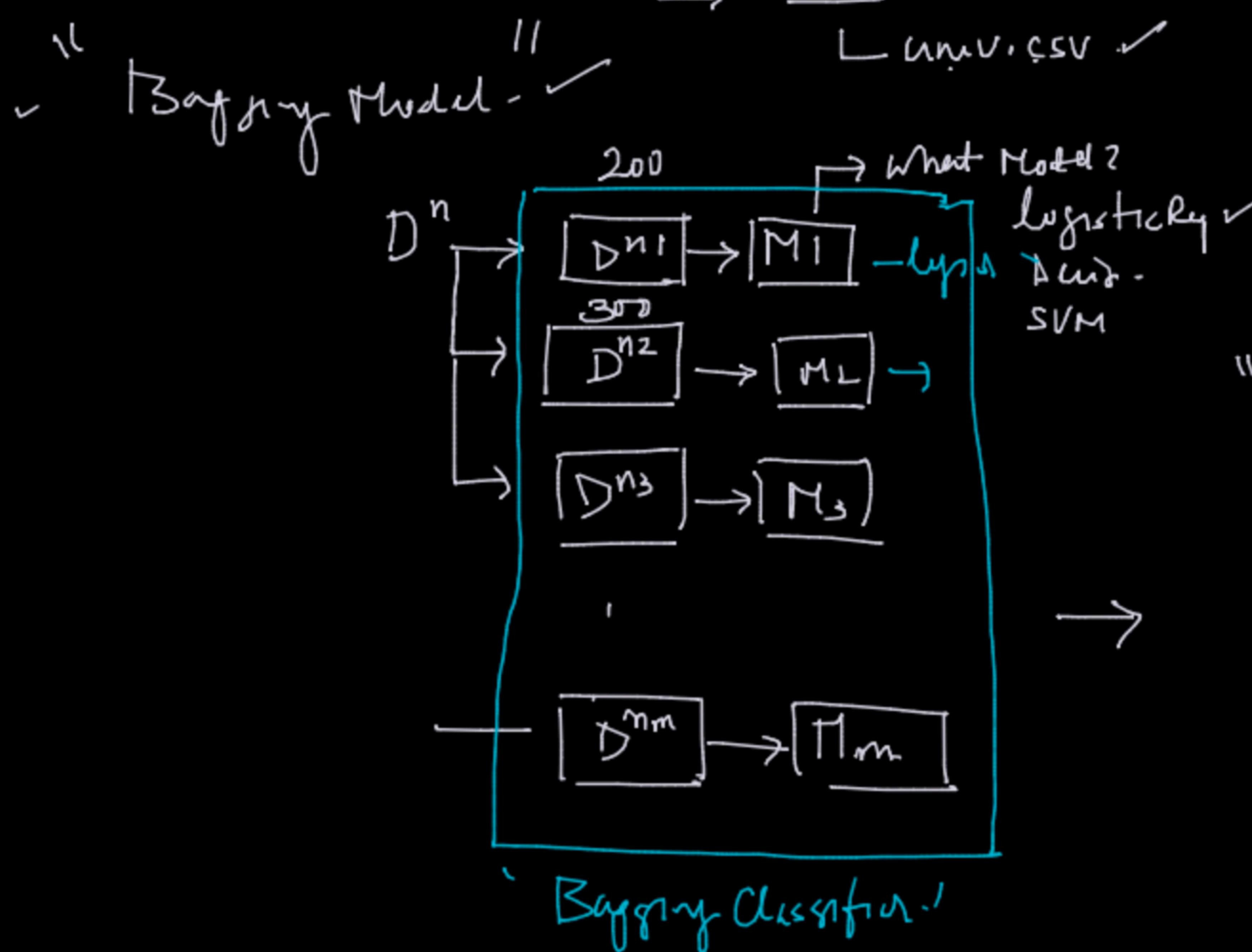


absolute path

`c:/Code/dataset/univ.csv`

.. /dataset/univ.csv

relative



(1) What is the base model → logReg

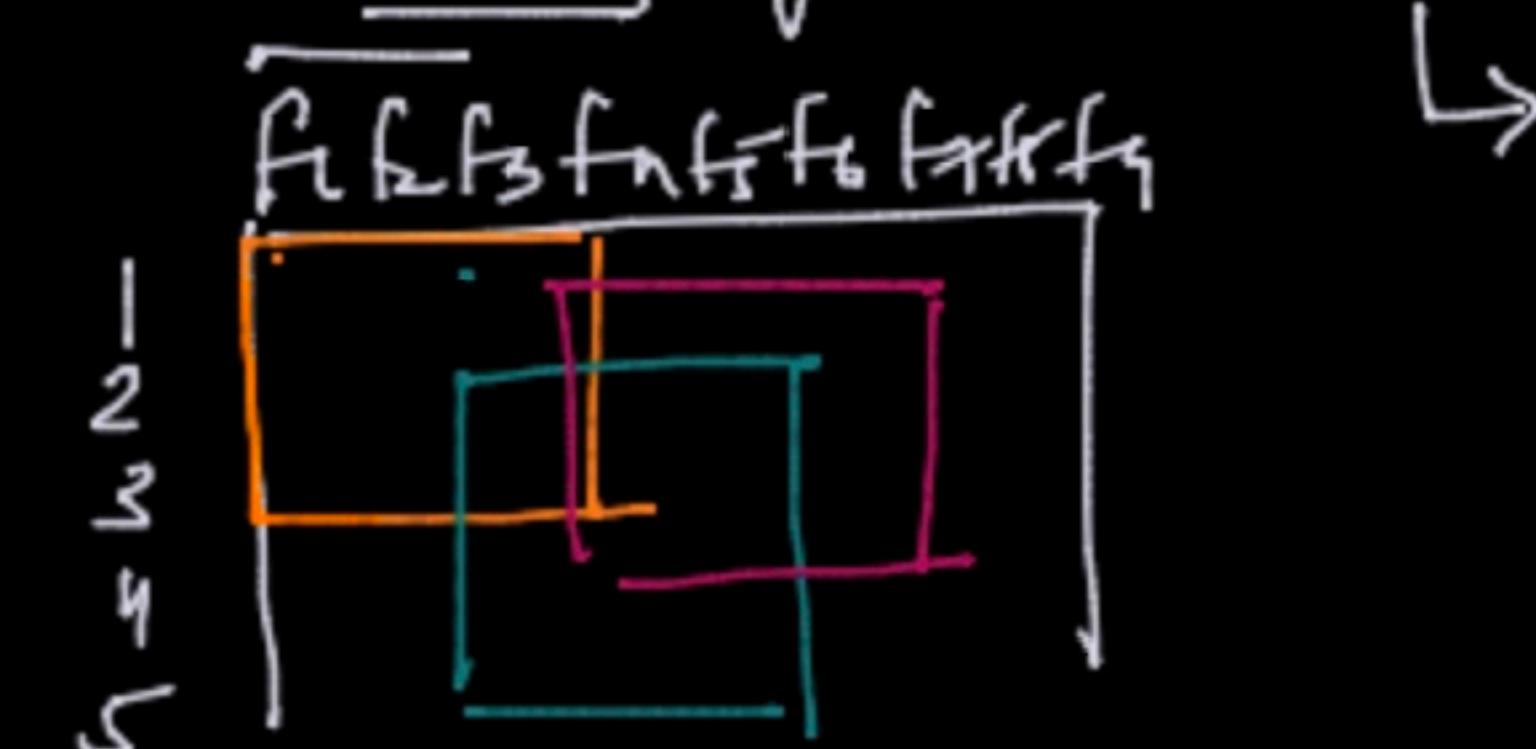
(2) No. of base Models → 100

"RandomForest" →

✓ (1) All base models are DT X

(2) No. of base models → 100

✓ (3) Row sampling + Column Sampling



$f_1 f_2 f_3 \quad y \quad y$

$$\begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{matrix} \quad \begin{bmatrix} \cdot & \cdot & \cdot \\ X_{1,1} & \dots & X_{1,n} \end{bmatrix} \quad \begin{bmatrix} \cdot \\ \vdots \\ \cdot \\ y_{\text{train}} \\ \cdot \\ \vdots \end{bmatrix} \quad \begin{bmatrix} \cdot \\ \vdots \\ \cdot \\ y_{\text{true}} \\ \cdot \\ \vdots \end{bmatrix}_{\text{true}} \rightarrow \begin{matrix} \text{training metrics} \\ \text{cm, clf-report} \end{matrix}$$

$$m \quad \begin{bmatrix} \cdot & \cdot & \cdot \\ X_{m,1} & \dots & X_{m,n} \end{bmatrix} \quad \begin{bmatrix} \cdot \\ \vdots \\ \cdot \\ y_{\text{true}} \\ \cdot \\ \vdots \end{bmatrix} \quad \begin{bmatrix} \cdot \\ \vdots \\ \cdot \\ y_{\text{true-test}} \\ \cdot \\ \vdots \end{bmatrix}_{\text{test}} \rightarrow \begin{matrix} \text{testing metrics} \\ \text{cm, clf-report} \end{matrix}$$

