Target variable:

n = 1000

0 → 988

1 - 12

Discrete variable

(0, 1, 2...)

(0 - 1)

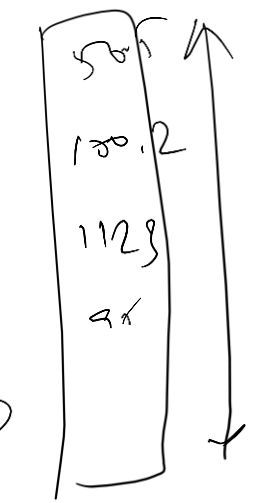
(Classification)

[Acc, CM, Class Report

Precision, Recall, F1-score]

Continuous variable (−∞ to ∞)

(Regression)



(Dis)

Classification

Logistic Regression

KNN classifier

NB classifier

D.T Classifier

RF "

(cont)

Regression

Linear Reg

KNN Regressor

D.T Regressor

RF "

(-∞ to ∞)

⇓

(0 to 1)

⇓

(0-1)

Log Reg

Bagging → Bootstrap Aggregating

→ Random subset of features

→ " " samples

→ " "

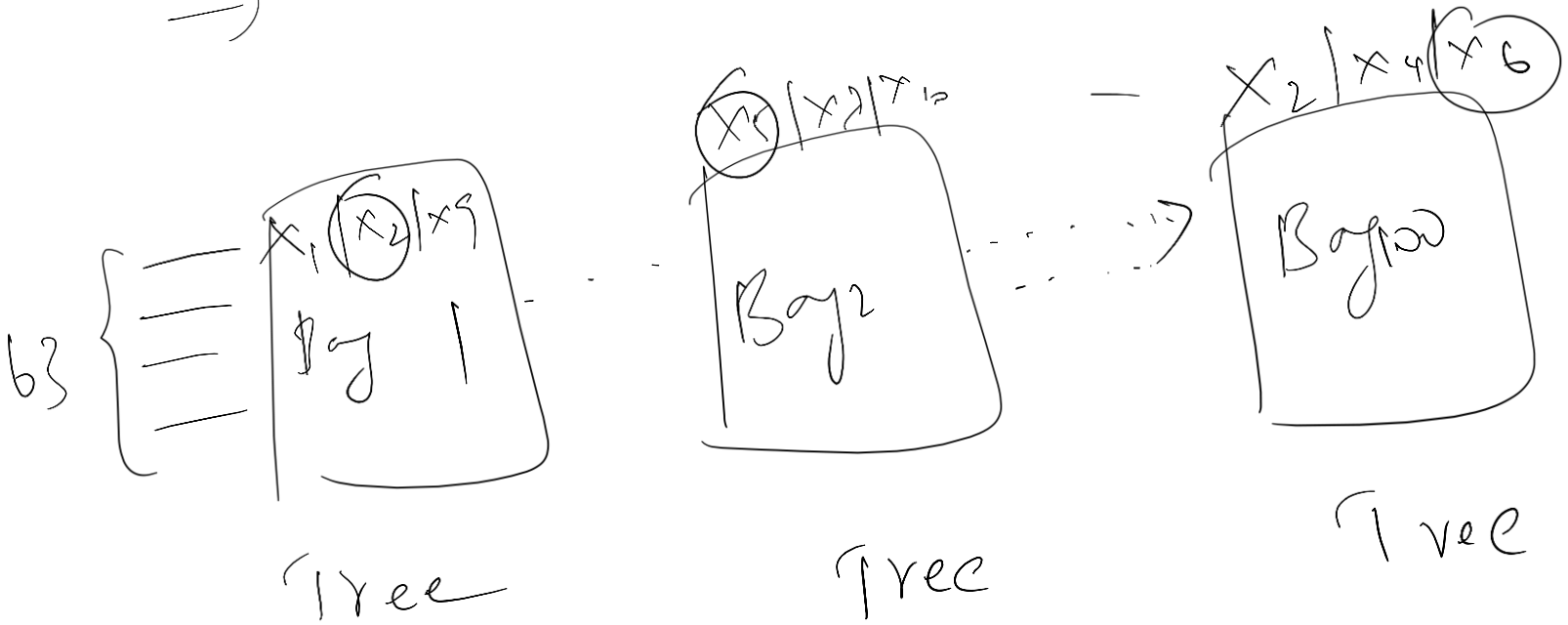
n = 100

(Training)

