2014

Deep Learning

AI Algo → Large computational power

↓

Large Data sets

2000

↓

100 machines

Chat GPT

↓

Transformer (4) → 2016

Generative AI

↓

2009

Inv

↓

Machine Reading Comprehension

1 GB of Data

Resources → Chap, highly available

Alphago → Servers

↓

Deep Net

↓

25 yrs

↓ ↑ i3

$1000

↓ i11

$100

10 yrs back → ISRO

↓

Supercomputer

today → Super computer by your own

Data availability

1956
IBM
5MB
HDD

Parallel Processing

Add 1000 numbers → 999 addition

1 Addition → 1 sec → 999 secs

→ 11 + 34 → 45 — 1 addition

→ 11 + 17 + 49 → 77 — → 2 addition

2	2
3	4

1000 → 0 sec — 999 → 1

999 → 108 r'

Data + Algo → CPU → 128 cores → n

3 days

128 operations in parallel

12 hours

1 GPU = 10,000 cores

3 sec

1 QPU → 1,000,000,000 cores

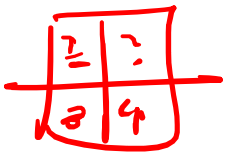
10 x 100

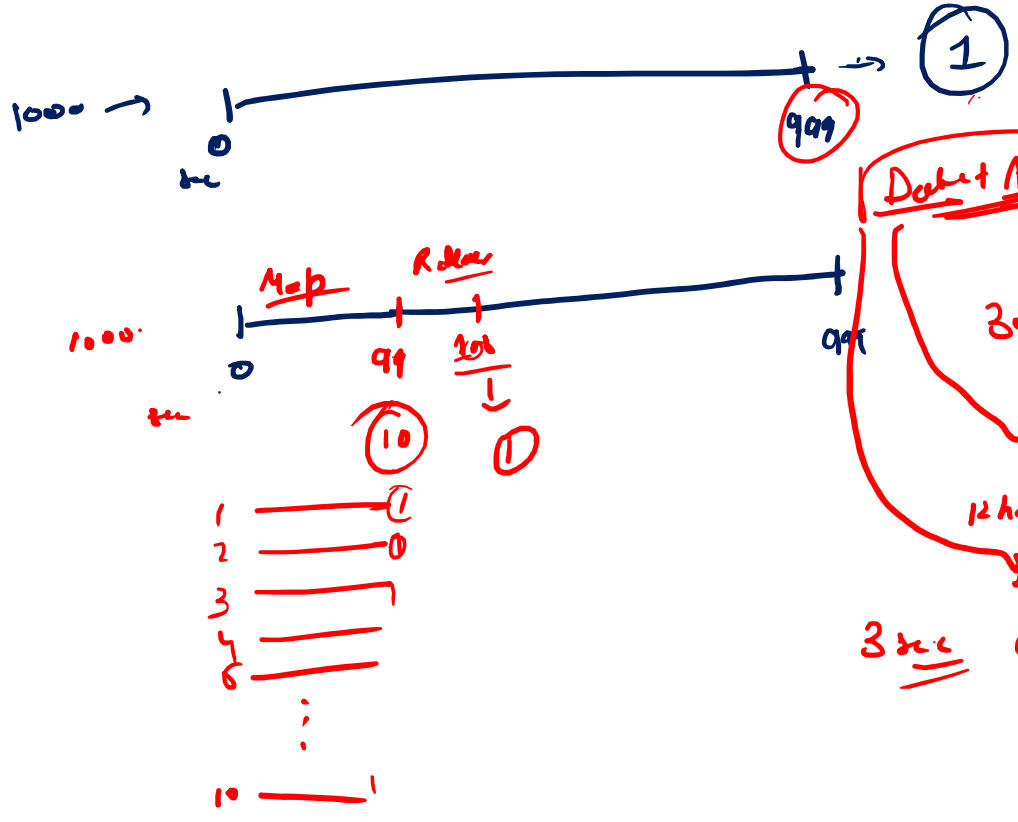
1000 sec — 0 — 99 — 108 — 999

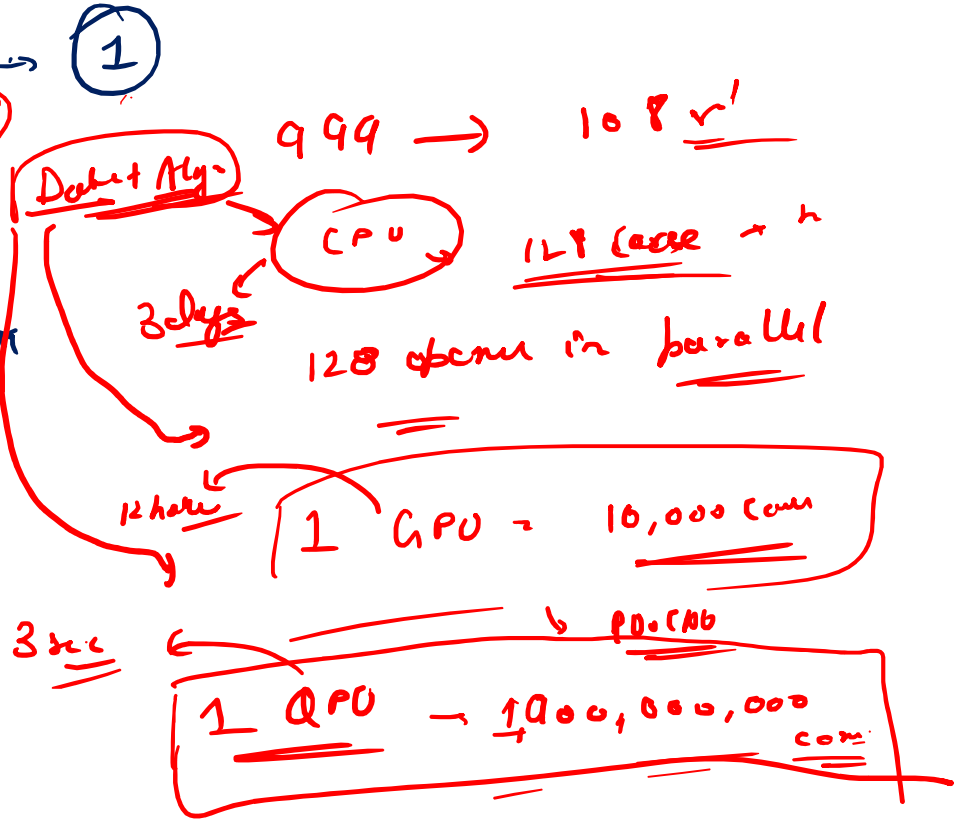
Map — Reduce

10 — 1

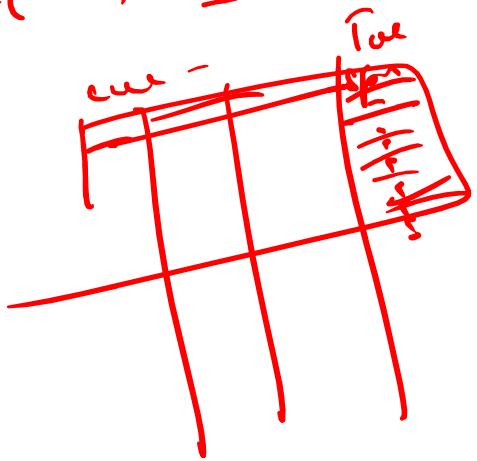
1 — 1
2 — 1
3 — 1
4 — 1
6 — 1
⋮
10 — 1







Email → sp / net



→ Alg → Model

→ Clust.