X  
98% ✓✓ / 60%  
75% 75% 58%

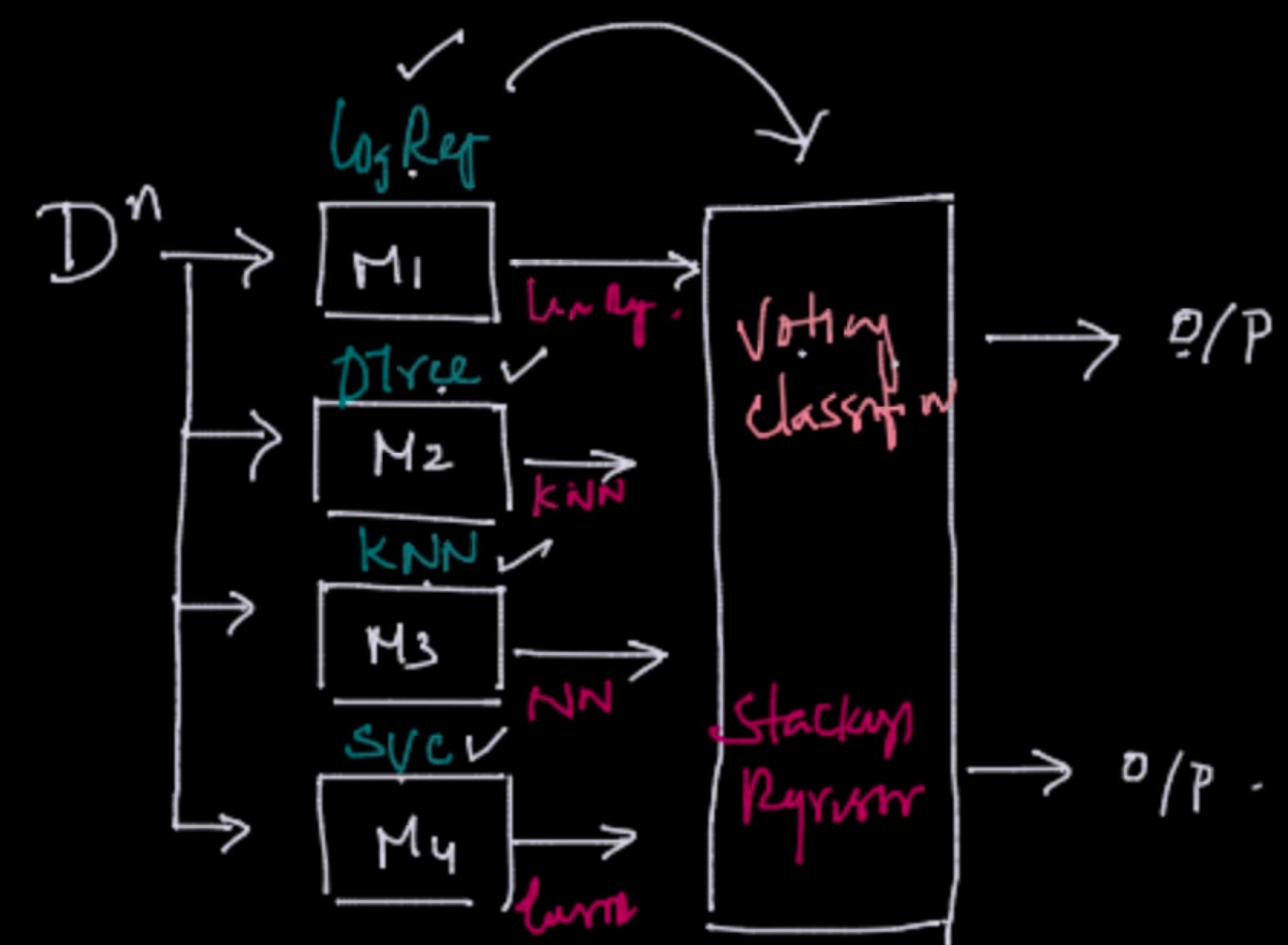


You are screen sharing



Stop share

Talking: Geethika



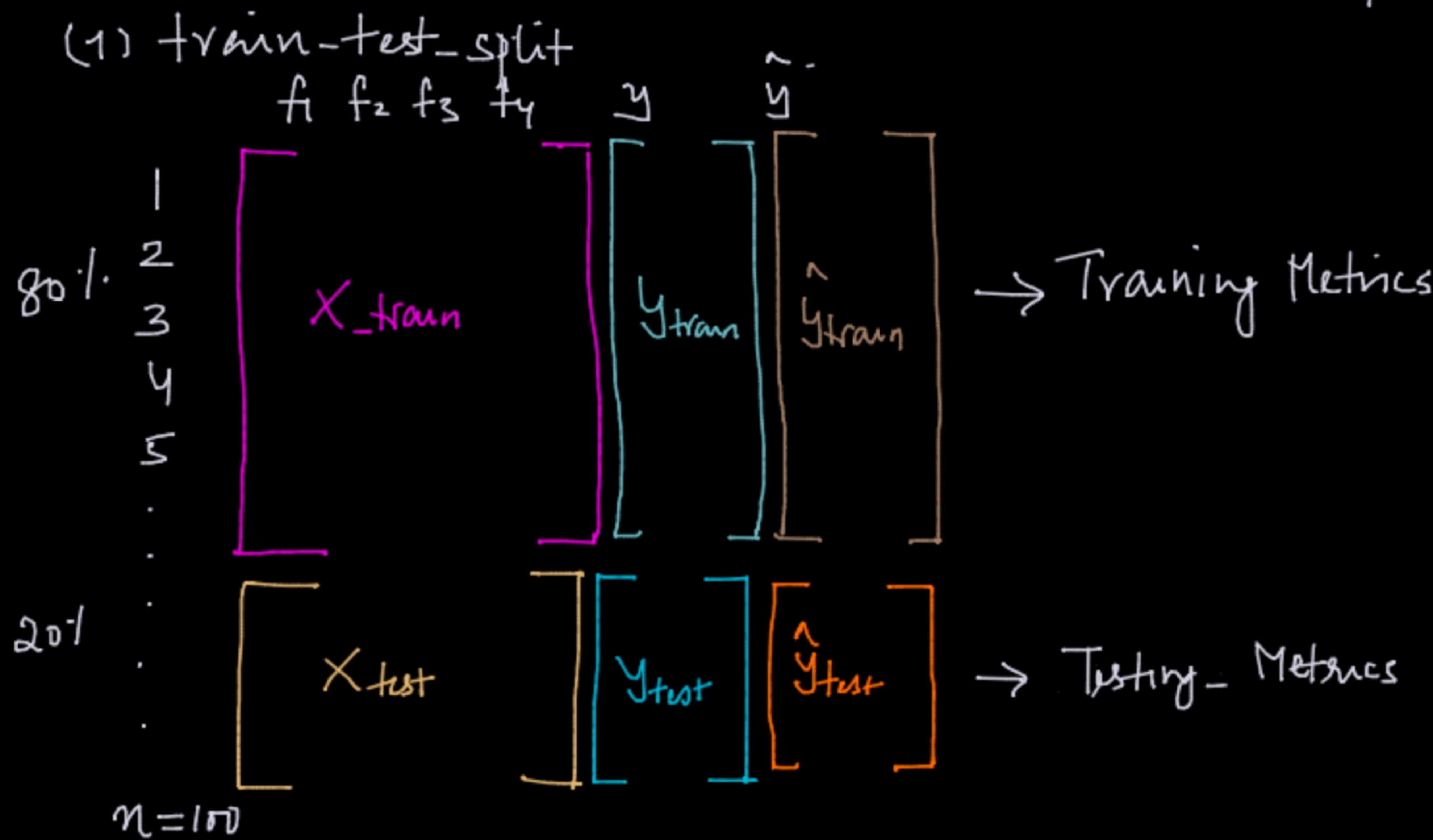
You are screen sharing

Stop share

# ← Model Validation Techniques →

Talking: Geethika

(1) train-test-split