63%

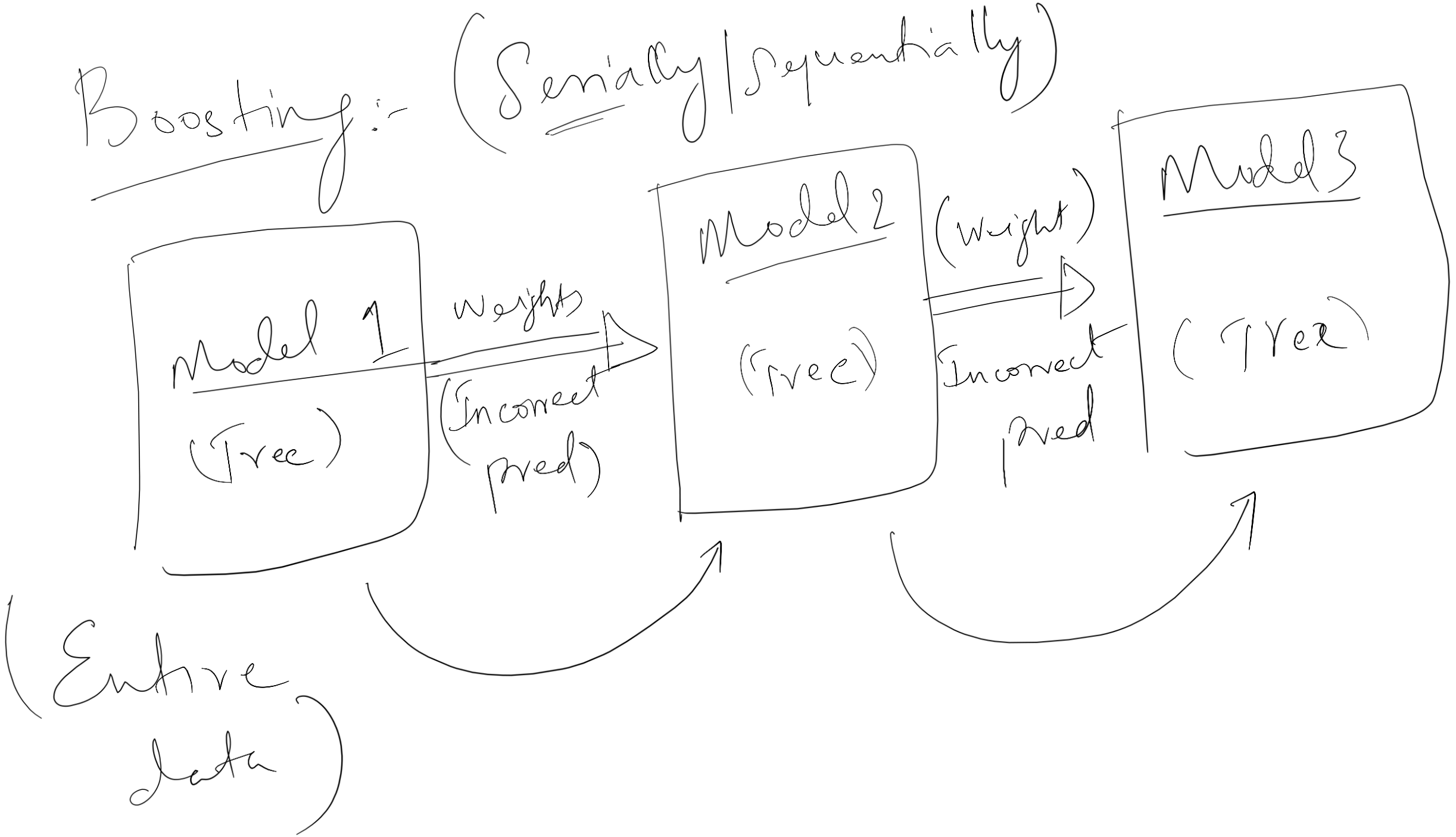
37%

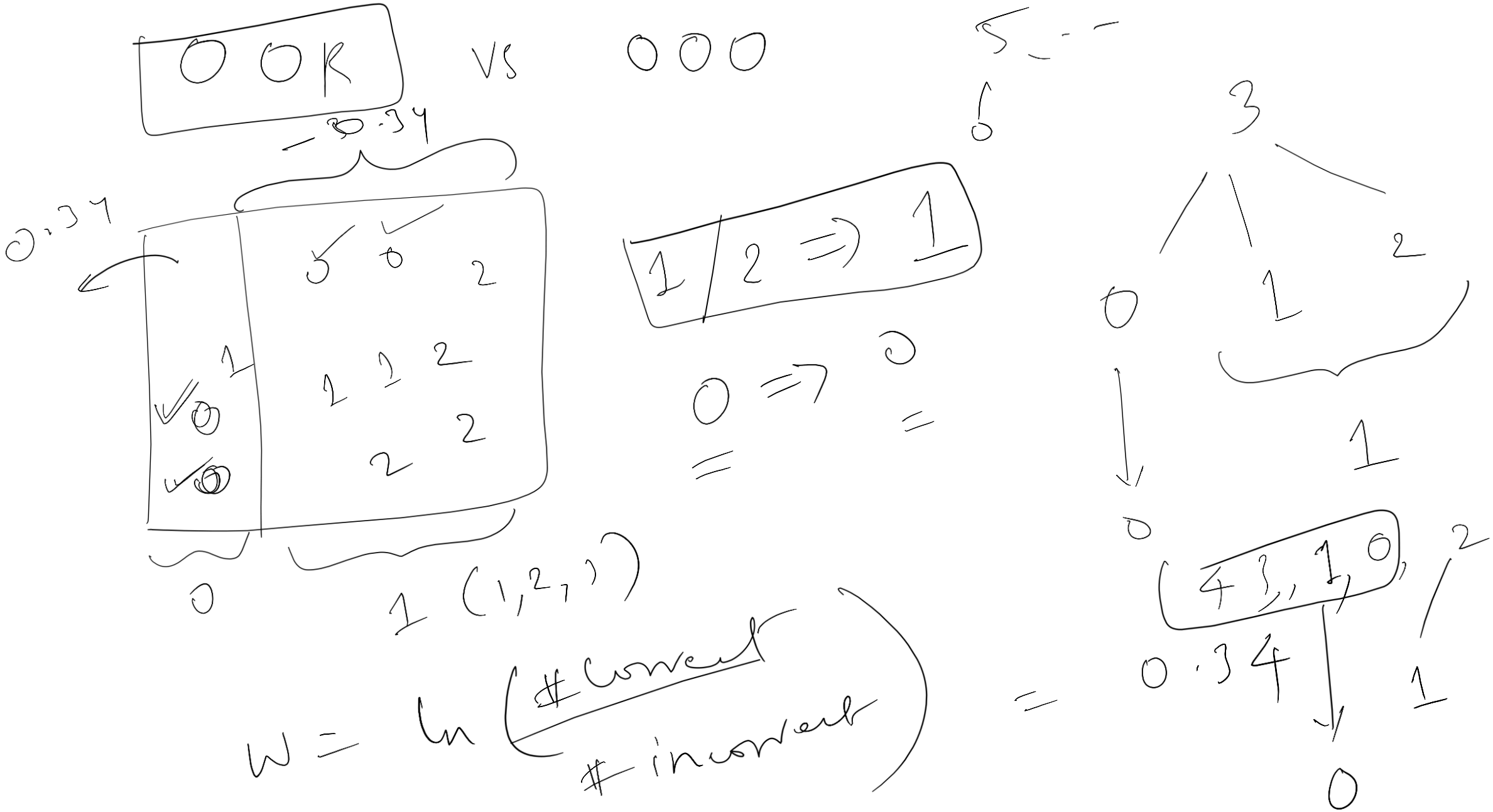
(oobscore)



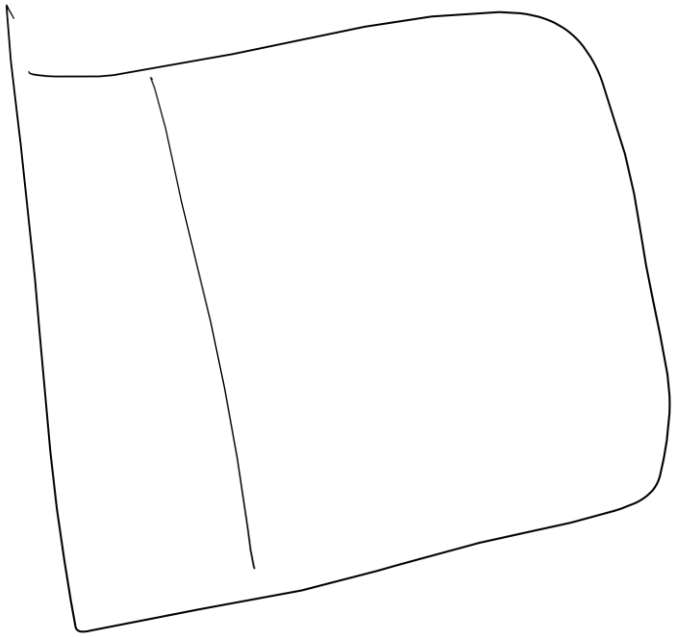
n = 10

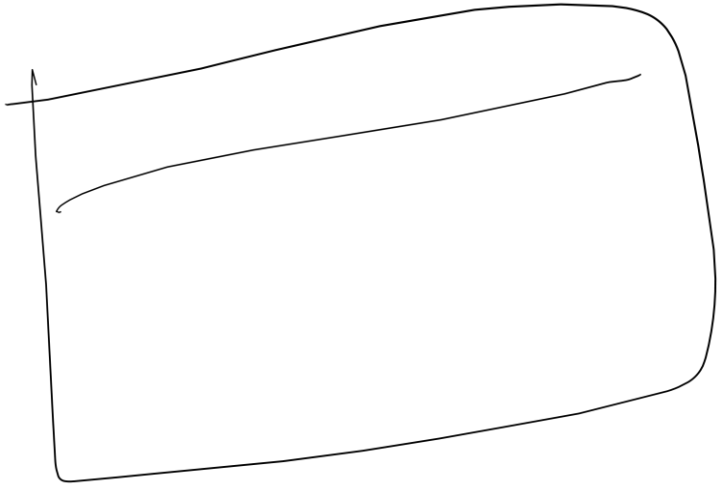
3

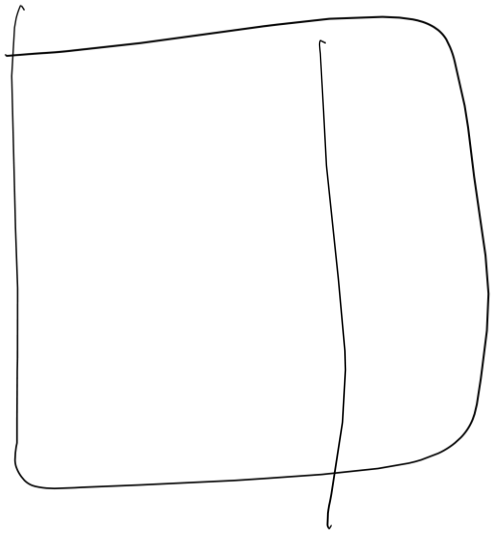

Boosting :-
(Serially | Sequentially)
Model 1
(Tree)
Weights
(Incorrect pred)
Model 2
(Tree)
(Weight)
Incorrect pred
Model 3
(Tree)
(Entire data)

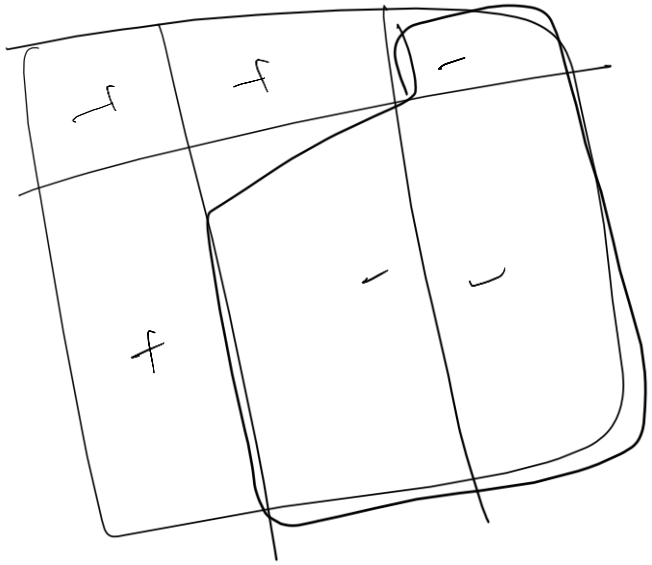


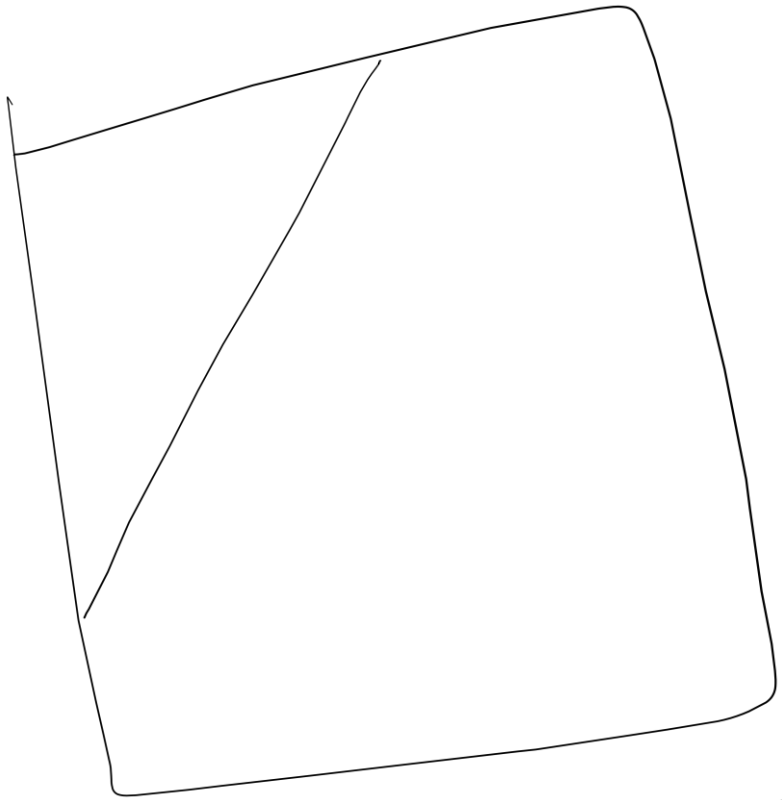
OOK vs OOO
0.34
0.34
1 2 ⇒ 1
0 ⇒ 0
1 (1,2,1)
W = ln (# correct / # incorrect)
= 0.34
4 3, 1, 0

