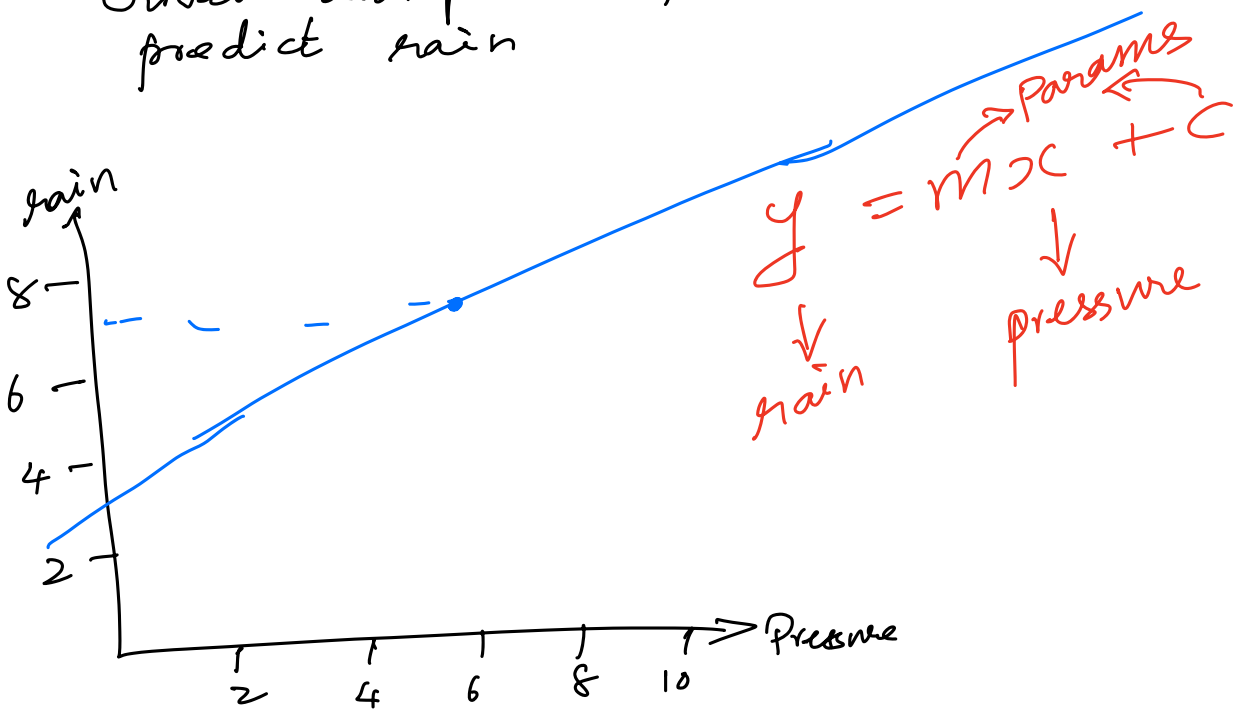
ML MODELS BASICS

Given air pressure, build a model to predict rain



STEPS TO BUILDING THIS MODEL:

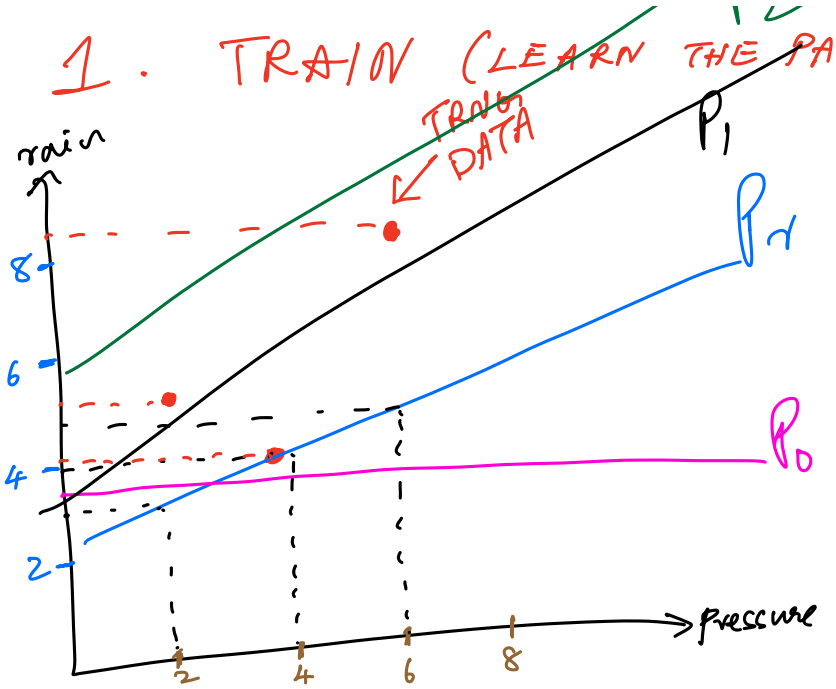
1) DATA SET PREP (TRAINING)
2) TRAIN (LEARN THE PARAMS) THE MODEL
3) TEST MODEL
4) DEPLOY & PREDICT WITH UNSEEN NEW DATA