City

Age

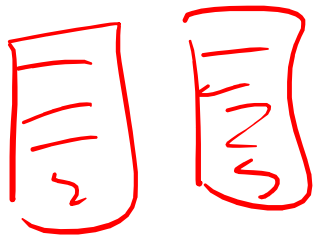
Hrs

Create th

↳ hidden

Walmart ⇒ US

→ Diapers → Beer



young fathers

Dup

→ incr

Data - Transformation

Target

Features

in_sf	beds	bath	price	year_built	sqft	price_per_sqft	elevation	
0	2	1	999000	1960	1000	999	10	→ height m
0	2	2	2750000	2006	1418	1939	0	
0	2	2	1350000	1900	2150	628	9	
0	1	1	629000	1903	500	1258	9	
0	0	1	439000	1930	500	878	10	
0	0	1	439000	1930	500	878	10	
0	1	1	475000	1920	500	950	10	
0	1	1	975000	1930	900	1083	10	
0	1	1	975000	1930	900	1083	12	
0	2	1	1895000	1921	1000	1895	12	→ hours 4
0	3	3	2095000	1926	2200	952	4	→ hours 12
0	1	1	999000	1982	784	1274	5	→ n B
0	1	1	999000	1982	784	1274	5	→ "
0	1	1	1249000	1987	826	1512	3	
0	0	1	1110000	2008	698	1590	5	
0	2	2	2059500	2008	1373	1500	5	
0	2	2	2000000	1928	1200	1667	10	
0	1	1	715000	1903	557	1284	3	
0	2	2	2498000	2005	1260	1983	2	

NY

1

SF

Usecase is to predict based on the features if a given house belongs to SF / NY

→ Classification - usecase

R

492 Rows → 250 Rows → Training Data

→ 242 Rows → Testing Data

House Classification

Ready Accessibility: Unavailable Display Settings 200%

(492 Rows)
Historical Data
↓
Excel file

→ Training Data (250 Rows)
→ Testing data (242 Rows)

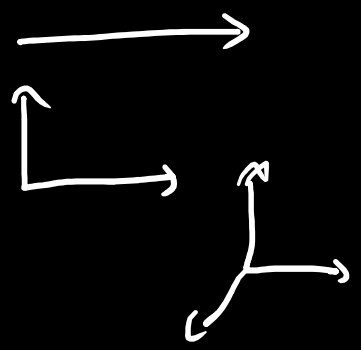
6-de

7-Features
↳ 7-dimension

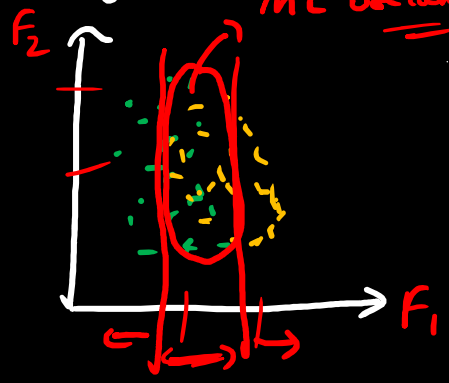
→ Machine Learning Algo

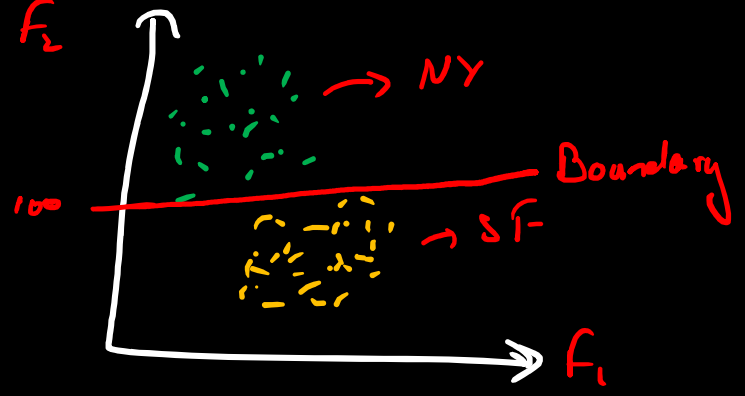
↓
Feasibility study
↓
Do we even need ML for this usecase?

