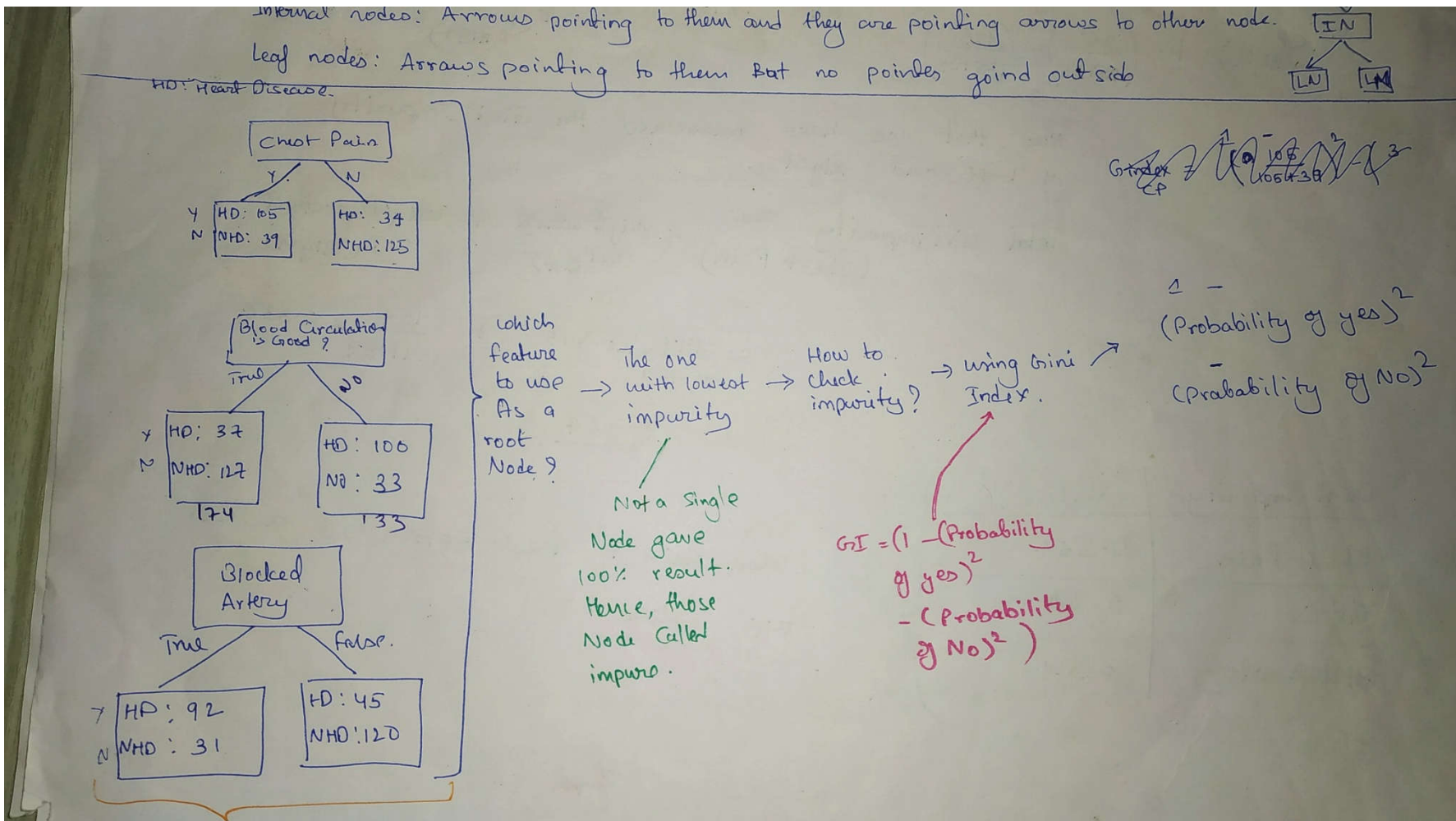


Decision Tree:

Root Nodes: no arrows pointing outwards

Internal Nodes: arrows pointing inwards and outwards

Leaf Nodes: Arrows pointing outwards



$GINI_{CP(Left)} = 1 - \left(\frac{105}{105+39}\right)^2 - \left(\frac{39}{105+39}\right)^2$

$GINI_{CP(Right)} = 1 - \left(\frac{34}{34+125}\right)^2 - \left(\frac{125}{34+125}\right)^2$