DATA SET PREP

	Pressure	Rain
1-1-15	2	3
2-1-15	3	8
⋮	⋮	⋮
1-1-23	5	11

Pg.

1. TRAIN (LEARN THE PARAMS) MODEL



TRAINING DATA SET

Pressure	RAIN
2	5
4	4
6	8

	P_2	P_0	P_1	P_2
m	17	20	21	22
c	45	30	25	21
MAE	1.6	1	0.3	0.7
MSE			WINNER	

Pressure → 2 4 6

exp 5 4 8

Pred (model) 3 4 5

ABS Error 2 0 3

4 0 9

AVG $= \frac{13}{3} = 4.3$

Pressure → 2 4 6

exp 5 4 8

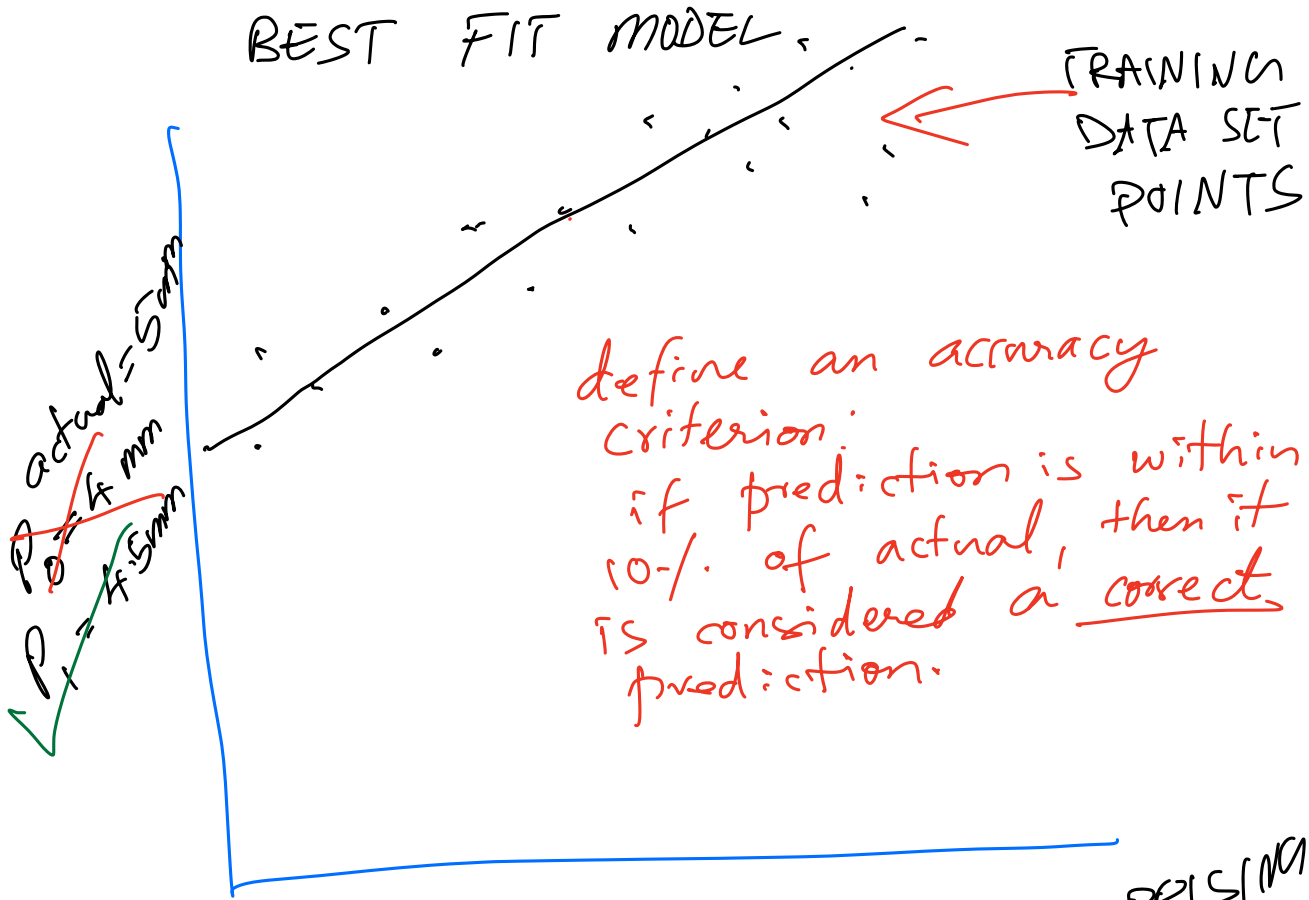
Pred 4 4 2

ABS Error 1 0 2

AVG $= \frac{3}{3} = 1$

DECLARE P_1 as WINNER

$y = 21x + 25$



define an accuracy criterion:
if prediction is within 10-/. of actual, then it is considered a correct prediction.

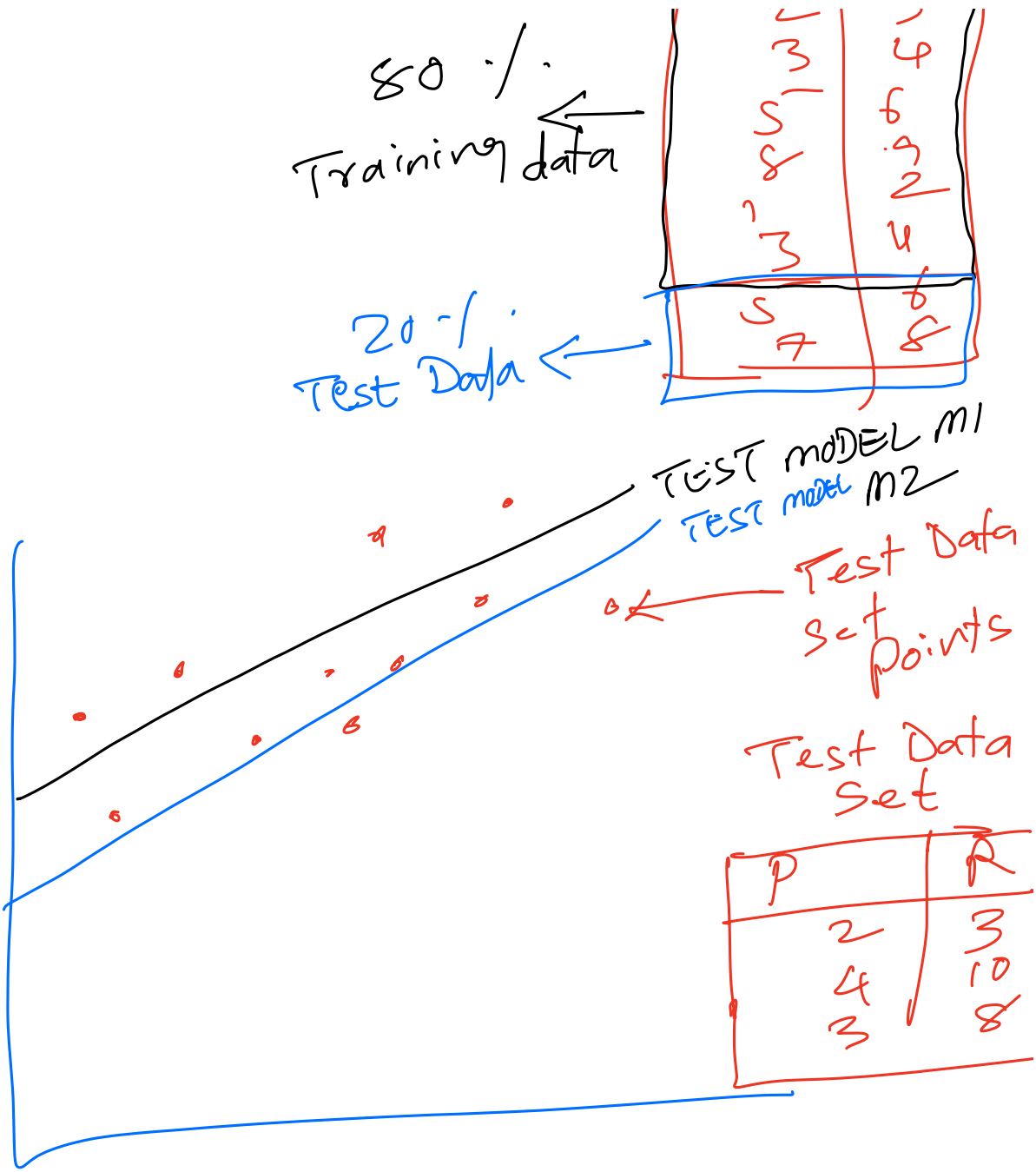
MODEL	TRAINING ACCURACY
m1	95-/.
m2	93-/.