4, 5, 6 ....



from the given data we need to divide/classify the rows into NY/SF houses





If any house has
$F_2 >= 100 \rightarrow NY$
$F_2 < 100 \rightarrow SF$

1, 2, 3 dimensions

1-D Boundary → $F_2 >= 100$ → NY (100%)

2-D Boundary → $F_2 >= 100$ & $F_1 > 50$ → NY (100%)

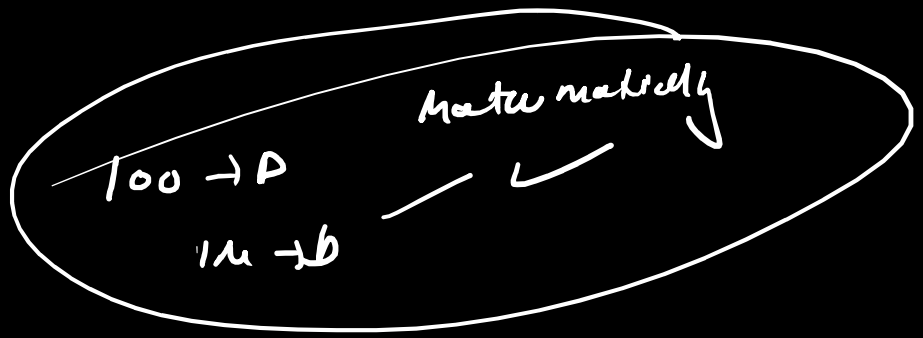
3-D Boundary → $F_1 > 50$, $F_2 \geq 100$, $F_3 < 10$ → NY (100%)

