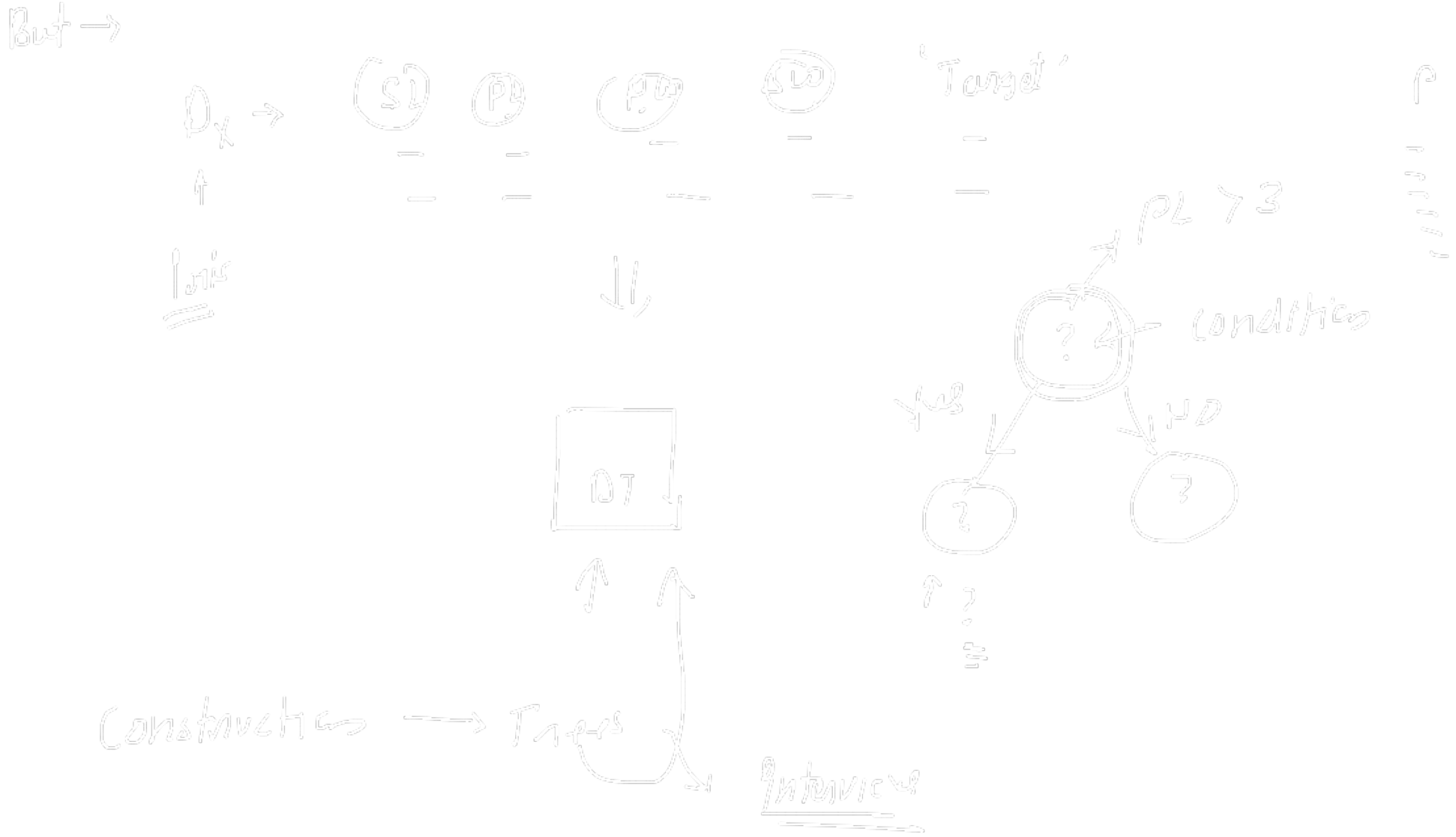
KNN → distance

LogReg → probabilistic model

But DT → Nested if else

Data → Iris SL PL SW PW Speed


Target
condition
Interview

⇒ Imp term

↳ (i) ENTROPY ⟶ Tennis Example

$Y \rightarrow Y_1\ Y_2 \text{ ---- } Y_k$

$$H(Y) = -\sum_{i=1}^{k} P(Y_i) * \log_2(P(Y_i))$$

$k = 2$

$$H(Y) = -\frac{9}{14}\log_2\left(\frac{9}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) - \boxed{\quad}$$