→ NOT SURPRISING IT'S EXPECTED CHECK ACCURACY

2. TEST MODEL

HISTORICAL DATA

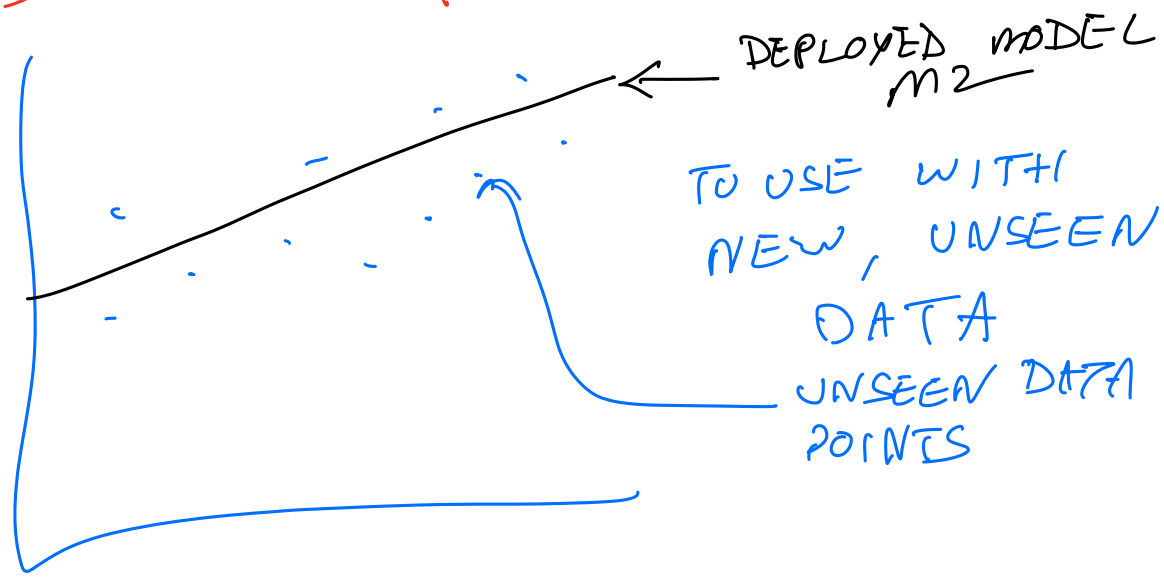
Pressure	Rain
?	?



