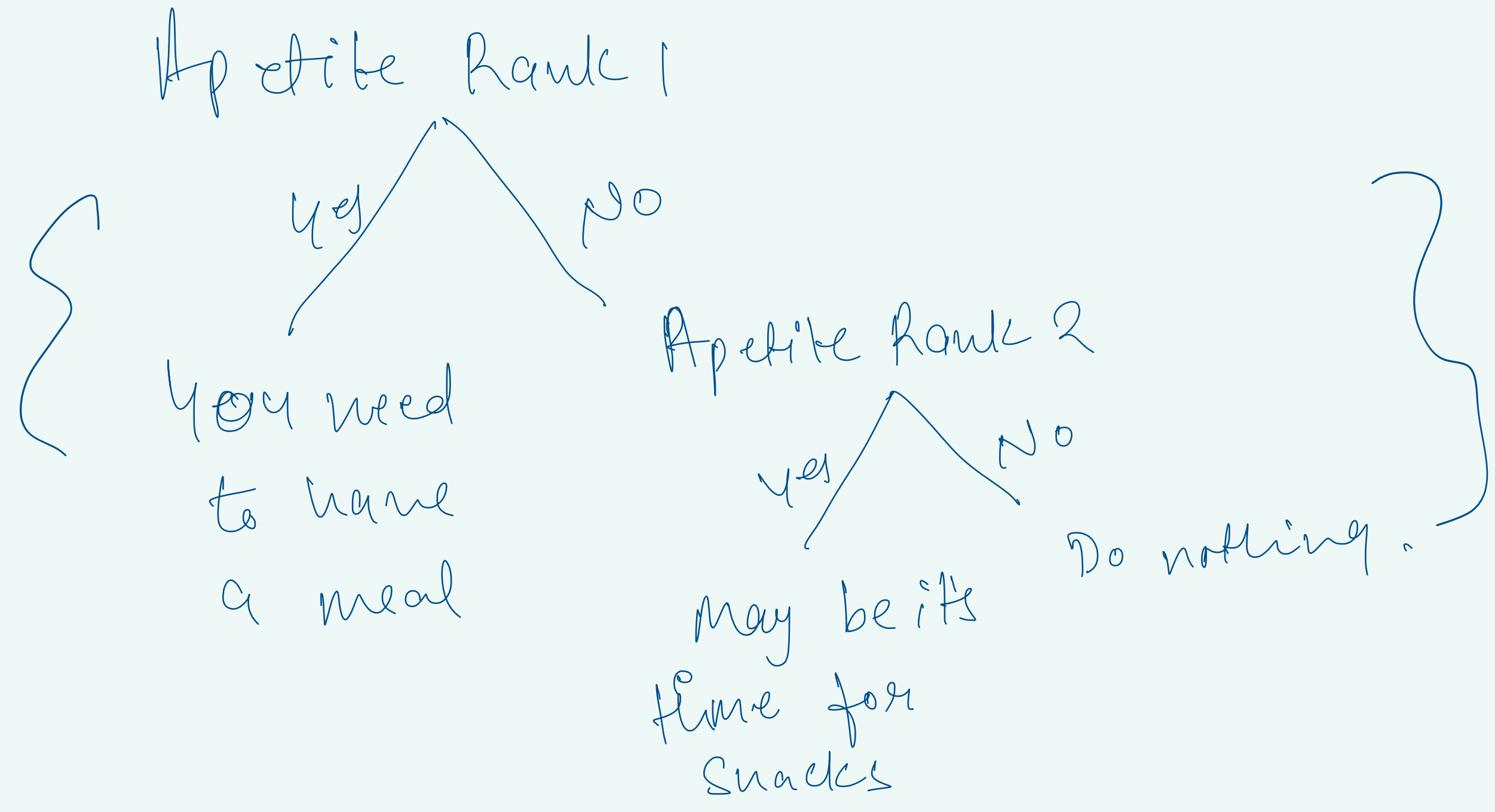
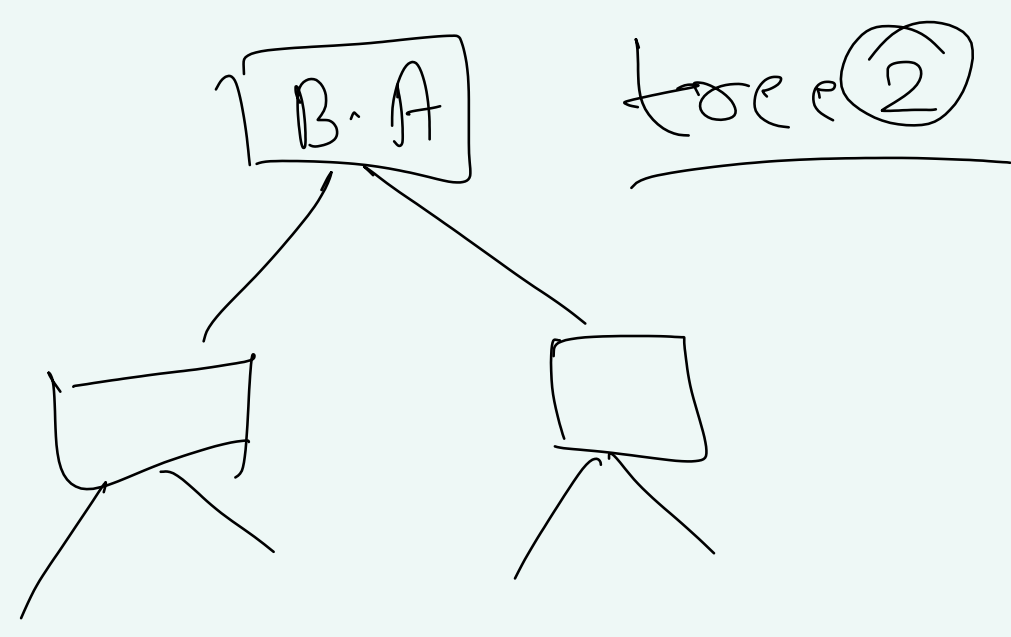
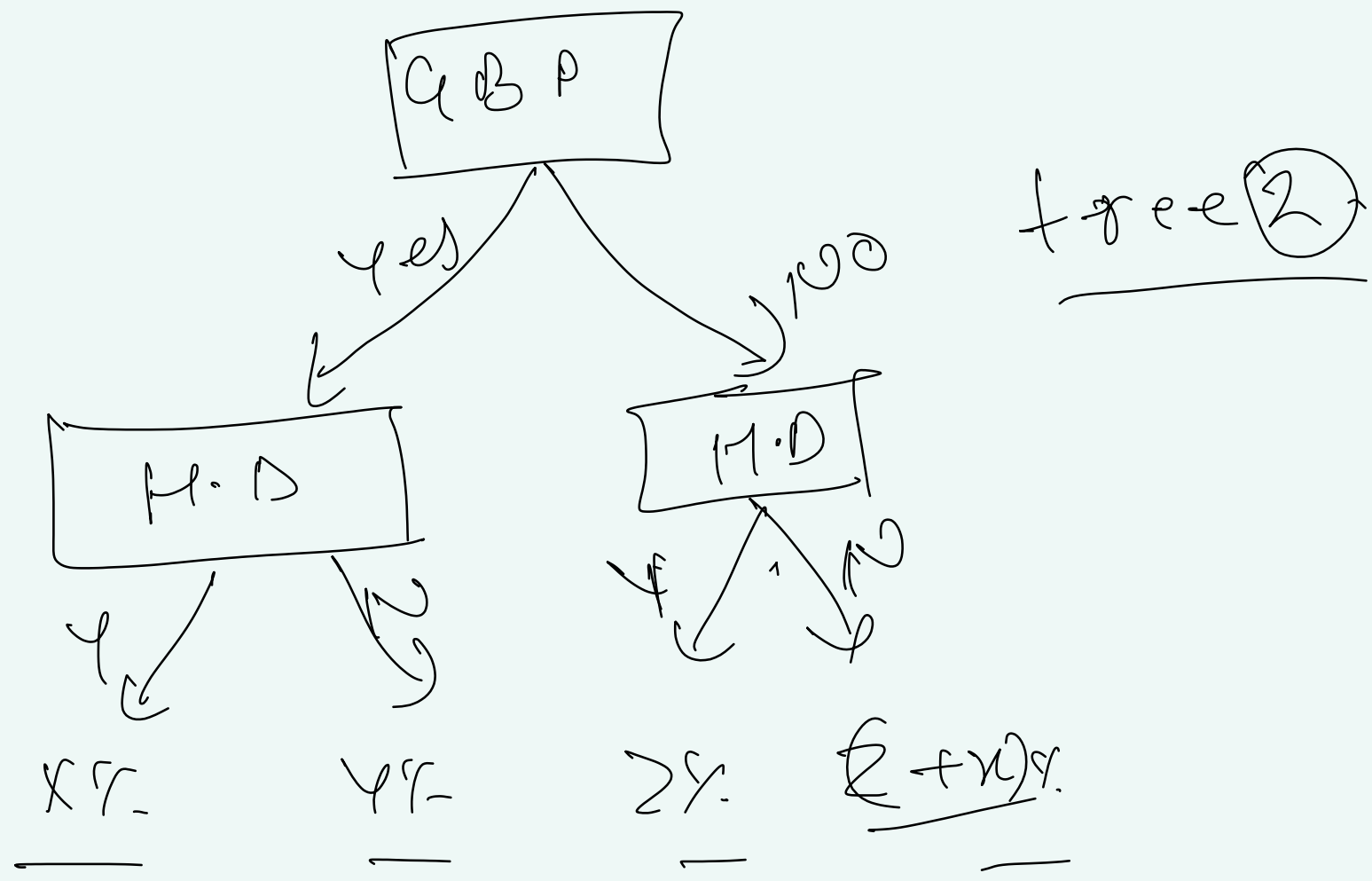
