Linearly depend → non-Linearly ✗

DT → $y \propto x$

$y \propto x^2$

independent → cat/cont

$y \to nL \to x$

dep → cat/con

Gender

Height

Class

Play ↳ -

NP

Split on Gender

Students =30
Play Cricket = 15 (50%)

Female

Male

Students =10
Play Cricket = 2 (20%)

Students = 20
Play Cricket = 13 (65%)

Split on Height

< 5.5 ft

>= 5.5 ft

Students = 12
Play Cricket = 5 (42%)

Students = 18
Play Cricket = 10 (56%)

Split on Class

Class IX

Class X

Students = 14
Play Cricket = 6 (43%)

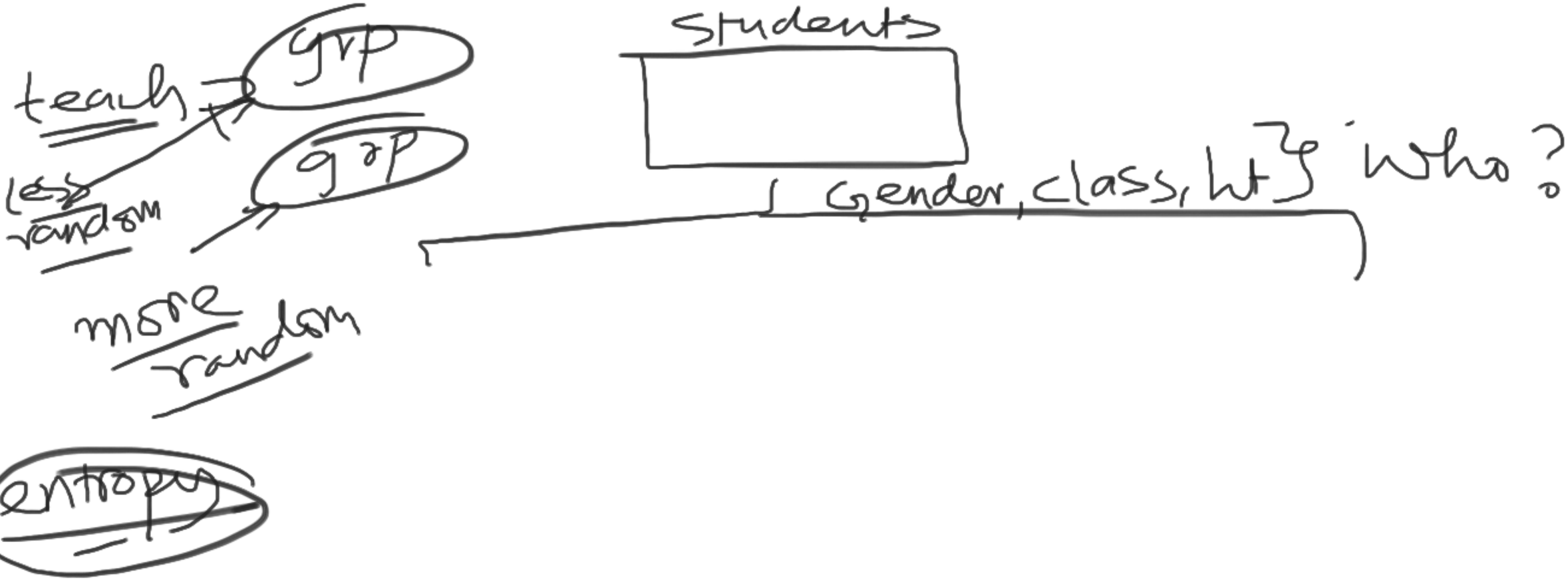
Students = 16
Play Cricket = 9 (56%)

3D

G_1	H_1	C_1
G_2	H_2	C_2
G_3	H_3	C_3

How?