MODEL	TEST ACCURACY	TRAINING ACCURACY
M1	75%	95%
M2	81%	93%

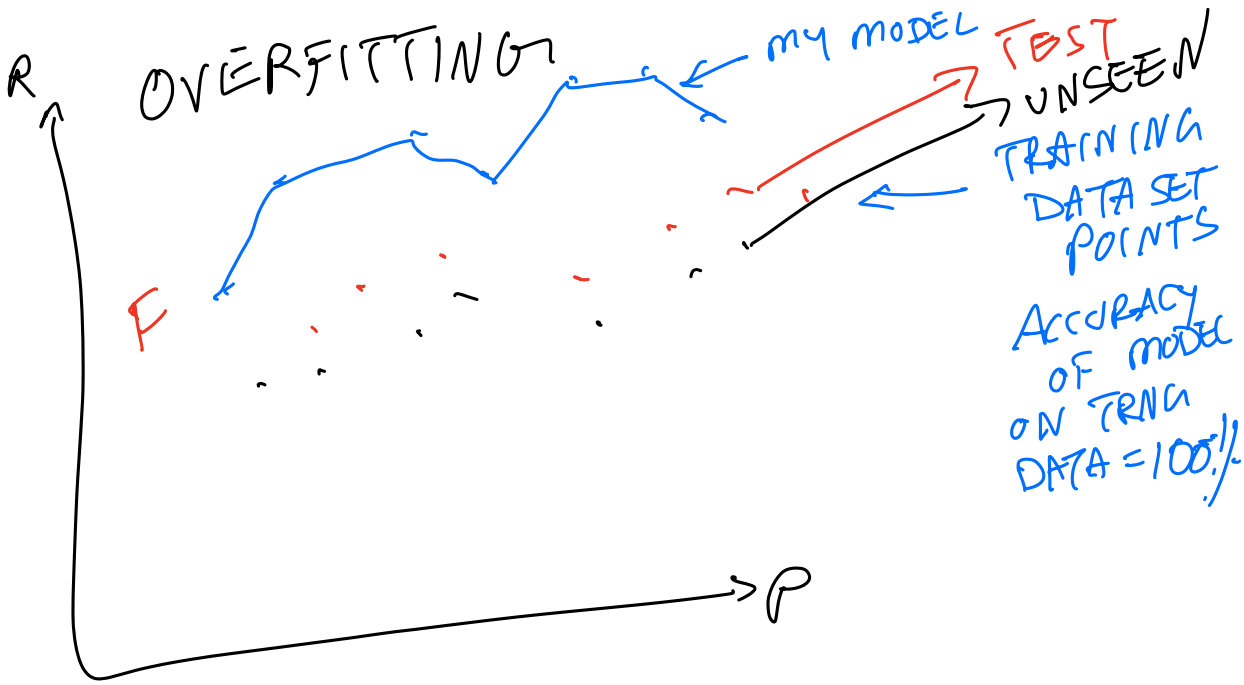
WINNER

3. DEPLOY & PREDICT



