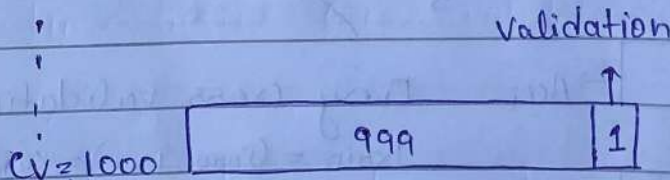
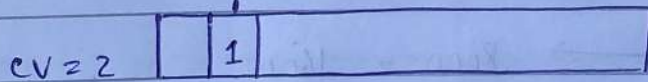
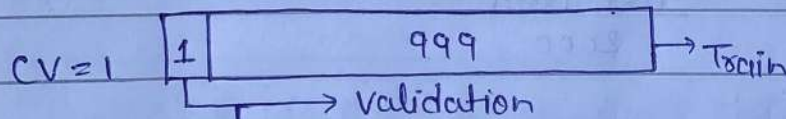
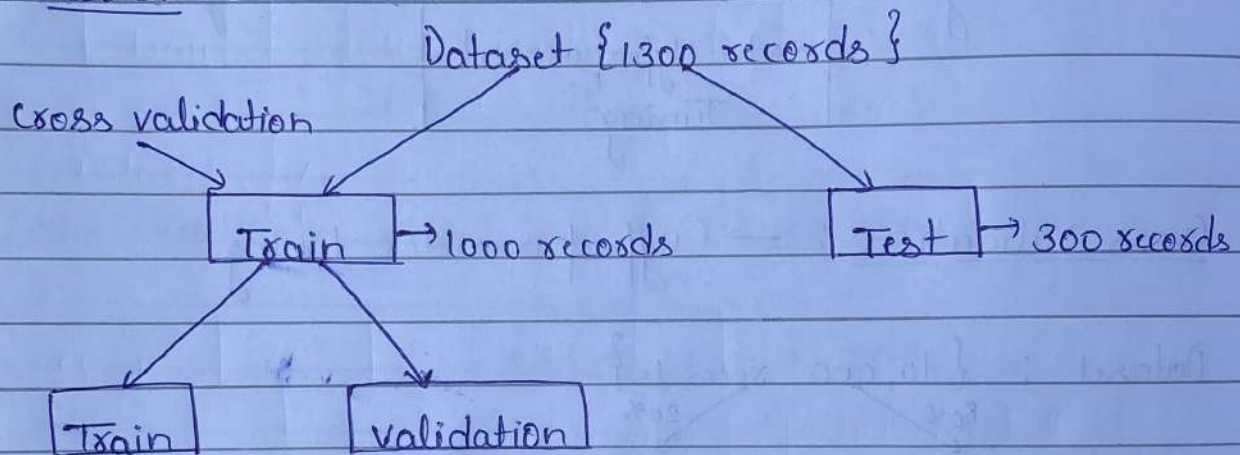


## Types of Cross Validation

- ① looCV (leave one out cross validation)
- ② Hold out cross validation
- ③ K-Fold cross validation
- ④ Stratified K-Fold cross validation
- ⑤ Time Series cross validation

\* looCV :

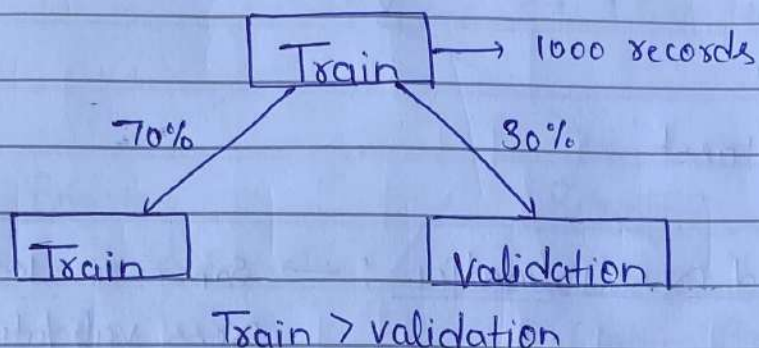


→ because 1000 records

→ Disadvantage :

- Time consuming task
- Overfitting

## \* Hold out cross validation :



Note : It will select train & validation data with random state.

Selected with random state value

70% 30% → Accuracy = 87%

30% → Accuracy = 84%

→ Disadvantage :

- Imbalance Dataset.

## \* k-Fold cross Validation : → Initialize k-value.

Train → { 900 records } → let  $k = 3 \Rightarrow \frac{900}{3} = 300$

CV = 1	300	700	→ Accuracy <sub>1</sub>	validation Data
CV = 2		300	→ Accuracy <sub>2</sub>	
CV = 3	700	300	→ Accuracy <sub>3</sub>	



$$\text{Average Accuracy} = \frac{\text{Accuracy}_1 + \text{Accuracy}_2 + \text{Accuracy}_3}{3}$$

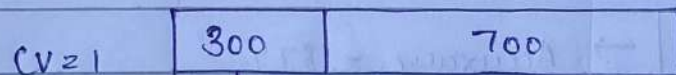
→ Disadvantage :

- Imbalance dataset

\* Stratified k-Fold Cross Validation : → Same as k-Fold cross validation

$$\boxed{\text{Train}} \rightarrow \{900 \text{ records}\} \rightarrow k=3 \rightarrow \frac{900}{3} = 300$$

$$\text{Dataset} = \begin{cases} A - 800 \\ B - 100 \end{cases}$$



→ Stratified {equal proportion for both class}

60-40% Ratio

70-30% Ratio

\* Time Series Cross Validation : → Sequential split based on time