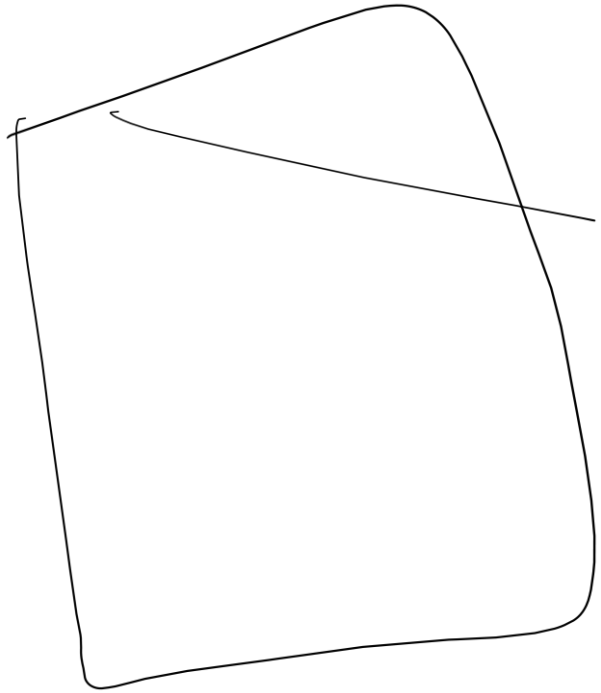


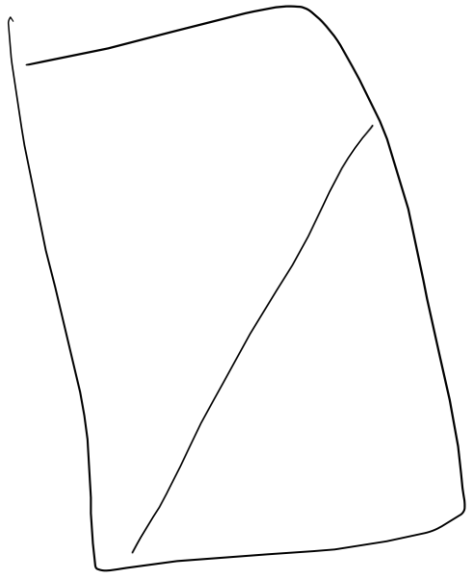


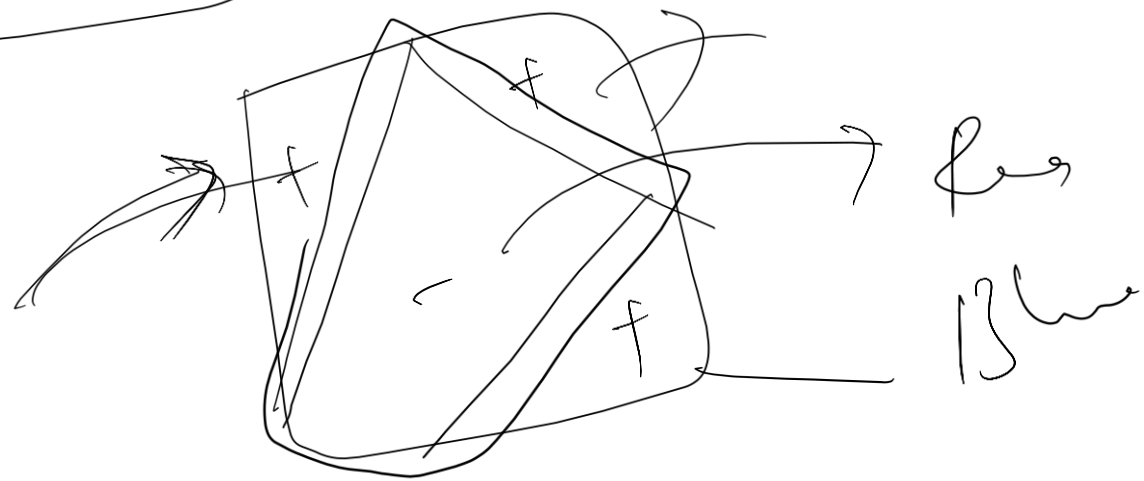


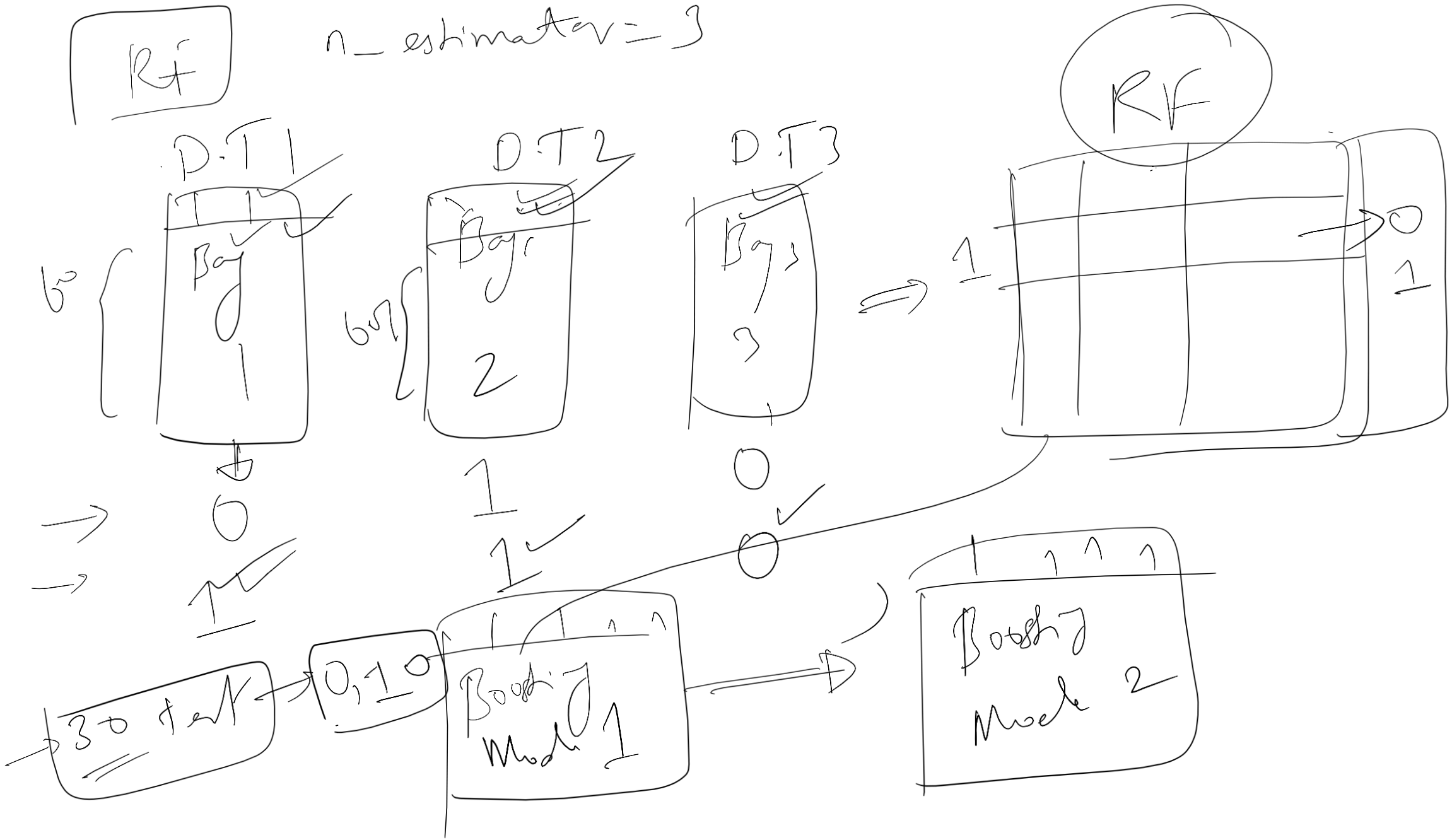


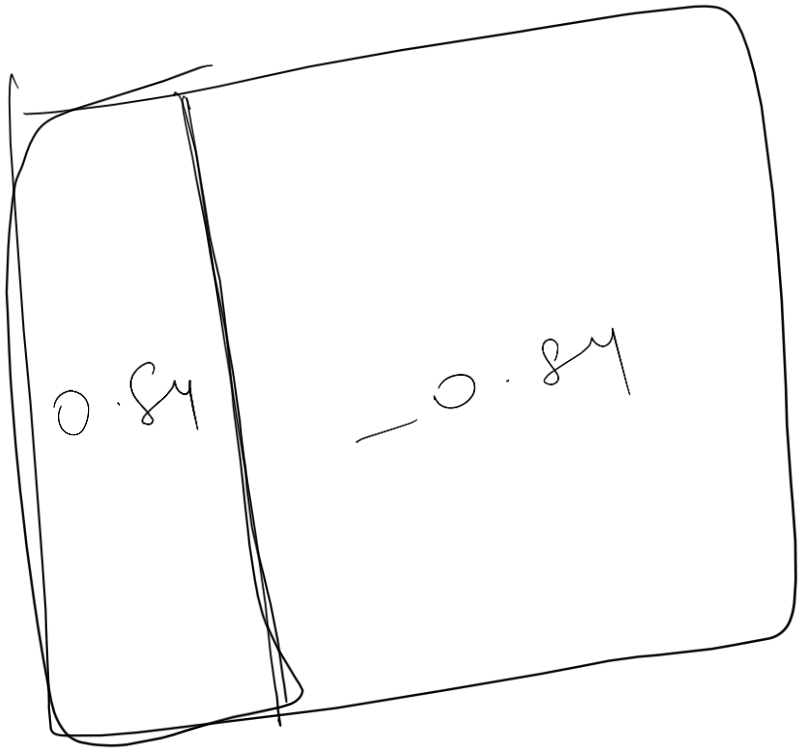


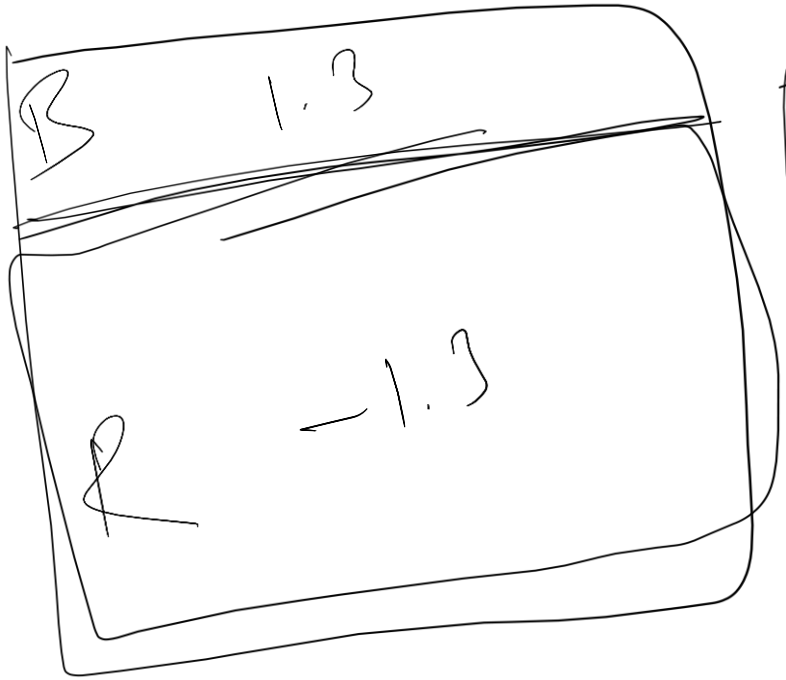


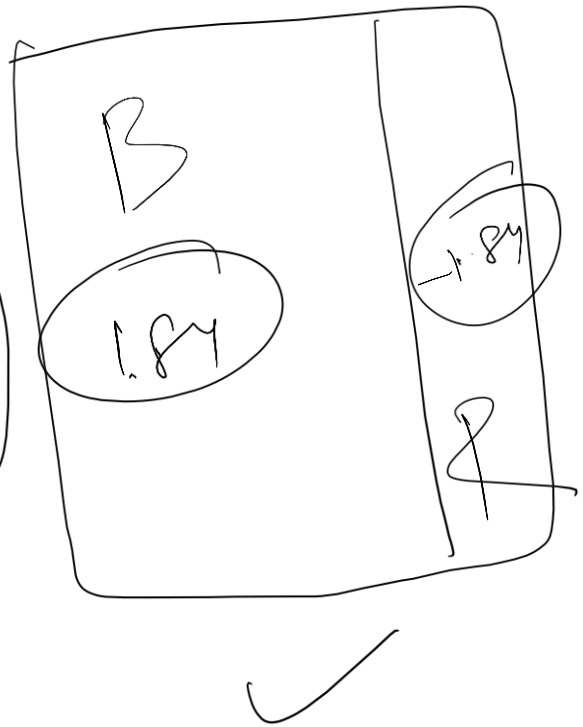



f
f
f
-

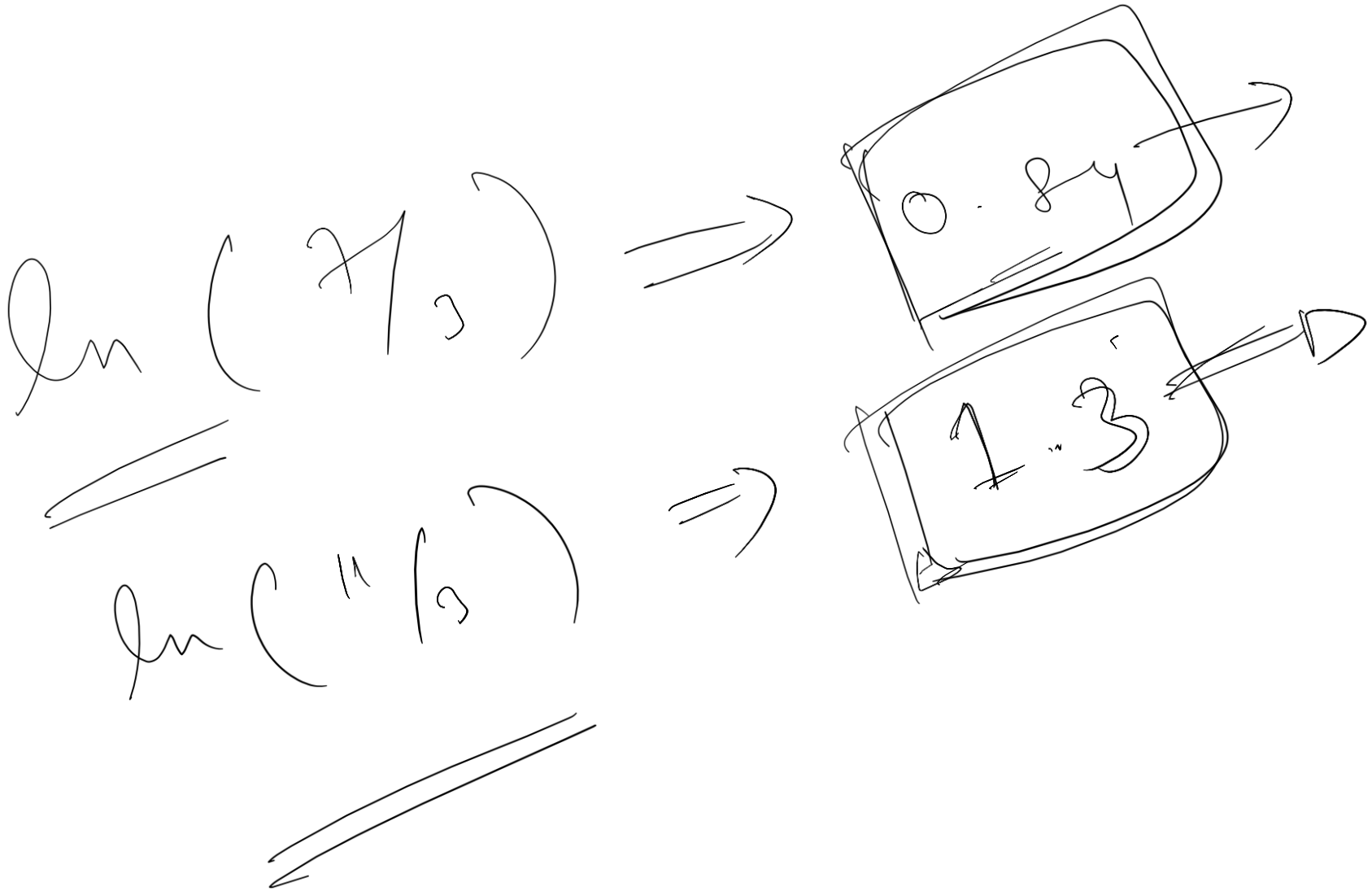

RF
n_estimator = 3
D.T 1
D.T 2
D.T 3
Bag 1
Bag 2
Bag 3
60
60
RF
1
0
1
0
1
0
1
1
0
30 test
0,1 0
Boosting Model 1
Boosting Model 2







1.84	1.84	−1.84
0.84 1.3	−0.84 1.3	−0.84 1.3
−1.3 0.84	−1.3 −0.84	−1.3 0.84



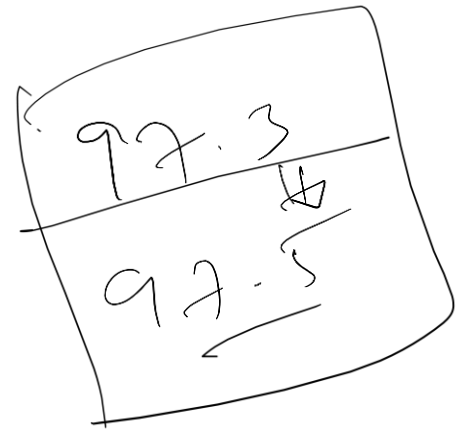
Voting: Classifier / Regressor

3 Models

1

2

3 AdaBoost



D.T

1 (5/6)