$P(Y) * \log_2\left(\frac{9}{14}\right)$

$H(Y) = 0.94$

Attribute → (9 - 1)

$Y \rightarrow Y_+ \quad Y_-$

(case i)

$Y_+ \rightarrow 99\%$

$Y \rightarrow$

$Y_- \rightarrow 1\%$

$H(Y) = -0.99 \log_2(0.99) - 0.01 \log_2(0.001)$

$H(Y) = 0.08065$

(case ii)

$Y_+ \rightarrow 50\%$

$Y \rightarrow$

$Y_- \rightarrow 50\%$

$H(Y) = -0.5 \log_2(0.5) - 0.5 \log_2(0.5)$

$H(Y) = 1$

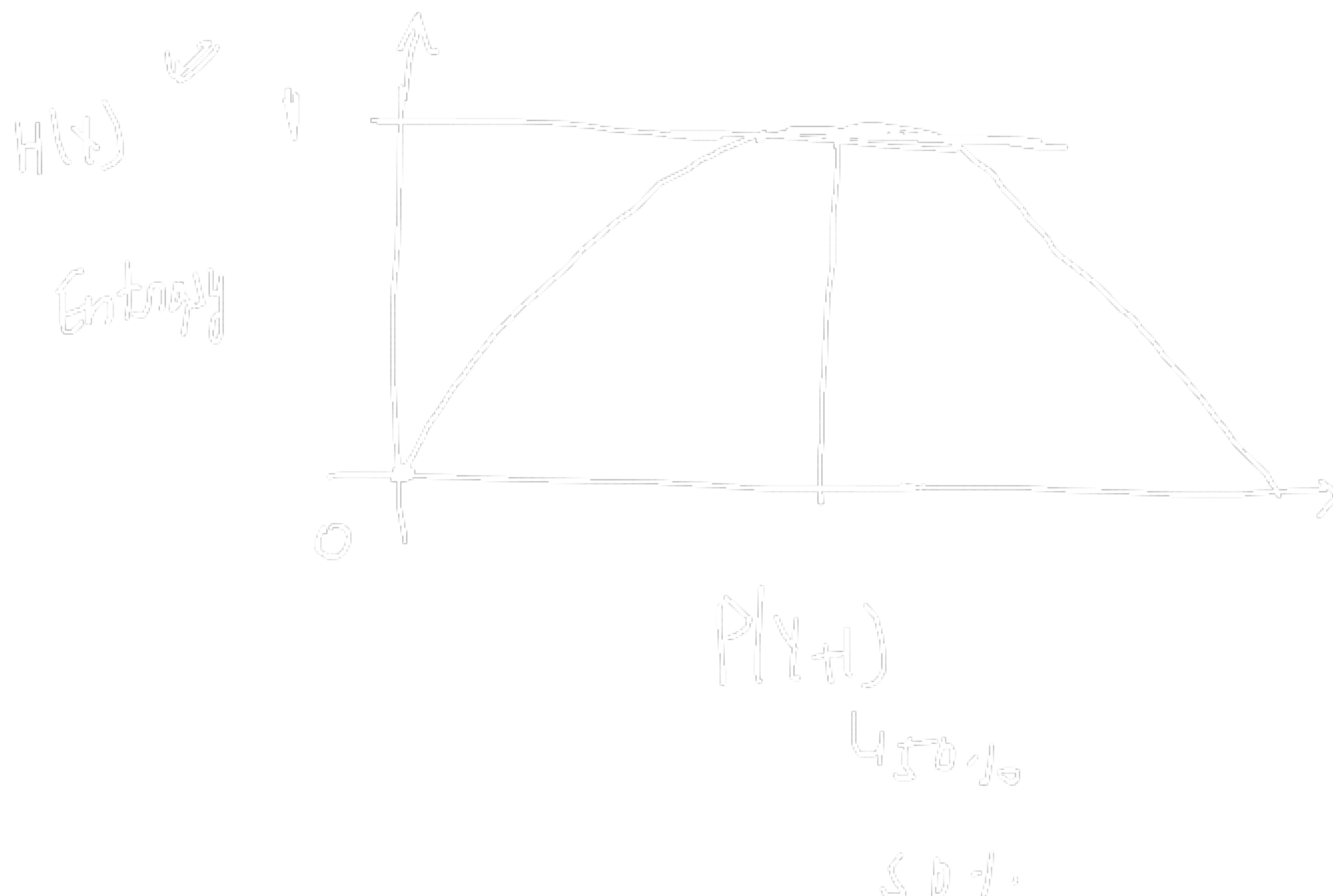
Case (iii)