$GINI_{CP(Left)} = 0.395$

$GINI_{CP(Right)} = 0.336$

Now, that we have measured the Gini Impurity of Left and Right,

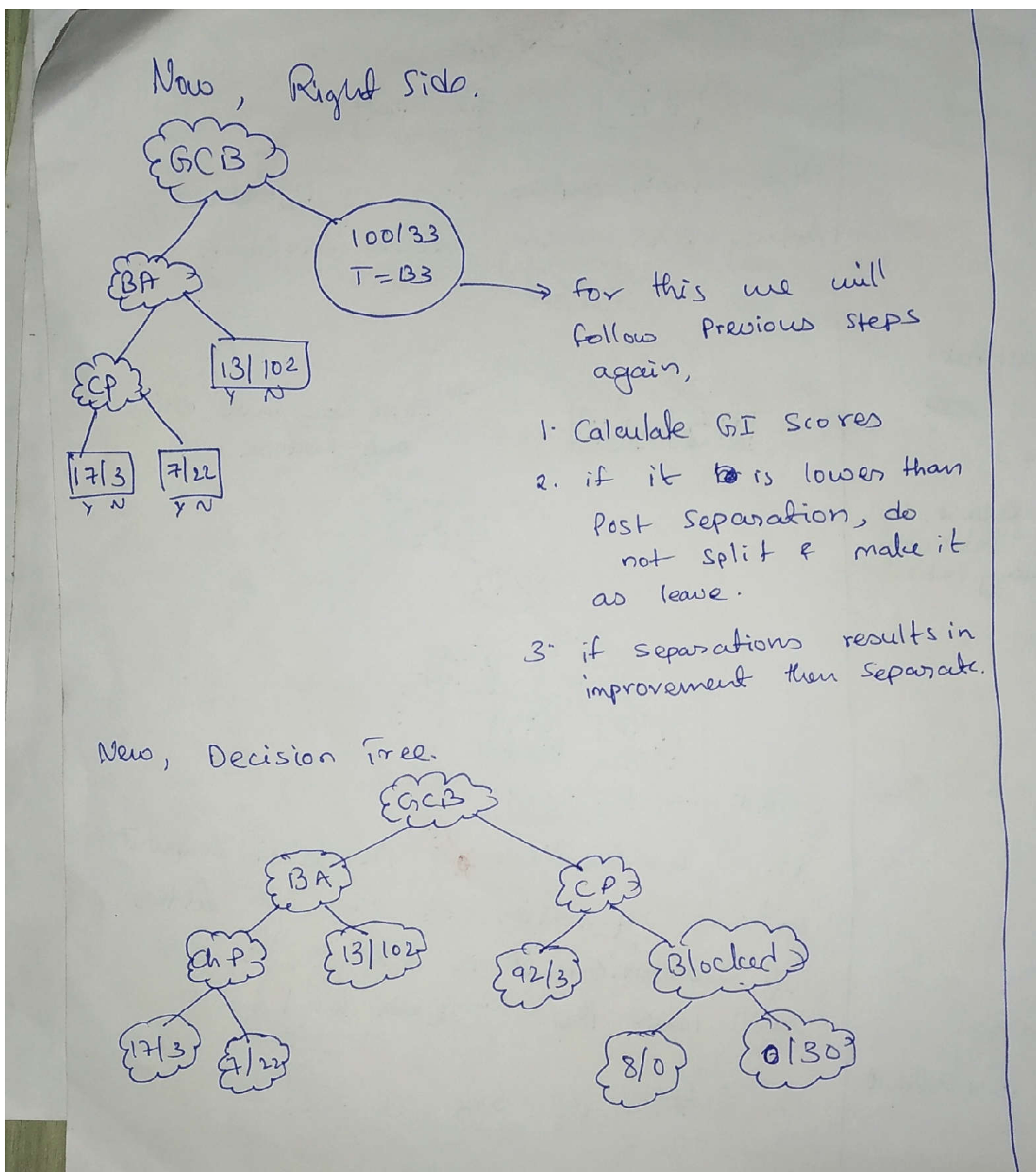
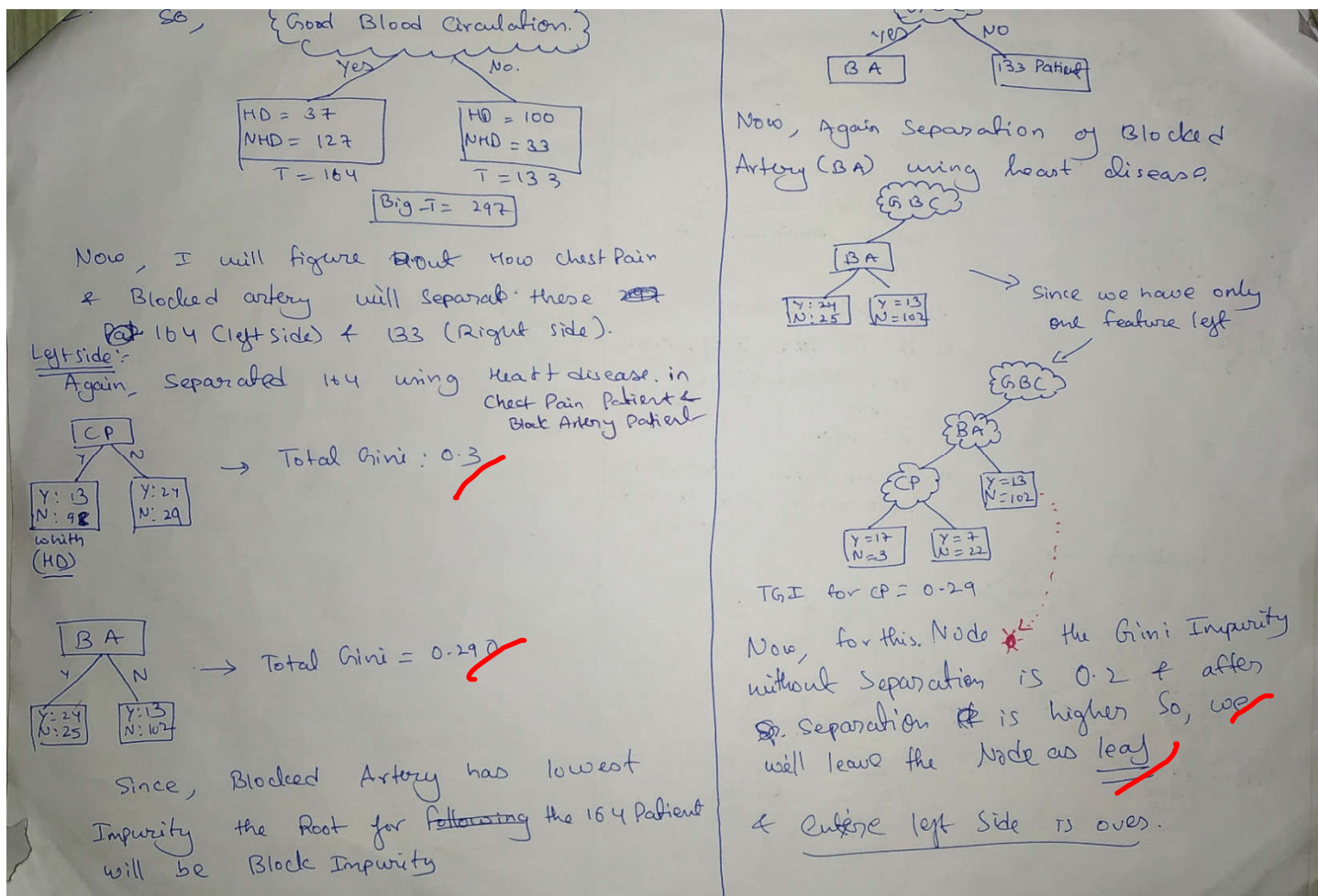
Total Gini Impurity (Chest Pain) = weight average $GINI_{Left}$ + weight average $GINI_{Right}$