DT → inspired → humans → decision



Split → less random impurity | more inf sub grp purity

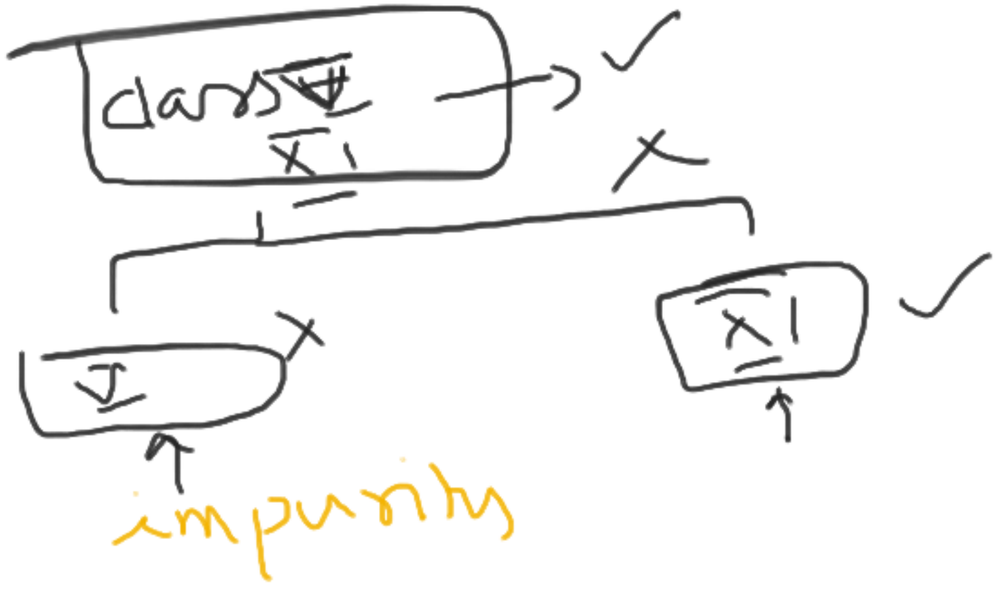
diff $\frac{dy}{dx}$

Entropy

IG

Gini Impurity

Gini



grid search

precision

density +ve

density -ve

Entropy $= -p \log_2 p - q \log_2 q$

$= -(p \log p + q \log q)$

$q = 0$

$p = 1$

$-(p \log p + q \log q)$

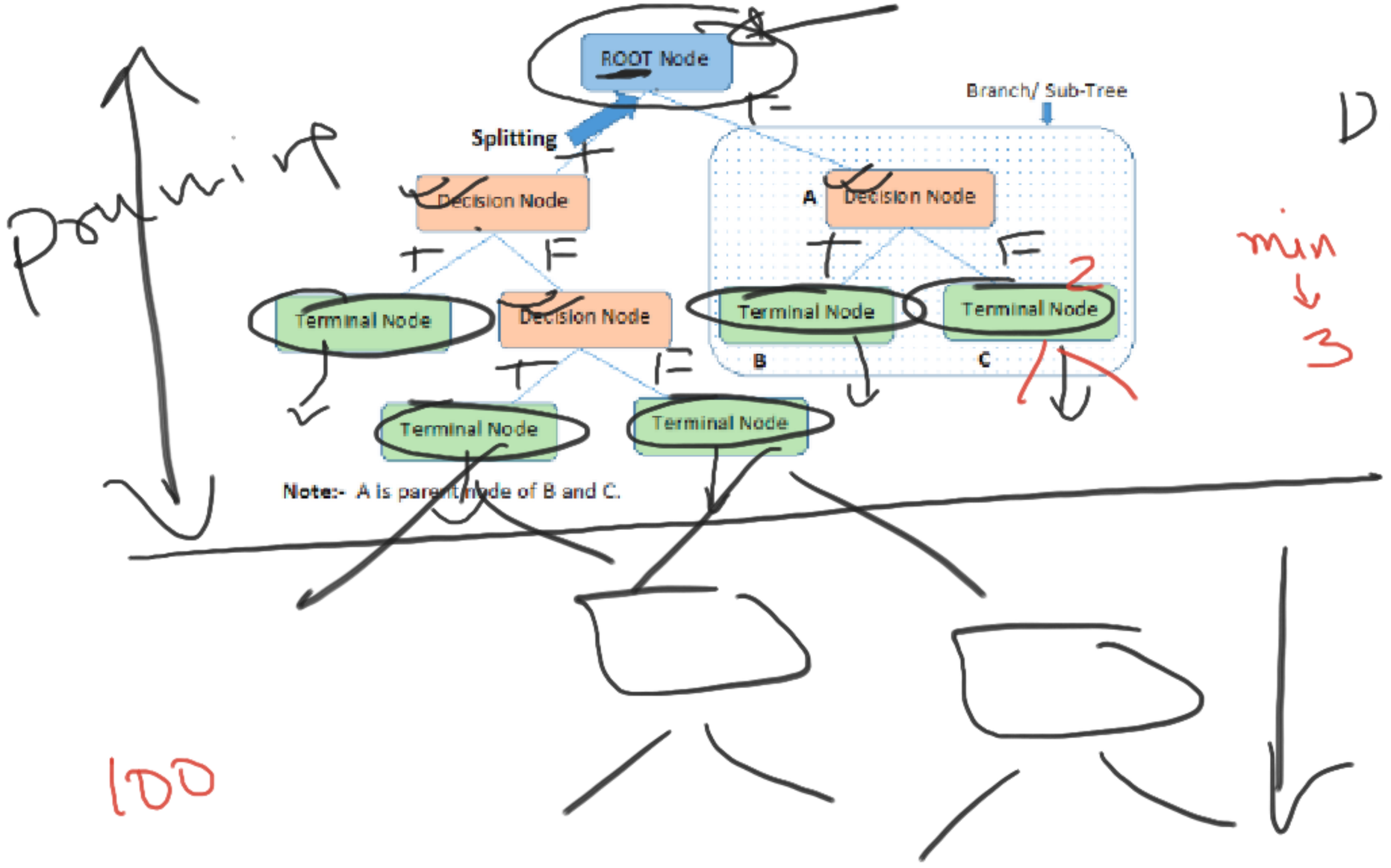
0 , 0

$q = 1 - p$

+ve . −ve

less

0.5



Decision

CART

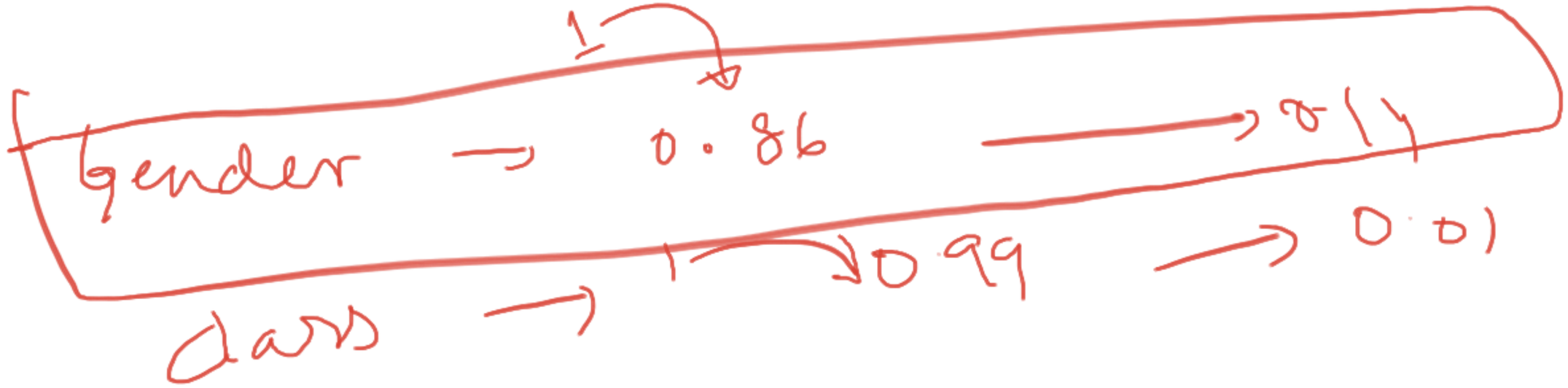
Binary

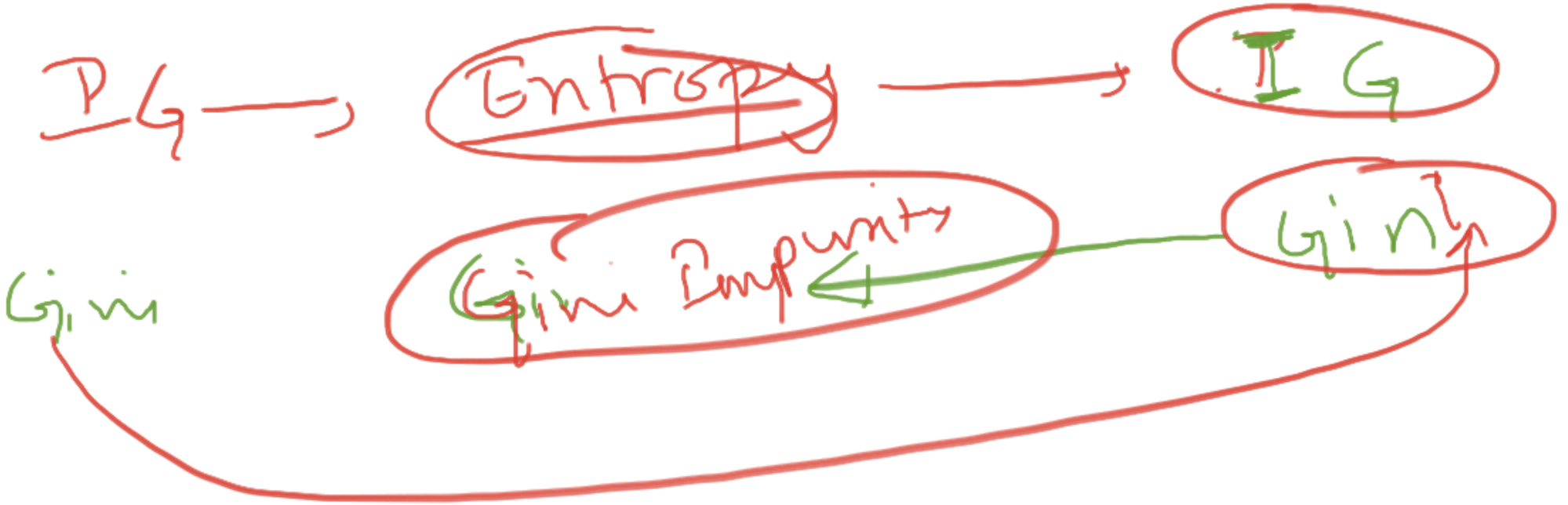
min

3

2

100


1
gender
0.86
0.14
0.99
0.01
days



Split on Gender

Students =30
Play Cricket = 15 (50%)

Female

Students =10
Play Cricket = 2 (20%)

Male

Students = 20
Play Cricket = 13 (65%)

Split on Class

Class IX

Students = 14
Play Cricket = 6 (43%)

Class X

Students = 16
Play Cricket = 9 (56%)

Split on Gender

Students =30
Play Cricket = 15 (50%)

Female

Students =10
Play Cricket = 2 (20%)

Male

Students = 20
Play Cricket = 13 (65%)

Split on Class

Class IX

Students = 14
Play Cricket = 6 (43%)

Class X

Students = 16
Play Cricket = 9 (56%)

