

Code ← cluster

mycode.ipynb

dataset

univ.csv

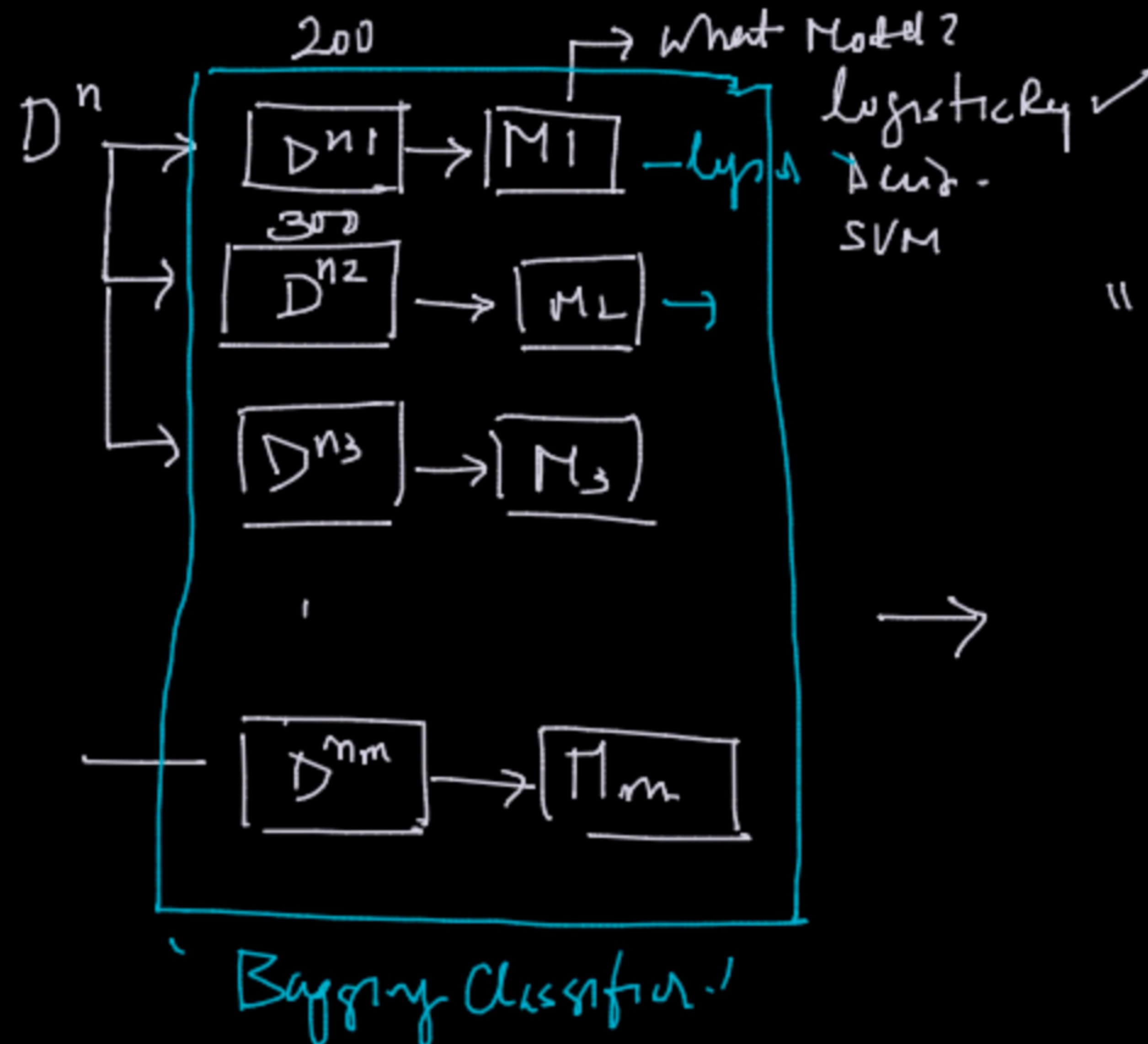
absolute path

c:/code/dataset/univ

read_csv('univ')

../dataset/univ.csv
relative

"Bagging Model"