↓

4-D
5-D ✗ } → half from ML

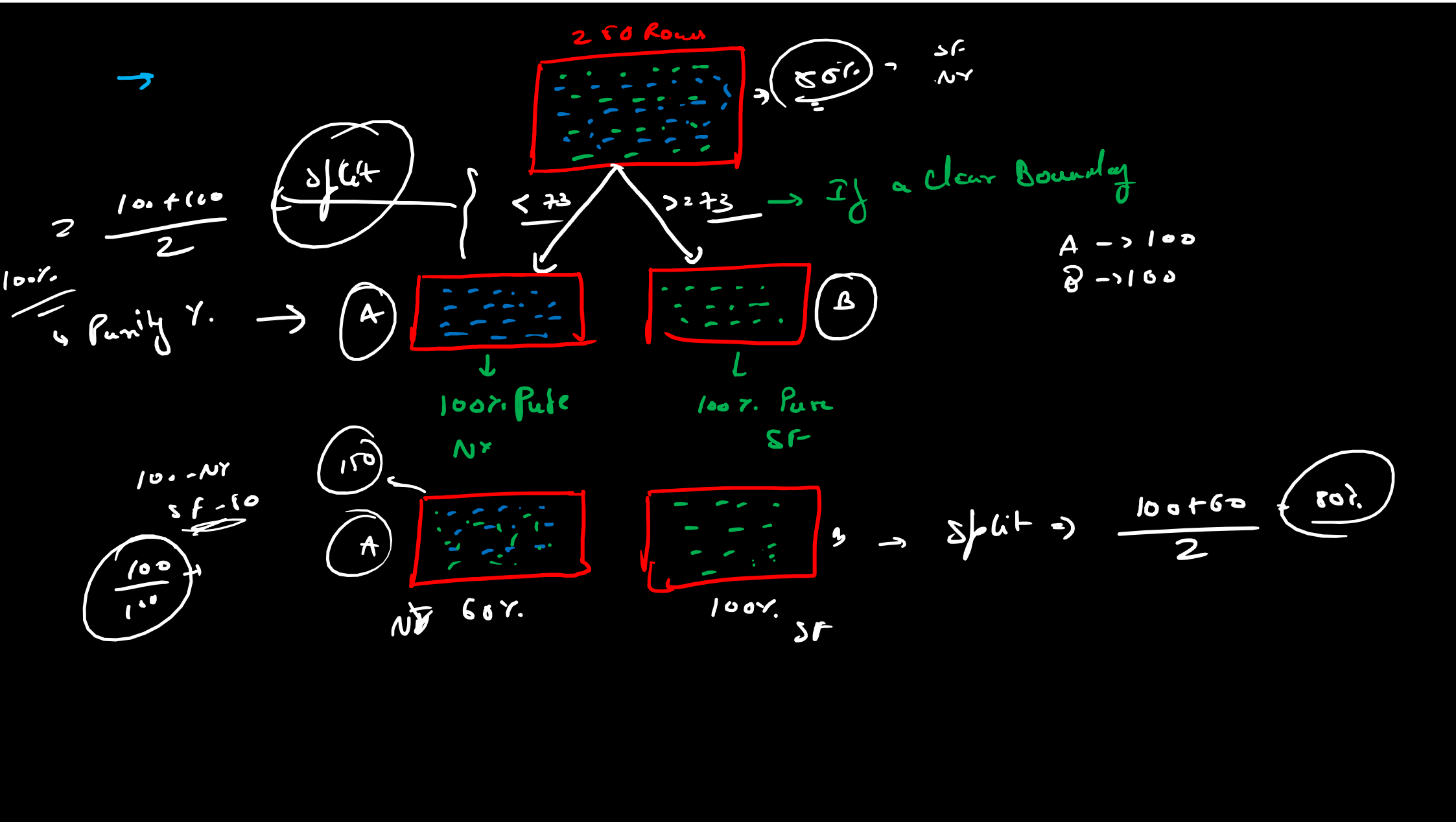
100 → D
1M → D

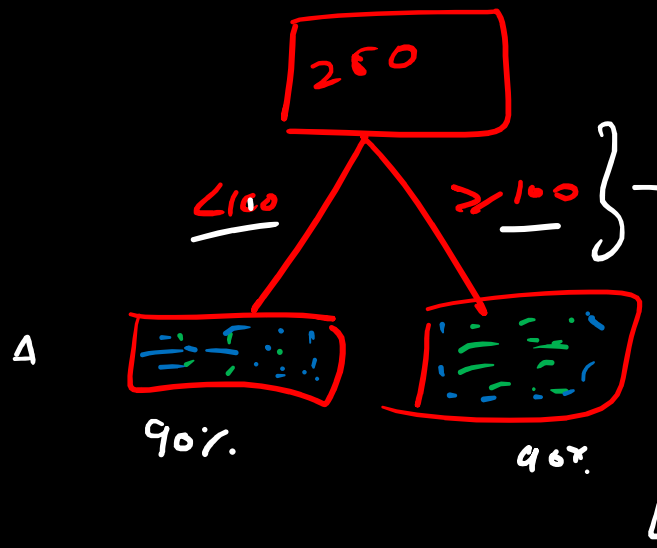
Mathematically ✓



visualization

✗


250 Rows
50%
SF
NY
split
< 73
> = 73
If a Clear Boundary
A -> 100
B -> 100
= (100+100)/2
100%
Purity %
A
B
100% Pure
NY
100% Pure
SF
100 - NY
SF - 50
100/100
150
A
NY 60%
100%
SF
split => (100+60)/2 = 80%

