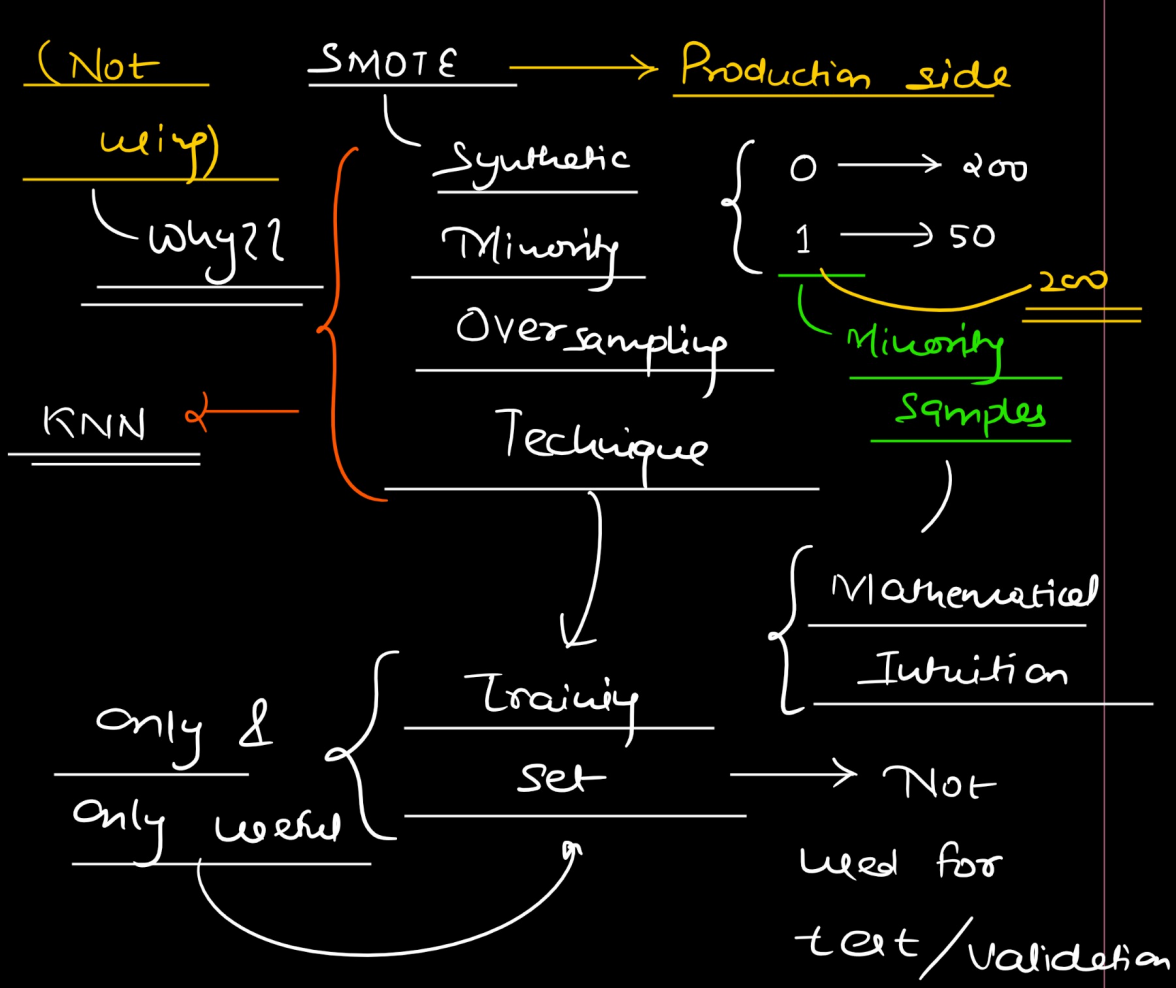