F-01

→ How does prediction on new dataset ✓

→ What does voting classifier ✓

→ Adv | Disadv

→ RF & XGBoost

Logistic → non-linearly

Linearity

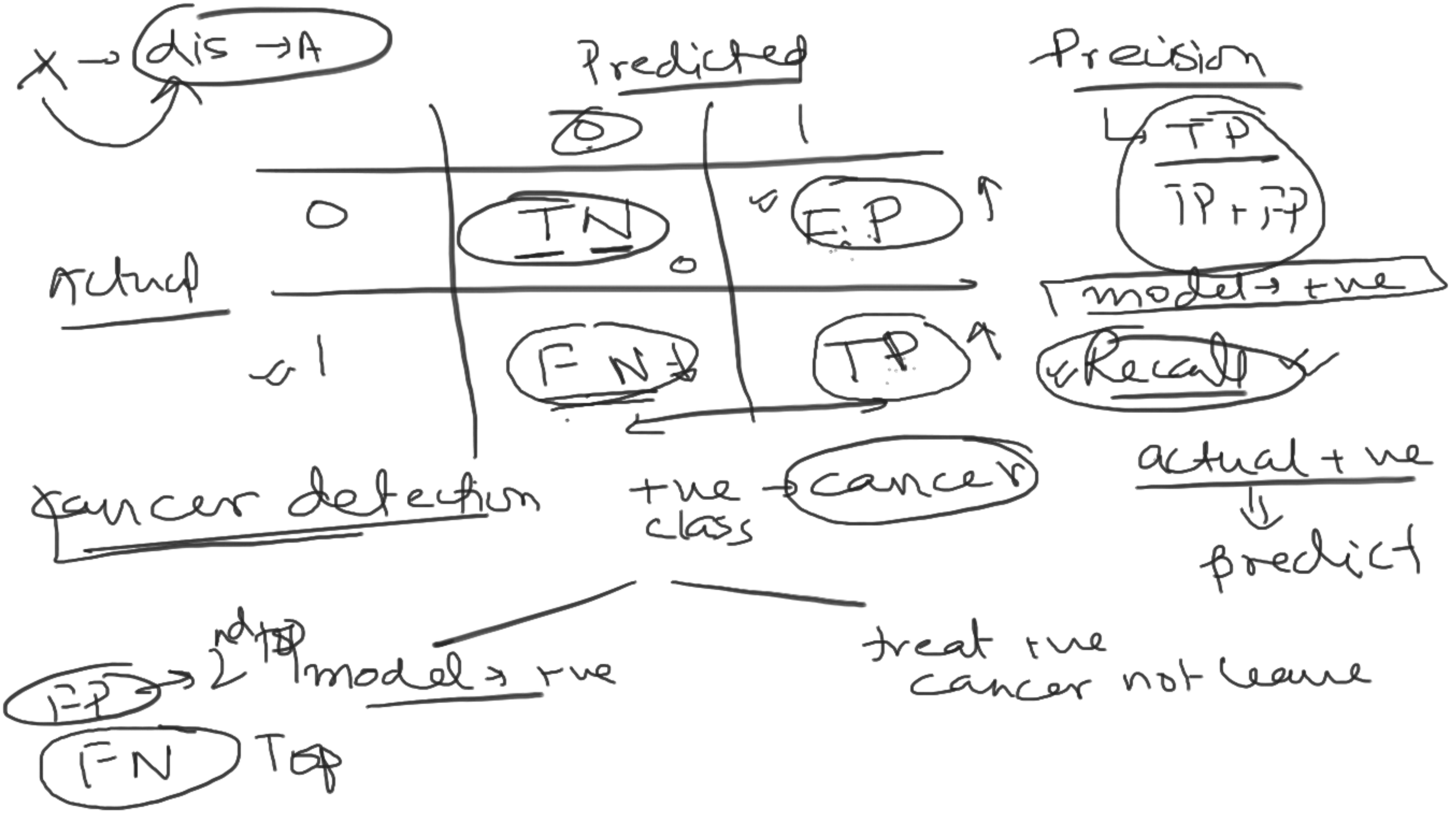
Precision Recall

Decision tree → criterion

split node

impu → entropy → IG

pure ← gini → Gini imp


X → dis → A
Predicted
0
1
0
TN
F.P
Actual
1
FN
TP
Precision
TP
TP + FP
model → +ve
Recall
cancer detection
+ve class → cancer
actual +ve
predict
FP
model → +ve
FN
TP
treat +ve cancer not cause

Precision (P)

Precise of model

Recall (R)