MODEL	TEST ACCURACY	TRAINING ACC	UNSEEN DATA ACC
M1	81.%	93.%	75%
M2	83%	95%	45%

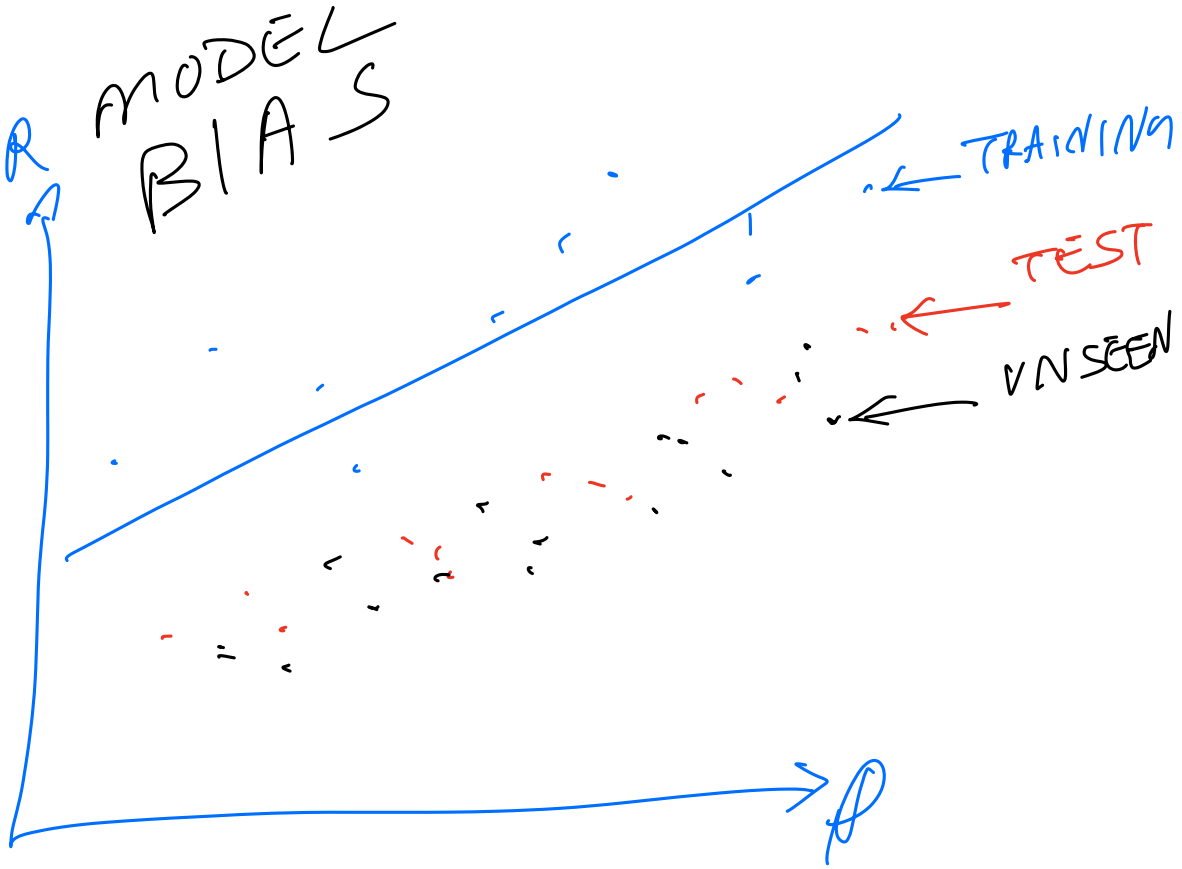
NOT OK