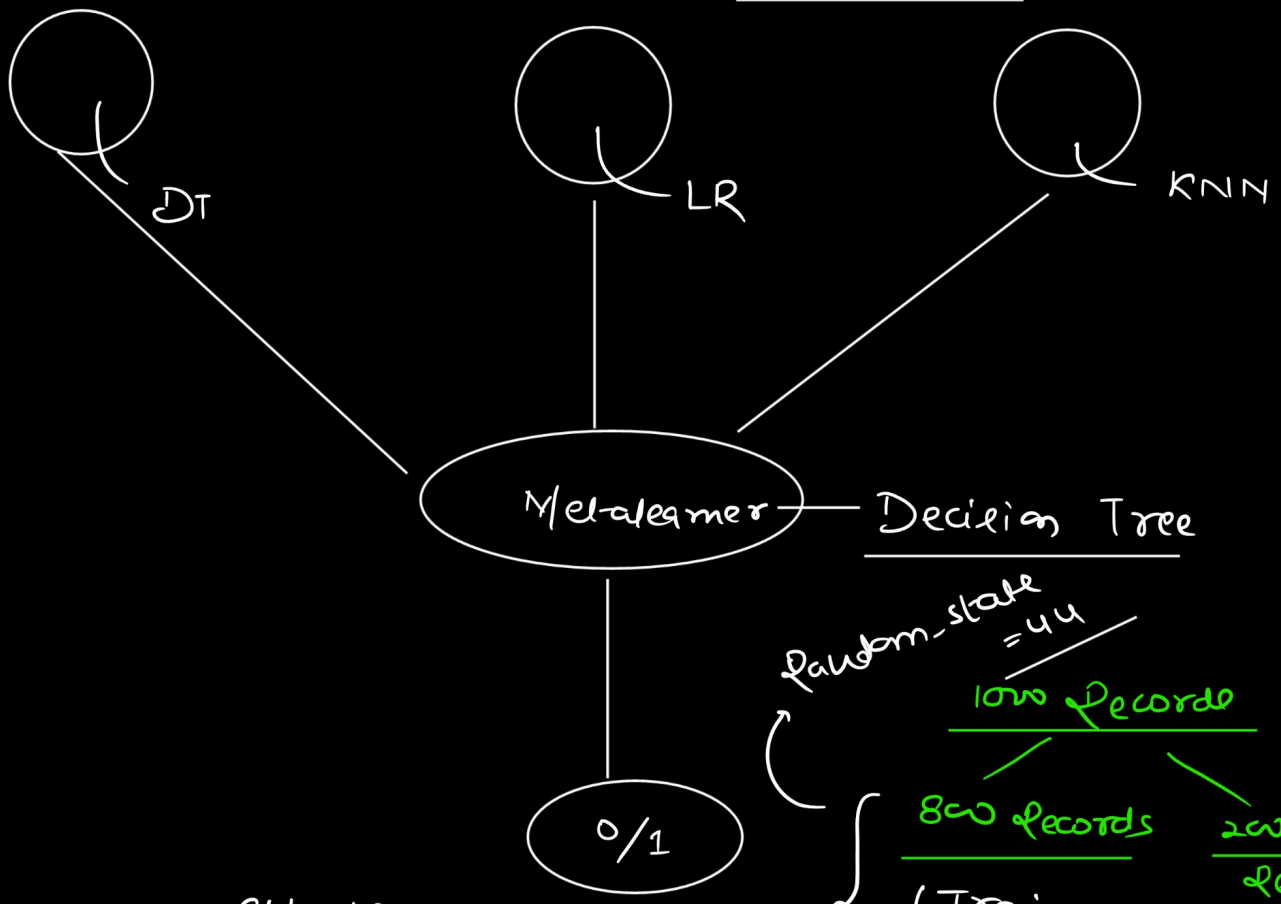