$= \frac{144}{144+159} \times 0.395 + \frac{159}{144+159} \times 0.336$

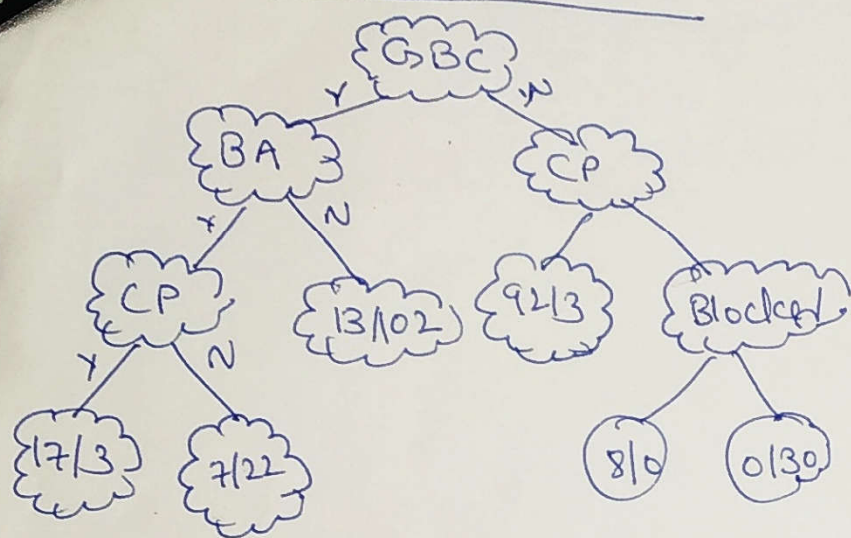
$= 0.364$

Gini Impurity of	value.
Chest Pain	0.364
GBC	0.360
Blocked artery	0.54

So, ~~GBC~~ GBC is Best option as a ~~Root~~ Root Node.

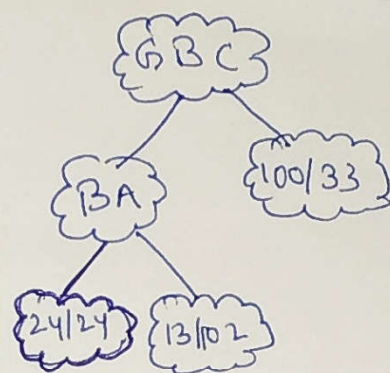


Feature Selection:-



ex: 1] if Patient has GBC, we checked Blocked artery?
if yes then chest pain?
if yes the high probability that he has Heart disease.

ex: 2] if someone has GBC But Not Blocked artery then we did not asked him for chest pain and considered it has leaf because Post-separation impurity was high compared to Pre-separation Impurity.



if this was the ^{case}, we would have never used Chest Pain to further separate the Patients.

* Decision Tree ~~are~~ have the downside of often being over fit.

It Requires large reduction in impurity to avoid the tree being over fit.