1

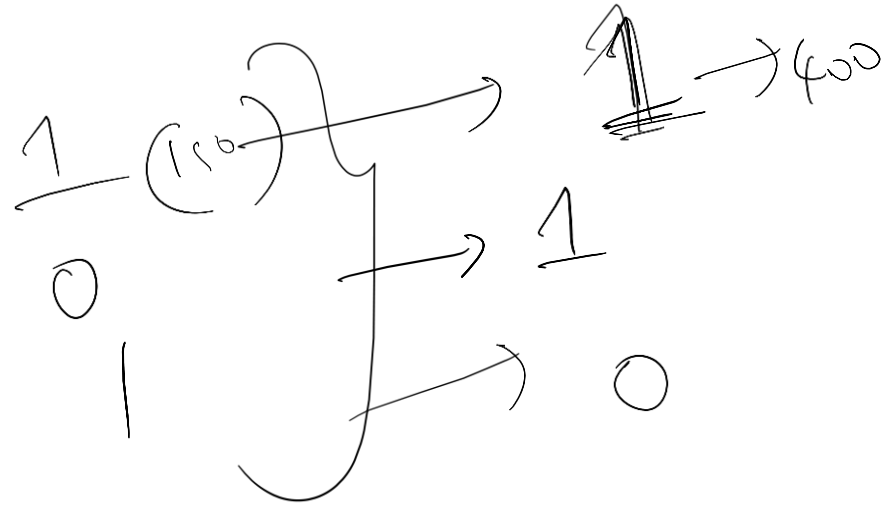
0

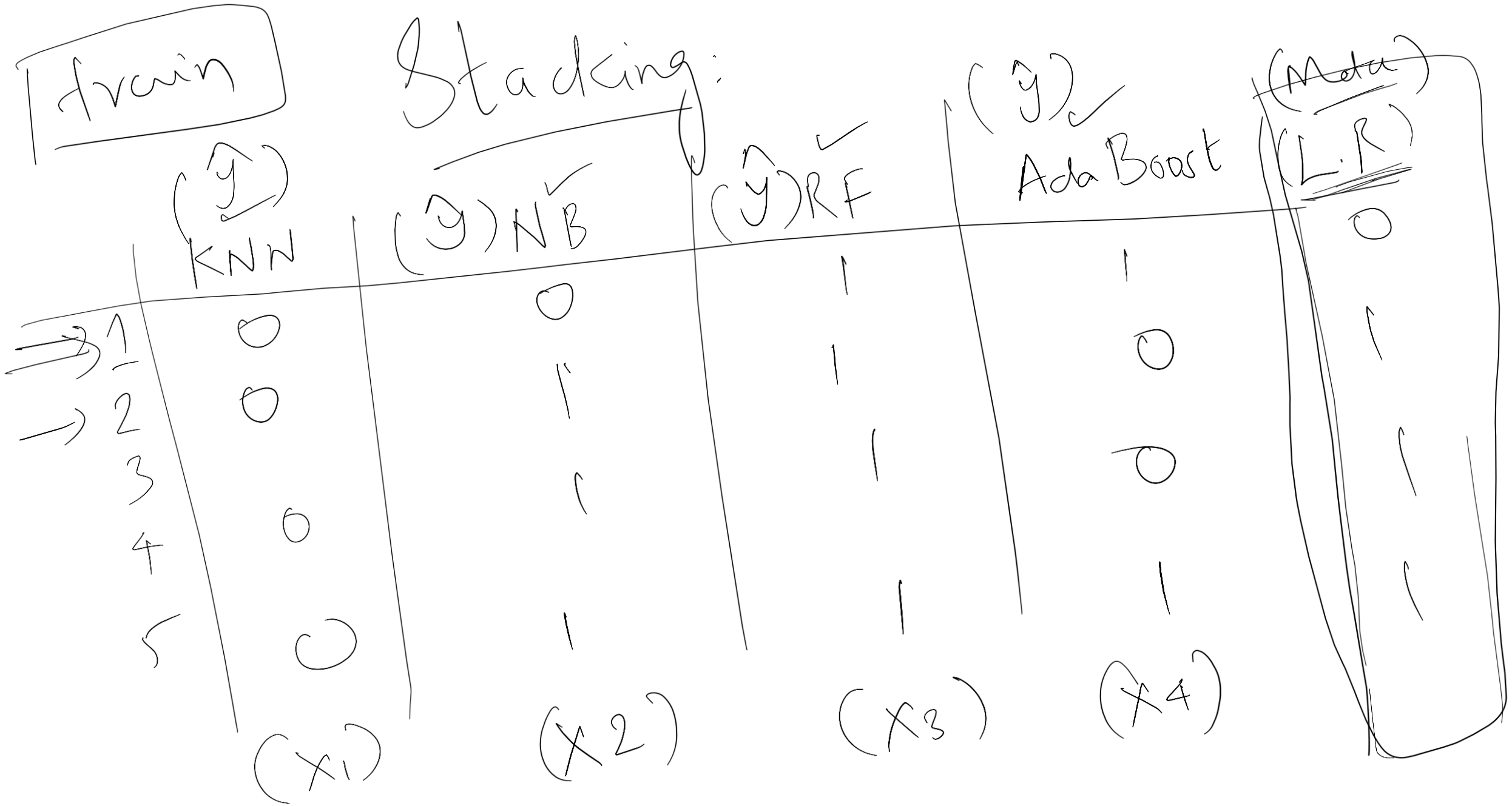
R.F

0 (3/6)

1

0





$\boxed{\text{train}}$

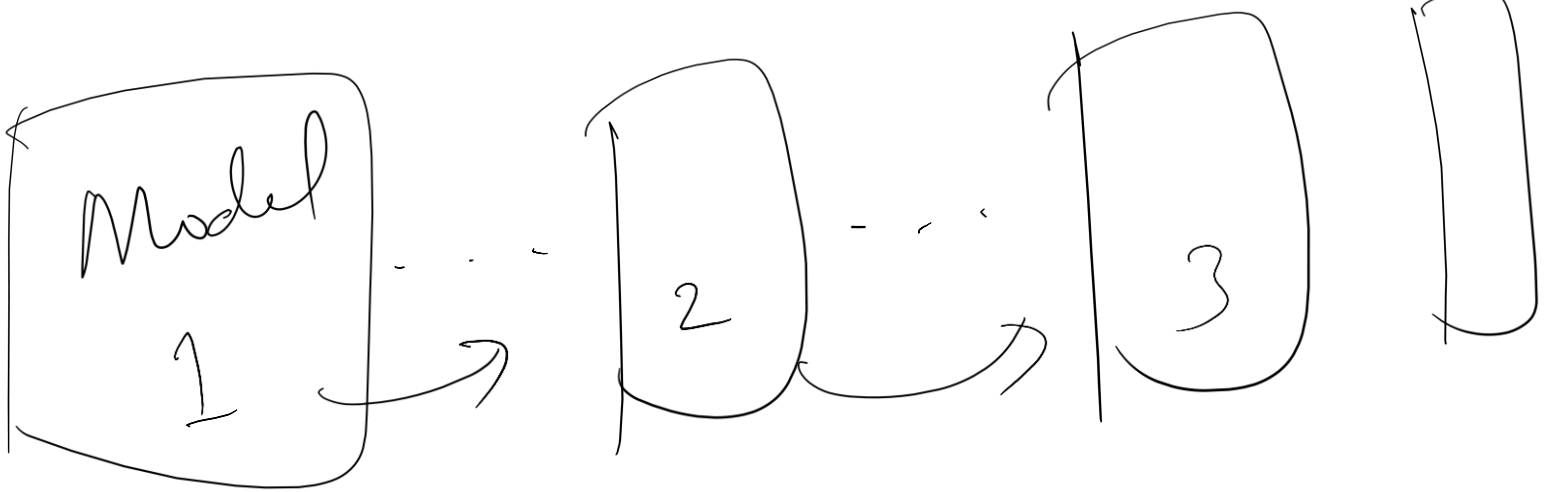
Stacking:

	$(\hat{y})$ KNN	$(\hat{y})$ NB	$(\hat{y})$ RF	$(\hat{y})$ AdaBoost	~~(Meta)~~ (L.R)
1	0	0	1	1	0
2	0	1	1	0	1
3		1	1	0	1
4	0				
5	0	1	1	1	1
	(x_1)	(x_2)	(x_3)	(x_4)	

Let's get back

at 7:50 P.M (I.S.T)

Ensemble :



Pred

Model

D.T

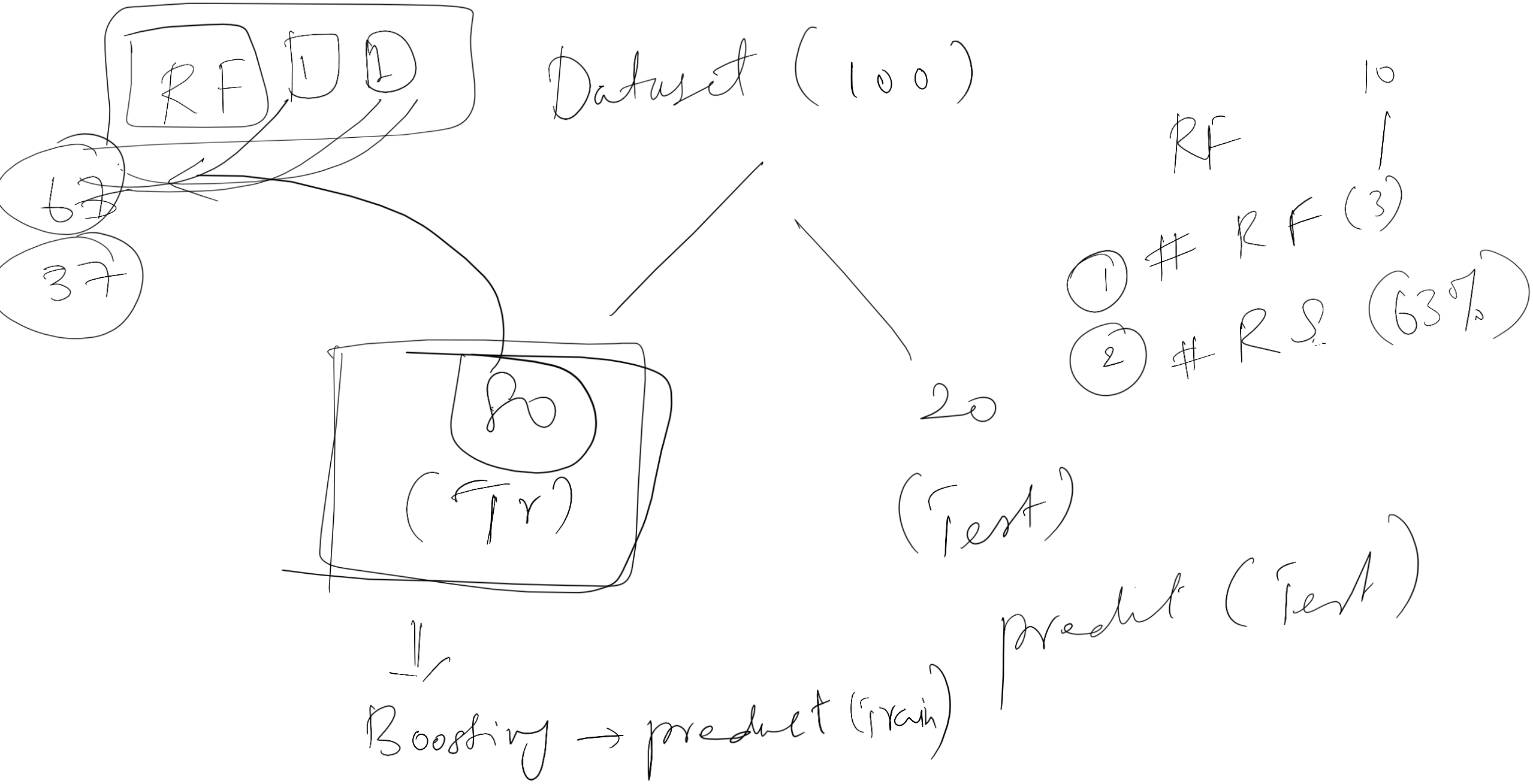
→ Bagging (RF)

→ Boosting (Ada, Xgboost ...)

