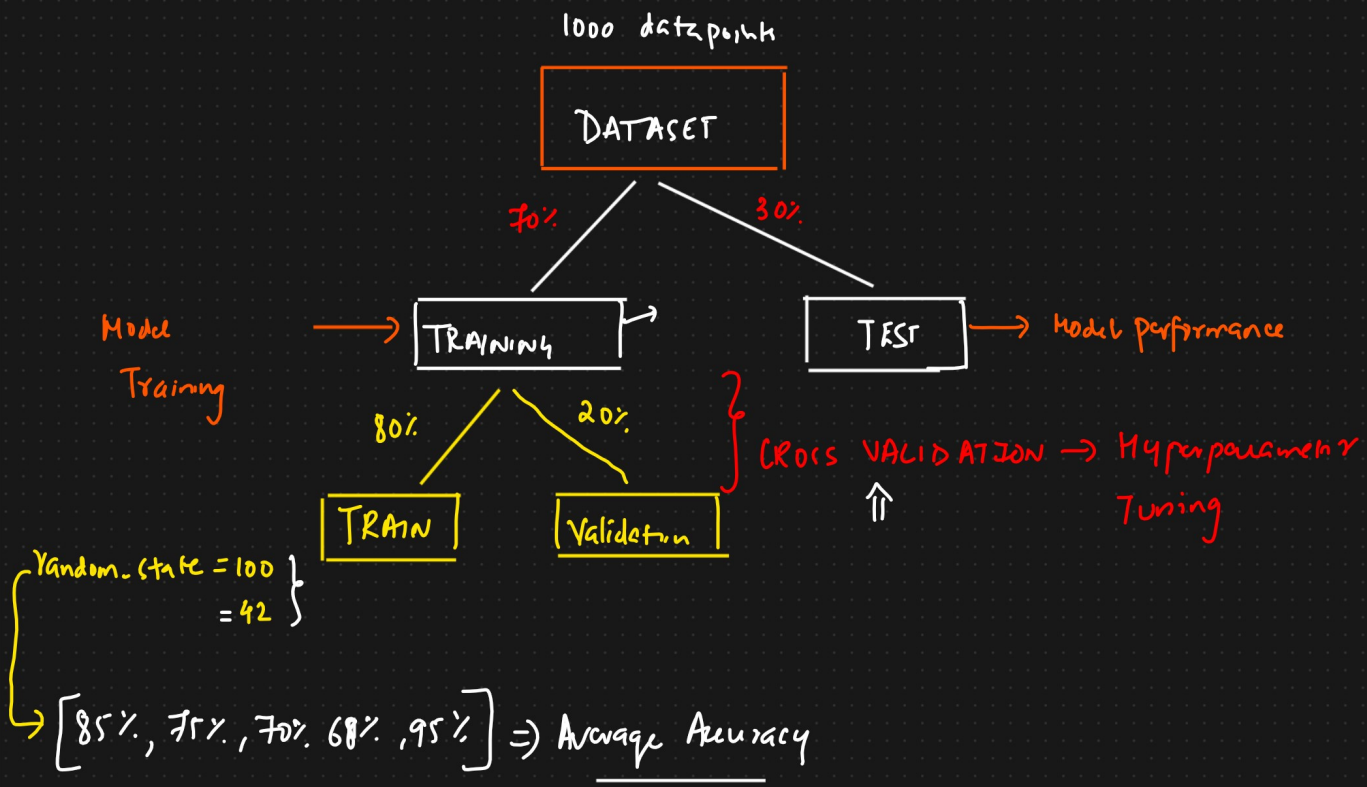
	EXP1	TRAIN	VALIDATION	Acc 1
TRAINING ⇒ CV = 5	EXP2	TRAIN	VALIDATION	Acc 2
	⋮			
	EXP	TRAN	VALIDATION	ACC 5

} Mean of Accuraci

Types of Cross Validation

$p = 10, 20, 100$

① Leave One Out Cross Validation (LOOCV)   Leave p out CV

TRAINING DATA → 500 datapoints. $p = $ Hyperparameter

Validation
↓

```
        Validation
           ↓
EXP 1   ┌─┬─────────────────────────┐
        │1│ 499  TRAINING DATA       │ ⇒ Acc 1 ⇒ 85%    } Average of all
        └─┴─────────────────────────┘                     Accuros.
      Validation
         │
```

Exp 2 | 1 | 2 | 498   TRAININ DATA | ⇒) Acc 2 ⇒) 80%.

!

Exp 500 | TRAIN. DATA | ⇒) Acc 500 ⇒) 82%.

Disadvantage

① Time Complexity is huge. for Training $\underset{\wedge}{\text{Big}}$ dataset

② Model Overfit ⇒ Training Acc ↑↑

New Data → Acc↓↓

③ K fold Cross validation.

K=5   n=500   Validation data $= \dfrac{500}{5} = 100$

Exp 1 | Vald | TRAIN | → Acc → • Training Acc
Exp 2 | train | Valt | Train | → Acc 2   Average of Accuracies.
!
Exp 5 | Train | Valt | → Acc 5

④ Stratified K Fold CV   {Imbalanced Dataset}

0's, 1's   K=5   n=500 $\dfrac{500}{5} = 100$ n=500   100

Exp 1 | 100 | TRAIN | Acc1   { 1 → 350 }   { 50 - 0's }
     validat↓ 100   { 0 → 150 }   { 50 → 1's }

Exp 2 | 0's&1's | Acc 2   { Average Accuracy.
!

Acc5