(Not
useful)
SMOTE
Production side
Why??
Synthetic
Minority
Oversampling
Technique
KNN
0 → 200
1 → 50
200
Minority
Samples
Mathematical
Intuition
Training
Set
Only &
Only useful
Not
used for
test/Validation

Base - learner

↳ Classification



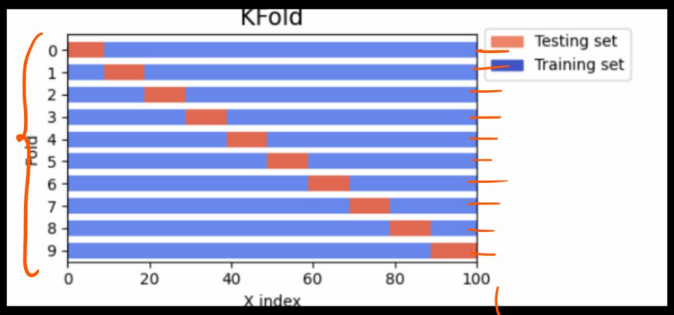
random_state = 44

1000 Records

800 Records (Train Set)

200 Records (Test Set)

CV = 10



CV = 10 — K = 10

10 - fold

Average → final evaluation metric

1000 Samples

CV = 5

20 samples — test set

20 - 40 samples — test set

40 - 60 samples — test set

to avoid the problem of overfitting

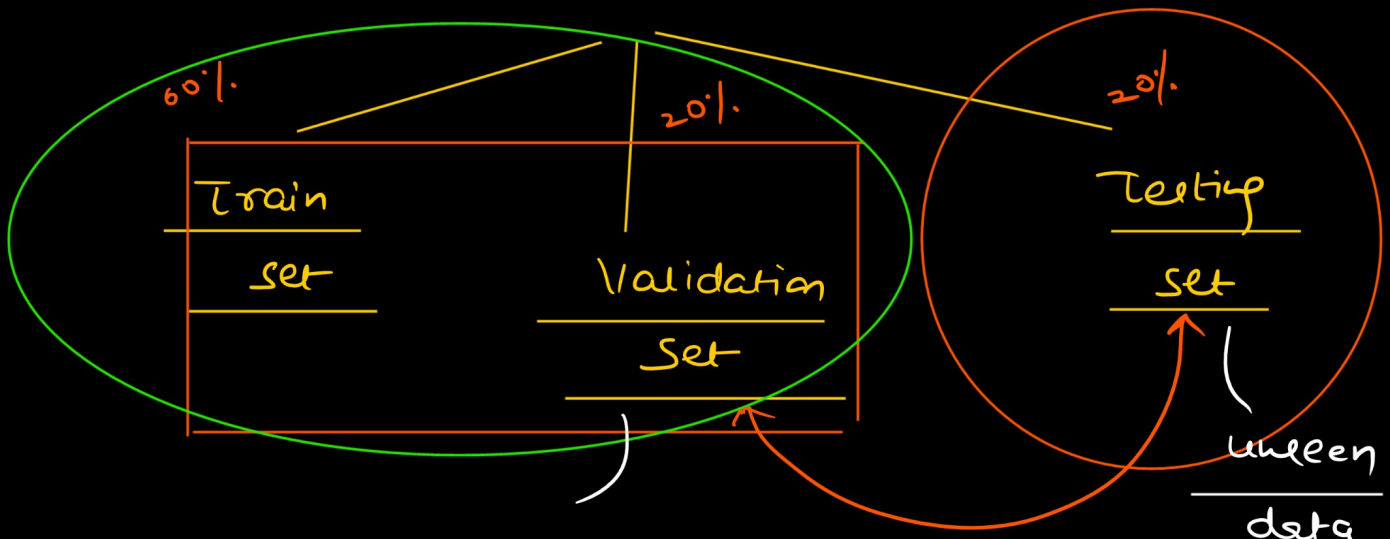
60 - 80 samples - test set

80 - 100 samples - test set

Imbalanced data

- 0 ⟶ 80%
- 1 ⟶ 20%

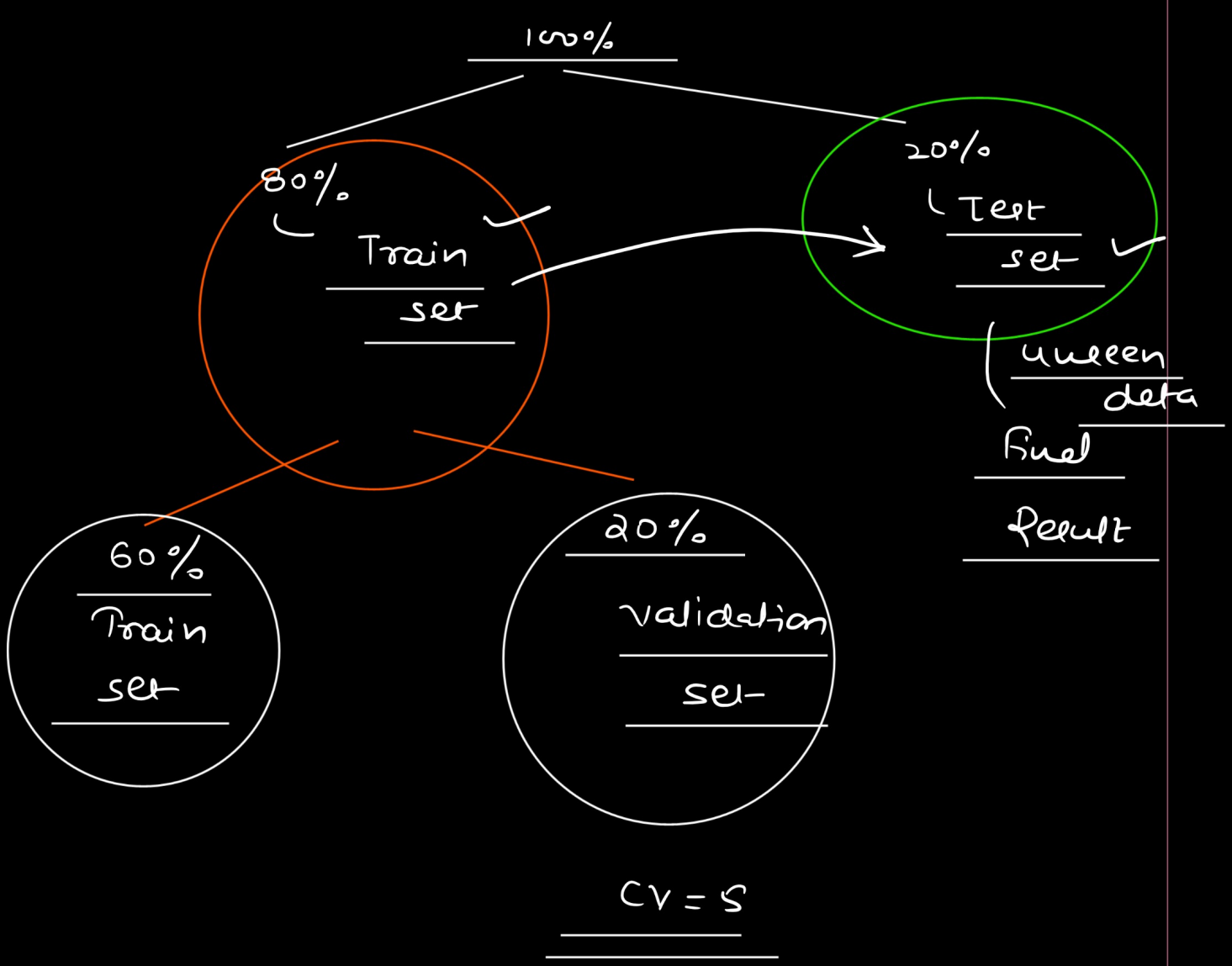
↳ Stratified K-fold