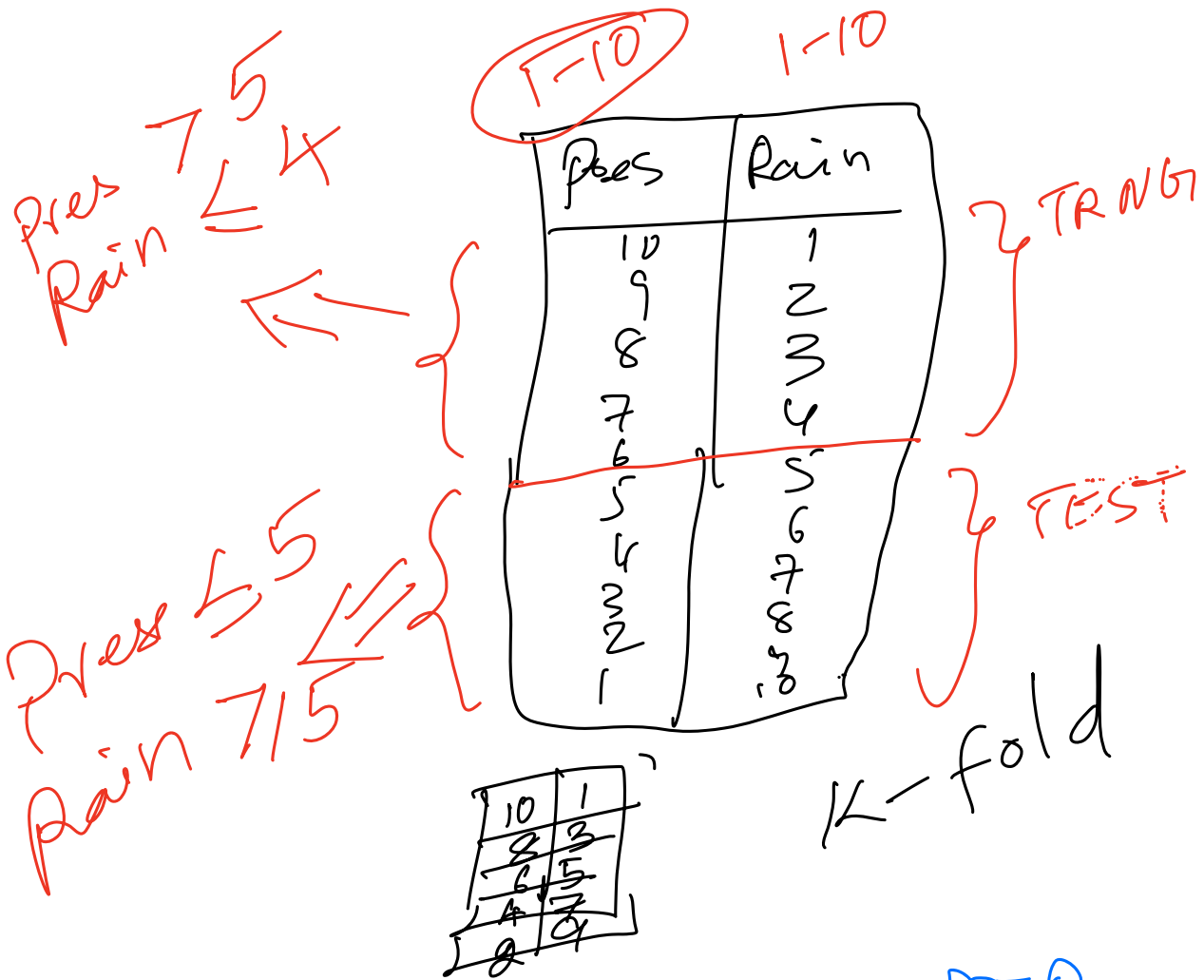
MODEL	TEST ACCURACY	TRAINING ACC	UNSEEN DATA ACC
F	70%	100%	65%
N	77%	87%	75%

WINNER

RED FLAG
HIGH RISK!