Eraser selected



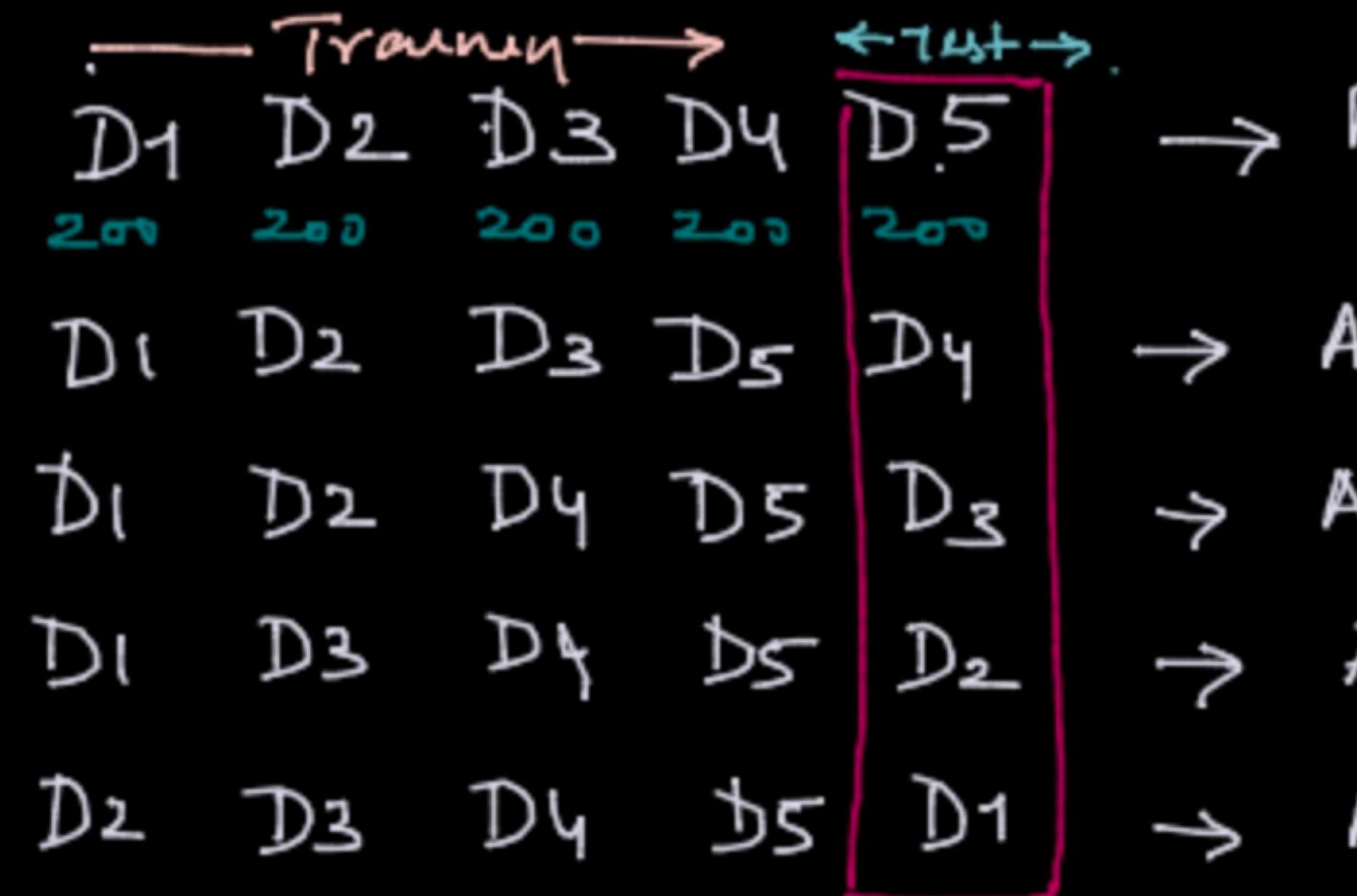
## (2) K-Fold Cross Validation

	Ht	Wt	y	$\hat{y}$
1	Tall	Heavy	-	-
2	Tall	thin	-	-
3	short	thin	-	-
4	short	thin	-	-
5	.	.	.	.
6	.	.	.	.
7	.	.	.	.
8	.	.	.	.
9	Tall	Heavy	-	-
10	Tall	thin	-	-

Train

 $\rightarrow$  No. of splits

$n = 1000$

 $\rightarrow A_1$  $\rightarrow A_2$  $\rightarrow A_3$  $\rightarrow A_4$  $\rightarrow A_5$  $\overline{Avg} \rightarrow$ 

## (3) Leave one out

	f <sub>1</sub>	f <sub>2</sub>	y	$\hat{y}$	Acc.
1	-	-	1	1	1
2	-	-	0	1	0
3	-	-			
4				1	
5				1	
.				.	
.				.	
10				.	

$\frac{5}{10} = 50\%$