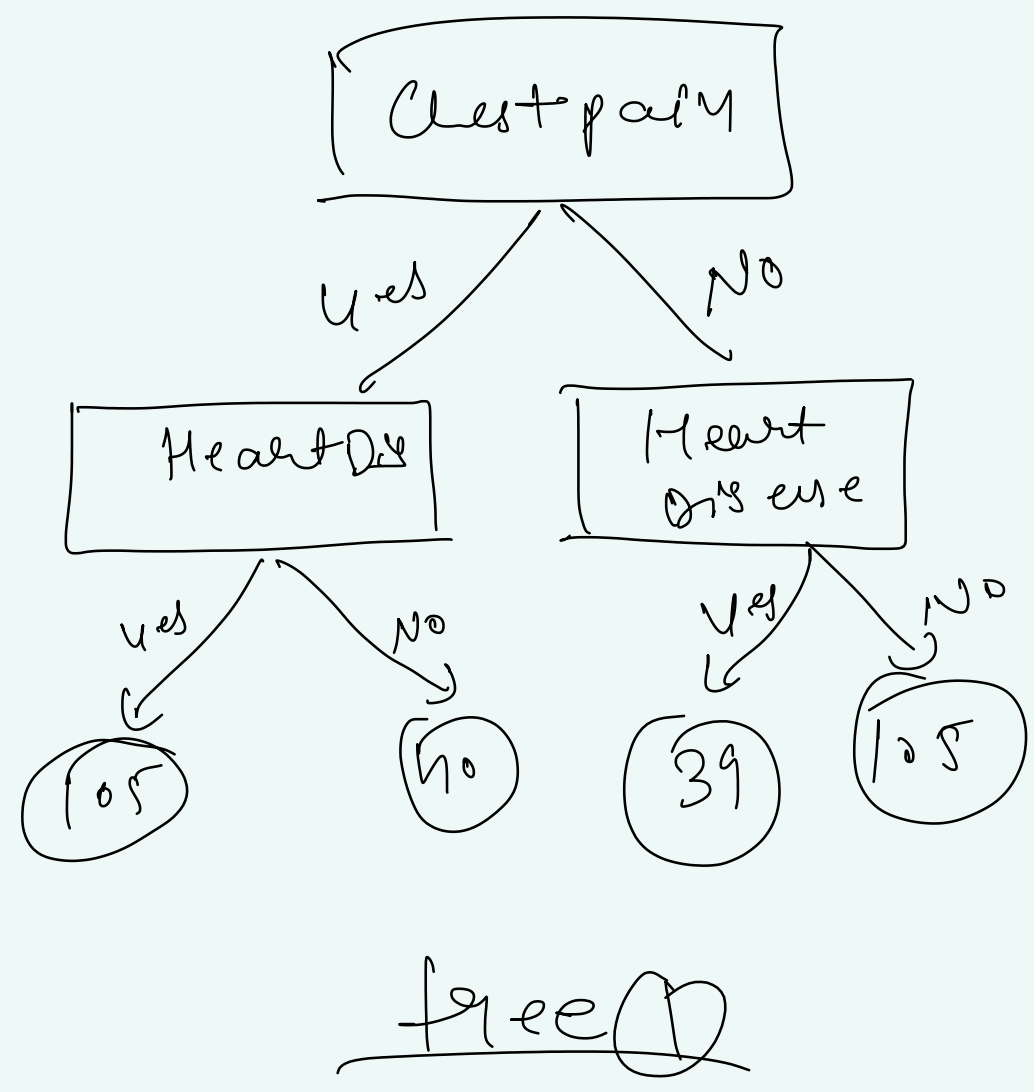


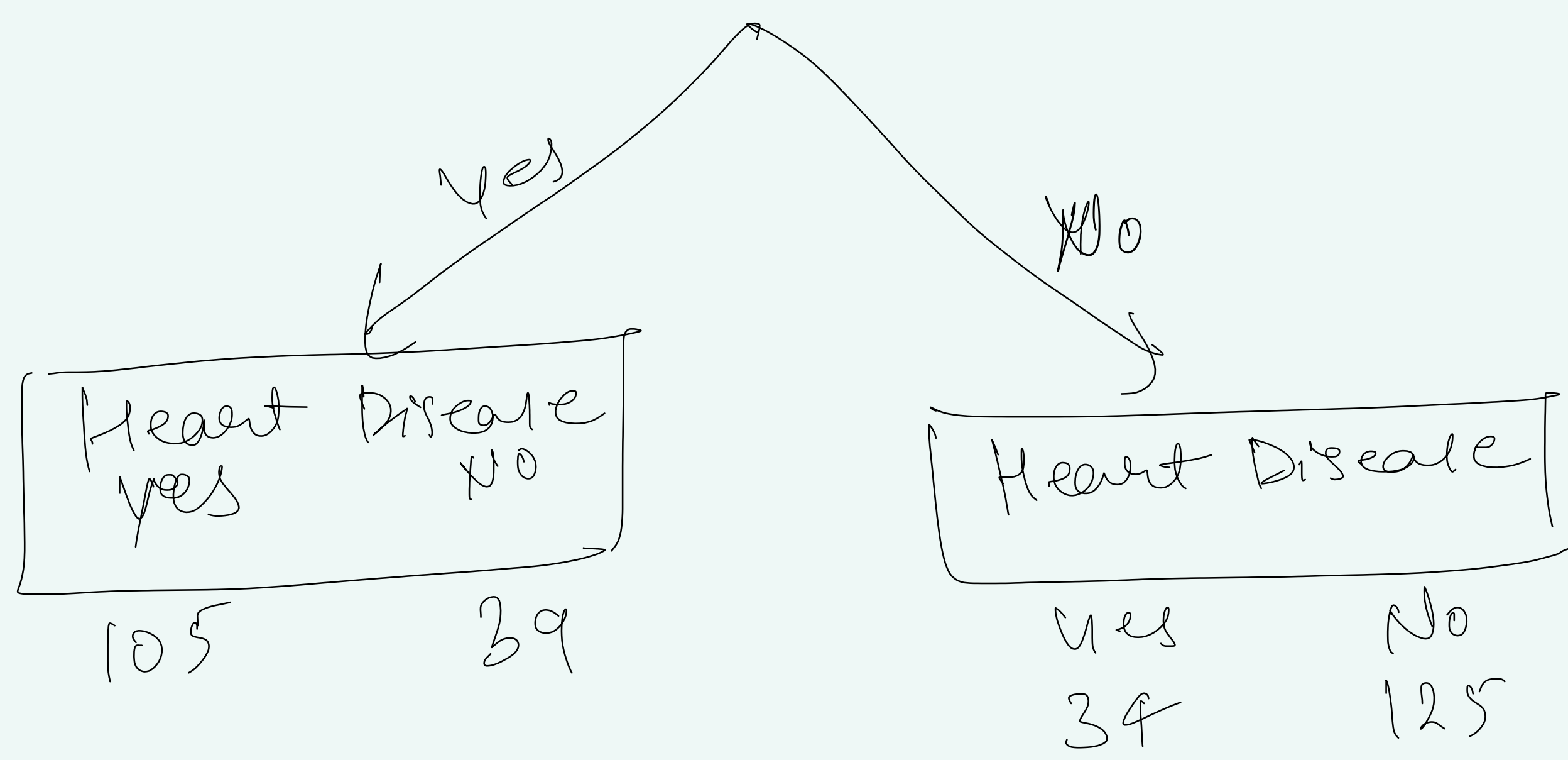
Consider the following table

Chest Pain	Good Blood Pres.	Blocked Arteries	Heart Disease
No	No	No	No
Yes	Yes	Yes	Yes
Yes	Yes	No	No
Yes	No	No	Yes



Which particular Independent variable should be considered as Root Node. and what would be the leaf and child node?