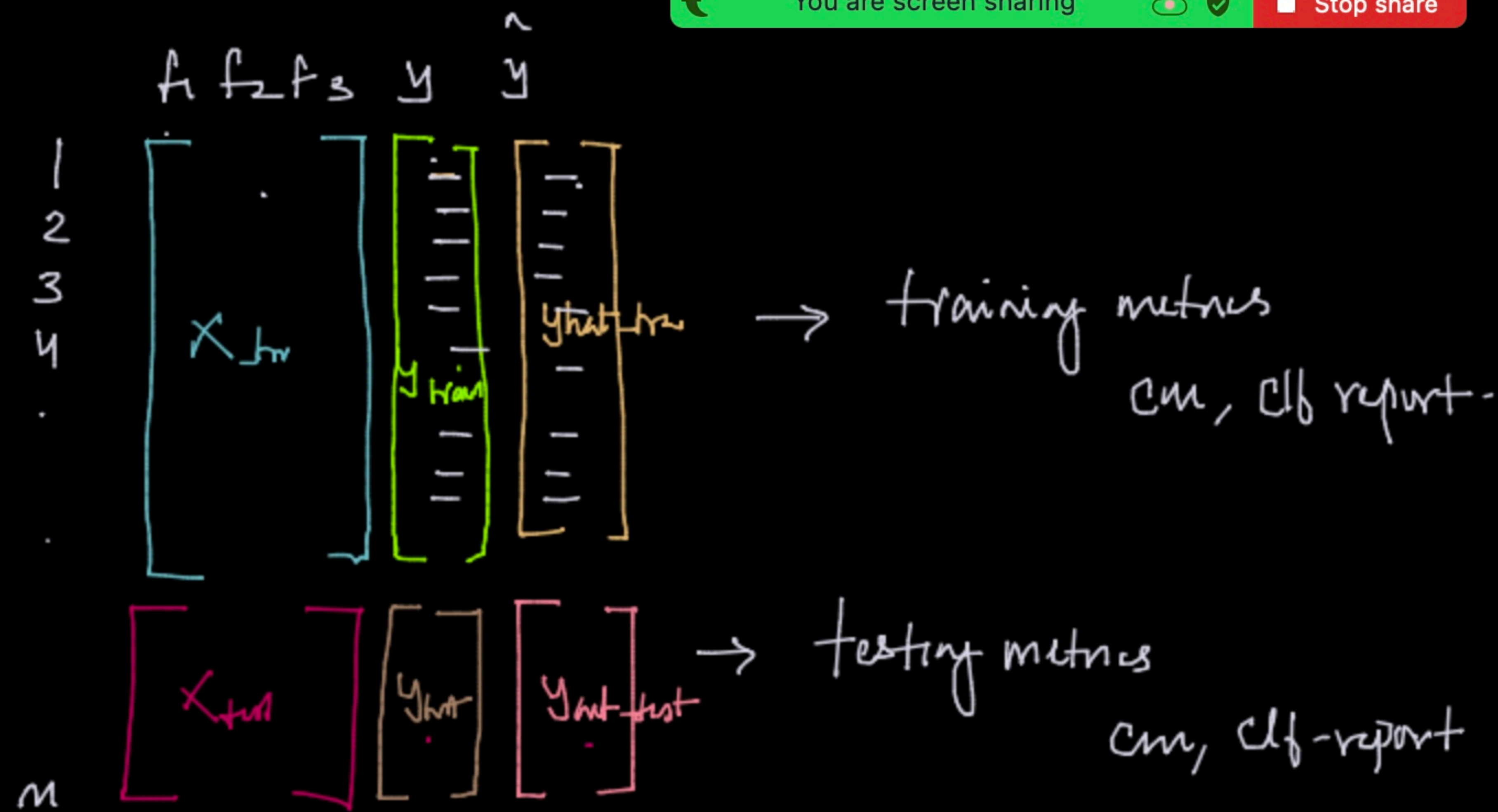


- (1) What is the base model → logpy
- (2) No. of base Models → 100

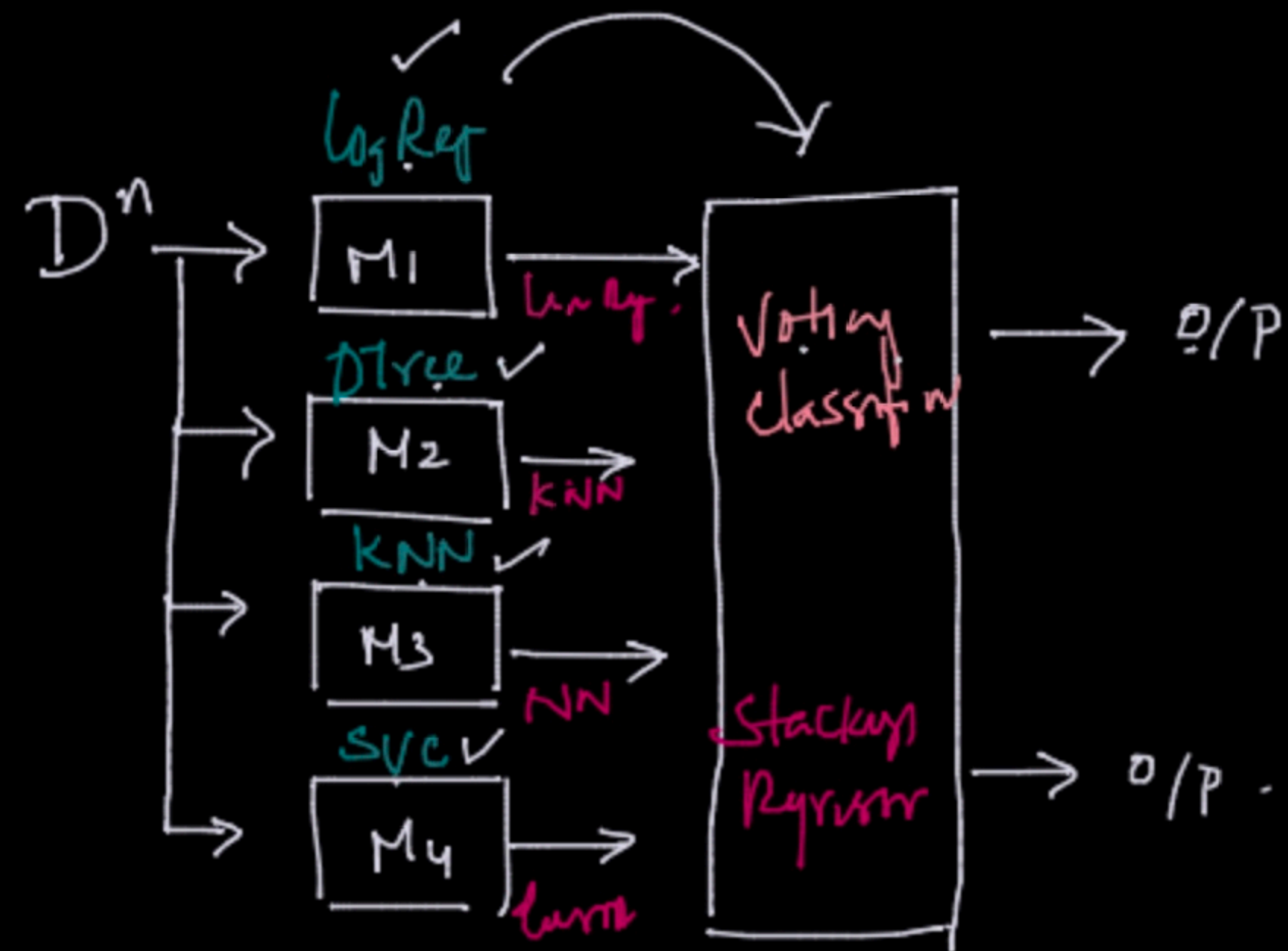
"Random Forest" →

- (1) All base models are DT X
- (2) No. of base models → 100
- (3) Row sampling + column sampling