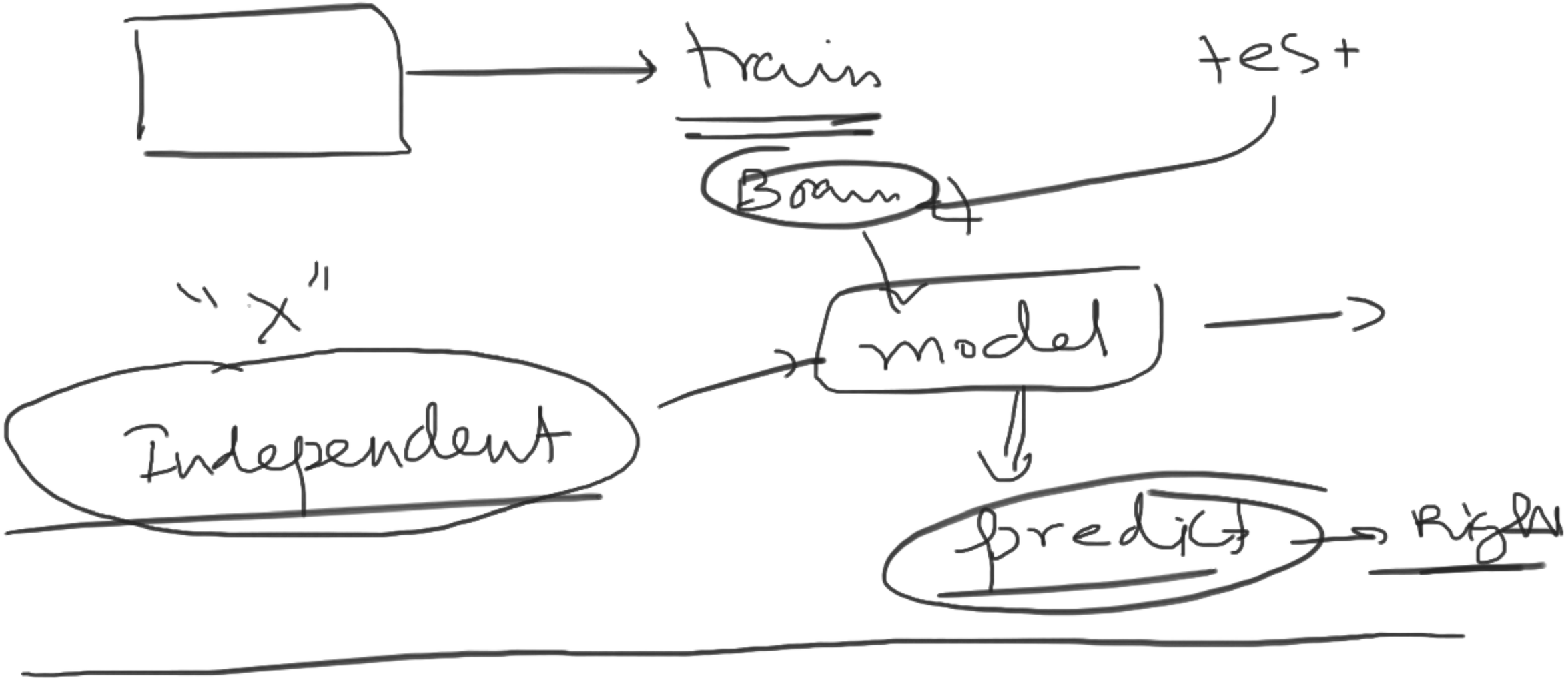
FOMO

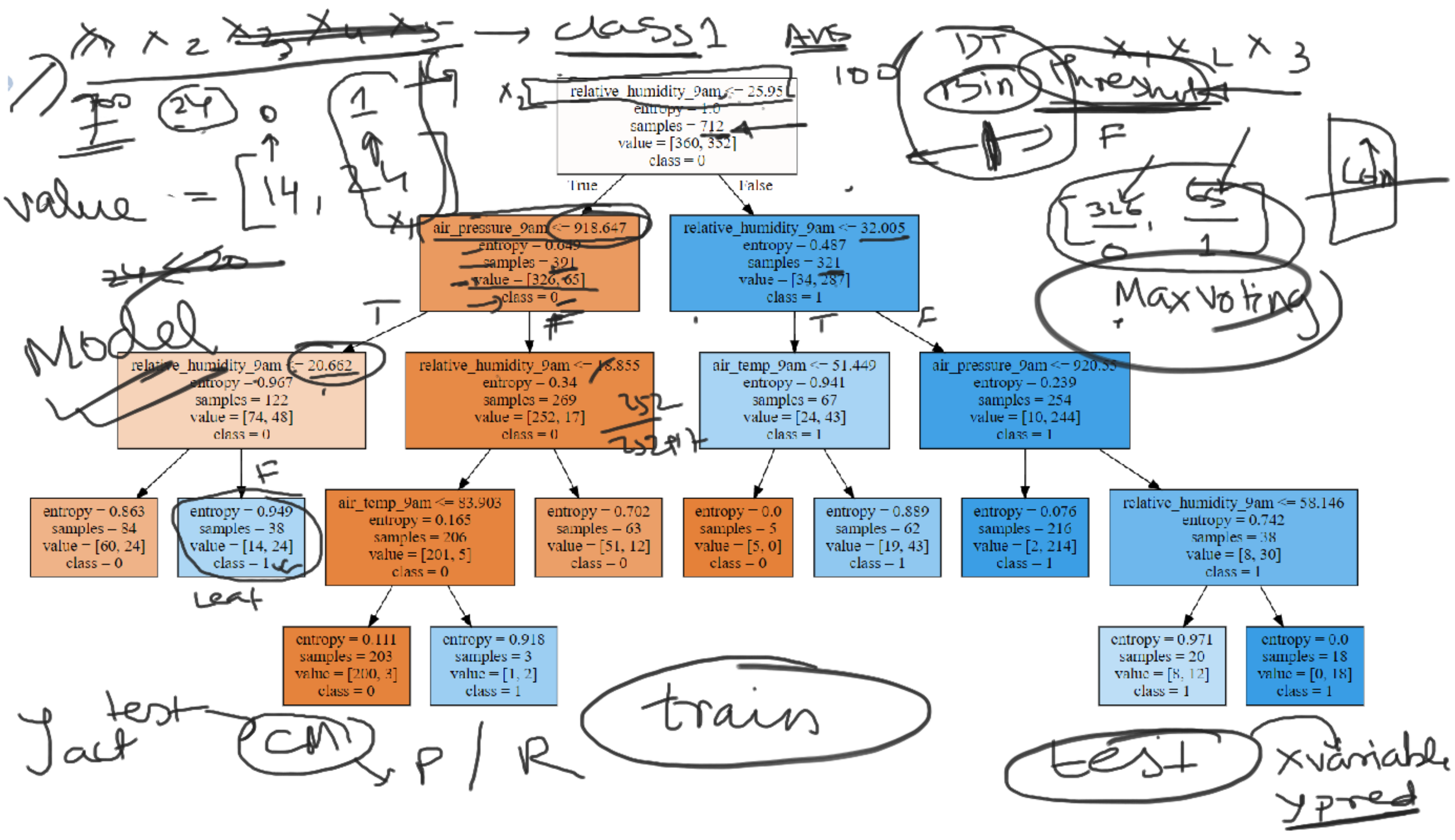
F1

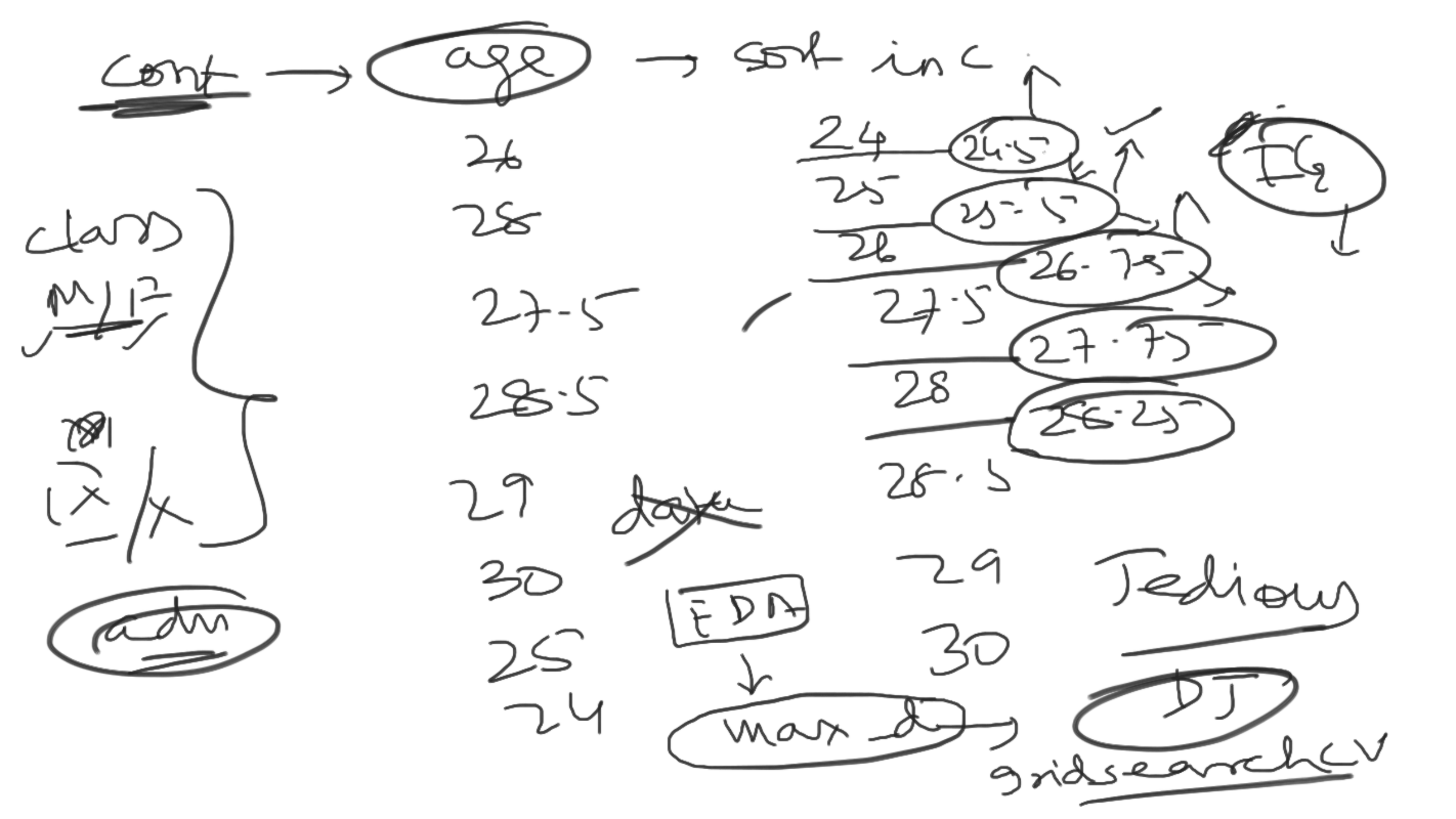
HM P&R

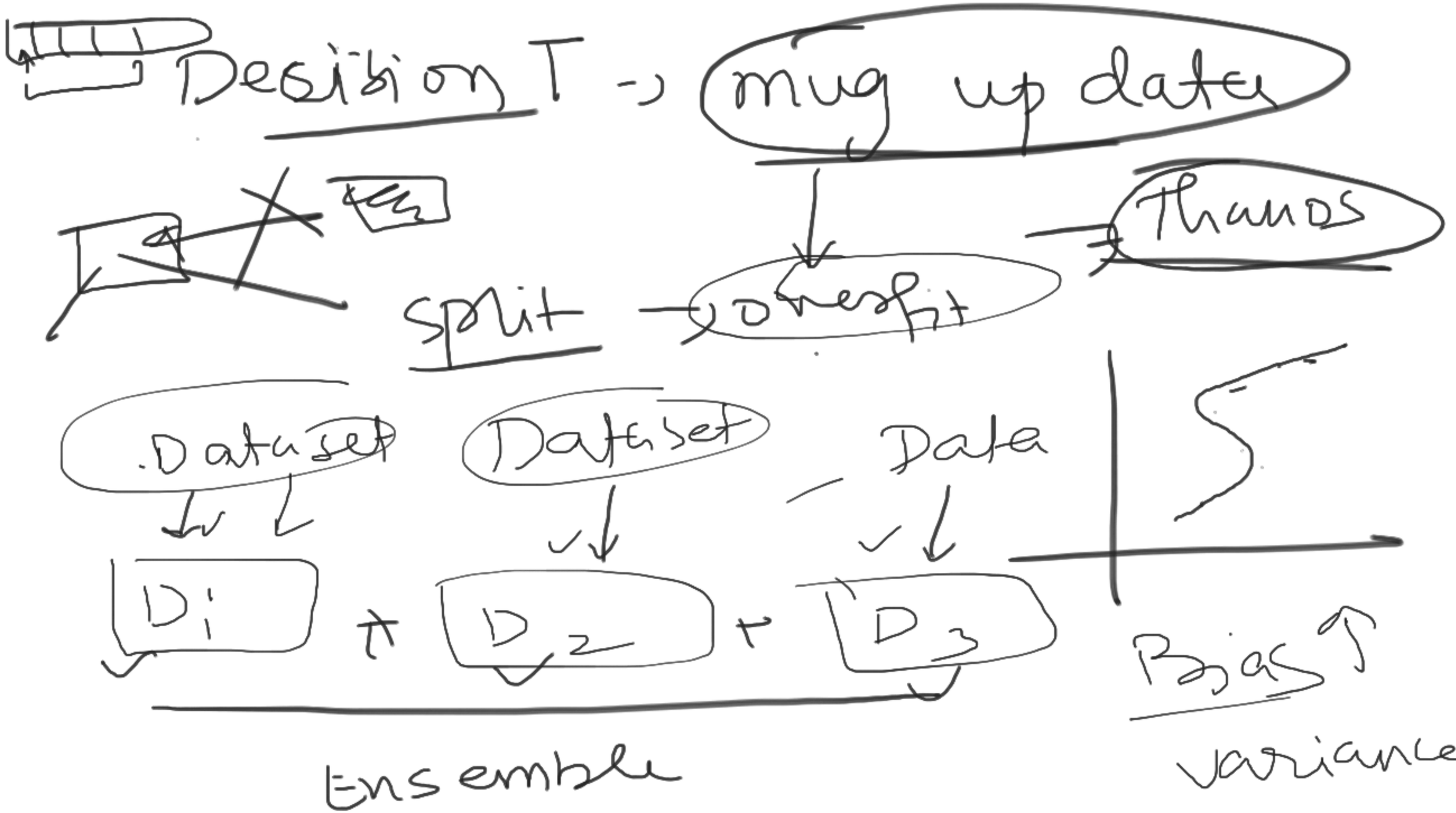
$$F1 = \frac{2PR}{P+R}$$

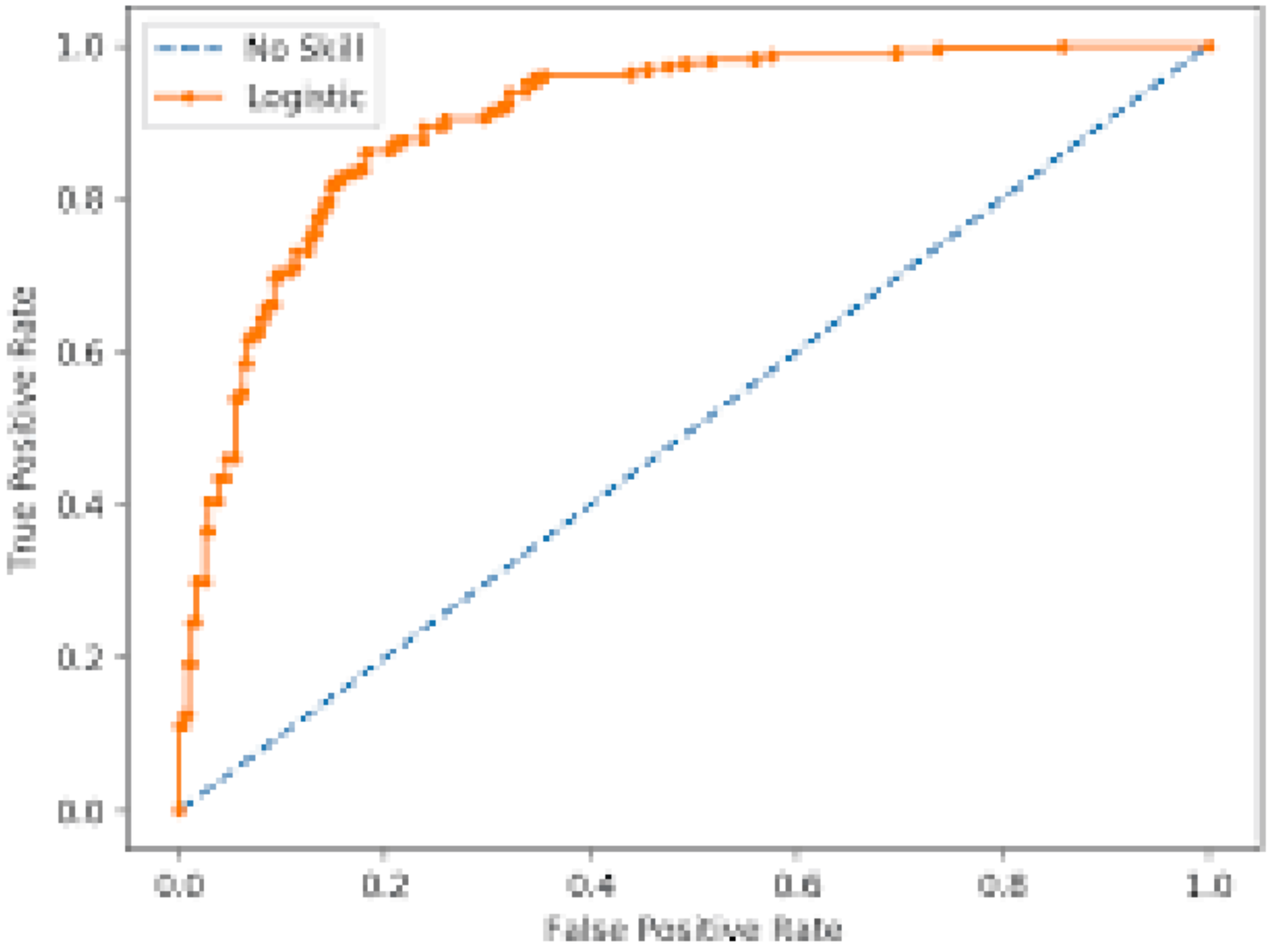
AUC ROC


train
test
Boom
"X"
Independent
model
predict


relative_humidity_9am <= 25.95
entropy = 1.0
samples = 712
value = [360, 352]
class = 0
True
False
air_pressure_9am <= 918.647
entropy = 0.649
samples = 321
value = [326, 65]
class = 0