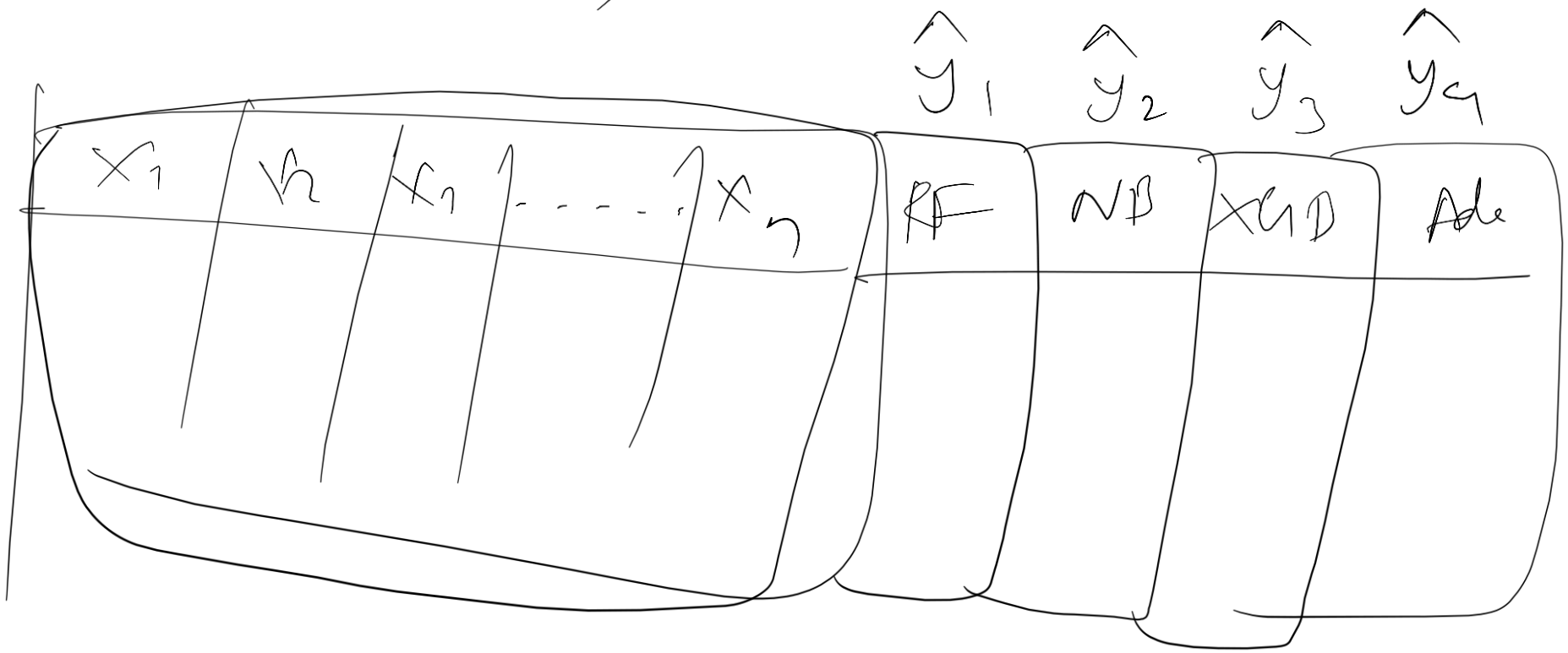
→ Voting → Votes of model pred

→ Stacking → Pred → I/p

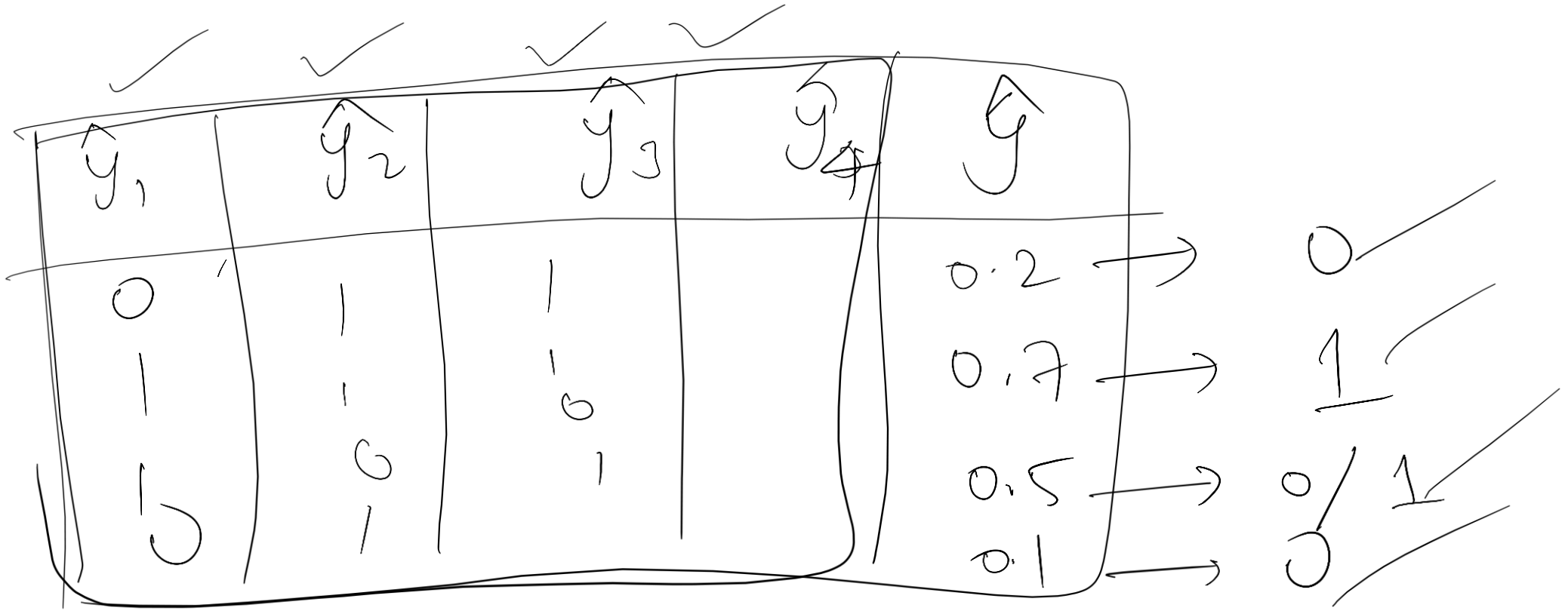
Meta $\xrightarrow{L.R}$ O/p


RF
1
1
63
37
80
(Tr)
Dataset (100)
20
(Test)
10
RF
1
RF (3)
1
RS (63%)
2
Boosting → predict (train)
predict (Test)