Blocked Arteries



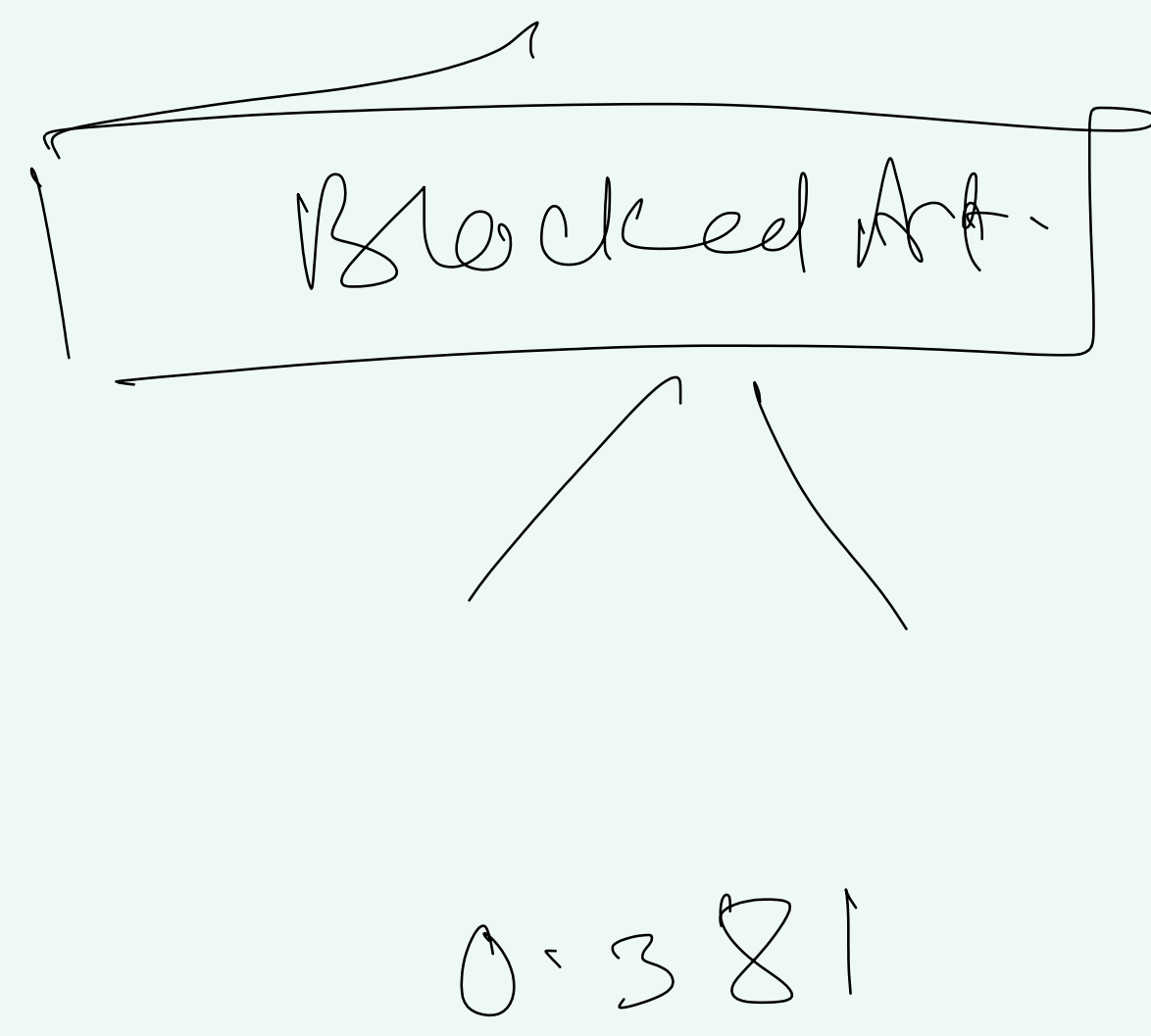
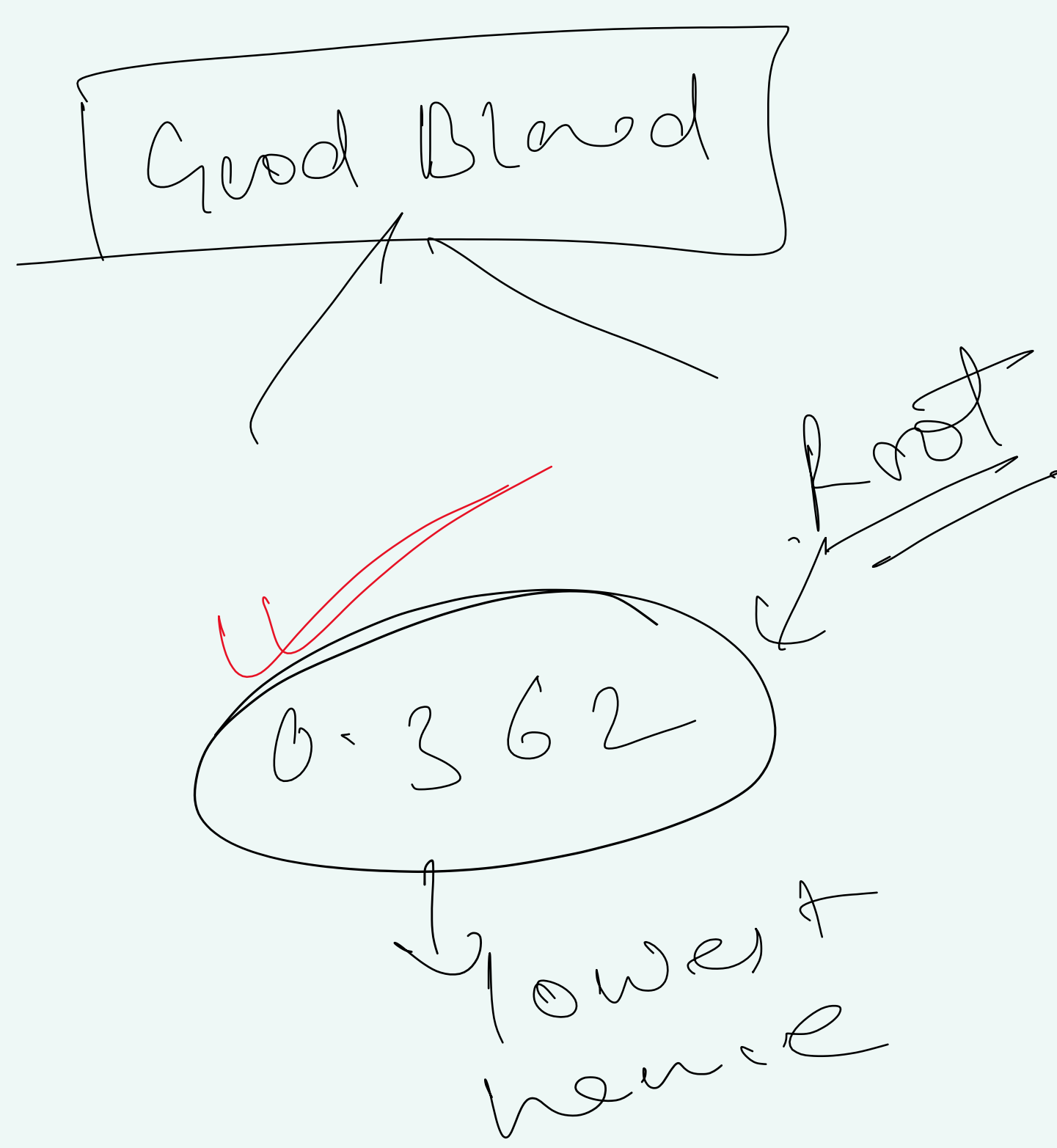
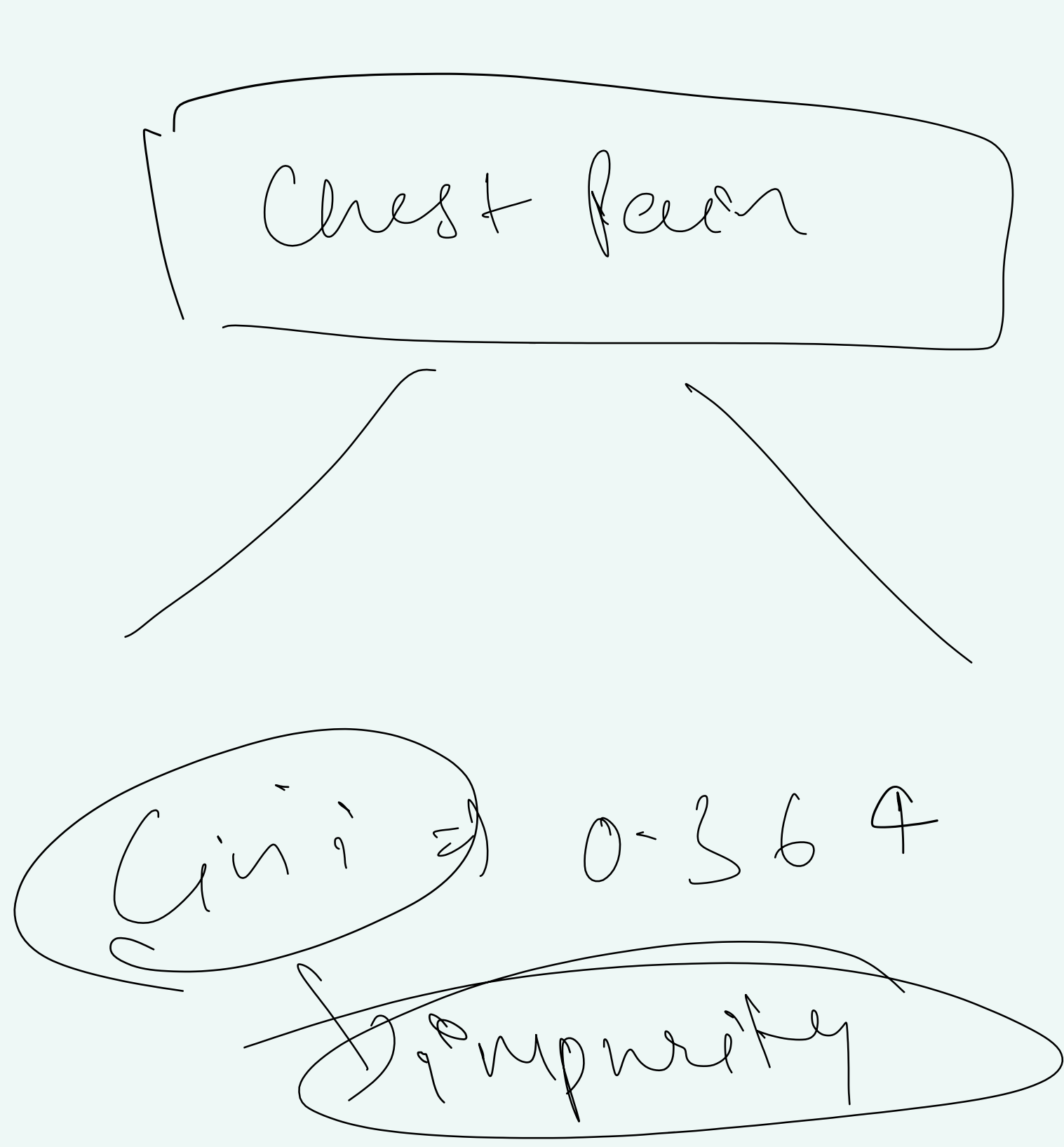
we decide either based on:

① Entropy $H(x) = -P(x_i) \log P(x_i)$

Binary $2) -P(A) \log(P(A)) - P(B) \log(P(B))$

② Gini Impurity $1 - P(A)^2 - P(B)^2$

(Entropy vs log value)



If the node itself has the lowest Gini score then there is no point in further splitting and it becomes a leaf.