MODEL	TEST ACCURACY	TRAINING ACC	UNSEEN DATA ACC
F	45%	90%	50%



DATA PREP

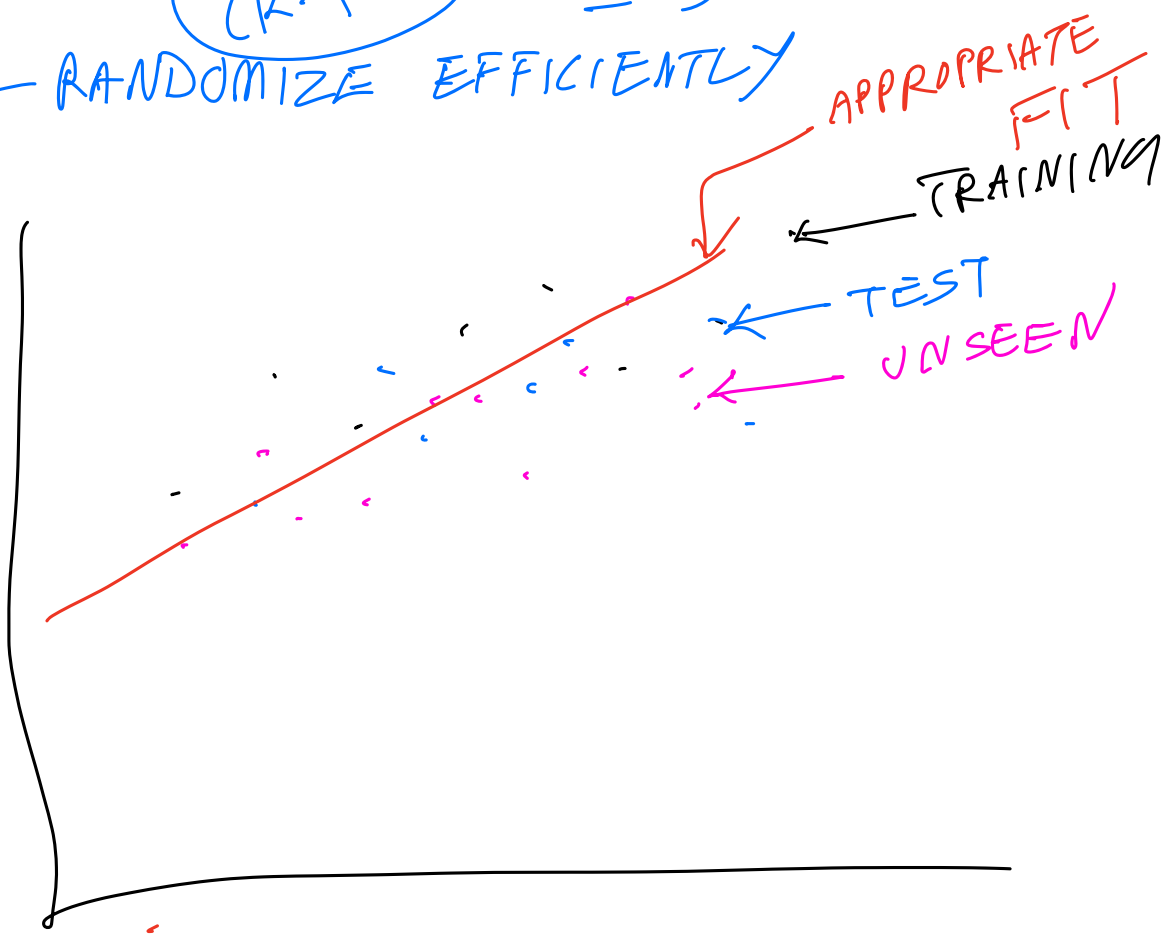
- CLEANSING (removing dups, NULL values)
- ELIMINATE UNNECESSARY (NOISY) FEATURES (e.g. the height of the table where the rain meter is placed)
- NORMALIZE

- SCALING (e.g. Rain is from 0 - 300 mm & Pressure 100 - 10,000)

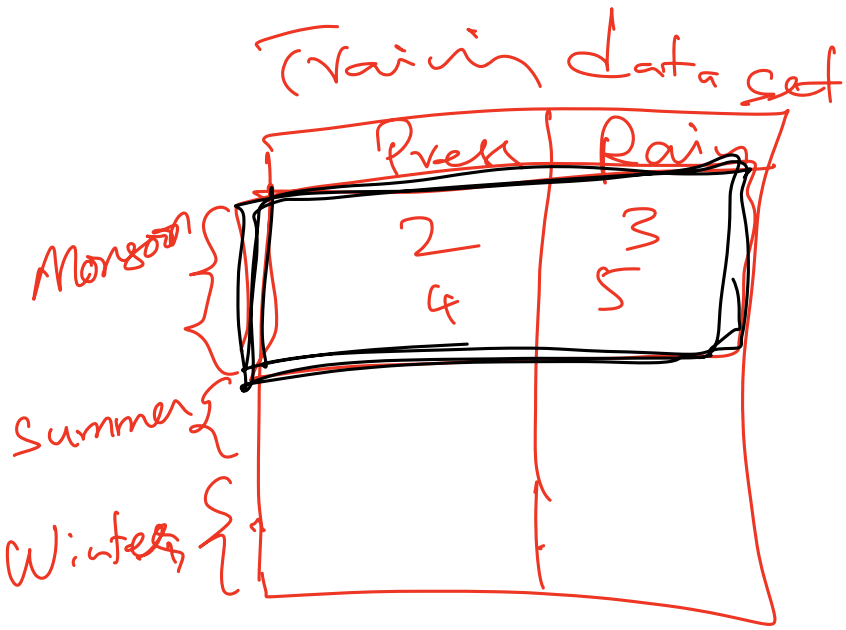
0 - 1 or
-1 - 1 or
-3 - +3

(K-fold)

- RANDOMIZE EFFICIENTLY



APPROPRIATE FIT	TRAIN ACC	TEST ACC	UNSEEN ACC
→	93.1 →	88.1 →	85.1


Training data set
Press
Rain
2
3
4
5
Monsoon
Summer
Winter