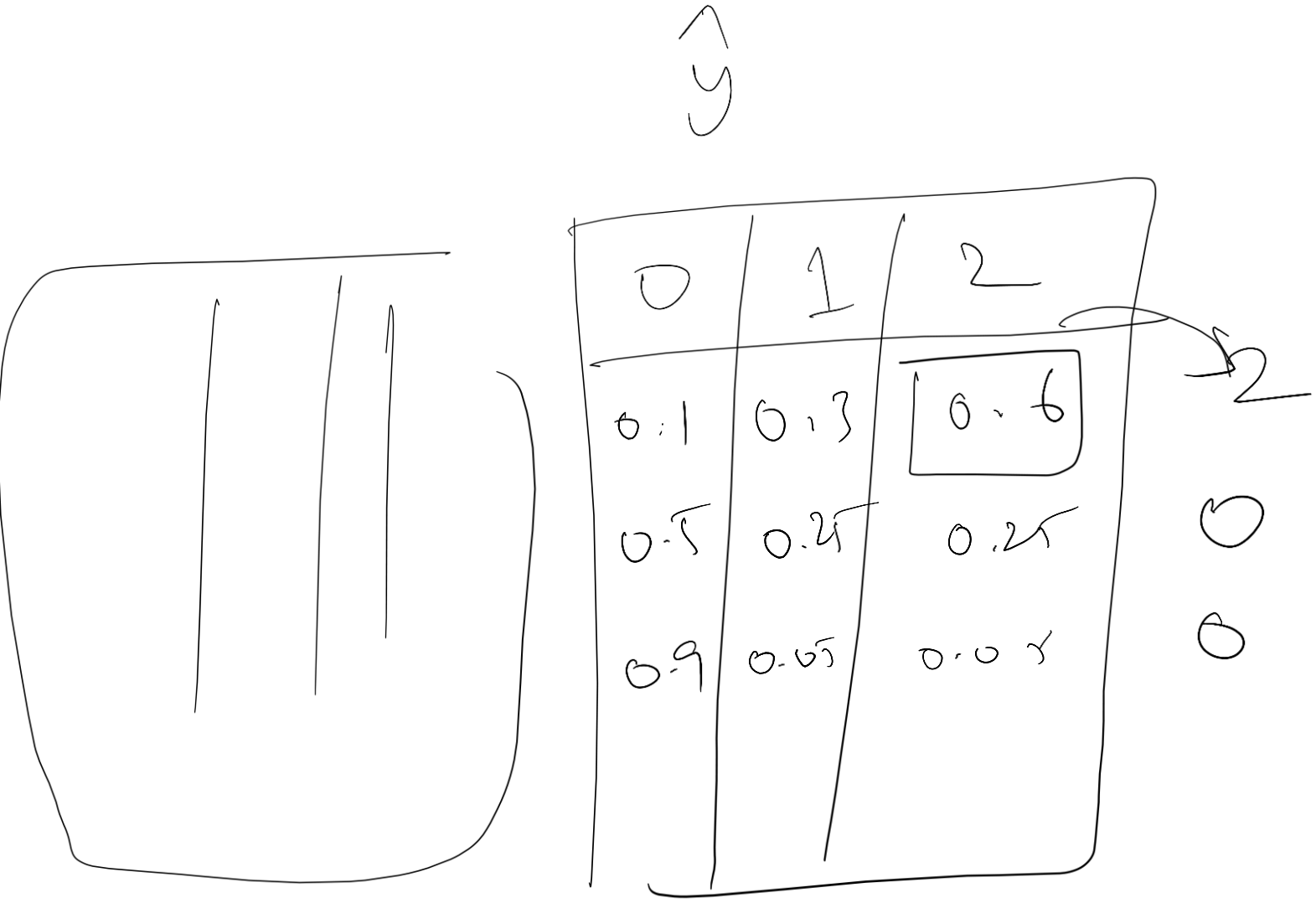
Model → RF, NB, XGB, Ada



0, 1, 2



Log Regression


ŷ
0
1
2
0.1
0.3
0.6
0.5
0.25
0.25
0.9
0.05
0.05
2
0
0

Br → 1 → Amazon

		Seventh
Br 1		L 1
Br 2		M 2
Br 3		H 3
Br 4		

0.7 ✓

0.1

0.1

0.1

4

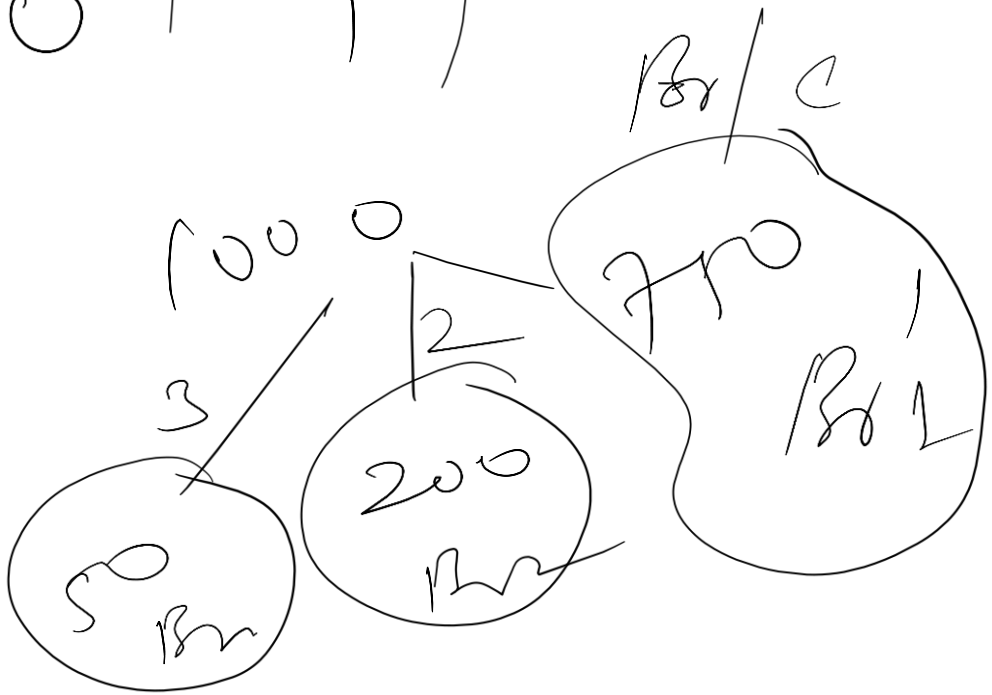
Brands

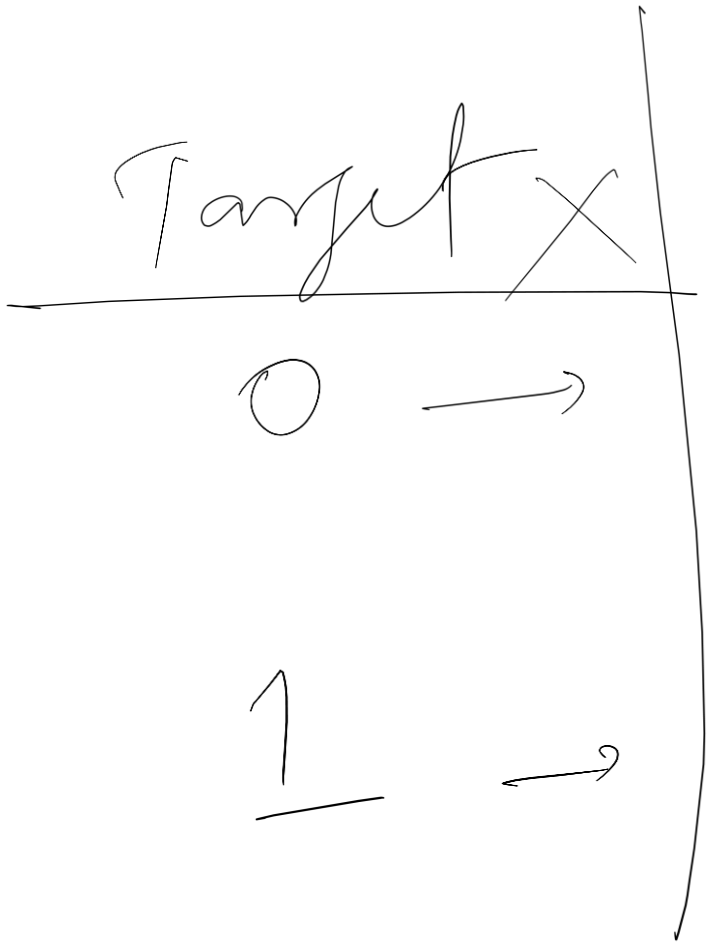
- Br 1 ⇒ 0.7
- Br 2 0.1
- Br 3 0.1
- Br 4 0.1