relative_humidity_9am <= 32.005
entropy = 0.487
samples = 321
value = [34, 287]
class = 1
relative_humidity_9am <= 20.662
entropy = 0.967
samples = 122
value = [74, 48]
class = 0
relative_humidity_9am <= 18.855
entropy = 0.34
samples = 269
value = [252, 17]
class = 0
air_temp_9am <= 51.449
entropy = 0.941
samples = 67
value = [24, 43]
class = 1
air_pressure_9am <= 920.55
entropy = 0.239
samples = 254
value = [10, 244]
class = 1
entropy = 0.863
samples = 84
value = [60, 24]
class = 0
entropy = 0.949
samples = 38
value = [14, 24]
class = 1
air_temp_9am <= 83.903
entropy = 0.165
samples = 206
value = [201, 5]
class = 0
entropy = 0.702
samples = 63
value = [51, 12]
class = 0
entropy = 0.0
samples = 5
value = [5, 0]
class = 0
entropy = 0.889
samples = 62
value = [19, 43]
class = 1
entropy = 0.076
samples = 216
value = [2, 214]
class = 1
relative_humidity_9am <= 58.146
entropy = 0.742
samples = 38
value = [8, 30]
class = 1
entropy = 0.111
samples = 203
value = [200, 3]
class = 0
entropy = 0.918
samples = 3
value = [1, 2]
class = 1
entropy = 0.971
samples = 20
value = [8, 12]
class = 1
entropy = 0.0
samples = 18
value = [0, 18]
class = 1


Cont
age
sort inc
24
25
26
27-5
28
28.5
29
30
24.5
25.5
26.75
27.5
27.75
28
28.25
28.5
29
30
class
MLP
adv
EDA
max
Tedious
DT
gridsearchCV


Decision T -> mug up data
split
overfit
Thanos
Dataset
Dataset
Data
D1
D2
D3
Bias ↑
variance ↓
Ensemble


No Skill
Logistic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0