$Y \to Y_+ \to 0\%$, $Y_- \to 100\%$ } $H(Y) = 0$

$P(Y_+) \to 0 \uparrow 50\%$

$P(Y_+) \to 60\%$
$P(Y_-) \to 40$ }

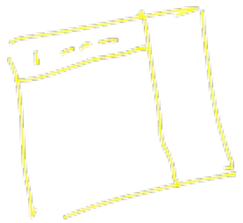
from (3) →



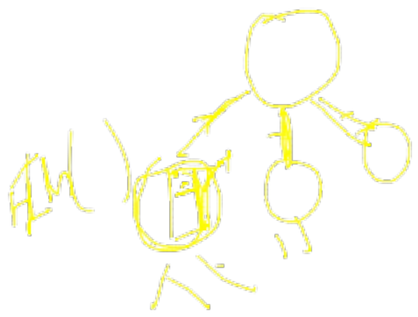
Information Gain

Parent

Outlook



H(Y)



D1

H(D1)

0.97

D2

H(D2)

D3

H(D3) = 0.97

Weighted Entropy → 0.69

IG(Y, Outlook)

IG(Y, Outlook) = (weighted entropy after breaking dataset into n parts) − (Entropy before breaking it down)

↓

IG(Y, Outlook) = 0.94 − Σ 5/14 × 0.97

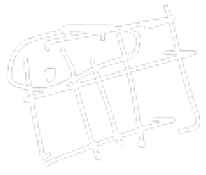
→ # Input features

$I_{th}(Y, \text{rainfall}) = ?$

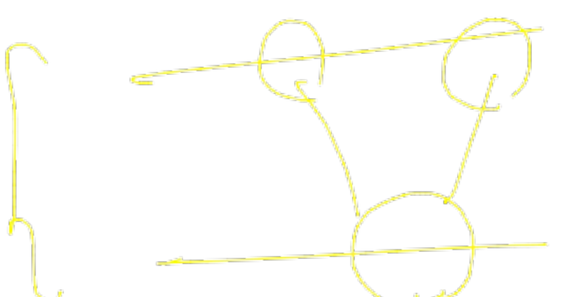
$I_{th}(Y, \text{temp}) = ?$

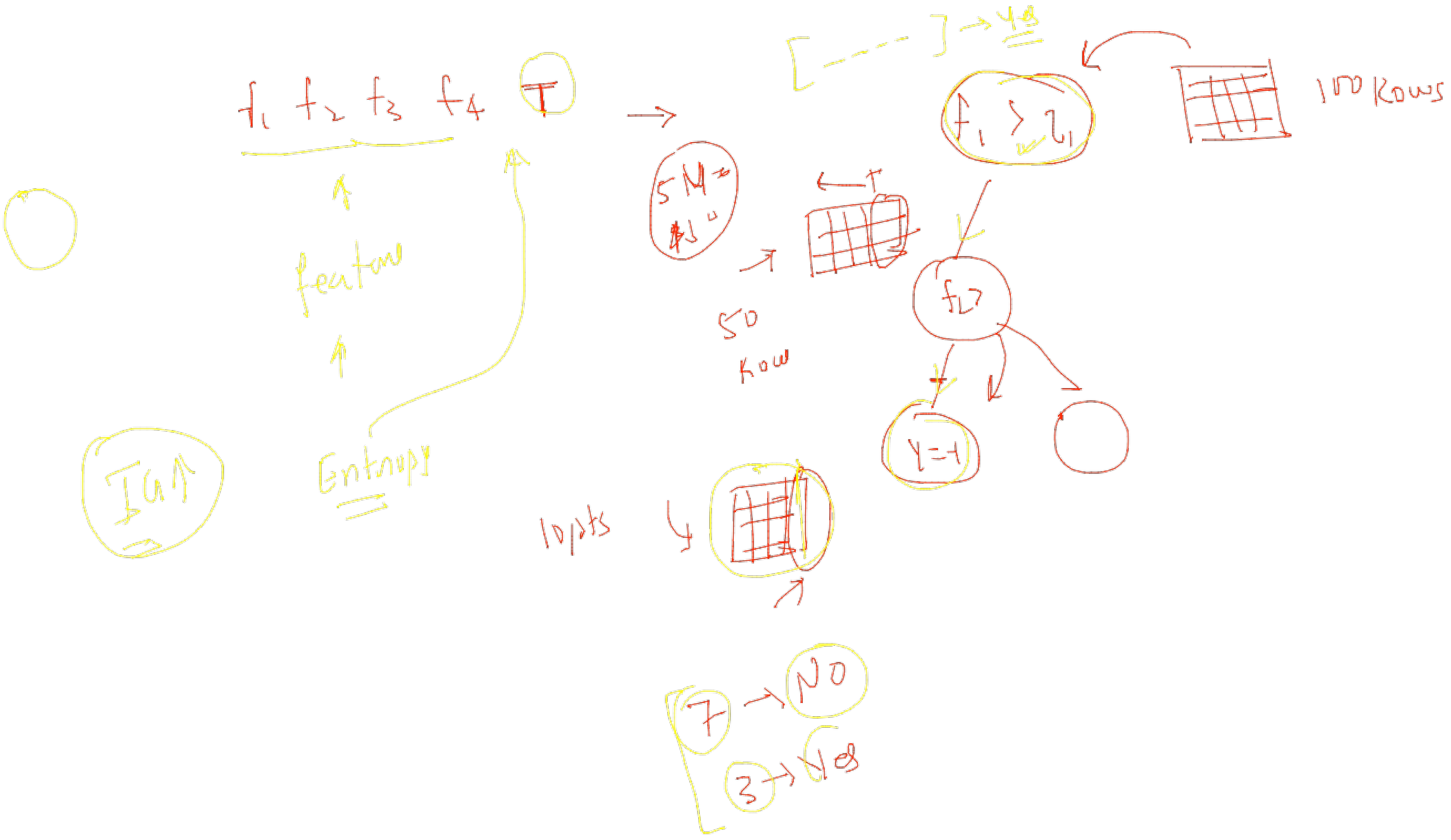
$I_{th}(Y, \text{wind}) = ?$

Based on MAY (Th)



?
20
less
less
Future Input
Other
DT
depth
min




features
Entropy
SMOTE
50 Row
100 Rows
Y=-1
10pts
7 → NO
3 → NO