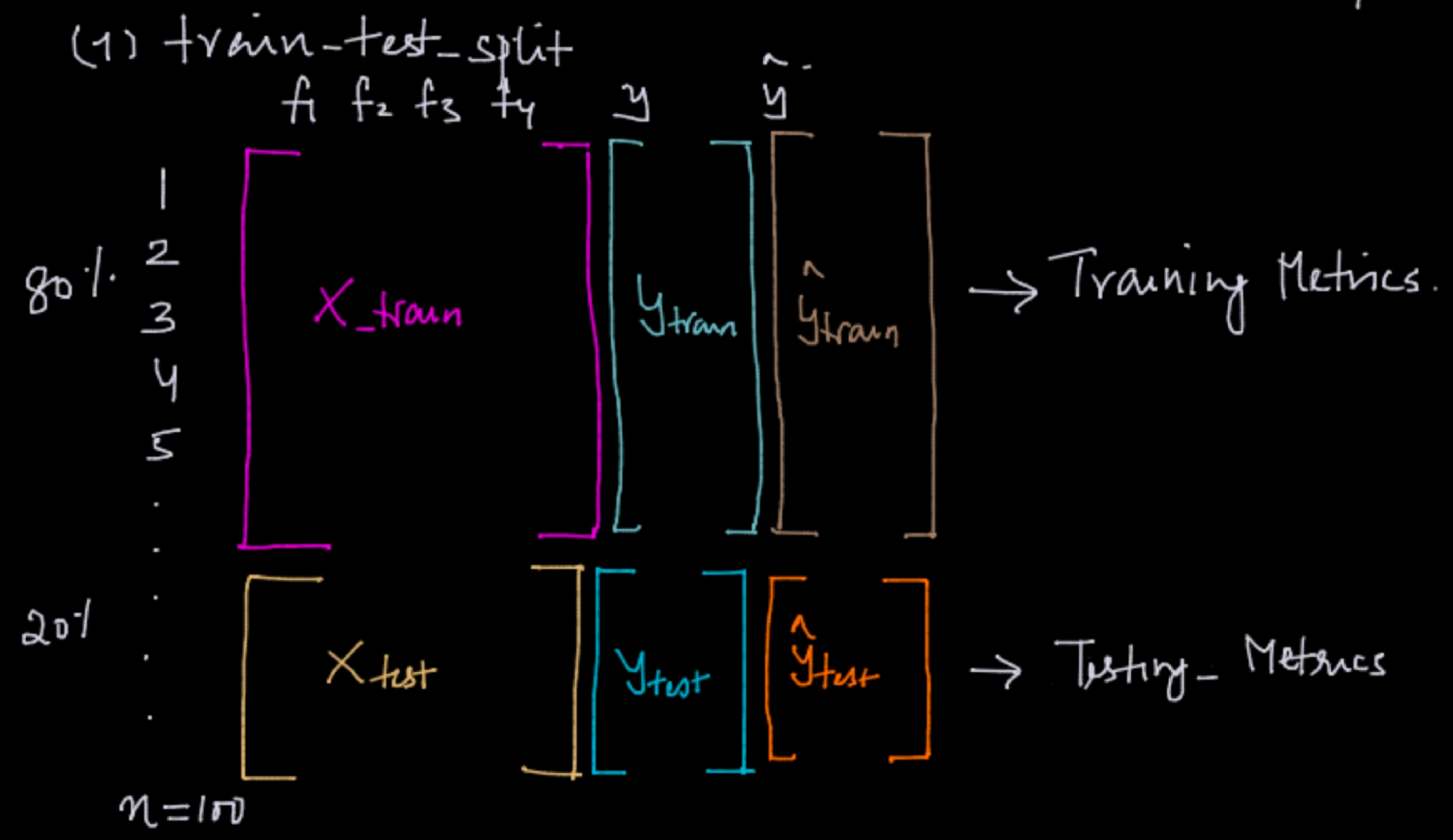


x	✓	x.
98%.	77%	60%.
75%.	75%.	58%.



← Model Validation Techniques →

Talking: Geethika



(2) K-Fold Cross Validation

You are screen sharing Stop share

Talking:

Train

	Ht.	Wt	y	\hat{y}
1	Tall	Heavy	-	-
2	Tall	thin	-	-
3	short	thin	-	-
⋮	⋮	⋮	⋮	⋮
8	⋮	⋮	⋮	⋮
9	Tall	Heavy	-	-
10	Tall	thin	-	-

→ No. of splits

$n = 1000$

Training				Test	
D1	D2	D3	D4	D5	→ A1
200	200	200	200	200	
D1	D2	D3	D5	D4	→ A2
D1	D2	D4	D5	D3	→ A3
D1	D3	D4	D5	D2	→ A4
D2	D3	D4	D5	D1	→ A5

Avg →

(3) Leave one out

→

	f1	f2	y	\hat{y}	Acc.
1	-	-	1	1	1
2	-	-	0	1	0
3	-	-			
4					1
5					1
⋮					⋮
⋮					⋮
10					

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