

---

---

---

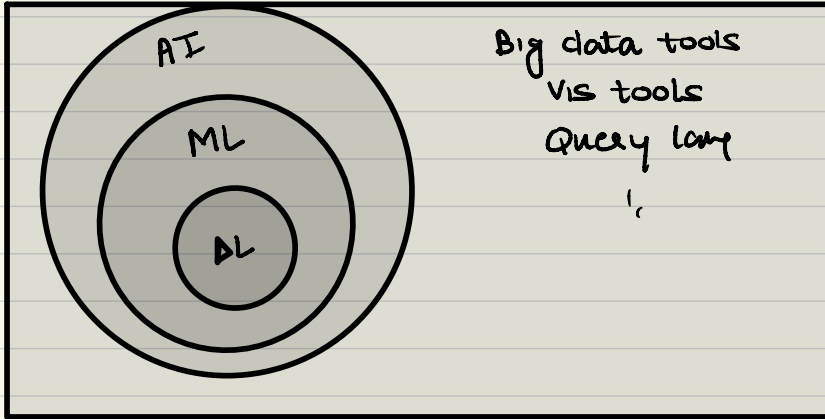
---

---

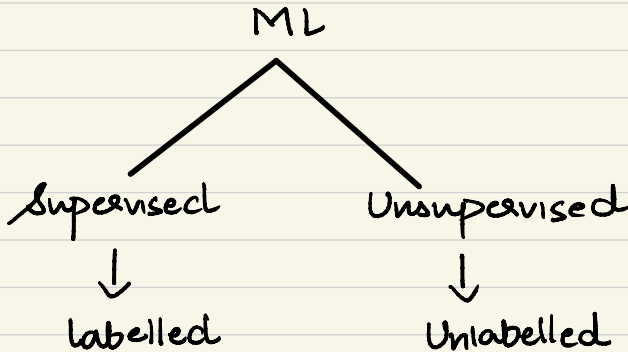


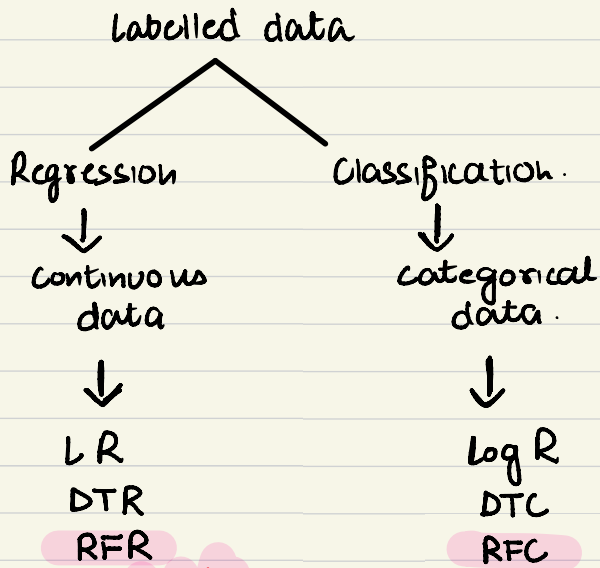
## Machine Learning

D.S



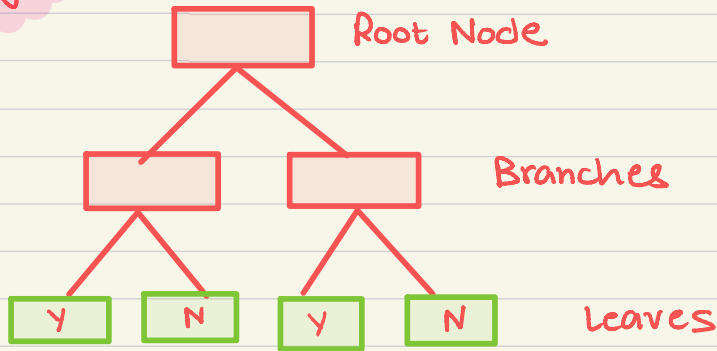
Labelled  $\rightarrow$  Features + Target  
unlabelled  $\rightarrow$  Features





~~Criteria~~  
Entropy / Information gain  
Gini's Impurity

### Decision Tree



Demerit : Overfitting

→ train data → good accuracy  
→ test data → bad accuracy

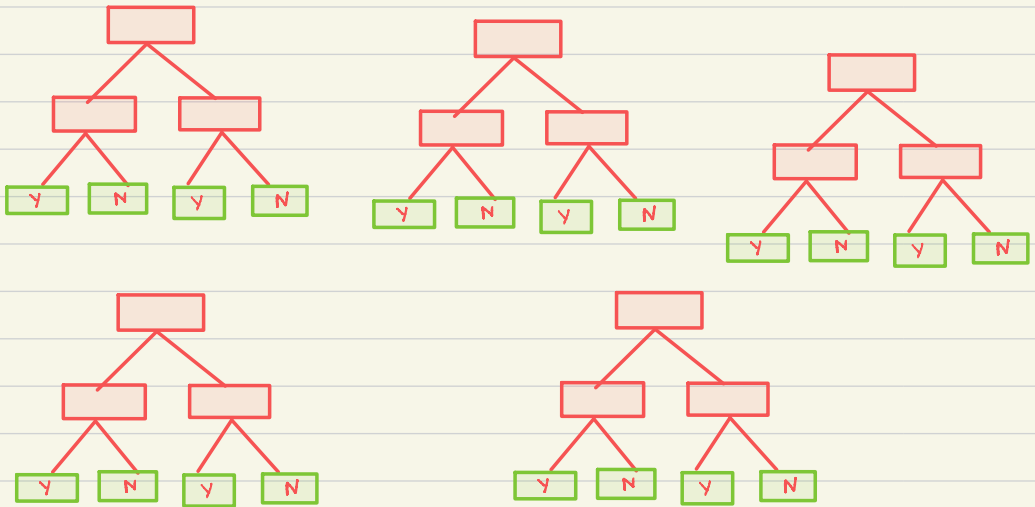
\* low bias high variance .

Underfitting  $\rightarrow$  In both bad accuracy.

high bias low variance.

Good model  $\rightarrow$  low bias low variance

### Random Forest



$\rightarrow$  R.F are built from DT

$\rightarrow$  R.F combines simplicity of DTs with flexible resulting in a vast improvement of accuracy.

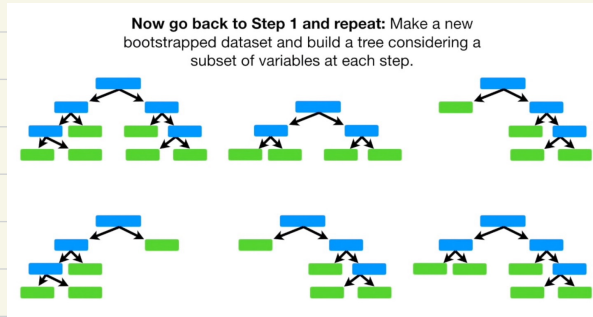
Step 1: Create a bootstrapped data

with replacement

Original Dataset				
Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
No	No	No	125	No
Yes	Yes	Yes	180	Yes
Yes	Yes	No	210	No
Yes	No	Yes	167	Yes

CP	BC	BA	W	HD
Y	Y	Y	180	Y
N	N	N	125	N
Y	Y	N	210	N
Y	Y	N	210	N

Step 2: Create a D-T using bootstrapped data.



Well, first we get a new patient...

Chest Pain	Good Blood Circ.	Blocked Arteries	Weight	Heart Disease
Yes	No	No	168	

Classification  
→ voting

Regression  
→ Average.