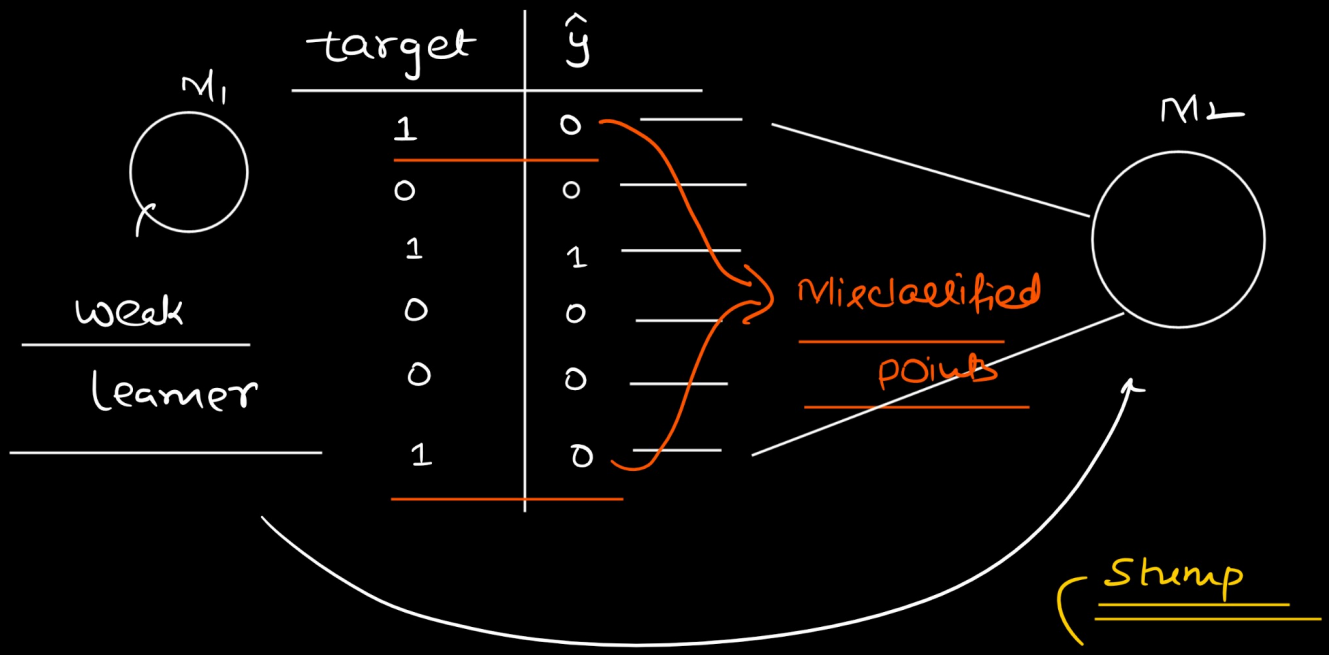
to set the best hyperparameters

(seen by the training set indirectly)

how the model actually behaves


100%
80%
Train set
20%
Test set
(unseen data
final Result
60%
Train set
20%
validation set
CV = 5

Boosting → Weak Learners



1st iteration → M_1 → Decision Tree — max_depth = 1

AdaBoost (Adaptive Boosting)

f_1	f_2	target	$\hat{y}$	weights = 1/7	updated weights	Normalized weights
		1	0	1/7 = 0.142	0.349 (increase)	0.349/0.697 = 0.5
		0	0	1/7	0.058 (decrease)	0.08
		1	1	1/7	0.058	0.08
		1	1	1/7	0.058	0.08
		0	0	1/7	0.058	0.08
		1	1	1/7	0.058	0.08
		1	1	1/7	0.058	0.08
				1	0.697	1

Total Error → Sum of weights of misclassified points

→ 1/7

$M_1 \rightarrow$ Performance of stump $\rightarrow \frac{1}{2} \log_e \left(\frac{1 - TE}{TE} \right)$

$\rightarrow \frac{1}{2} \log_e \left(\frac{1 - 1/7}{1/7} \right)$

$\rightarrow \frac{1}{2} \log_e 6$

$\rightarrow 0.895$

(Q) New weights (Misclassified point) $\rightarrow$ previous weight $* e^{\text{perfo}}$

$\rightarrow 1/7 * e^{0.895}$

$\rightarrow 0.349$

New weight (Correctly classified points) $\rightarrow$ previous weight $* e^{-\text{perf.}}$

$\rightarrow 1/7 * e^{-0.895}$

$\rightarrow 0.0583$

2nd iteration → M2

Random (Dataset creation

f1 f2

Weights Normalized

Range (Buckets)

0.349/0.697 = 0.5 → 0 – 0.5

0.08 → 0.5 – 0.58

0.08 → 0.58 – 0.66

0.08 → 0.66 – 0.75

0.08 → 0.75 – 0.83

0.08 → 0.83 – 0.92

0.08 → 0.92 – 1

Range is huge (Probability of appearing 1st record is higher)

misclassified points

1 record → frequency is higher