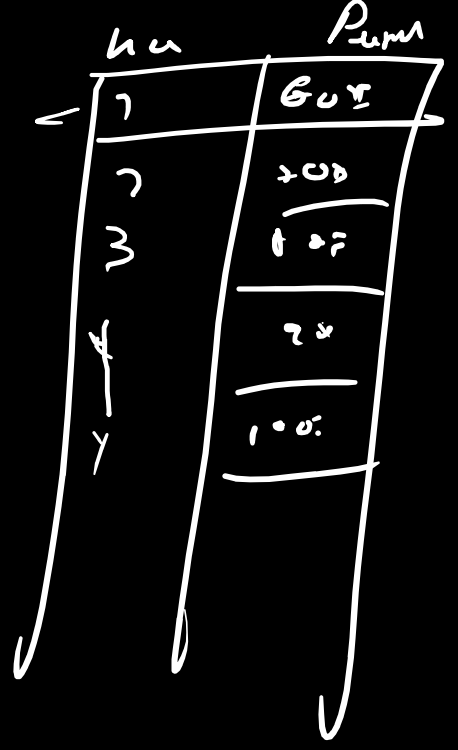


→ split → $\frac{90+90}{2}$ = 90%

250 → 100 unique height values

whatever value gives the max purity of split

→ Best split



hieght = 73 m → 63%

hieght = 0 m ⇒ 56%

$\boxed{\text{hieght} = 91.9\text{ft} \Rightarrow 82\%}$ → Best split on hieght

Interview? → 1 comparison

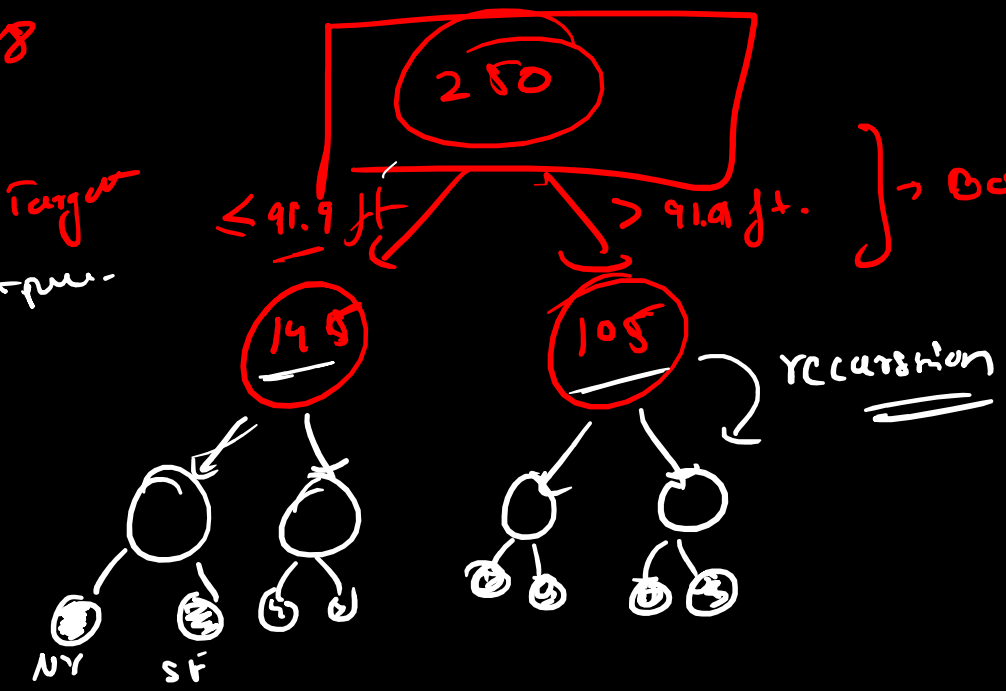
→ A → 40%

B → 10%

Columns → 8

7 features

1 Target



250

≤ 91.9 ft

> 91.9 ft.

→ Best split on height

(F = 90%)