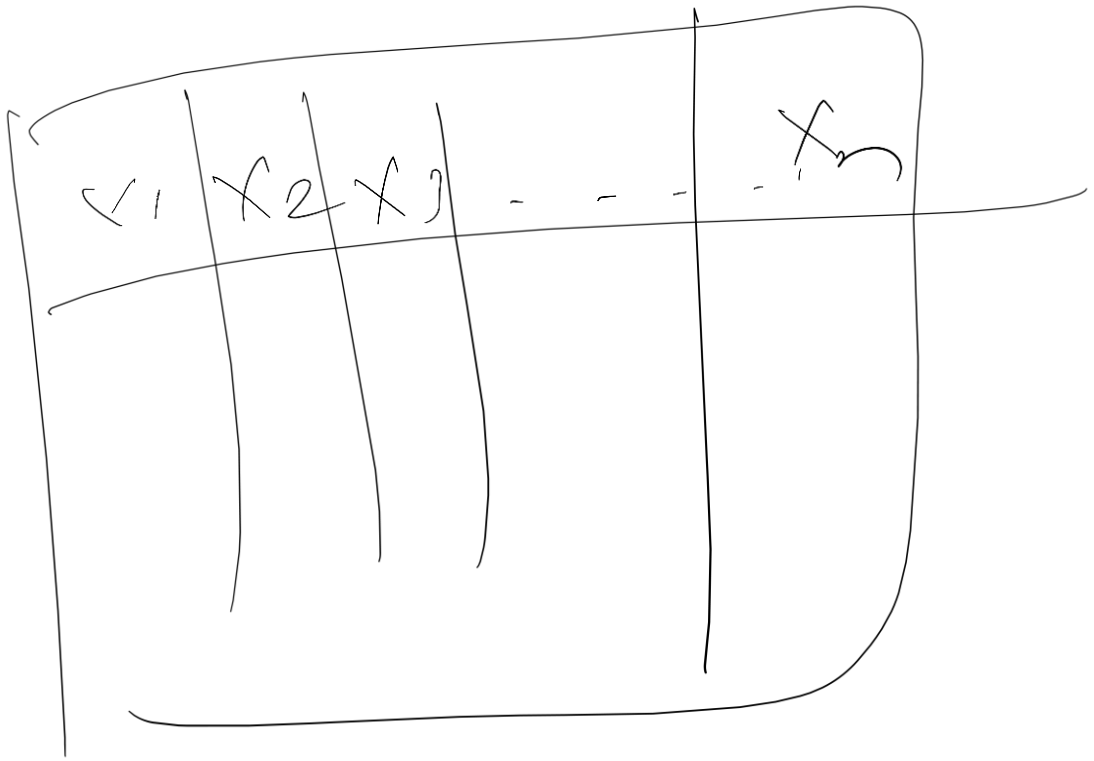
1000

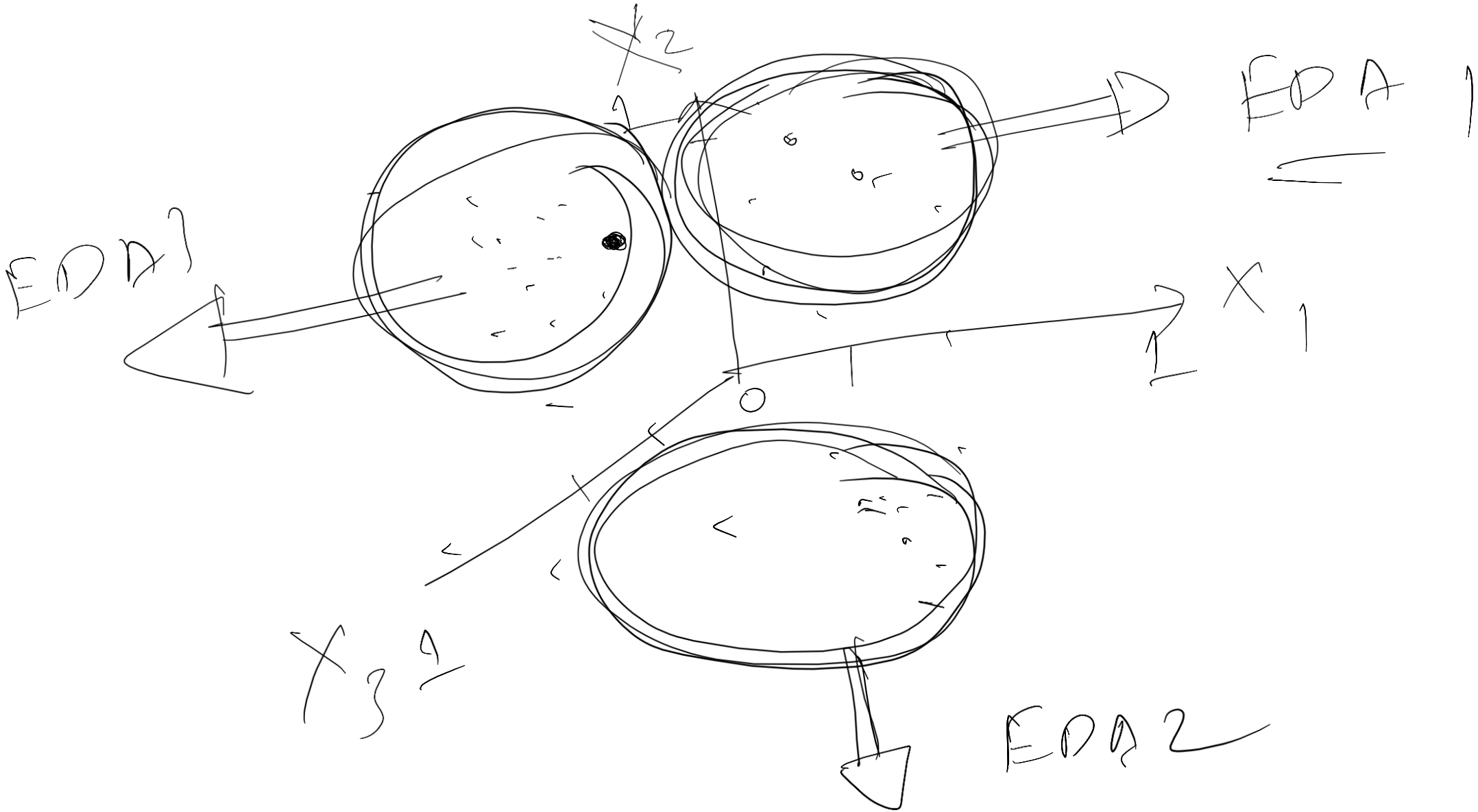
Brand	
Am	0.7
Fl	0.2
Zom	0.1

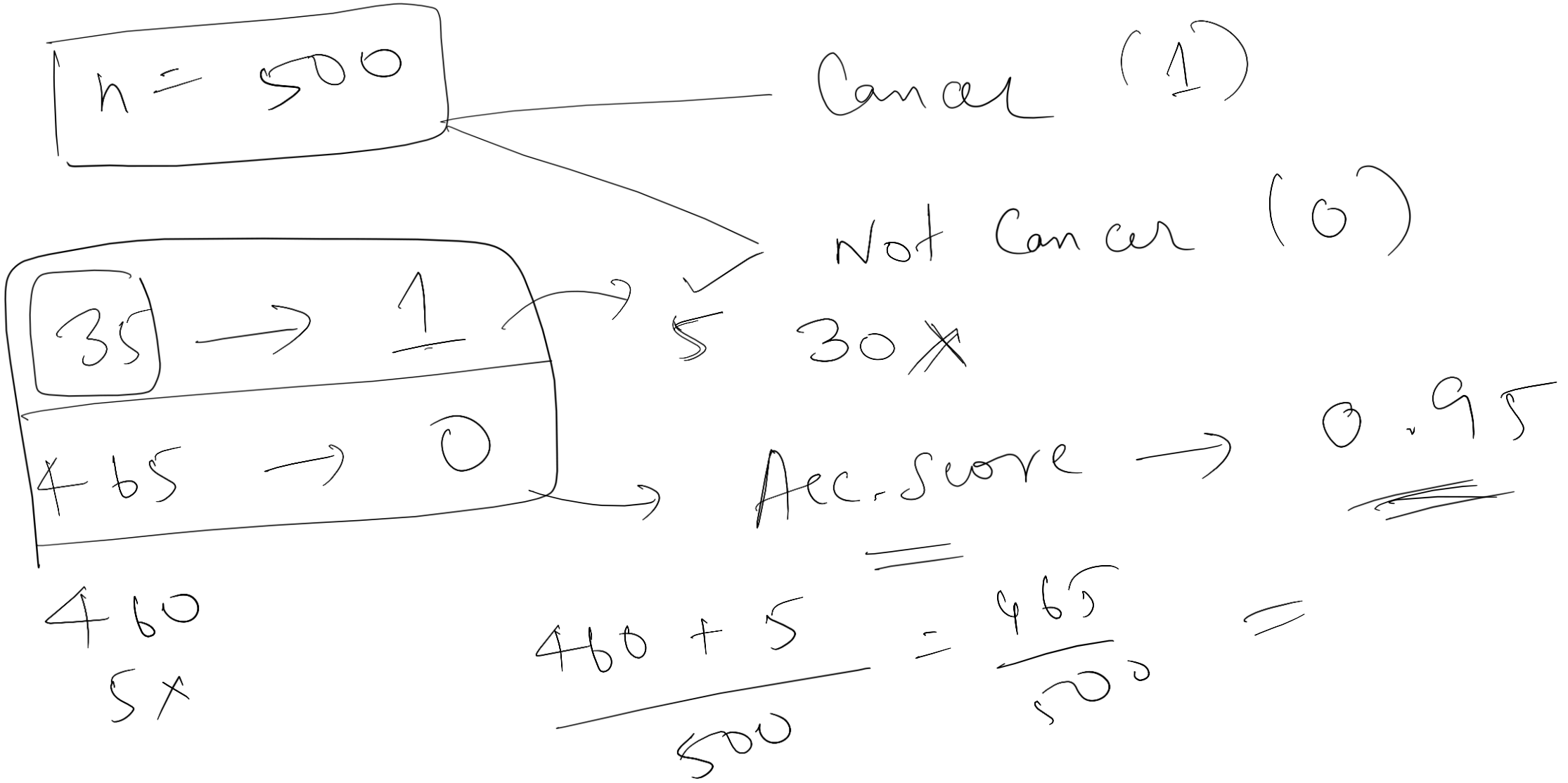
Am (1)	Fl (2)	Zom (3)
1	0	0
0	1	0
0	0	1



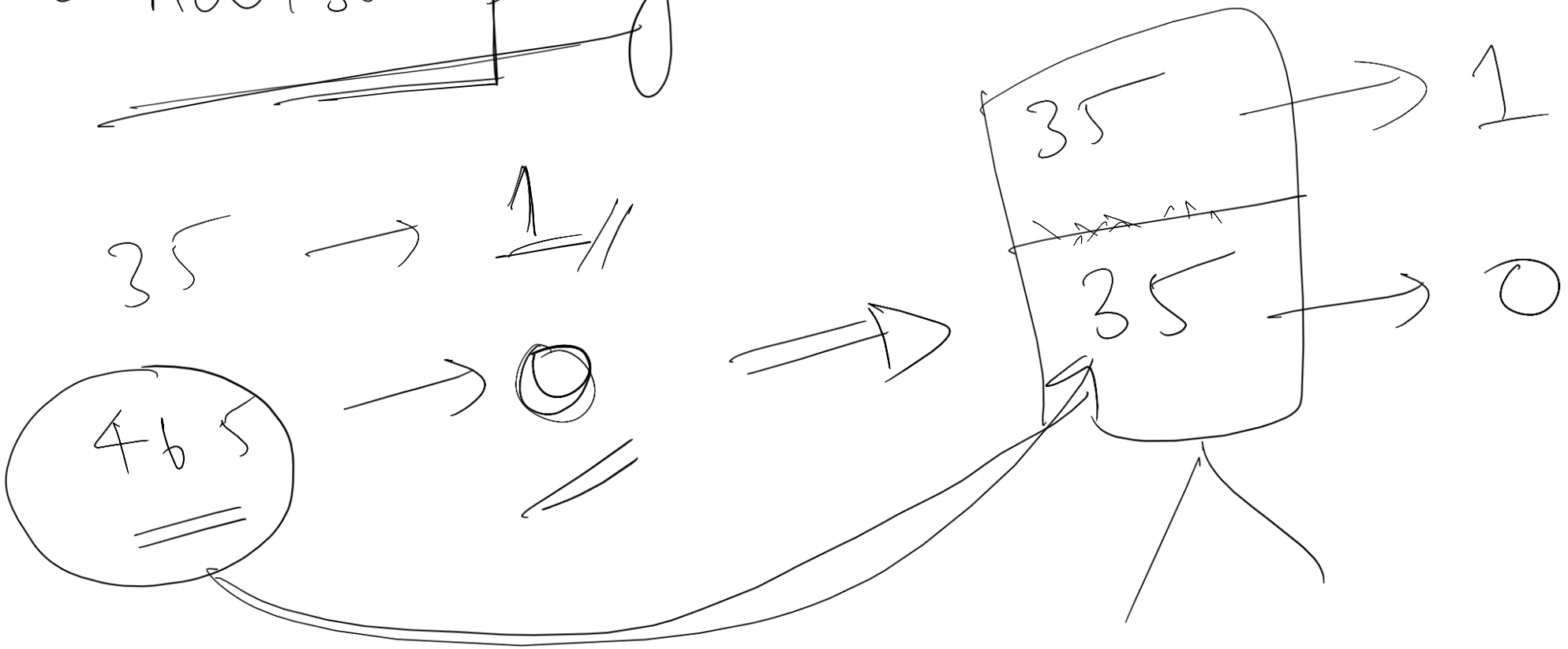

Target x
0 →
1 →


x1 x2 x3 - - - - xn


EDA1
X2
EDA 1
X1
O
X3
EDA2


n = 500
Cancer (1)
Not Cancer (0)
35 → 1
465 → 0
5
30 ✗
460
5 ✗
Acc. Score → 0.95
460 + 5 / 500 = 465 / 500 =

Undersampling:



465 → 1

465 → 0