145

105

recursion

NY

SF

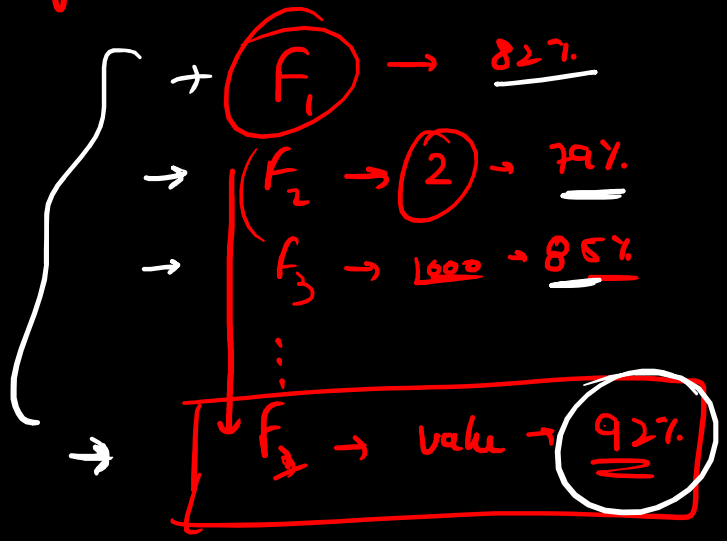
→ F_1 → 82%

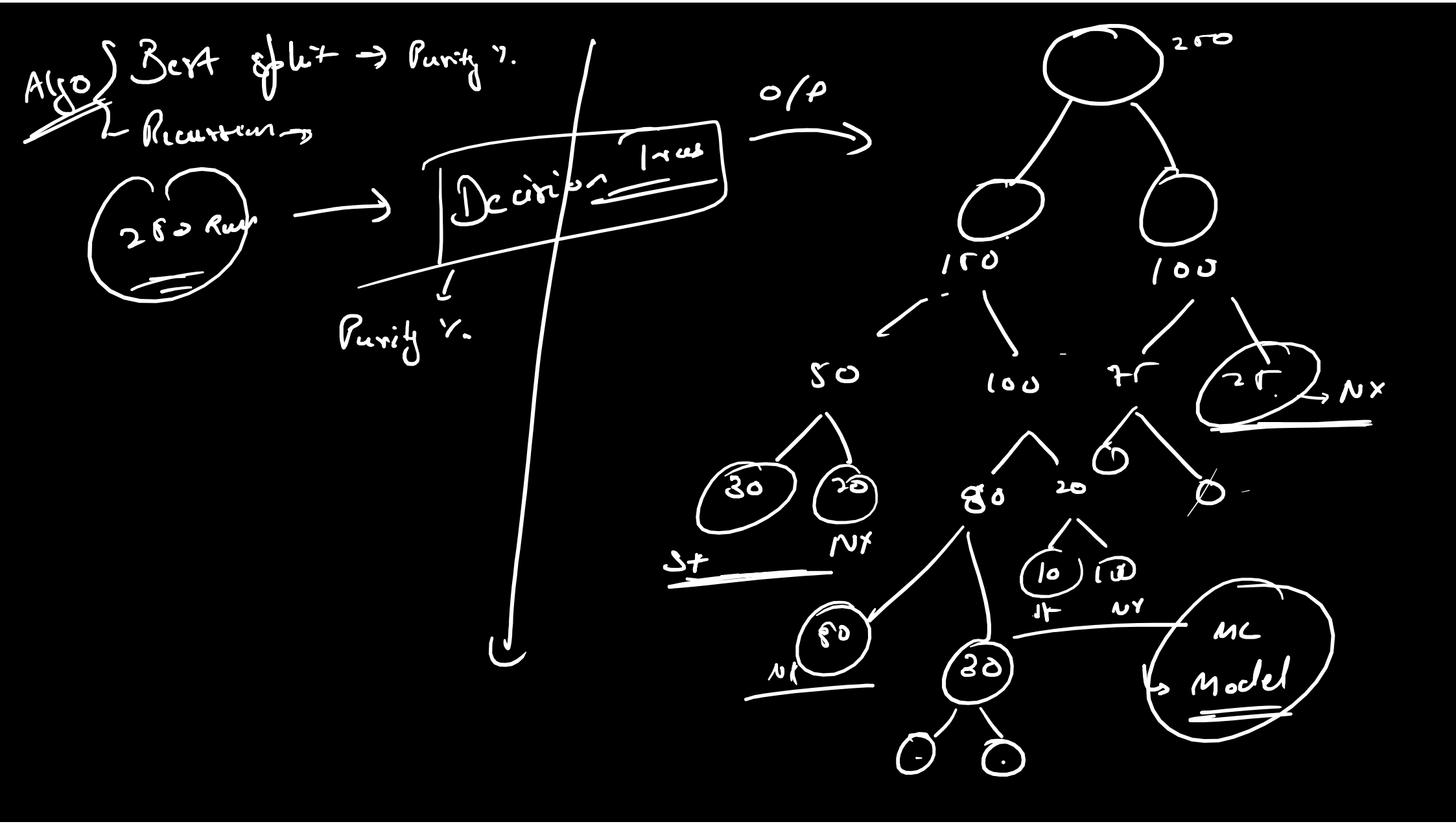
→ F_2 → 2 → 79%

→ F_3 → 1000 → 85%

⋮

→ F_3 → value → 92%




Algo
Best split → Purity %
Recursion →
250 Rows
Decision Trees
Purity %
O/P
250
150
100
50
100
75
25
NX
30
20
St
NX
80
20
10
10
St
NX
80
NX
30
MC
Model